Economic Quarterly—Volume 96, Number 1—First Quarter 2010—Pages 33–58

On the Fundamental Reasons for Bank Fragility

Huberto M. Ennis and Todd Keister

Over the course of the recent financial crisis, several large financial institutions experienced sudden, massive withdrawals of their usual funding sources. In the U.K., for example, depositors lost confidence in the bank Northern Rock and started a *run* of withdrawals that ended with the bank being taken into state ownership. In the United States, the investment bank Bear Stearns and the commercial bank Wachovia both experienced a rapid loss of funding and were taken over by other institutions to avoid their outright failure. This same phenomenon affected other types of institutions as well, including a large part of the money market mutual fund industry, which experienced heavy withdrawals following the failure of the Reserve Fund in September 2008.

These episodes are only the most recent examples of a phenomenon that has been a recurrent theme in the history of banking. Banking panics, with massive withdrawals often leading to widespread bank failures, were a regular occurrence in the United States prior to the advent of government-sponsored deposit insurance in 1933. Developing economies have also experienced runs on their banking system, including episodes in Ecuador (1999), Argentina (2001), and Russia (2004).

Observers of these episodes often claim that there is an important self-fulfilling component to the behavior of depositors and/or investors. In this view, each depositor fears that the withdrawals of *other* depositors will cause the bank to fail and rushes to withdraw her funds before this failure occurs. Collectively, these actions validate the original belief that a wave of withdrawals will cause the bank to fail. During the height of the Panic of 1907 in the United States, J.P. Morgan was reported in the *New York Times* to have

We would like to thank Borys Grochulski, Ned Prescott, and Juan Sánchez for comments on a previous draft. The views expressed here do not necessarily represent those of the Federal Reserve Bank of New York, the Federal Reserve Bank of Richmond, or the Federal Reserve System. E-mails: huberto.ennis@rich.frb.org; todd.keister@ny.frb.org.

said, "If the people would only leave their money in the banks instead of withdrawing it...everything would work out all right."[1] In other words, Morgan claimed that it was the behavior of the depositors themselves that was placing the largest strain on the banking system. If this strain were removed, individuals would be willing to leave their money deposited and a superior outcome would obtain.

This view of events implies that banks and other financial intermediaries are inherently *fragile*, in the sense of being susceptible to a self-fulfilling run by their depositors. The degree to which one accepts this view has strong implications for public policy. The desirability of government-provided deposit insurance, for example, and of other public interventions in the banking system depends in large part on whether banking crises do indeed have an important self-fulfilling component or whether they instead result from other, more fundamental causes.

A substantial economic literature has developed that attempts to identify the essential components that would justify a self-fulfilling interpretation of events. Bryant (1980) and Diamond and Dybvig (1983) provided the first steps in the development of a coherent theory along these lines. Subsequently, various authors have tried to understand if the set of elements included in these early contributions is sufficient to explain banking and the fragility of banks and other financial intermediaries, and which other elements, if any, may be missing.

The approach taken in this literature has been to specify a complete physical environment and to study economic outcomes that agents in such an environment could achieve without imposing any artificial restrictions on their ability to enter mutually beneficial arrangements. In following this approach, the literature has become fairly technical and intricate. In this article, we aim to provide an informal discussion of the issues and the results produced so far in this literature. We hope that our endeavor will make the lessons obtained from this body of work more readily accessible to readers who may be less inclined to endure over the many technical issues involved in the subject.

We begin our discussion by reviewing the key theoretical contribution of the seminal work by Diamond and Dybvig (1983). We discuss the basic elements of their banking theory and how subsequent researchers have addressed the technical difficulties involved in designing an equilibrium concept that allows for the possibility of a bank run. As will become clear in the discussion, one essential element of the theory is the existence of a first-come, first-served (or *sequential service*) constraint. In Section 2, we discuss how the literature has handled the specification of an explicit sequential service constraint. Several important recent contributions in this literature have resulted from the

[1] *New York Times*, October 26, 1907, "Bankers Calm; Sky Clearing."

efforts to combine explicitly modeled sequential service with the presence of aggregate uncertainty about the fundamental need for liquidity in the system. We review those contributions and how they relate to each other in detail. In Section 3 we discuss some potentially fruitful directions for further research and, finally, we close the article with some brief concluding remarks.

1. THE DIAMOND-DYBVIG MODEL

This section presents an overview of the seminal contribution by Diamond and Dybvig (1983) and sets the stage for the discussion of the more recent literature that explores the fundamental reasons for bank fragility. In Diamond and Dybvig's theory, banks play an essential role in the process of *maturity transformation*: they issue short-term (deposit) liabilities in order to finance long-term productive investment. While maturity transformation may happen through other channels in the economy, Diamond and Dybvig identify two other essential features of banking arrangements: the fact that agents' demands must be dealt with on a *first-come, first-served* basis, and the fact that agents' true liquidity needs remain *private information*. These three elements constitute the foundations of Diamond and Dybvig's theory of banking and are also the source for the potential of bank fragility in their model.

The Physical Environment

Diamond and Dybvig (1983) consider an environment where a large number of agents face idiosyncratic uncertainty about their intertemporal desire to consume. Agents have an initial endowment of goods and there is a technology that can be used to transform these goods into (potentially more) goods in the future. If investment is left in place long enough to mature, the net returns are positive. However, some agents will discover that they are *impatient* and need to consume before the investment matures. Other agents are *patient* and able to consume after investment has matured.

Investment takes place before agents discover their intertemporal preference for consumption. To the extent that the idiosyncratic desire to consume early is not perfectly correlated among agents, there are insurance possibilities to be exploited in this environment. In particular, there exists a clear social benefit from pooling resources ex ante, before preferences are realized, investing in the long-term technology, and then making payments ex post to agents, contingent on their needs.

Diamond and Dybvig (1983) assume that an agent's realized preference type (patient or impatient) is private information. Any attempt to provide consumption to agents in a way that depends on their intertemporal preference for consumption must, therefore, rely on reports from agents. This fact could complicate matters in two ways. First, the ex-post payments to agents must be arranged in such a way as to create the right incentives for each individual agent to not misrepresent her consumption needs. Second, private information opens the door to the possibility of a coordinated misrepresentation by agents, which may be interpreted as a *run* to withdraw from the pool. The insurance possibilities associated with a pooling arrangement depend crucially on its ability to avoid these two types of misrepresentation.

In principle, it would be beneficial to collect as much information as possible about the total demand for withdrawals before making any payments from the resource pool. However, Diamond and Dybvig (1983) assume that agents who decide to withdraw early place their demands sequentially, and that payments from the pool must be made at the time each demand is placed. In other words, payments ought to respect a first-come, first-served rule, which they call a *sequential service constraint*. Diamond and Dybvig argue that this kind of restriction is a realistic description of how banks operate.[2]

Resource Allocation and Optimality

Diamond and Dybvig's simple environment provides a natural setup to think about the institution of banking. In the model, agents initially deposit their endowments in a pool, which can be interpreted as a "bank." In exchange for her deposit, an agent receives a claim to future consumption from this bank. After deposits are made, the bank invests in the long-term technology. Finally, agents discover their consumption needs and contact the bank sequentially to withdraw resources and consume. The bank makes payments to agents, on demand, in a pre-arranged manner.

