Federal Reserve Bank
of Minneapolis

Fall 1990

Quarterly Review

Federal Reserve Bank of Minneapolis

Quarterly Review

Vol. 14, No. 4 ISSN 0271-5287

This publication primarily presents economic research aimed at improving policymaking by the Federal Reserve System and other governmental authorities.

Produced in the Research Department. Edited by Preston J. Miller, Kathleen S. Rolfe, and Martha L. Starr and typeset by Joseph R. Piepgras. Graphic design by Barbara Birr, Public Affairs Department.

Address comments and questions to the Research Department, Federal Reserve Bank, Minneapolis, Minnesota 55480 (telephone 612-340-2341).

A Banking Model in Which Partial Suspension Is Best

Neil Wallace
Adviser
Research Department
Federal Reserve Bank of Minneapolis
and Professor of Economics
University of Minnesota

Bank suspensions occurred in the United States several times in the second half of the 19th century and the early part of the 20th century, and they were widely viewed as both undesirable and avoidable. The view that bank suspensions should and could be avoided has played a decisive role in determining the nature of our current banking system. This view was the driving force behind the creation of the Federal Reserve System, a system which helped determine the course and severity of banking difficulties during the Great Depression. The experience of the Great Depression, in turn, produced our current financial system and our deposit insurance system with its attendant problems.

In this paper I question the view that bank suspensions are undesirable by providing a model of an economy in which a system with occasional partial suspension achieves the best possible outcome. By a *partial suspension* I mean a situation in which some depositors receive less than some other depositors merely because they withdraw later, after a (partial) suspension has been declared. This was a widely reported feature of the banking panic of 1907 and in one form or another must have been a feature of other suspension episodes. In the model I describe, the best possible arrangement has this feature. There is simply no better way to deal with randomness in withdrawal demands—which in this model arises exogenously, that is, naturally—than to declare a partial suspension.

The model developed here is a variant of an earlier one (Wallace 1988), which itself was a variant of the well-known Diamond-Dybvig (1983) model of banking. These models are ones in which individuals are uncertain at the time they make deposits when they will want to make withdrawals. Deposits exist because they provide better withdrawal options than individuals could achieve on their own—but, under the assumptions made, only if the banking system takes on an illiquid portfolio. Diamond and Dybvig's article and my earlier article discussed two versions of our models, one in which individual uncertainty about the timing of withdrawals exactly averages out across people, a version without aggregate risk, and one in which it does not completely average out, a version with aggregate risk.

The no-aggregate-risk version has the defect that a very simple banking arrangement achieves the best possible outcome, but despite banking system illiquidity, it is an arrangement that does not resemble the banking suspensions that we have from time to time experienced. The aggregate-risk version is more promising in this regard, but in general is very complicated. In Wallace 1988, I established only a limited property of desirable arrangements for the aggregate-risk version: In the best arrangement, withdrawal options are somewhat dependent on when individuals attempt to withdraw. Here I look at a special case of the aggregate-risk version with a small amount of aggregate risk, all of which is limited to a group who show up last. Although this case is somewhat special, it is easy to describe the best arrangement for it. Under the assumptions that imply banking system illiquidity, the best arrangement has partial suspension: when the late-to-show-up group want to withdraw, they get less than those who withdrew earlier. Moreover, as I will discuss later, arrangements that resemble partial suspension are likely to emerge from less special specifications with aggregate risk.

In my model, the cause of a bank run and a partial suspension is exogenous—an aggregate shock to tastes that makes the number of people wanting to withdraw unusually large. No attempt is made here to argue that this was the main cause of bank runs and suspensions in the historical episodes mentioned above. Rather, the argument here is that the crucial elements of the model are plausible features of actual economies. Since these plausible features by themselves imply that the best arrangement has partial suspensions, partial suspensions per se should not be grounds for indicting a banking system.

The Model

The model has three main ingredients. One ingredient is that at the time that individuals make decisions about what assets to hold, individuals are uncertain about when they will want to dispose of their assets in order to spend. A second ingredient is that individuals are naturally isolated from each other in a way which forces banks and the business sector to deal with customers sequentially. The third ingredient is that real investment opportunities have relatively good returns if they are not interrupted. Although these ingredients seem straightforward, it is not easy to put them together so that their implications for what is possible and desirable can be deduced. Doing that requires that these fairly plausible ingredients fit into a picture of an entire economy that we can understand. That, in turn, requires some drastic simplifications which I now describe.

Suppose there are N people, where N should be thought of as large, and three time periods, labeled 0, 1, and 2. There is a single good at each period, which is the economy's consumption good. The economy starts out with some of this good at period 0. Denote the per-capita amount of this by y. A technology is also available which allows this good at period 0 to be used to produce consumption at period 1 and period 2. If x is the period 0 input into this technology, then R_1x is the period 1 output, R_1 being the return between periods 0 and 1. Any output removed from the production process during period 1—call it λR_1x for some λ between 0 and 1—must be consumed or it is lost. The part of R_1x not removed from the production process, $(1 - \lambda)R_1x$, becomes $R_2(1 - \lambda)R_1x$ at period 2. The marginal return between period 1 and period 2, R_2, is assumed to be relatively high in a sense to be described later.

Also suppose that the preferences or tastes of each of the N people are identical as of period 0 and are such that individuals are uncertain about whether they will want to consume at period 1 or, relative to the return R_2, will want to wait and consume at period 2. No one wants to consume at period 0. Each person learns at period 1 whether he or she will want to consume at period 1, or whether he or she will want to wait and consume at period 2. This is how uncertainty about when people will want to spend is put into the model. Some fraction of the N people will turn out to want to consume at period 1, so we label them *type 1* or the *impatient* type. The remaining fraction will turn out to want to consume at period 2, so we label them *type 2* or the *patient* type. (To be precise, for any amount x, the patient people are happy to trade x units of period 1 consumption for R_2x units of period 2 consumption, where, recall, R_2 is the marginal return of the technology between periods 1 and 2.) In the no-aggregate-risk version of the model, these fractions are known to everyone at period 0. In the aggregate-risk version, which is the primary focus here, there is uncertainty about these fractions, the details of which will be spelled out later.

