No. 45

REGULATION OF FINANCIAL INSTITUTIONS -- A FUNCTIONAL ANALYSIS

John Chant

Bank of Canada
Technical Reports

Rapports techniques
Banque du Canada

January, 1987

Technical Report 45

REGULATION OF FINANCIAL INSTITUTIONS
-- A FUNCTIONAL ANALYSIS

by

John Chant

Dr. Chant is a professor of economics at Simon Fraser University and prepared this report while acting as a consultant for the Bank of Canada. The views expressed in this report are those of the author; no responsibility for them should be attributed to the Bank of Canada.

ACKNOWLEDGEMENTS

I am indebted to the Bank of Canada for the opportunity to carry out this study under ideal conditions. At the Bank, Charles Freedman, Clyde Goodlet, Paul Jenkins, John Murray and Gordon Thiessen gave valued support through their encouragement, questions and comments at various stages of the project. In addition, Keith Acheson, Tom Courchene, John McManus, Jim Pesando and Henri-Paul Rousseau have all assisted me with their helpful comments on the study. Paul Gomme supplied helpful research assistance and Jill Moxley and Lea-Anne Solomonian provided valued editorial assistance.

CONTENTS

ABSTRACT

This study has three main objectives: First, to examine the reasons for the existence of financial institutions that simultaneously lend to one group and borrow from another; second, to analyze the reasons for the special treatment of these institutions in terms of the regulation that governs their activities; and third, to assess the types of regulation that can be explained as a consequence of the reasons for regulation. The analysis suggests that deposit-taking institutions that offer fixed-value liabilities to their customers are best explained either through their advantages in monitoring and enforcement of investments or through the economies inherent in being payments intermediaries. A case for the regulation of financial institutions arises because of the nature of the assets of financial institutions and the transactions costs required to coordinate their many creditors. The forms of regulation suggested by the analysis include minimum capital ratios, eligibility requirements for investments, prohibitions on the joint ownership of financial institutions and other enterprises, and limitations on the terms and scale of transactions between associated enterprises.

RÉSUMÉ

La présente étude comporte trois volets. Dans le premier, l'auteur cherche à expliquer pourquoi il existe des institutions qui font à la fois des opérations de prêt et d'emprunt. Dans le second, il étudie les raisons pour lesquelles il faut soumettre les activités des institutions financières à une réglementation spéciale. Dans le troisième, il évalue les types de réglementation qui, à ses yeux, découlent de ce besoin. Il ressort de cette étude que la meilleure explication de l'existence d'institutions de dépôt offrant à leurs clients des placements à valeur fixe est fournie par les avantages dont jouissent ces dernières sur le plan de la surveillance de leurs placements et du recouvrement des sommes qui leur sont dues ou encore par les économies d'échelle inhérentes à leur rôle d'intermédiaire. On peut justifier la réglementation des rapports entre les institutions financières avec leur rôle d'intermédiaire. On peut justifier les réglementation des rapports entre les institutions financières avec leur clientèle -- qui diffèrent beaucoup des relations entre d'autres types de créanciers et de débiteurs -- par la nature des avoirs de ces institutions et par les coûts de transactions auxquels celles-ci doivent faire face pour assurer la coordination entre leurs nombreux créanciers. Différentes formes de réglementation sont proposées dans l'étude, notamment l'imposition de taux minimums de capitalisation, la soumission des investissements à des conditions d'admissibilité, l'interdiction de posséder conjointment des institutions financières et d'autres entreprises et les restrictions sur les modalités et l'ampleur des transactions entre entreprises associés.

INTRODUCTION

Financial institutions that borrow from ultimate lenders and lend to ultimate borrowers are a prominent feature of any modern economy. These institutions are typically scrutinized and regulated by governments much more comprehensively than other sectors in the economy. The approach taken to the regulation of financial institutions has recently been subject to question and re-examination in a number of countries. Any reassessment of the present regulatory system requires a prior understanding of both the functions of financial institutions and the ways in which they perform these functions.

The purpose of this study is to contribute to the understanding of the workings of financial institutions and of the institutional framework that governs their operation. Three issues relating to financial institutions and their regulation are examined. What functions do financial institutions perform? What do these functions and the way that they are performed imply about the need for regulation? What forms of regulation are justified by these functions?