From a theoretical point of view, it is appealing to abstract from institutional details and focus instead on allocations of consumption that are achievable while respecting the constraints imposed by the physical environment and the structure of information. Much of what is done in Diamond and Dybvig's (1983) article is consistent with this strategy. Following the basic principles in the theory of mechanism design, the way to proceed is to set up a planning problem that consists of choosing a (contingent) consumption allocation to maximize the ex ante expected utility of agents subject to incentive compatibility, sequential service, and resource feasibility constraints.[3] We will call this allocation the *constrained-efficient allocation*.

[2] An important component of a formal sequential service constraint is the specification of whether or not agents who decide to not withdraw early still contact the pool at that time. Diamond and Dybvig (1983) implicitly assume that only agents who are attempting to withdraw contact the pool. We return to this issue later in this article.

[3] Going back to the interpretation of the theoretical constructions in terms of the institutions of banking, it can be demonstrated that under certain conditions the solution to this planning problem is equivalent to the outcome that would obtain when profit-maximizing banks compete for deposits.

To understand the implications of agents possibly misrepresenting their consumption needs, it is useful to solve the same planning problem, but without imposing the incentive constraints. We will call the solution to this modified problem the *unconstrained-efficient allocation*.[4]

In general, the incentive compatibility constraint for an individual agent in this environment depends on the assumed behavior of the rest of the agents. While it may be incentive compatible for an agent to not misrepresent her consumption needs when all the other agents are also not misrepresenting, the situation may be different when the other agents are expected to misrepresent. This payoff complementarity is important because it creates the potential for strategic coordinated responses that may result in substantial inefficiencies.

The strategic interaction among agents takes place in the *withdrawal game* induced by a given contingent consumption allocation, i.e., a complete payment scheme. In the withdrawal game, agents decide when to contact the resource pool (the bank) to demand payment. An allocation is *implementable* (under truthful representation) if there is a Nash equilibrium of the induced withdrawal game in which all impatient agents withdraw early and all patient agents wait until the investment matures. An implementable allocation is often also called *incentive feasible*, in the sense that it satisfies the incentive compatibility constraint for each individual agent given that all the other agents are not misrepresenting their consumption needs. If the equilibrium of the withdrawal game is unique, we say that the allocation is *strongly* (or *fully*) *implementable*. As we will see, implementable allocations in the Diamond-Dybvig model are sometimes not strongly implementable. In those cases, there exists another Nash equilibrium of the withdrawal game in which some patient agents misrepresent their need to consume and attempt to withdraw early, in effect running to obtain payment from the pool before its resources are exhausted.

Diamond and Dybvig (1983) make some additional simplifying assumptions that turn out to have significant implications for their results. In particular, they assume that there is a continuum of agents in the economy and that preference types (patient or impatient) are independent and identically distributed (i.i.d.) across agents. The combination of these two assumptions and the law of large numbers implies that the total need for early consumption is completely predictable. In other words, if the bank believes that only impatient agents will withdraw before investment matures, then it knows the total demand for liquidity even before agents begin placing their requests.

---

[4] The unconstrained-efficient allocation is the best allocation that can be attained when preferences of agents are observable. Since we consider the sequential service constraint a reflection of a feature of the physical environment (Wallace 1988), the unconstrained-efficient allocation must satisfy sequential service in the same way that it must satisfy resource feasibility.

Diamond and Dybvig (1983) show that the unconstrained-efficient allocation is actually implementable in their environment.[5] Hence, the constrained-efficient allocation is equal to the unconstrained-efficient allocation, and the fact that agents' preferences are private information imposes no restrictions in terms of what is implementable in this environment (i.e., the incentive constraints in the planning problem are not binding at the solution). Furthermore, the fact that agents withdraw from the resource pool sequentially has no implications for the choice of the constrained-efficient allocation. In other words, the sequential service constraint is also nonbinding at the solution to the planning problem.

Under certain conditions on the relative risk aversion of agents, Diamond and Dybvig also show that in the unconstrained-efficient allocation, agents withdrawing early receive more than what they initially deposit at the bank. In other words, the best allocation provides some degree of insurance against the contingency that the agent becomes impatient and cannot wait for the investment to mature. This finding is important to understand the fundamental reasons for the possibility of bank fragility in the model.

Deposit Contracts and the Possibility of Runs

Interestingly, there are many possible payment schemes that can be used to implement the unconstrained-efficient allocation. One such scheme specifies that each agent, after depositing her resources in the pool, is entitled to a fixed payment if she withdraws early and a different fixed payment if she withdraws late. This arrangement resembles a simple demand deposit contract, commonly used in practice, in which agents experience a penalty for withdrawing early but their payment is otherwise not contingent on information that the bank might receive as (sequential) withdrawals occur. We call this scheme the *optimal simple demand deposit contract*.

This demand deposit contract must respect important restrictions imposed by the physical description of the environment. First, it must obviously conform with resource feasibility. This unavoidable constraint implies that payments will be fixed only as long as the bank does not run out of resources. Second, the contract must be consistent with the assumption that agents have private information about their own preferences. In particular, payments cannot be made contingent on the true preference of agents, since these are unknown to the bank.

When only impatient agents withdraw early, the bank does not run out of resources when following this contract and the payments generate the

[5] Making early payments is costly for the pool since it removes resources from investment before it has had time to mature. For this reason, it is always optimal to give agents who are withdrawing late at least as much utility as those withdrawing early and, hence, the unconstrained-efficient allocation always satisfies the incentive constraints.

unconstrained-efficient allocation. However, the demand deposit contract does not strongly implement the unconstrained-efficient allocation. There is another equilibrium of this withdrawal game in which all agents attempt to withdraw early and the bank runs out of resources before paying some agents.[6] This equilibrium resembles a self-fulfilling run on the bank.

What is the logic behind this run equilibrium? Agents have, sequentially, an opportunity to withdraw early from the bank. Those agents who become impatient have no real decision to make: they place a demand to withdraw when their turn comes. Patient agents, on the other hand, need to decide whether to try to withdraw early or wait until investment matures. If all patient agents expect that all other patient agents will try to withdraw early, then they also expect that the bank will run out of resources before all agents have been paid. Waiting until investment matures in such circumstances is pointless, since the bank's resources will be depleted before then. Hence, all patient agents attempt to withdraw early, fulfilling their beliefs and making this outcome consistent with equilibrium.