Because each person at period 0 is uncertain about whether he or she will be impatient or patient at period 1, each cares about what he or she will be able to consume at period 1 if impatient and also what he or she will be able to consume at period 2 if patient. From the point of view of period 0, turning out to be impatient or patient is analogous to experiencing an auto accident or not experiencing an auto accident. Before the event, a person cares about both contingencies, even though only one will occur; a person will either have or not have an accident. As I proceed, I will be making additional assumptions about the way that each person cares about the two contingencies: namely, what the person will be able to consume at period 1 if impatient and what the person will be able to consume at period 2 if patient.

Before describing the remaining features of the model, it is helpful to introduce the device I use to describe what is possible and what is desirable. Suppose that there is a benevolent planner (not one of the N people), whose goal is to set up an arrangement to make each of the N people as well off as possible as of period

0, subject to treating them identically as of period 0. Suppose also that this planner has control of all the period 0 resources, Ny units of the period 0 good, and has access to the technology described above. Very roughly speaking, what the planner will do is invest all the resources in the technology, withdraw some during period 1 to allow those who turn out to be impatient to consume then, leave the rest in the technology, and distribute the proceeds at period 2 to those who turn out to be patient. After describing the best planning arrangement, I will show that this arrangement can be interpreted as a mutual banking system that would be voluntarily entered into at period 0 by all the N people, if each person starts with y units of the period 0 good and has access to the technology.

Two difficulties remain that complicate the problems of the planner. First, although the planner is assumed to know at period 0 how each person values consumption at period 1 if impatient and consumption at period 2 if patient, the planner will not know at period 1 whether a person turned out to be patient or impatient unless the person reveals it. That is, whether a person turns out to be type 1 or type 2 is private information. Second, at period 1, the planner will encounter the N people in a sequential, random order and must determine a period 1 consumption for each person when that person is encountered. This means that if some person is the kth person to encounter the planner at period 1, then the planner can ask whether he or she is patient or impatient and can give that person some amount of period 1 consumption based on that person's response and on the responses of the $k - 1$ people the planner has already encountered. But the planner cannot delay determining the kth person's period 1 consumption until he or she gets responses from those not yet encountered. Moreover, it is assumed that the N people do not have any contact with each other during period 1. This second feature is how sequential service is put into the model.

Before describing the aggregate-risk version of the model in detail, it will be helpful to review the motivation for a kind of banking arrangement in the simpler setting without aggregate risk because the same motivation carries over to the aggregate-risk version. This discussion will also allow me to describe some important additional assumptions about preferences, about the way people value consumption at period 1 if impatient and consumption at period 2 if patient.

No Aggregate Risk

Since the first difficulty facing the planner, private information, will turn out to be unimportant, it is convenient to begin by ignoring this difficulty. That is, I will first find the best arrangement under the assumption that the planner can identify types at period 1. I will then describe how the planner achieves this arrangement despite having to rely on people to truthfully reveal their type. Also assume that it is desirable to give impatient people consumption only at period 1 and patient people consumption only at period 2. (See the Appendix for the argument.)

The planner could give people of the same type who are encountered at different times during period 1 different amounts of period 1 consumption and of period 2 consumption. From the point of view of people at period 0, there is uncertainty about when they will encounter the planner. I assume that people are risk averse so that, all else equal, they would prefer not to have the amounts they get to consume depend on the random order in which they encounter the planner. (A person is *risk averse* if the person rejects fair gambles, or buys actuarially fair insurance when given the opportunity.) Since the planner wants to make each person as well off as possible, the planner chooses an arrangement where the amount people get to consume depends on the order in which they encounter the planner only if that serves some purpose. Here it does not. Therefore, I will look at all pairs of period 1 consumption (for those who turn out to be impatient) and period 2 consumption (for those who turn out to be patient) that are consistent with the resources and the technology.

The technology limits the total amount of period 1 consumption, denoted C_1, and the total amount of period 2 consumption, denoted C_2, in the following way. Given that the planner invests the total resources, Ny, at period 0, at period 1 the planner has R_1Ny. If the amount C_1 is withdrawn for period 1 consumption, then what is left is $R_1Ny - C_1$, which accumulates to $R_2(R_1Ny - C_1)$ in period 2 and becomes the amount available for consumption at period 2. That is, the resources and technology allow for total consumption pairs, (C_1,C_2), that satisfy

(1) $C_2 = R_2(R_1Ny - C_1)$.

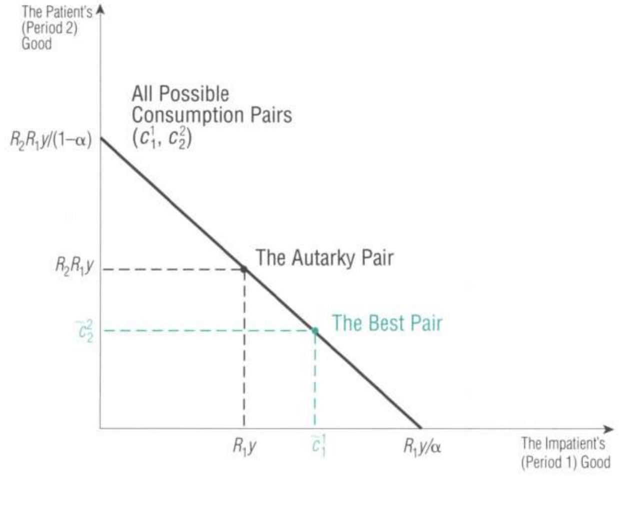
In this no-aggregate-risk version, let α be the fraction of the N people who will turn out to be impatient, type 1, and let $1 - \alpha$ be the remaining fraction who will turn out to be patient, type 2. Whatever are C_1 and C_2, the planner wants to divide the former equally among the impatient and the latter equally among the patient. For the moment, let c_i^h be the consumption at period i of type h, so that c_1^1 denotes the consumption at period 1 of each impatient person and c_2^2 denotes the consumption at period 2 of each patient person. Given that the planner devotes all of C_1 to the impatient and all of C_2 to the patient, $C_1 = N\alpha c_1^1$ and $C_2 = N(1 - \alpha)c_2^2$. If these expressions are substituted into equation (1), and given that types can be identified, the pairs (c_1^1, c_2^2) that are attainable are those that satisfy

$$(2) \quad (1 - \alpha)c_2^2 = R_2(R_1 y - \alpha c_1^1).$$

All the (nonnegative) pairs (c_1^1, c_2^2) that satisfy (2) are shown as the downward-sloping line in Figure 1.