In Chapter 1, the question is posed -- why do financial institutions exist? Many different theories have been used to explain the functioning of financial institutions. These different theories exist, in part, because they explain different types of institutions. Some relate to brokers who expedite transactions between direct borrowers and direct lenders while others explain mutual fund intermediaries that offer a proportionate share of their portfolio of assets to their customers. A final category of theories explains deposit-taking institutions that offer their customers fixed-value claims that are independent of the performance of the institution's portfolio.

This study focuses on the deposit-taking institution. This form of financial institution corresponds to the chartered banks, the trust and mortgage loan companies, and the credit unions and caisses populaires in Canadian financial markets. The emphasis on these institutions is justified because many current issues in financial market policy concern their regulation.

The second issue addressed in this study concerns the general case for the regulation of financial institutions overall, and deposit-taking institutions in particular. In Chapter 2, it is asked whether the functions performed and the ways in which institutions carry out these functions provide a rationale for their regulation. The analysis suggests that too great a distinction should not be drawn between the framework of contract law required for the conduct of everyday business and the regulations that govern the activities of financial institutions. Rather, the analysis suggests that a distinction be made between remedial measures and preventive measures for the protection of creditors. Remedial measures are found to be less suitable i) the more specific the capital used in any activity, and ii) the more numerous the creditors. The analysis concludes that the nature of financial intermediation makes it more suited to control by preventive, as distinct from remedial, measures.

The appropriate form of regulation to be applied to financial institutions is the final issue considered. Chapter 3 examines the differences in the interests of the depositors and the operators of financial institutions. Their interests are likely to differ with respect to i) the riskiness of the investments undertaken by the financial institution, ii) the choice that institution makes between the holding of debt and equity securities, and iii) the extent of transactions made between the institutions and other enterprises with associated ownership. Suggestions are made with respect to the types of regulations that could be used to overcome or control these conflicts between operators of financial institutions and their customers. The forms of regulation suggested correspond to a number of regulations that either currently govern or have been proposed to govern the activities of deposit-taking institutions.

Chapter 1
FUNCTIONS OF FINANCIAL INSTITUTIONS

The literature explaining the existence and functions of financial institutions is extremely diverse. No single explanation of the phenomenon of financial intermediation has emerged. One source of diversity arises from the various types of activity taking place. While explanations for several types of financial institutions are reviewed in this chapter, primary emphasis is on the deposit-taking institution that offers fixed-value claims to its customers. This emphasis is justified because these deposit-taking institutions are a predominant feature of our financial system and much concern is currently being directed towards the formulation of policies to govern their operations.

1.1 Types of Intermediaries

The essence of financial intermediation is the presence of a third party between the ultimate borrower and ultimate lender in the saving-investment process. The simplest form of intermediary is the broker who facilitates transactions between the lender and the borrower without acquiring the debt of the borrower or issuing claims to the lender. In contrast, the "true" financial intermediary becomes a party to the financing activity by simultaneously holding claims on the borrower and issuing claims to the lender. Among true intermediaries a distinction can be made between mutual funds and deposit-taking entities. A mutual fund offers investors a proportionate claim on a portfolio of assets. This claim fluctuates in value with the underlying collection of assets that make up the intermediary's portfolio. The deposit-taking intermediary, the focus of this study, offers investors a claim, the value of which is stated by contract to be independent of the value of the portfolio held by the intermediary.

A distinction should be made between the functions performed by financial institutions and the reasons why these functions are performed

through financial institutions. In other words, it is not adequate to describe the function of financial institutions; it must be shown that the use of an intermediary permits lenders and borrowers to perform the function differently than would be possible on their own. Moreover, use of the financial intermediary must dominate direct transactions between the lender and borrower.

The following analysis starts with a simple economy with perfect certainty and the complete absence of transactions costs. In such an economy, there would be no need for financial institutions. The analysis proceeds by removing the simplifying assumptions to determine the conditions under which financial intermediaries could be expected to emerge. A number of questions are asked under each set of assumptions. Does an economic problem exist for lenders and borrowers that cannot be resolved in a mutually beneficial way through direct contract between them? Under what assumptions can this problem be resolved through financial intermediation? What type of financial intermediation is required to resolve the problem?