The possibility of this type of self-fulfilling run is a direct consequence of the presence of the sequential service constraint. Without sequential service, the bank could wait until all agents have placed their withdrawal requests before making any payments. Since the bank knows the number of impatient agents in the population, once requests pass this threshold the bank would be able to clearly identify that a run is taking place. Importantly, the bank would then know about the run before making any payments to agents. It is not hard to see that, once a run has been identified, the payment scheme in the simple demand deposit contract is no longer optimal. Because agents are risk averse and everyone is attempting to withdraw, the best way to allocate existing resources is to distribute them evenly among agents. In this case, however, patient agents would actually prefer not to participate in the run. By waiting and leaving their funds in the bank, patient agents will be able to receive a higher payment after the investment matures. In summary, the lack of sequential service would be sufficient to rule out runs as possible equilibrium phenomena.

Another critical assumption in the run situation described above is that only those agents who intend to withdraw are expected to contact the bank. If this were not the case, then it would be easy for the bank to realize that a run is taking place before any significant portion of agents have attempted to withdraw. In general, when no run is taking place, the bank would expect withdrawal demands to be scattered among nonwithdrawal demands. If every agent contacting the bank places a demand for withdrawal, the bank can quickly infer that a run is taking place. In the case of the continuum of

[6] This is a consequence of the provision of insurance in the unconstrained-efficient (which, in this case, is also equal to the constrained-efficient) allocation.

agents, this logic is extreme and the run could be identified before any significant payments have been made.[7] In a sense, with a continuum of agents the sequential service constraint is only relevant when not all agents contact the bank in the early period. When there is a finite number of agents, however, things are different. As we will see in the next section, the sequential service constraint can be meaningfully specified in either way in this case, with differing implications for bank fragility.

An unsettling characteristic of the run situation under the optimal simple demand deposit contract is that once the number of withdrawals surpasses the number of impatient agents in the population, which is nonstochastic, the bank is certain that a run is underway. This information could potentially be used to design a more robust payment scheme. In fact, there is a payment scheme that strongly implements the unconstrained-efficient allocation by modifying only payments that will then lie off the equilibrium path of play and, hence, never be made. This payment scheme involves a suspension of convertibility clause, which says that after a certain number of withdrawals the bank will suspend payments and wait until investment has matured. If this suspension is designed to take place only after the number of withdrawals is larger than the number of impatient agents in the population, but not too much after that, then it will never occur in equilibrium; the expectation that it would occur if needed is sufficient to rule out a possible run on the bank.

In summary, even when payments are required to be made sequentially, there is a scheme involving an (off-equilibrium) suspension of convertibility that rules out runs and strongly implements the unconstrained-efficient allocation. In that sense, the presence of sequential service does not change the configuration of equilibrium outcomes in the benchmark version of the Diamond-Dybvig model.

One crucial feature that allows the suspension of payments to work so effectively, without any cost, is the absence of aggregate uncertainty about the total number of impatient agents in the economy. In other words, the model described above has no uncertainty about the total *fundamental* need for early liquidity. If the bank were unsure about the true aggregate need for early liquidity, it would be much more difficult to choose the right time to suspend payments. Suspending too soon may leave some impatient agents without precious resources at the time that they truly need to consume. Suspending too late may leave resources sufficiently depleted to make the run consistent with equilibrium. Diamond and Dybvig (1983) recognize this important limitation in their analysis and give some preliminary steps in the direction of relaxing the assumption of no aggregate uncertainty.

---

[7] De Nicoló (1996) exploits this idea to design a contract that strongly implements an allocation arbitrarily close to the constrained-efficient allocation.

The presence of aggregate uncertainty, together with sequential service, significantly complicates the analysis. Not only is solving and characterizing the unconstrained-efficient allocation a much more complex problem, but studying the strategic interaction among agents also involves more sophisticated techniques and logic. Diamond and Dybvig (1983) only hint at these issues in their seminal analysis. They abstract from incentive compatibility and sequential service constraints to solve for a benchmark allocation under aggregate uncertainty.[8] They then demonstrate that this benchmark allocation is not implementable under private information and sequential service. Whether the unconstrained-efficient allocation (which takes into account sequential service) could be implemented and/or strongly implemented in the presence of aggregate uncertainty was left as an open question in the literature for a long time. Only 20 years later was the first detailed analysis of this question in the Diamond-Dybvig framework provided by Green and Lin (2003). We discuss their contribution in Section 2.

Runs and the Equilibrium Concept

Before we conclude our discussion of Diamond and Dybvig's (1983) initial contribution, it is worth mentioning some important issues related to the formal treatment of bank fragility that originated in their work.[9] It is easy to see that in the Diamond-Dybvig model a bank run can happen only if the agents and the bank are not certain ex ante that one will occur. If the bank is certain that a run will happen, it will make payments to agents without providing insurance, which makes the run strategy of agents inconsistent with equilibrium. If the bank believes that a run will not occur, but the agents are certain that one will, then agents will not choose to deposit their resources at the bank. Hence, runs can occur in equilibrium only if they are expected to happen with some probability strictly less than unity. Formally, this kind of uncertainty can be captured by introducing an *extrinsic random variable* in the model, which allows agents to condition their behavior on the realization of such a variable. This modeling strategy was suggested by Diamond and Dybvig (1983) and subsequently formalized by Cooper and Ross (1998) (see also Peck and Shell [2003]).[10]

---

[8] Note that without aggregate uncertainty, this strategy delivers the unconstrained-efficient allocation. With aggregate uncertainty, however, this is no longer the case.

[9] Postlewaite and Vives (1987) propose a related model that does not rely on multiplicity of equilibria as an explanation for bank runs. In their model, there is aggregate uncertainty about agents' preferences over intertemporal consumption and, in some cases, agents strategically rush to withdraw their funds before they have a true need to consume. Postlewaite and Vives do not have a sequential service constraint in their analysis.

[10] Gu (forthcoming) studies the case when different groups of agents observe the realization of different extrinsic random variables. She constructs run equilibria in which only a subgroup of the patient agents chooses to misrepresent preferences and withdraw.

As discussed above, suspension of convertibility rules out runs altogether in the standard Diamond-Dybvig framework. For this reason, Cooper and Ross (1998) restrict the possible set of banking contracts to those that take the form of a demand deposit contract without a suspension clause. Given this restriction, they show that the optimal demand deposit contract is consistent with the possibility of runs if the probability of a run is small enough. The extrinsic random variable in their model acts as a coordinating device. Agents observe the realization of the random variable and, for some realizations, play the run action. Interestingly, this modeling procedure works only if the bank does not observe the realization of the random variable. In this way, the bank remains uncertain about the motivation of the initial group of agents who attempt to withdraw: they may need to consume or they may be part of a run. If the probability that the bank assigns to experiencing a run is small enough, it will make fairly generous payments to early withdrawers, compromising the availability of resources for payment to those who wait. It is the anticipation of this situation by patient agents that, in turn, makes the run strategy consistent with equilibrium.[11]

While the findings of Cooper and Ross (1998) are quite interesting, their restriction to demand deposit contracts without a suspension clause is unsatisfactory when trying to identify the fundamental reasons for bank fragility. In Cooper-Ross' model, as in Diamond-Dybvig's, the fully unrestricted optimal banking contract rules out runs. Even if one does not go so far as to rule out runs completely, it is easy to see how their demand deposit contract without suspension would clearly be suboptimal and, hence, unlikely to materialize. At some point in the withdrawal process, the bank should be expected to realize that a run is taking place. In this (predictable) contingency, the simple demand deposit contract is easily seen to be suboptimal. Reducing the amount of resources paid to early withdrawers after that point would allow the bank to spread consumption more evenly among the remaining withdrawers, which would clearly improve the allocation (compared to keeping the payment constant and then running out of resources before some agents have been paid). As it turns out, this type of "partial suspension" (Wallace 1988, 1990) is also a feature of the optimal banking contract when there is uncertainty about the aggregate need for early liquidity in the economy, as we discuss in the next section.