In interpreting Figure 1, it is important to remember that c_1^1 and c_2^2 are contingent commodities. No one person will end up consuming the amount c_1^1 at period 1 and the amount c_2^2 at period 2. Rather, each person will end up consuming one or the other depending on whether the person turns out to be impatient or patient. Since each person, as of period 0, does not know which type he or she will turn out to be and therefore cares about both eventualities, it is relevant to look at the possible combinations. One of the pairs singled out in the diagram is the pair $(R_1 y, R_2 R_1 y)$. This is the pair attainable if each person owns y units of the period 0 good and acts alone. In that case, if the person turns out to be impatient, the person does best to withdraw everything from the production process and consume $R_1 y$ at period 1. If, instead, the person turns out to be patient, then the person does best to leave the production process undisturbed and to consume the result, $R_2 R_1 y$, at period 2. No matter what α is, this pair, called the *autarkic outcome*, is one that is always possible. The planner, or people acting together, can also achieve other outcomes. At one extreme, for example, the planner could decide to devote all the resources to the fraction α of people who turn out to be impatient, giving each one $R_1 y/\alpha$. At the other extreme, the planner could give nothing to those who turn out to be impatient and give $R_2 R_1 y/(1 - \alpha)$ to each person who turns out to be patient. Figure 1 shows these extremes (on the axes) and all other (c_1^1, c_2^2) attainable pairs. Notice that the higher is α—that is, the greater is the fraction who are impatient—the steeper is the line through the point $(R_1 y, R_2 R_1 y)$ that depicts the (c_1^1, c_2^2) attainable pairs.

Figure 1
Consumption Without Aggregate Risk



The Best Arrangement

What should the planner do? As of period 0, all N people prefer the same (c_1^1, c_2^2) pair from among all those on the line in Figure 1. That is, subject to being treated the same as of period 0, there is no conflict of interest among them. So, the planner should simply choose the (c_1^1, c_2^2) pair on the line in Figure 1 that all the N people most prefer. Assume that the most preferred pair is southeast of the autarkic outcome, $(R_1 y, R_2 R_1 y)$. This says that relative to the autarkic outcome, people prefer more period 1 consumption and less period 2 consumption. This assumption, which will be maintained throughout, can also be interpreted as people wanting to buy some actuarially fair insurance against turning out to be impatient relative to where they would be under the autarkic outcome. The most preferred pair is denoted $(\tilde{c}_1^1, \tilde{c}_2^2)$.

I now describe how the planner achieves this most preferred pair even though people are encountered in order and at random and even though the planner does not know anyone's type. The planner at period 0 announces the following scheme. All the resources have been invested in the technology at period 0. As people are encountered during period 1, they will be asked their type. Those who say they are type 1, impatient, will be given $\tilde{c}_1^1$; those who say they are type 2, patient, will be told that they will divide equally at period 2 what is left after the period 1 withdrawals have been made. The planner also announces that, at most, the total amount $\alpha N\tilde{c}_1^1$ will be withdrawn at period 1. This last feature should be thought of as a *threatened suspension* scheme.

Another aspect of the best possible no-aggregate-risk arrangement is how people care about these preferred pairs. One characteristic of the way that people care about different pairs (c_1^1, c_2^2) is that the pair $(\tilde{c}_1^1, \tilde{c}_2^2)$ is *incentive compatible* in the following sense. Those who turn out to be type 1, impatient, prefer receiving $\tilde{c}_1^1$ at period 1 to receiving $\tilde{c}_2^2$ at period 2, and those who turn out to be type 2, patient, prefer receiving $\tilde{c}_2^2$ at period 2 to receiving $\tilde{c}_1^1$ at period 1. This property and the threatened suspension were shown by Diamond and Dybvig (1983) [and Wallace (1988)] to imply that each person wishes to announce his or her true type to the planner no matter what the person thinks others will announce. The threatened suspension plays an important role here because it assures the patient people that no matter what other people announce, there will be enough resources available at period 2 to give those who have not received a period 1 payment at least $\tilde{c}_2^2$.

Interpretation as a Banking Arrangement

It is straightforward to interpret this planning arrangement as a banking arrangement that the N people voluntarily enter into at period 0 if I assume that each one starts with y units of the period 0 good and has access to the intertemporal technology previously described. The banking arrangement allows people who deposit y at period 0 either to withdraw $\tilde{c}_1^1$ at period 1 or to get a pro rata share of what is left at period 2, and it has the threatened suspension feature that no more than $\alpha N\tilde{c}_1^1$ will be paid out during period 1. Note that the bank is like a cash machine, except that it dispenses the consumption good, not cash. People contact it in a random order, just as they are assumed to encounter the planner. Since the bank dispenses the consumption good instead of cash and since production occurs within it, the machine represents the consolidated banking-business sector for our model economy.

The same reasoning used for the planning solution implies that any depositor who tells the truth actually receives $\tilde{c}_1^1$ if impatient and $\tilde{c}_2^2$ if patient, no matter what others attempt to do during period 1. By construction, such a deposit is preferred at period 0 to what a person can achieve on his or her own—namely, R_1y if impatient and R_2R_1y if patient. It follows that everyone wants to participate in the deposit scheme. Since $\tilde{c}_1^1/y$ exceeds R_1 and $\tilde{c}_2^2/y$ is less than R_2R_1, the deposit offers a higher short-term return and a lower long-term return than people can achieve on their own. The scheme also produces an illiquid portfolio for the banking-business sector in the following sense. As of period 0, each person has the same deposit, a deposit which permits the person either to withdraw $\tilde{c}_1^1$ at period 1 or to withdraw $\tilde{c}_2^2$ at period 2. Therefore, from the point of view of an outsider at period 0, the banking-business sector has a potential period 1 obligation equal to $N\tilde{c}_1^1$, while having total resources at that time equal to the lesser amount, NR_1y.

Thus, this no-aggregate-risk version of the model displays some of the main properties of actual economies with banking systems. People hold deposits because they pay relatively good returns if held for short periods, and this liquidity of depositors is accompanied by an illiquid portfolio for the banking-business sector. Although a threatened suspension is part of this system, it nonetheless cannot be claimed that suspensions actually occur. In particular, no one ever regrets showing up to withdraw later than others. To get that property to be part of a best outcome, a slightly different model is needed. The aggregate-risk version of the model, which is the focus of the rest of the paper, displays this additional property.