In summary, Diamond and Dybvig (1983) identify three basic elements of a plausible theory of banking and bank fragility: (1) maturity transformation; (2) private information; and (3) sequential service. Uncertainty about the agents' total need for early liquidity could also be an important ingredient of a successful theory. Studying an explicit model of banking that incorporates

[11] Ennis and Keister (2006) clarify some aspects of the analysis in Cooper and Ross (1998) and derive additional results in their framework.

these components has proved to be a challenging task. Only recently has there been significant progress in understanding the implications for banking and bank fragility of combining all four components. We will review this research next.

2. TAKING SEQUENTIAL SERVICE SERIOUSLY

In this section, we discuss a series of papers that study versions of the Diamond-Dybvig model and in which special attention is devoted to the explicit specification of the sequential service constraint. We highlight (i) the interaction of aggregate uncertainty with the details of the environment that motivate the sequential service constraint and (ii) the implications of these assumptions for the possibility of bank fragility.

The Wallace Critique

In an influential article, Jacklin (1987) clarifies the role of trading restrictions in the Diamond-Dybvig model. He demonstrates that if agents are allowed to interact in a market after they discover the timing of their consumptions needs, there is an alternative arrangement that implements the unconstrained-efficient allocation without any possibility of runs. In this mechanism, agents initially buy shares in a firm that invests in the long-term technology. After discovering their consumption needs, impatient agents trade their shares with patient agents in exchange for consumption. Jacklin (1987) shows that this arrangement is capable of delivering the unconstrained-efficient allocation in the Diamond and Dybvig model, leaving no essential role for the institution of banking.

The market arrangement in Jacklin (1987), however, requires that the sequential service constraint be considered a restriction on the banking mechanism rather than a feature of the environment. The basic logic that allows the market arrangement to work requires that agents wait until all of them have discovered their consumption needs before they trade and consume. Under such a specification, however, it is not clear why a bank should be subject to sequential service. In principle, the bank could also wait before making any payments. In a way, assuming that banks make payments sequentially, as they do in real life, seems ad hoc in Jacklin's version of the Diamond-Dybvig model.

Wallace (1988) argues that the sequential service constraint should be considered a direct consequence of some frictions in the environment. If this were not the case, Jacklin's results imply that we should expect to see maturity transformation taking place solely in market-based arrangements and not in banks. Wallace interprets the fact that banks do perform a significant amount of maturity transformation as clear evidence of fundamental frictions that prevent markets from playing this role. He describes an environment in which agents are isolated from each other when the early consumption opportunities arise and cannot meet to trade in a market. Agents are, however, able to contact the bank and they do so sequentially. Wallace assumes that all agents contact the bank before investment matures: some agents make an early withdrawal and others inform the bank that they will not withdraw until after investment has matured.[12]

These assumptions could be regarded, a priori, as fairly restrictive. The key to understanding their role is to realize that without these (or similar) assumptions, the Diamond-Dybvig model is unable to explain banking, illiquidity, or excess fragility. In a sense, these assumptions are necessary to have a successful theory of banking in the Diamond-Dybvig tradition. With this stipulation in mind, we can consider the isolation assumption a reasonable approach to capture, in a stylized manner, the fact that agents often have limited access to financial and asset markets when consumption opportunities arise. Banks, then, help agents overcome this kind of financial friction by providing a more reliable source of on-demand liquidity.

Wallace (1988) also emphasizes that once the sequential service constraint is considered a feature of the environment, it implies that payments to agents cannot be recalled at a later time. One can imagine that when a payment is made, the agent consumes these resources immediately. This approach implies that the type of deposit insurance scheme discussed by Diamond and Dybvig (1983) is infeasible in an environment with sequential service. Diamond and Dybvig assume that the government can tax agents after the opportunities to withdraw from the bank have passed. Wallace argues that if such taxation is possible, then agents must not need immediate access to their funds and the bank could wait until it has received all of the withdrawal requests before making any payments. If the sequential service constraint is truly a feature of the environment, it must apply to the government as well as to private institutions.

As we mentioned before, solving for the constrained-efficient allocation in the presence of an explicit sequential service constraint and aggregate uncertainty is a complicated matter. Wallace (1988, 1990) identifies some relevant features of such a solution. The basic insight is that each payment can only be contingent on information revealed up to the point when this payment is made. While the probability distribution over the possible values of the aggregate need for (early) liquidity is known a priori, the actual realization must be inferred from the withdrawal demands of agents. In other words,

[12] In the Diamond-Dybvig tradition, the order in which agents get an opportunity to withdraw is assumed to be exogenously given (generally determined by a random draw). In other words, agents in the model are not allowed to take explicit actions to change their order of arrival. This assumption is, of course, extreme and, unfortunately, not much is known so far about the case where it is not made.

the allocation must reflect the gradual process of information revelation that results from an explicit sequential service constraint.

Wallace (1988) shows that the constrained-efficient allocation under aggregate uncertainty must have early payments that depend on the order in which they occur. As more agents place withdrawal demands, the probability that the final number of impatient agents is large increases and the size of the payment to early withdrawers tends to decrease. This adjustment in the size of payments is the upshot from the fact that higher aggregate need for early liquidity implies less investment left to mature and, hence, a smaller total amount of resources available to distribute. Wallace (1990) calls the decreasing size of early payments a "partial suspension of convertibility."

Wallace (1990) studies a particular case of aggregate uncertainty that, at the cost of appearing somewhat artificial, provides a clear illustration of the forces influencing the determination of the efficient allocation. In particular, he considers a situation in which there are two groups of agents: one group that contacts the bank first (still sequentially) and has a known proportion of patient and impatient members, and another group that contacts the bank afterward and has either all patient or all impatient agents. This second group is the driver of aggregate uncertainty in the model.

Wallace demonstrates that the optimal payments to the first group of agents do not depend on the order in which the agents are paid (as in the Diamond-Dybvig model without aggregate uncertainty). However, once the first agent of the second group reveals his preferences, the efficient payment to him, and the payments to the rest of the agents that have not yet withdrawn from the bank, adjust significantly. The reason for this adjustment is that when the first agent of the second group contacts the bank, he reveals crucial information about the aggregate state, and this new knowledge renders necessary an adjustment to the pattern of payments. In more general (and, perhaps, realistic) cases of aggregate uncertainty, a similar logic applies: Payments to subsequent agents adjust if the information provided by the new agent contacting the bank reveals substantial information about the realization of the aggregate state.

Note that these articles, and indeed the entire literature we review here, do not explicitly consider a deposit insurance system. As mentioned above, Wallace's specification of the sequential service constraint prevents the government from being able to finance deposit insurance by taxing agents who have already withdrawn. In line with the mechanism design literature, one way to interpret the exercise in these articles is by asking: What is the optimal way to distribute whatever resources are available in the economy given the constraints imposed by the physical environment (and, in particular, sequential service)? Wallace's results suggest that complete deposit insurance is unlikely to be optimal; when there is an unusually large number of early withdrawals, the efficient allocation gives less consumption to those depositors who are relatively late in the order induced by sequential service.

The Green-Lin Model

In an influential article, Green and Lin (2003) pick up, basically, where Wallace leaves off. They write down an environment in the Diamond-Dybvig tradition with a finite number of agents and i.i.d. preference shocks, and they study the possibility of banking fragility in such a setup. They first study an environment without sequential service and show that the unconstrained-efficient allocation is strongly implementable.[13] This result is not very surprising, but confirms the need to deal with sequential service if the theory is to have any hope of addressing issues associated with the possibility of bank fragility.

After dealing with the simple case with no sequential service, Green and Lin (2003) specify a Wallace-style, explicit sequential service constraint and prove a remarkable result. They show that the unconstrained-efficient allocation (which takes into account sequential service but not incentive compatibility) is also strongly implementable. In other words, under their specification of the environment (including a specific form for the sequential service constraint), there is no room for bank fragility in the model.

The details of the sequential service constraint specified by Green and Lin are important for our discussion. Following Wallace (1988, 1990), Green and Lin assume that agents are isolated from each other during the early period and cannot observe other agents' actions during that time. Furthermore, as in Wallace, all agents contact the bank during the early period (i.e., before investment has had time to mature), either to demand a withdrawal or to inform the bank of their decision not to withdraw. Lastly, Green and Lin introduce a novel element into the picture: They assume that the order in which agents contact the bank is known to them with some degree of accuracy; in the extreme and simplest case, each agent exactly knows his or her place in the sequence of contacts with the bank. In the more complicated case, agents observe their "time" of arrival to the bank, which allows them to estimate their approximate position in the order. As it turns out, nothing of substance is lost from adopting the extreme case of perfect knowledge of the position in the order (see Green and Lin [2000]).

Several important implications arise from the particular assumptions used by Green and Lin in their specification of the explicit sequential service constraint. We briefly discuss these implications here since they help one appreciate the nature of the results and the way those results change when alternative specifications of the environment are used.

The combination of a finite population with i.i.d. preference shocks allows aggregate uncertainty to play a significant role in the determination of the outcomes in the model. In fact, the i.i.d. assumption implies that all possible partitions of the set of agents between patient and impatient occur with positive

[13] To prove this result, the i.i.d. assumption is actually not needed.

probability and that the bank can never fully discover the aggregate state until all agents have had a chance to withdraw. In other words, as each new agent contacts the bank, additional information becomes available that must be taken into account in designing the optimal allocation. As a result, the sequential service constraint is always binding in the unconstrained-efficient allocation in their environment.

Even though the sequential order of withdrawals gives the environment a certain degree of "dynamics" during the early period, the isolation assumption implies that the withdrawal game played by agents is a simultaneous-move, *static* game. Agents simultaneously decide on their strategies that, in combination with the particular realization of agents' preferences, will determine the final allocation of resources across the population. A strategy for an agent in the withdrawal game is a contingent plan that specifies whether or not to withdraw when contacting the bank in the early period, depending on the agent's realized preferences and (expected) place in the order of arrivals. The simultaneous-move, static nature of the game eliminates several technical complications like the need to specify off-equilibrium beliefs or to consider the possibility that agents would want to influence the decisions of other agents that come later in the order of withdrawals.

The remarkable result in Green and Lin (2003) relies on a type of backward-induction logic that comes into play once the agents receive reliable information about their order of withdrawal. Consider an agent who knows she will be the last one to contact the bank. By the time her opportunity to withdraw arrives, all of the other agents will have already taken their actions. Suppose, for example, that all of these agents have chosen to withdraw early. Then this last agent knows that if she chooses to withdraw early, she will receive whatever resources are left in the bank.[14] If she chooses to wait, however, she will receive the matured value of these assets in the later period, which is larger. Hence, if she is patient, she is strictly better off waiting to withdraw.

Now consider the penultimate agent to contact the bank. From the reasoning above, he knows that the agent who comes after him will only withdraw if she is truly impatient. He does not know her preferences, of course, but he knows the probability of her being impatient. The unconstrained-efficient allocation has the property that this agent will always be strictly better off waiting if he is patient. The heart of Green and Lin's proof that the unconstrained-efficient allocation is strongly implementable consists of showing that this property holds in general: If any agent believes that all agents whose opportunity to withdraw arrives after hers will report truthfully, she strictly prefers to report truthfully herself, regardless of the reports of those who contact the

[14] Green and Lin show that, because all sequences of preference types are possible and agents' marginal utility of consumption is assumed to be unbounded at zero, the resources available for the last agent are always strictly positive, even if all previous agents have chosen to withdraw.

intermediary before her. It is important to note that the unconstrained-efficient allocation is not chosen to satisfy this property, and hence the reasons why this property holds are far from straightforward. Once this property is established, however, their main result follows from using iterated deletion of strictly dominated strategies to arrive at the strategy profile in which all agents report truthfully.

Green and Lin (2003) conclude from their analysis that something is missing in the Diamond-Dybvig theory of banking fragility. In their specification of the model, a bank can ensure that resources are allocated efficiently across depositors without introducing the type of fragility highlighted by Diamond and Dybvig.

Extensions and Clarifications

Andolfatto, Nosal, and Wallace (2007) study a modified version of the Green-Lin model in which they allow for a more general class of utility functions and clarify the importance of the i.i.d. assumption for obtaining the strong implementation result. They also (implicitly) change the sequential service constraint so that it differs in important ways from the one used by Green and Lin (2003). In the Green-Lin model, an agent does not observe the actions of those agents that have contacted the bank before her. Andolfatto, Nosal, and Wallace (2007) instead assume the bank informs each agent of the complete profile of actions taken by the agents before her, which allows an agent's action to be contingent on the actions of (a subset of) the other agents. This change in the environment makes the incentive compatibility constraints stronger, in the sense that fewer allocations are implementable.