Aggregate Risk

Now suppose that the fraction of people who will turn out to be impatient is not known at period 0. Assume instead that that fraction will take on one of two magnitudes in the following way. As of period 0, the order in which people will encounter the planner during period 1 is completely random, which means that all orders are equally likely. Now suppose that among the fraction p who encounter the planner first, exactly the fraction α are impatient and exactly the fraction $1 - \alpha$ are patient. Also suppose that among the fraction $1 - p$ who encounter the planner last, there are two possibilities. With probability q all these people are impatient, and with probability $1 - q$ all are patient. Assume that $1 - p$ is positive, but near zero. Therefore, the total fraction who will turn out to be impatient will be either $p\alpha + (1 - p)$ or $p\alpha$. The former occurs with probability q and the latter with probability $1 - q$.

The assumption about the order in which people meet the planner is to be understood as follows. The planner first meets the fraction p among whom it is known that the fraction of impatient people is α. This could be thought of as the *normal demand* for period 1 consumption. Then the planner encounters the aggregate risk; all the remaining people encountered are either impatient or patient. Note that I am continuing to assume that all N people are identical as of period 0. That is, each person holds the following views with regard to his or her chances of being a particular type and encountering the planner at a particular place in line. Each person thinks there is a probability p of being among the first fraction p to encounter the planner, and conditional on being in this first group, each person thinks there is a probability α of being impatient and a probability $1 - \alpha$ of being patient. Each person also thinks there is a probability $1 - p$ of being among the last fraction $1 - p$ to encounter the planner, and conditional on that each person thinks there is a probability q of being impatient and a probability $1 - q$ of being patient.

The assumption that the total fraction who are impatient is random seems plausible. The assumption that randomness occurs only among the group who encounter the planner last is admittedly special. It is adopted because it simplifies the analysis. Quite naturally it is assumed that the only way for the planner to learn the aggregate state, whether the total fraction of impatient people is $p\alpha + (1 - p)$ or $p\alpha$, is to infer it from the responses of the people the planner encounters. The critical implication of the assumption that aggregate randomness arises from the group the planner encounters last is that the planner can infer nothing about the aggregate state from the responses of the fraction p of people encountered first, people who themselves do not know the aggregate state. It is also helpful that if the responses of people in the fraction $1 - p$ whom the planner encounters last are truthful, then the aggregate state is implied by the response of any one of them.

The Best Arrangement

Next I will describe the qualitative features of the best arrangement for the aggregate-risk version. This best arrangement will resemble a partial suspension in that those who encounter the planner last if impatient get less during period 1 than the impatient who encounter the planner first. Moreover, the period 2 consumption of the patient is smaller when those who encounter the planner last are impatient (when a partial suspension occurs) than when those people are patient (when no partial suspension occurs). All the assertions made in this section are proved in the Appendix.

I describe the best arrangement using a notation that describes period 1 and period 2 consumption for each combination of two contingencies. One contingency is the *aggregate state*—whether those who encounter the planner last are all impatient, labeled *state 1*, or patient, *state 2*. The other contingency is whether a person is among the first p to encounter the planner or among the last $1 - p$ to encounter the planner. The notation is set out in Table 1, which is a 2×2 contingency table.

In Table 1, the second superscript on c denotes when the person encounters the planner—1 for the first p and 2 for the last $1 - p$—while the number in parentheses denotes the aggregate state. As in the earlier notation, the first superscript on c denotes the type—1 for impatient, 2 for patient—and the subscript denotes the period. The probability (prob) attached to each consumption pair is the probability viewed from period 0 that a person will end up with that pair of contingencies.

Table 2 is similar, except that it specifies some characteristics of the best arrangement. *Best* means most preferred as of period 0, subject to treating all N people identically as of period 0 and subject to all the constraints set out above, including sequential service and private information regarding type. Each person at period 0 cares about all four contingencies because each person is faced with the probabilities given in Table 1.

In the first row of Table 2, one characteristic is that period 1 consumption of impatient people who are among the first p to encounter the planner is not dependent on the aggregate state. This is a feature of the best arrangement because it is forced on the planner by the sequential service difficulty; the planner does not know the aggregate state when period 1 consumption for these people is determined and therefore is simply unable to make their period 1 consumption dependent on the aggregate state. Another feature of the best arrangement is the zeros in the second row. When all the last $1 - p$ people to encounter the planner are impatient, which is aggregate state 1, the best arrangement does not give them any period 2 consumption, and when they are all patient, aggregate state 2, the best arrangement does not give them any period 1 consumption. This is a consequence of the preference and technology assumptions.

Table 1
Notation for Consumption Pairs in the Aggregate-Risk Version

	Aggregate State (Type of Last $1-p$ People to Meet the Planner)	
Place in Line	1 (Impatient)	2 (Patient)
First p	$c_1^{11}(1), c_2^{21}(1)$ prob pq	$c_1^{11}(2), c_2^{21}(2)$ prob $p(1-q)$
Last $1-p$	$c_1^{12}(1), c_2^{22}(1)$ prob $(1-p)q$	$c_1^{12}(2), c_2^{22}(2)$ prob $(1-p)(1-q)$

Table 2
Some Characteristics of the Best Possible Consumption Pairs

	Aggregate State (Type of Last 1–p People to Meet the Planner)	
Place in Line	1 (Impatient)	2 (Patient)
First p	$c_1^{11}, c_2^{21}(1)$	$c_1^{11}, c_2^{2}(2)$
Last 1–p	$c_1^{12}(1)$, 0	0, $c_2^{2}(2)$

One more feature of Table 2 shows up in the second column. In the aggregate state 2, all patient people get the same period 2 consumption. This is a consequence of the assumption that people in the model are risk averse.