Andolfatto, Nosal, and Wallace (2007) show that, in this modified environment, any allocation that is implementable is also strongly implementable. The logic of their proof is simple but powerful. In order for an allocation to be implementable in their environment, it must be the case that an agent, following *any* sequence of reports by the agents who have preceded her, prefers to report truthfully when all other agents report truthfully. Suppose now that an agent believes that some of the agents who preceded her have lied about their types, but that all agents who come after her will report truthfully. Under the assumption that preference types are i.i.d., the fact that some agents may have lied has no impact on her payoffs—all that matters is the sequence of actual reports. The fact that the allocation is implementable, therefore, implies that an agent will prefer to report truthfully as long as she believes that those who follow her will also report truthfully. Given this fact, the same type of backward-induction argument used by Green and Lin (2003) can be used to show that the allocation is strongly implementable. Since the constrained-efficient allocation is, by definition, incentive compatible and, hence, implementable, a corollary to the main result in Andolfatto, Nosal, and

Wallace is that the constrained-efficient allocation is strongly implementable and that there is no room for fragility in their model.

If preferences are of the type used by Diamond and Dybvig (1983), then Green and Lin's (2003) proof of their main result is actually powerful enough to establish the strong implementability of the unconstrained-efficient allocation even when the sequential service constraint is specified as in Andolfatto, Nosal, and Wallace. While the analysis presented by Andolfatto, Nosal, and Wallace is more general in that it allows for a wider range of preferences than does the analysis by Green and Lin, it does not focus on the unconstrained-efficient allocation; the results only apply to implementable allocations.

An important clarification should be made at this point. Green and Lin (2003) find the constrained-efficient allocation by first solving an auxiliary problem without the incentive compatibility constraints and then showing that the solution is, actually, incentive compatible. For the general class of utility functions considered by Andolfatto, Nosal, and Wallace (2007), the incentive compatibility constraints are likely to be binding in many cases, even if agents' preference shocks are independent. For this reason, the methodology employed by Green and Lin (2003) to find the constrained-efficient allocation is likely to fail in many of the cases considered by Andolfatto, Nosal, and Wallace (i.e., the solution to the planning problem without the incentive constraints may not be incentive compatible). Finding the constrained-efficient allocation, then, may involve additional complications like identifying which incentive compatibility constraints are likely to be binding and then "reshaping" the payment scheme to minimize the distortions induced by the incentive compatibility requirement.

Ennis and Keister (2009a) modify the Green-Lin model in a different way by relaxing the assumption that preference types are independent across agents. All other elements of the model, including the specification of the sequential service constraint, are exactly as in the Green and Lin analysis. Under the assumption that preferences exhibit constant relative risk aversion, they derive the unconstrained-efficient allocation in closed form, which allows them to calculate examples with more agents than had been done in the previous literature. They present a series of examples that show how the results of Green and Lin (2003) can break down when types are correlated. In these examples, there exists an equilibrium of the withdrawal game in which some, but not all, agents run on the bank by withdrawing early regardless of their true consumption needs.

The logic used by Green and Lin (2003) to show that the last agent to contact the bank has no incentive to misreport her type still holds in this setting. For this reason, there cannot be an equilibrium in which all agents run on the bank. The equilibria constructed in Ennis and Keister (2009a) have the property that those agents who have a relatively early opportunity to withdraw choose to run, while those who are relatively late withdraw only if they have a true consumption need. An essential feature of these examples is that the key property identified by Green and Lin fails to hold—an agent who believes that everyone who arrives after her will report truthfully may nevertheless prefer to misrepresent her type. These results show that the strong-implementability result of Green and Lin (2003) relies on more than a simple use of backward-induction logic; it depends critically on properties of the unconstrained-efficient allocation that may not hold when agents' preference types are not i.i.d.

Alternative Approaches to Sequential Service

The Green-Lin formulation of the sequential service constraint is appealing in several dimensions. To begin with, it is clearly specified and helps the reader view the allocation problem in the Diamond-Dybvig model in terms of the standard theory of mechanism design. In addition, their specification shows how important "dynamic" features of bank runs can be captured in a model without bringing in the complications associated with dynamic games. Several subsequent articles have investigated how much the particular assumptions Green and Lin made matter for their strong implementation result. Peck and Shell (2003) modify the Green and Lin environment in two ways, considering both a more general specification of agents' preferences and a different specification of the sequential service constraint. They find that the strong implementation result of Green and Lin goes away under this alternative set of (also reasonable) assumptions.

With respect to agents' preferences, Peck and Shell allow the marginal utility of impatient agents to differ from that of patient agents. When impatient agents have a high marginal value of consumption, the bank will want to give relatively large payments to those agents who withdraw early. If this effect is strong enough, the incentive constraint for patient agents will be binding in the constrained-efficient allocation, something that could not happen in the Green-Lin model. The relatively large payments made on early withdrawals increases the incentive of patient agents to misrepresent their type if they expect others to do so.

The second change introduced by Peck and Shell is in the way the sequential service constraint is specified. The agents in Peck-Shell do not observe any information about their position in the order of arrival at the bank before making their withdrawal decision. Instead, each agent views the positions as being randomly assigned after withdrawal decisions have been made. Under this approach, the backward-induction logic used by Green and Lin cannot be applied since no agent is confident that she will be the last one to contact the bank.

These two changes—in preferences and in the specification of sequential service—are both important for the examples of run equilibrium constructed by Peck and Shell. The change in the sequential service constraint enlarges the set of implementable allocations relative to the Green-Lin model, since now there is a single incentive compatibility constraint rather than a separate constraint following each possible history of reports leading up to an agent's decision. The change in preferences implies that the constrained-efficient allocation in the Peck-Shell setting may not be implementable in the Green-Lin specification of sequential service and, in fact, the examples in Peck and Shell have this feature. It remained an open question whether both of these elements were needed to overturn the strong implementation result of Green and Lin.

Ennis and Keister (2009a) answer this question by constructing examples of run equilibria in which the environment is identical to that in Green and Lin's article except that agents do not know their position in the order of arrival at the bank. There is no change in preferences and, as a result, the constrained-efficient allocation is exactly as in Green and Lin's model. These results show that it is the change in the sequential service constraint, and not the nonstandard specification of preferences, that is at the heart of the Peck-Shell result.

Peck and Shell make another interesting change to the sequential service constraint, although they show it is not important for their results. Green and Lin assume that all agents contact the bank during the first round of withdrawals, regardless of whether the agent wishes to withdraw or not. Peck and Shell, instead, assume that only agents who wish to withdraw contact the bank. This change results in a more coarse information structure for the bank. In particular, the bank only observes withdrawals and, as a consequence, the efficient allocation is less responsive to the type realizations of those agents who are early in the line. In the Green and Lin setup, when the bank observes that the first agent in the line is impatient, it adjusts the constrained-efficient allocation by reducing the early payments. In Peck and Shell, information arrives to the bank more slowly, leading the bank to make fewer adjustments to the allocation early on in the process. However, Peck and Shell show that their same result obtains when all agents report to the bank regardless of whether they want to withdraw or not (see their Appendix B).

The banking contract that implements the optimal allocation in the Green-Lin setup generally involves payments to agents that are highly contingent on the information collected by the bank up to the point of actually making the payment. This feature is a consequence of the combination of aggregate uncertainty with sequential service and seems counter to common practice in banking where the face value of deposits is respected under most circumstances. This counterfactual implication of the Diamond-Dybvig theory was, in fact, recognized since its inception (see, for example, Postlewaite and Vives [1987]).