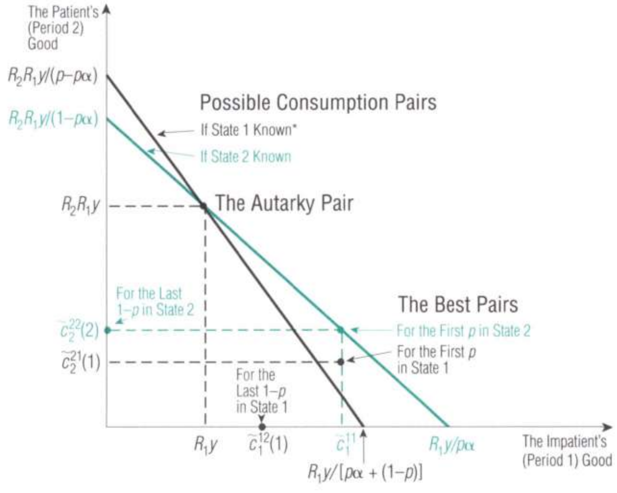
The remaining qualitative features of the best arrangement are described with the aid of Figure 2. Figure 2 contains two lines, each like the line in Figure 1. The steeper line gives the combinations of period 1 consumption for all impatient people and period 2 consumption for all patient people that would be possible if the planner knew at the beginning of period 1 that the aggregate state was state 1. The flatter line gives the combinations that would be possible if the planner knew at the beginning of period 1 that the aggregate state was state 2. If the planner had that information, then the best arrangement would be two points, one on each line. Both points would be southeast of the autarkic point [the combination (R_1y, R_2R_1y) where the two lines cross] and one would be northeast of the other. That is, if the planner knew the aggregate state early enough, then the best arrangement would have both period 1 consumption for the patient and period 2 consumption for the impatient dependent on the aggregate state and would have nobody's consumption dependent on when the planner is encountered.

Such an outcome is not possible, however. Instead, period 1 consumption of those who are among the first p to meet the planner must be the same in both aggregate states. That is, in terms of Figure 2, the combinations of period 1 consumption if impatient and period 2 consumption if patient in the two aggregate states of the fraction p who first encounter the planner must align vertically—one must be directly above the other. Since the best arrangement does not make period 2 consumption of the patient dependent on when the planner was encountered during period 1, the best outcomes in state 2 have those who encounter the planner first with a pair like that shown in Figure 2 on the flatter line and those who encounter the planner last with the pair on the vertical axis at the same height. Now what is interesting is where the other two consumption pairs are located qualitatively relative to the state 2 outcomes—the pair for those who encounter the planner first and the pair for those who encounter the planner last if the state is 1.

The best arrangement is the best way of coping with the aggregate risk, given that the planner has no way of knowing the aggregate state until period 1 consumption of the impatient among the first fraction p the planner encounters has been determined. In order to avoid having all the uncertainty implied by the aggregate risk fall on those who turn out to be patient, the best arrangement has consumption of the impatient depend somewhat on their place in line. Since people are risk averse, making consumption of the impatient depend on the random order in which they encounter the planner is by itself undesirable; however, having some such uncertainty is better than having all the uncertainty implied by the aggregate risk fall on the period 2 consumption of those who turn out to be patient.

The best arrangement makes period 1 consumption of some of the impatient people dependent on the aggregate state by having the period 1 consumption of the

Figure 2
Consumption With Aggregate Risk



*In *state 1*, all who meet the planner last are *impatient*; in *state 2*, they are *patient*.

$1 - p$ who encounter the planner last in state 1 be less than the period 1 consumption of the impatient among those who encounter the planner first. This is the feature I identify with partial suspension. (In Figure 2, the dot representing the consumption pair for those who encounter the planner last in aggregate state 1 is on the horizontal axis to the left of the first period consumption of the impatient who are among the first fraction p to encounter the planner.) Giving out less at period 1 in aggregate state 1 preserves additional resources for period 2 and permits the second period consumption of the patient in aggregate state 1 to be higher than it could otherwise be. (In Figure 2, the dot representing the consumption pair for those who encounter the planner first in aggregate state 1 is above the steeper line; it would have to be on the steeper line if all the impatient were given the same period 1 consumption.)

I now describe how the planner achieves this most preferred outcome even though people are encountered in order and at random and even though the planner does not know anyone's type. The planner at period 0 announces the following scheme. All the resources have been invested in the technology at period 0. As people are encountered during period 1, they will be asked their type. Those who say they are type 1, impatient, will be given $\tilde{c}_1^{11}$ until a total of $Np\alpha\tilde{c}_1^{11}$ is disbursed. At that point partial suspension will go into effect; any additional people who claim to be impatient will be given $\tilde{c}_1^{12}(1)$, which, as explained above, is less than $\tilde{c}_1^{11}$. The planner will continue to make disbursements to those who claim to be impatient until total disbursements reach $Np\alpha\tilde{c}_1^{11} + N(1 - p)\tilde{c}_1^{12}(1)$. At that point a total suspension will go into effect and no further period 1 disbursements will be made. Those who say they are type 2, patient, will be told that they will divide equally at period 2 what is left after the period 1 withdrawals are made.

This scheme obviously is compatible with sequential service and with the planner's information. It also has the feature that each person has an incentive to announce his or her true type no matter what the person thinks others will do. Under the preference assumptions made, which are essentially those of the original Diamond-Dybvig model, I need to assume that $1 - p$ is near zero to get such a truth-telling result. This truth-telling result requires, among other things, that each patient person among the fraction p who first encounter the planner has an incentive to reveal his or her type no matter what the person thinks others will do. If the person thinks all other patient people will withdraw, then the person must anticipate getting the lower state 1, period 2 consumption for sure rather than getting the higher state 2, period 2 consumption with probability $1 - q$ and the lower state 1 consumption with probability q. The assumption that $1 - p$ is small makes these two consumptions similar. (Notice that if $1 - p$ is small, then the two lines in Figure 2 are near each other.) A more precise description of the role of that assumption is given in the Appendix, where I also report the results of some numerical examples exploring how near to zero $1 - p$ must be. Total suspension plays the same role it played in the no-aggregate-risk version. As in that version, it is a threatened suspension that never really occurs in the sense that no one is ever actually turned away as a result of that suspension. The partial suspension, in contrast, occurs with probability q.

Interpretation as a Banking Arrangement

In a sense, adding uncertainty to the model economy does not change the interpretation of this planning arrangement as a banking system. Despite the partial suspension aspect and the riskiness of deposits—deposits do not pay one amount for sure if withdrawn during period 1 and another amount for sure if withdrawn during period 2—they are preferred at period 0 to not participating in the production process and getting the sure autarkic outcome. These risky deposits should be interpreted as deposits in a mutual, as opposed to a stock, banking organization. As in the no-aggregate-risk version, the attractive feature of these deposits is their relatively high return if held only until period 1. Also, as in that version, the banking-business sector must take on an illiquid portfolio.