A plausible modification of the details involved in the specification of Green and Lin's sequential service constraint may move the theory closer to reality in this respect. In particular, some preliminary results from our own research (see Ennis and Keister [2008]) suggest that when the bank only observes withdrawals as they occur (following the specification in Peck and Shell [2003]), but obtains no information about the realized preferences of agents who do not intend to withdraw, the constrained-efficient allocation more closely resembles a demand deposit contract. This result holds even when agents know their place in the order at the time of their early withdrawal decision (an assumption made by Green and Lin [2003] that was not present in Peck and Shell [2003]). Interestingly enough, when this modification to Green and Lin's specification of the sequential service constraint is introduced, the efficient allocation may no longer be strongly implementable for some parameter configurations and the possibility of bank fragility reappears in the model.

As we have seen, one of the main differences among the alternative specifications of the sequential service constraint lies on the amount of information that an agent has at the time of deciding whether or not to withdraw. In the version studied by Green and Lin (2003) the agent knows if she is patient or impatient and her place in the order of sequential contacts with the bank. In the version of Andolfatto, Nosal, and Wallace (2007) the agent knows more (the actions of those agents prior to her in the line) and in the Peck-Shell version the agent knows less (only whether she is patient or impatient).

In a recent article, Nosal and Wallace (2009) propose an alternative interpretation of the various specifications of the sequential service constraint in this dimension. In particular, they assume that the agent directly receives information only about his preferences, and that the bank can communicate to the agent (before he chooses whether or not to withdraw) information that it may have about the agent's place in the order and what the other agents before him have done. This way of thinking about the model provides a unified way of viewing the alternative specifications that have been studied in the literature, each corresponding to a different assumption about the amount of information the bank is revealing to the agent.

A natural question to ask under this approach is how much information the bank would reveal to agents if it were allowed to choose. Nosal and Wallace (2009) study this question when the bank is a benevolent entity (a planner). An interesting complication arises at this point. The set of implementable allocations is strictly larger when agents do not have any information about the order, which would in principle give the planner more flexibility in designing the payoff schedule. However, as Peck and Shell (2003) have shown, the constrained-efficient allocation may not be strongly implementable in this case. Nosal and Wallace show that if the planner is only concerned with implementation (but not with strong implementation) then, under some parameter values, it will not want to reveal information about the order to the agents.

This finding has important implications for the possibility of a bank run in the model. If the bank believes a run is very unlikely to occur, even when one is consistent with equilibrium, then it may choose not to reveal information that could rule out the possibility of a run. This happens because, by not revealing information, the bank improves the outcome that obtains when agents do not run. In other words, there is a tradeoff in the model between efficiency when a run does not occur and eliminating the possibility of a run altogether.

3. OTHER POSSIBLE INGREDIENTS

The Green-Lin model and the modifications of it that we have discussed so far describe a very basic environment that abstracts from many other features that are typically associated with the workings of banking institutions. A natural question, then, is to ask whether or not there might be additional ingredients that are relevant to explain banking and bank fragility in models within the Diamond-Dybvig tradition. In this section, we discuss three possibilities that have been recently examined: self-interested bankers, limited commitment, and investment restrictions.

Self-Interested Bankers and Moral Hazard

In all of the discussion above, the bank is operated with the objective of maximizing the welfare of its depositors. It seems more in line with reality, however, to explicitly model situations in which the banker does not always act in depositors' best interests. In Green and Lin's environment, the banker centralizes the information about the aggregate state as it is gradually revealed by the sequential decisions of agents. While the banker may be able to commit to a payment contract, the contract may give the banker incentives to manipulate the information provided to the remaining agents after each withdrawal. Based on this logic, Andolfatto and Nosal (2008) illustrate how a self-interested banker in charge of delivering a contract of the type studied by Green and Lin may want to misrepresent the situation and artificially reduce payouts to depositors.

After establishing this fact, Andolfatto and Nosal investigate alternative schemes that could be used to give proper incentives to the banker. To benefit from a banking arrangement depositors must, eventually, be able to compare the claims of the banker with some relevant information about the true aggregate state. As a consequence, new assumptions are needed about the accessibility to information by agents and the banker. Unfortunately, there is no clear natural way to proceed in formalizing this issue. Andolfatto and Nosal pick one particular configuration—agents can convene after investment matures and collect information about their actual preferences. In this case, Andolfatto and Nosal show that the payments in the best contract delivered by a self-interested banker may be less sensitive to the aggregate state than the Green-Lin contract and, hence, may appear more in line with the type of demand deposit contracts that are common in real-world banking. However, this result depends on parameters, and in certain cases the contract actually becomes more complex than the Green-Lin contract (with new contingencies and positive early payments to patient depositors).

Andolfatto and Nosal also study the implications for financial fragility of considering explicitly the incentives of the banker. They conclude that it may be harder to construct equilibria in which patient agents misreport their types. However, their analysis is far from conclusive. Overall, their work demonstrates that analyzing the effects of bankers' agency problems in the Green-Lin model, while potentially important, is not a straightforward task. This line of inquiry, though, seems to us potentially very fruitful and deserving of further attention.

Limited Commitment

Another ingredient that may be important for explaining financial fragility is limited commitment on the part of the bank or on the part of policymakers more generally. Ennis and Keister (2010) study a version of the Diamond-Dybvig model in which the bank cannot commit to a plan of action; rather, the payment to each agent is only decided when that agent arrives to withdraw. The key aspect of this lack of commitment power is that it prevents the bank from being able to credibly use a suspension of convertibility clause to uniquely implement the constrained-efficient allocation.

In the environment studied by Ennis and Keister (2010), there is no aggregate uncertainty and the bank knows precisely how many agents will be impatient.[15] The sequential service constraint follows Peck and Shell (2003) in assuming that only those agents seeking to withdraw contact the bank. Once the number of withdrawals passes a certain threshold, therefore, the bank will know for sure that a run is underway. However, once this situation is reached, it will not be ex-post optimal for the bank to follow through with a suspension, since doing so would imply giving no consumption to some agents who are truly impatient (as in Ennis and Keister [2009b]). Instead, the bank continues making payments to some of the withdrawing agents, compromising resource availability in the later period.

Ennis and Keister (2010) demonstrate that when the bank initially believes that a run is unlikely, it will in some cases choose payments that make a run

[15] An interesting avenue for future research would be to apply the techniques developed in Ennis and Keister (2010) to the model with a finite number of agents and aggregate uncertainty.

consistent with equilibrium. In other words, bank fragility is possible in the Ennis-Keister version of the Diamond-Dybvig model with limited commitment, even though there is no fundamental source of aggregate uncertainty in the model. Interestingly, the equilibria of the model have a natural "dynamic" structure, which derives from the fact that agents have information about their position in the order of early withdrawal opportunities (as in Green and Lin [2003]). An equilibrium bank run consists of an initial wave of withdrawals, which is followed by a reaction from policymakers. Following this reaction, the run may end or it may continue with another wave of withdrawals taking place, which would lead to another reaction from policymakers, and so on. This interplay between the withdrawal decisions of agents and the reaction of policymakers seems to be an important feature of real-world banking crises.

Investment Restrictions

There is a long tradition in policy of regulating the activity of banking. One common approach has been to restrict the type of investments that banks are allowed to undertake. For example, for more than 50 years, banks in the United States that accepted deposits from the public were prohibited from engaging in certain asset management activities, which were reserved for a different set of institutions called investment banks. These restrictions were imposed partly as a way to address the possibility of bank fragility. When those policies were designed, a formal theory of banking was not available. Diamond and Dybvig (1983) and the literature that followed have provided such theory and, hence, it is natural to ask how this kind of policy influences outcomes in the models within this tradition. Peck and Shell (2010) address this question.

Peck and Shell (2010) consider an environment with an indivisibility in consumption, which is aimed at capturing the payment function of demand deposits: A check written for a purchase, for example, either pays the bearer at par or may not be useful for exchange. Peck and Shell consider an environment with two investment technologies: one technology is as in the standard Diamond-Dybvig model and the other has higher long-run return but is completely illiquid in the short run. They analyze two regulatory systems for banks—a unified system and a separated system. In the unified system, banks are allowed to invest in both technologies on behalf of agents. In the separated system, however, banks cannot invest in the illiquid technology and agents do that directly. Somewhat surprisingly, Peck and Shell show that runs can happen in the separated system but not in the unified system. They conclude that policies that impose restrictions on the investment strategies of banks can actually have unexpected, counterproductive effects by inducing fragility in the system.

4. CONCLUSION

Understanding the root causes of the banking crises that have been observed around the world is an extremely difficult task. Some commentators claim that self-fulfilling behavior on the part of depositors and investors plays a critical role, while others emphasize more fundamental factors related to the value of banks' assets. Banking crises are complex phenomena that typically occur in conjunction with a variety of unfavorable financial and macroeconomic factors, making it difficult to determine the true underlying cause of an event. In spite of these difficulties, progress has recently been made in several directions. This article reviews the progress in one of these directions.

The literature we have discussed shows that it is possible to provide an internally consistent explanation for the self-fulfilling interpretation of bank runs. However, this literature also shows that the details of the environment are important. In other words, the fragility of banks in these models is the result of physical and informational frictions, but only specific combinations of these frictions lead to fragility. In particular, information about the actions of agents must not flow too quickly, so that the bank makes a significant amount of payments to depositors before discovering whether or not a run is underway. In addition, some feature of the environment must make suspension of convertibility clauses in deposit contracts either undesirable or ineffective.

How important are self-fulfilling factors in the explanation of observed crises? It may very well be the case that the types of frictions described in this paper were present in the real economy and that observed financial crises have had a considerable self-fulfilling component. If these theories are a useful reflection of reality, however, it is important to realize that natural changes in the way information flows in the economy (because of, for example, technological innovation) could have substantial implications for bank fragility in the future. In addition, it seems important to recognize that our understanding of the issues involved remains fairly limited. Identifying appropriate policies to deal with bank fragility, then, must be an ever-evolving activity that takes into account changes in the structure of the financial system as well as further developments in our understanding of the issues. The theories we have discussed here provide a solid foundation for pursuing these important and pressing issues.

REFERENCES

Andolfatto, David, and Ed Nosal. 2008. "Bank Incentives, Contract Design, and Bank Runs." *Journal of Economic Theory* 142 (September): 28–47.

Andolfatto, David, Ed Nosal, and Neil Wallace. 2007. "The Role of Independence in the Green-Lin Diamond-Dybvig Model." *Journal of Economic Theory* 137 (November): 709–15.

Bryant, John. 1980. "A Model of Reserves, Bank Runs, and Deposit Insurance." *Journal of Banking and Finance* 4 (December): 335–44.

Cooper, Russell, and Thomas W. Ross. 1998. "Bank Runs: Liquidity Costs and Investment Distortions." *Journal of Monetary Economics* 41 (February): 27–38.

De Nicoló, Gianni. 1996. "Run-Proof Banking Without Suspension or Deposit Insurance." *Journal of Monetary Economics* 38 (October): 377–90.

Diamond, Douglas W., and Phillip H. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy* 91 (June): 401–19.

Ennis, Huberto M., and Todd Keister. 2006. "Bank Runs and Investment Decisions Revisited." *Journal of Monetary Economics* 53 (March): 217–32.

Ennis, Huberto M., and Todd Keister. 2008. "Run Equilibria in a Model of Financial Intermediation." Federal Reserve Bank of New York Staff Report No. 312 (January).

Ennis, Huberto M., and Todd Keister. 2009a. "Run Equilibria in the Green-Lin Model of Financial Intermediation." *Journal of Economic Theory* 144 (September): 1,996–2,020.

Ennis, Huberto M., and Todd Keister. 2009b. "Bank Runs and Institutions: The Perils of Intervention." *American Economic Review* 99 (September): 1,588–607.

Ennis, Huberto M., and Todd Keister. 2010. "Banking Panics and Policy Responses." *Journal of Monetary Economics* 57 (May): 404–19.

Green, Edward J., and Ping Lin. 2000. "Diamond and Dybvig's Classic Theory of Financial Intermediation: What's Missing?" Federal Reserve Bank of Minneapolis *Quarterly Review* 24 (Winter): 3–13.

Green, Edward J., and Ping Lin. 2003. "Implementing Efficient Allocations in a Model of Financial Intermediation." *Journal of Economic Theory* 109 (March): 1–23.

Gu, Chao. Forthcoming. "Partial Bank Runs Triggered by Noisy Sunspots." *Macroeconomic Dynamics*.

Jacklin, Charles J. 1987. "Demand Deposits, Trading Restrictions, and Risk Sharing." In *Contractual Arrangements for Intertemporal Trade*, edited by Edward C. Prescott and Neil Wallace. Minneapolis: University of Minnesota Press, 26–47.

Nosal, Ed, and Neil Wallace. 2009. "Information Revelation in the Diamond-Dybvig Banking Model." Federal Reserve Bank of Chicago Policy Discussion Paper Series (December).

Peck, James, and Karl Shell. 2003. "Equilibrium Bank Runs." *Journal of Political Economy* 111 (February): 103–23.

Peck, James, and Karl Shell. 2010. "Could Making Banks Hold Only Liquid Assets Induce Bank Runs?" *Journal of Monetary Economics* 57 (May): 420–7.

Postlewaite, Andrew, and Xavier Vives. 1987. "Bank Runs as an Equilibrium Phenomenon." *Journal of Political Economy* 95 (June): 485–91.

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.

Wallace, Neil. 1990. "A Banking Model in which Partial Suspension is Best." Federal Reserve Bank of Minneapolis *Quarterly Review* 14 (Fall): 11–23.