Consistency Between the Model and U.S. Banking History

Since the model was motivated by U.S. banking suspensions in the second half of the 19th century and the beginning of the 20th century, it is appropriate to point out how the model is consistent with aspects of that period and with subsequent U.S. banking history.

The model is consistent with people being willing to hold deposits even though they know that a partial suspension may occur. It is also consistent with unhappiness, on average, after such a partial suspension actually occurs. In the model, those who withdraw prior to the partial suspension are unaffected; those who withdraw during period 1 after the partial suspension regret having shown up late (a matter of chance); and those who are patient get less in state 1 than they do in state 2, when a partial suspension does not occur.

However, such unhappiness after suspensions does not by itself imply that people automatically search for remedies—as in fact they did, following historical incidents of panics or suspensions.

To account for a search for remedies, and at the same time maintain that the model is correct in crucial respects, I must take the view that people did not recognize that the economy as a whole faced the same kind of illiquidity problem that individual banks faced. This is not unreasonable. Until recently, there was simply no available theory to suggest that the best outcome for the economy as a whole could be one in which the entire banking-business sector is in an illiquid position. The remedies adopted in the first part of this century—the Aldrich-Vreeland Act in 1908 and the Federal Reserve Act in 1914, which created the Federal Reserve System—were attempts to provide liquidity by centralizing the system. If the model is correct, those remedies would not work.

That, in turn, can help us explain the banking difficulties of the early 1930s. Without claiming that the banking difficulties of the early 1930s were the result of the kind of aggregate risk in the model, it is possible, by adopting a view of those banking difficulties much like that of Friedman-Schwartz (1963), to reconcile them with the model. Suppose people living in the world of the model, and still not having a clear view of the overall constraints, are lulled into the view that the Federal Reserve System with its so-called lender of last resort powers will deal with bank runs—both natural bank runs of the type in the model and panic bank runs in which everyone tries to withdraw during period 1 because they fear that others will. But when people learn that those powers are not sufficient to provide liquidity to the economy as a whole, and that they are not protected by timely, threatened suspensions, then bank runs of the panic type can occur.

The model can also be reconciled with the period of banking system stability that began with the creation of the deposit insurance system and extends, one might say, until our recent financial system difficulties. One way to get the aggregate-risk version of the model to display this stability—stability which makes deposits have sure payments—is to put regulation into it. Two kinds of regulations tend to produce such stability. One is interest ceilings which limit returns on early withdrawals. The other is the presence of bank capital, interpreted as a separate bank liability, which receives a return only at period 2 after all other claims are paid. It is certainly possible, in the presence of aggregate risk, to structure the payments on deposits and on bank capital so that all the risk is borne by the latter and, hence, by period 2 consumers. In the model I have described, however, it is not so easy to have that kind of bank liability structure be preferable to autarky. One reason this is so is that it gives some period 2 consumption to everyone in state 2, including those who turn out to be impatient. A less extreme model, one in which turning out to be impatient does not mean that one wants no period 2 consumption, would make it easier to have this asset structure be better than autarky while still implying that such an outcome is not the best possible outcome. The best outcome would continue to have the partial suspension feature, which means that some of the uncertainty implied by the aggregate risk is borne by consumers during period 1.

The model as it stands is not well-suited to describe the recent savings and loan association difficulties. One view of those difficulties is that they are the result of the deposit insurance instituted in the 1930s and the relaxation of regulation instituted in the 1980s. Together, these two aspects of the legal environment allowed savings and loans and banks to take risks with other people's money. (See, for example, Kareken 1981.) It would not be difficult to amend the model to be consistent with this view, but it would require significant changes that are not germane to showing the desirable role of partial suspensions.

Conclusion

A model of the sort I have described has fairly plausible features—individual uncertainty about the desired timing of withdrawals, uncertainty which does not completely average out over people, and sequential service—and is not inconsistent with some of the main features of U.S. banking history, including the overall illiquidity of the banking-business sector. I therefore take seriously the model's implication that a good banking system would have the partial suspension property.

Although the desirability of this property has been demonstrated only for a very special assumption regarding the order in which people attempt to withdraw early, something similar is likely to emerge from settings with less special specifications of aggregate risk. If the planner encounters people in a completely random order, then the planner will try to learn the aggregate state from the responses of everyone encountered. Finding the best arrangement for such a specification is complicated because the arrangement has to be directed both to making people encountered well off and to learning the aggregate state. If N is very large and there is no relationship between the type and the order in which the planner is encountered, then it is conceivable that the planner can learn the aggregate state from encounters with a small fraction of the N people. If so, then the planner would be in a position to adopt for the rest something like the arrangement that would be best if the planner knew the aggregate state at the beginning of period 1. That arrangement resembles partial suspension across possible outcomes of the aggregate state. When the planner decides that the number of impatient people is large, the first period consumption of the subsequently encountered impatient is less than when that number is small. In addition, all subsequently encountered depositors are worse off than when the number of impatient people is small. Here *partial suspension* does not mean that a depositor gets less because he or she shows up later than someone else. Rather he or she gets less simply because there are many others who are impatient. Because the resource constraint and the preferences dictate that good arrangements give the impatient more consumption in period 1 than they get under autarky, the occurrence of a large number of impatient people is an unfavorable outcome. That fact dictates that any good arrangement will display some aspects of a partial suspension.

Finally, it should be noted that the result that good arrangements display something like partial suspension comes from a model, essentially the Diamond-Dybvig model, which was not designed to explain partial suspensions. This model was designed primarily to show that a coherent economy could be one in which it is desirable for the banking-business sector to be illiquid. However, it turns out that the preference and technology assumptions which imply such illiquidity also imply that the best arrangement in a version with aggregate risk displays something resembling partial suspension. This new result matches some historical observations in that it explains why depositors would prefer a return pattern on deposits that includes the risk of partial suspension to all other possible return patterns. This new result from what is essentially the Diamond-Dybvig model therefore gives me additional confidence in the model and its implications.

Appendix
Proofs for the Best Aggregate-Risk Arrangement

Here I describe precisely the assumptions made about preferences and prove the assertions made about the best arrangement in the aggregate-risk section of the preceding paper. I also include some examples that allow us to judge how stringent is the assumption that $1 - p$ is near zero.

I first describe the outcome I get by maximizing period 0 expected utility subject only to the resource constraints and to sequential service. Then I show that this outcome is attainable by the planner even though type is private information. In particular, I show that with $1 - p$ near zero, each person has an incentive to reveal his or her true type no matter what the person thinks others will do.

The Best Outcome With Sequential Service

I assume that each person at period 0 judges outcomes according to

$$\text{(A1)}\quad U = p\Big[\alpha\textstyle\sum_{s=1}^{2} q_s u^1(c_1^{11}, c_2^{11}(s)) + (1-\alpha)\textstyle\sum_{s=1}^{2} q_s u^2(c_1^{21}, c_2^{21}(s))\Big] + (1-p)\Big[q_1 u^1(c_1^{12}(1), c_2^{12}(1)) + q_2 u^2(c_1^{22}(2), c_2^{22}(2))\Big]$$

where q_1 is the probability that the aggregate state is state 1, the q defined in the text, and q_2 is the probability that it is state 2, $1 - q$ in the text. The functions u^i are assumed to satisfy the following conditions: $u^i(x,y) = g(x + \gamma_i y)$, where the function g has a positive first derivative and a negative second derivative and satisfies $g'(z) > rg'(rz)$ for all z and all $r > 1$, g' being the derivative of g and being such that g' goes to infinity as its argument goes to zero. It is also assumed that $\gamma_1 = 0$ and $\gamma_2 > 1/R_2$. These are Diamond and Dybvig's (1983) preference assumptions.

This formulation builds in sequential service by making the period 1 consumption of the proportion p who first meet the planner independent of the aggregate state s. It also supposes that the only dependence on when the planner is encountered occurs between the proportion p who first encounter the planner and the proportion $1 - p$ who encounter the planner last. Any other dependence can be shown to be undesirable.

I want to describe the consumption allocation that maximizes U, subject to the following resource constraints:

$$\text{(A2)}\quad C_2(s) = R_2[NR_1y - C_1(s)]$$

$$\text{(A3)}\quad C_1(s)/N = p[\alpha c_1^{11} + (1-\alpha)c_1^{21}] + (1-p)\textstyle\sum_{h=1}^{2} \delta_{sh} c_1^{h2}(h)$$

$$\text{(A4)}\quad C_2(s)/N = p[\alpha c_2^{11}(s) + (1-\alpha)c_2^{21}(s)] + (1-p)\textstyle\sum_{h=1}^{2} \delta_{sh} c_2^{h2}(h)$$

for $s = 1$ and 2 and where $\delta_{sh} = 1$ if $s = h$ and 0 otherwise.

The solution to this problem is unique, and some components of the solution are easily shown to be zero. First, $c_2^{11}(s) = 0$ for $s = 1, 2$ and $c_2^{12}(1) = 0$. These follow because second period consumption is not valued at all by those who turn out to be impatient. Second, $c_1^{21} = c_1^{22}(2) = 0$. This follows because the intertemporal technology can be used to shift consumption from period 1 to period 2 in a way that is preferred by those who turn out to be patient. It is also evident from the concavity of g that the solution satisfies $c_2^{21}(2) = c_2^{22}(2)$, the common magnitude of which is hereafter denoted $c_2^2(2)$.

In terms of the four nonzero unknowns—c_1^{11}, $c_1^{12}(1)$, $c_2^{21}(1)$, and $c_2^2(2)$—U can be written as

$$\text{(A5)}\quad U = p\Big\{\alpha g(c_1^{11}) + (1-\alpha)\Big[q_1 g(\gamma_2 c_2^{21}(1)) + q_2 g(\gamma_2 c_2^2(2))\Big]\Big\} + (1-p)\Big[q_1 g(c_1^{12}(1)) + q_2 g(\gamma_2 c_2^2(2))\Big]$$

and the resource constraints as

$$\text{(A6)}\quad R_2[p\alpha c_1^{11} + (1-p)c_1^{12}(1)] + p(1-\alpha)c_2^{21}(1) + R_2R_1y$$

$$\text{(A7)}\quad R_2[p\alpha c_1^{11}] + [p(1-\alpha) + (1-p)]c_2^2(2) + R_2R_1y$$

for $s = 1$ and $s = 2$, respectively. The assumptions about g imply that the allocation that maximizes U subject to (A6) and (A7) is given by the solution to (A6) and (A7) and the following first-order conditions:

$$\text{(A8)} \quad c_1^{11}: \quad g'(c_1^{11}) = (\lambda_1 + \lambda_2)R_2$$

$$\text{(A9)} \quad c_1^{12}(1): \quad q_1 g'(c_1^{12}(1)) = \lambda_1 R_2$$

$$\text{(A10)} \quad c_2^{21}(1): \quad q_1 \gamma_2 g'(\gamma_2 c_2^{21}(1)) = \lambda_1$$

$$\text{(A11)} \quad c_2^{2}(2): \quad q_2 \gamma_2 g'(\gamma_2 c_2^{2}(2)) = \lambda_2$$

where λ_1 is the LaGrange multiplier associated with (A6) and λ_2 is that for (A7). I now show that the solution to (A6)–(A11) satisfies the qualitative claims displayed in Figure 2—namely, $c_1^{12}(1) < c_1^{11} > R_1 y$, the first inequality being partial suspension and the second being banking system illiquidity.

Eliminating the λ's in (A8) using (A10) and (A11), we get

$$\text{(A12)} \quad g'(c_1^{11}) = R_2\gamma_2\left[q_1 g'(\gamma_2 c_2^{21}(1)) + q_2 g'(\gamma_2 c_2^{2}(2))\right].$$

Substituting λ_1 from (A10) into (A9), we get

$$\text{(A13)} \quad g'(c_1^{12}(1)) = R_2\gamma_2 g'(\gamma_2 c_2^{21}(1)).$$

Subtracting (A13) from (A12) and using $q_2 = 1 - q_1$, we get

$$\text{(A14)} \quad g'(c_1^{11}) - g'(c_1^{12}(1)) = R_2\gamma_2 q_2\left[g'(\gamma_2 c_2^{2}(2)) - g'(\gamma_2 c_2^{21}(1))\right].$$

It follows from (A14) and $g'' < 0$ that

$$\text{(A15)} \quad \text{Sign}[c_1^{11} - c_1^{12}(1)] = \text{Sign}[c_2^{2}(2) - c_2^{21}(1)].$$

Next subtract (A7) from (A6) to get

$$\text{(A16)} \quad p(1-\alpha)[c_2^{21}(1) - c_2^{2}(2)] + (1-p)[R_2 c_1^{12}(1) - c_2^{2}(2)] = 0.$$

Now suppose by way of contradiction that the left side of (A15) is not positive. Then, by (A15) and (A16),

$$\text{(A17)} \quad R_2 c_1^{12}(1) \le c_2^{2}(2) \le c_2^{21}(1).$$

By (A17) and $g'' < 0$,

$$\text{(A18)} \quad g'(\gamma_2 c_2^{21}(1)) \le g'(R_2\gamma_2 c_1^{12}(1))$$

or

$$\text{(A19)} \quad R_2\gamma_2 g'(\gamma_2 c_2^{21}(1)) \le R_2\gamma_2 g'(R_2\gamma_2 c_1^{12}(1)).$$

Then by (A13),

$$\text{(A20)} \quad g'(c_1^{12}(1)) \le R_2\gamma_2 g'(R_2\gamma_2 c_1^{12}(1)).$$

Since $R_2\gamma_2 > 1$, this contradicts $g'(z) > rg'(rz)$ for all z and all $r > 1$.

Suppose next, by way of contradiction, that $c_1^{11} \le R_1 y$. Given the partial suspension property just established, (A6) and (A7) imply respectively that

$$\text{(A21)} \quad R_2[p\alpha R_1 y + (1-p)R_1 y] + p(1-\alpha)c_2^{21}(1) > R_2 R_1 y$$

$$\text{(A22)} \quad R_2 p\alpha R_1 y + [p(1-\alpha) + (1-p)]c_2^{2}(2) \ge R_2 R_1 y.$$

These, in turn, imply that $c_2^{21}(1) > R_2 R_1 y$ and $c_2^{2}(2) \ge R_2 R_1 y$. These inequalities and our presumed contradiction imply that $g'(\gamma_2 c_2^{21}(1)) < g'(\gamma_2 R_2 R_1 y)$, $g'(\gamma_2 c_2^{2}(2)) \le g'(\gamma_2 R_2 R_1 y)$, and $g'(c_1^{11}) \ge g'(R_1 y)$. These inequalities and (A12) imply that $g'(R_1 y) < \gamma_2 R_2 g'(\gamma_2 R_2 R_1 y)$, which again contradicts $g'(z) > rg'(rz)$ for $r > 1$.

Achieving the Best Outcome With a Suspension Scheme

I now argue that if $1 - p$ is near zero, then the suspension scheme described in the text gives each person an incentive to truthfully reveal his or her type no matter what the person thinks others will do. In Wallace 1988, this was proven for $p = 1$, a no-aggregate-risk version of the current model. That $p = 1$ conclusion and continuity of the above solution in the parameter p yields the result.

To see this, first note that impatient people always truthfully reveal their type because they do not value period 2 consumption at all. As for patient people, no matter when they confront the planner the suspension scheme assures that those who truthfully reveal their type will get at least $\tilde{c}_2^{21}(1)$ at period 2, where I now use a tilde to denote the above solution. Since $\tilde{c}_1^{11} > \tilde{c}_1^{12}(1)$, a sufficient condition for all patient people to be truthful is $\gamma_2\tilde{c}_2^{21}(1) \ge \tilde{c}_1^{11}$. From Figure 2 it is clear that $\tilde{c}_2^{21}(1)$ approaches $\tilde{c}_2^{2}(2)$ as $1 - p$ approaches 0. Therefore, by continuity of the solution in p, as $1 - p$ approaches 0, $\tilde{c}_2^{21}(1)$ approaches second period consumption of the patient in the no-aggregate-risk version—call it $\tilde{c}_2^{2}$—and $\tilde{c}_1^{11}$ approaches period 1 consumption of the impatient in the no-aggregate-risk version—call it $\tilde{c}_1^{1}$. Since, as was shown in Wallace 1988, $\gamma_2\tilde{c}_2^{2} > \tilde{c}_1^{1}$, I am assured that for $1 - p$ sufficiently close to zero, $\gamma_2\tilde{c}_2^{21}(1) \ge c_1^{11}$.

To buttress this continuity argument, numerical solutions for a class of examples were calculated. The examples have the following specification: $y = 1$, $R_1 = 1$, $g(z) = -z^{-0.5}$, $\gamma_2 = 1$, $\alpha = 0.5$, $q_1 = 0.1$, and all combinations of $1 - p$ from the set {0, 0.05, 0.10, 0.15, 0.20, 0.25} and of R_2 from the set {1.01, 1.04, 1.07, 1.10}. In every case, the inequality $\gamma_2\tilde{c}_2^{21}(1) \geq \tilde{c}_1^{11}$ was satisfied. The results for $R_2 = 1.10$ are shown in the accompanying table.

Example Solutions When the Marginal Return Between Periods 1 and 2 (R_2) Is 1.10

The Last Fraction to Meet the Planner	Consumption			
$1 - p$	$\tilde{c}_1^{11}$	$\tilde{c}_1^{12}(1)$	$\tilde{c}_2^{21}(1)$	$\tilde{c}_2^{2}(2)$
.00	1.0159	1.0159	1.0825	1.0825
.05	1.0165	1.0137	1.0802	1.0835
.10	1.0172	1.0118	1.0782	1.0845
.15	1.0178	1.0102	1.0764	1.0855
.20	1.0185	1.0087	1.0749	1.0865
.25	1.0191	1.0074	1.0735	1.0874

References

Diamond, Douglas W., and Dybvig, Philip H. 1983. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy* 91 (June): 401–19.

Friedman, Milton, and Schwartz, Anna Jacobson. 1963. *A monetary history of the United States, 1867–1960.* Princeton, N.J.: Princeton University Press (for National Bureau of Economic Research).

Kareken, John H. 1981. Deregulating commercial banks: The watchword should be caution. *Federal Reserve Bank of Minneapolis Quarterly Review* 5 (Spring–Summer): 1–5.

Wallace, Neil. 1988. Another attempt to explain an illiquid banking system: The Diamond and Dybvig model with sequential service taken seriously. *Federal Reserve Bank of Minneapolis Quarterly Review* 12 (Fall): 3–16.