Emphasis, as already mentioned, is placed on the ability of any theory to explain the need for deposit-taking institutions. Any such theory must integrate three prominent features that characterize deposit-taking institutions. First, they issue liabilities that have by contract a predetermined money value regardless of the performance of the intermediary's portfolio. Second, while some of their liabilities are in the form of chequable deposits, the majority are not. For example, only a small proportion of the total deposit liabilities of Canadian chartered banks and trust and mortgage loan companies are chequable. Finally, deposit-taking institutions differ from other intermediaries not only in terms of form but also with respect to the markets in which they operate. Brokers and mutual funds deal most commonly in so-called marketable securities whereas deposit-taking institutions have a large proportion of non-marketable securities (including loans) in their portfolios.

1.2 The Approach to Explanation

The cost of performing transactions is a central element of the existing theory of financial institutions. Indeed Benston and Smith (1976) argue that "the raison d'être for this industry is the existence of transactions costs." A simple theory of financial institutions could be derived on the argument that these institutions have access to superior technology for overcoming transactions costs. Such an explanation carries the danger of being tautological to the extent that it fails to advance understanding of the sources of advantage from which financial institutions arise.[1]

In the discussion that follows, intermediaries are assumed to have the same technology in all activities as other members of the general public. This assumption does not deny that the use of intermediation can have a cost advantage over direct finance. As Gurley and Shaw (1960) state

> Financial intermediaries exploit economies of scale in lending and borrowing. On the lending side the intermediary can invest and manage investments in primary securities at unit costs far below the experience of most individual lenders. (p. 194)

Any cost advantage is assumed for the present to arise from the organization of the activity through an intermediary that uses existing technology and not from the assumption of any inherent advantage of those who choose to be intermediaries.

The assertion that transactions costs explain the existence of financial institutions is not very useful unless the nature of these costs is specified. In some cases, financial intermediation provides the means for overcoming the problem of transactions costs; in others, some other form of contractual arrangement may serve the interest of lenders and borrowers to a greater degree. The heterogeneity of transactions

---

1. For a similar discussion about using differences in tastes to explain phenomena, see Stigler and Becker (1977). For an example in which financial institutions are assumed to exist for unspecified cost advantages, see Pyle (1971).

costs is recognized in the literature. At various points Benston and Smith (1976) identify fixed and differential costs; costs associated with individual securities; transportation and inconvenience costs; costs of barter; administrative, monitoring and processing costs; and documentation, information and monitoring costs. Each cost poses an additional problem for the lender and the borrower beyond those of the "perfect" economy in which transactions are costless.

1.3 Complete Certainty

The simplest economy to consider is one characterized by complete certainty. Households are assumed to know their incomes and consumption demands for all times in the future. Moreover, production technologies and supplies of productive inputs are also known with certainty.

1.3.1 Costless transactions

Households in this economy receive income from supplying production services over a number of periods. Any lack of correspondence between the household's pattern of income and its pattern of consumption over time means the household must sell or purchase assets. In a pure exchange economy, the household can either exchange claims for consumption at different times with other households or hold consumption goods. In a production economy, households can acquire durable capital goods and sell them to finance their consumption at later dates. Initially, all transactions can be carried out without cost. Moreover, transactions over time can be enforced without cost.

The production technology may require people to use productive capital in amounts different from the amount that they own. In the absence of transactions costs the ownership and use of productive assets can be separated without consequence. The terms of the arrangements by which one household permits another to use its production assets are negotiated, monitored, and enforced without resource costs.

The household's choice of assets would not matter in the absence of uncertainty and transactions costs. The return for holding each asset

would be fully known to the household at the time of acquisition. Assets would be priced so as to equalize their rates of return. As Benston and Smith (1976) observe

> it should be obvious that in a perfect market, a market with no frictions such as transactions costs, information costs or indivisibilities, financial intermediaries would not exist. (p. 217)

1.3.2 Costly transactions

The assumption of perfect certainty eliminates many types of transactions costs used to explain the existence of financial institutions. Under this assumption, neither the lender nor the borrower needs to incur any search costs in finding others with whom to trade financial claims. Also, the outcomes of investments over time are known so that investors are obliged neither to screen investments in advance nor to determine their subsequent outcomes.

A case for intermediation cannot be established under these conditions even if significant transactions costs are assumed to exist. In the absence of uncertainty, an investor would need to hold more than one asset only to the extent that his portfolio exceeded the scale of the borrowers' needs. With such a mismatch, the number of transactions between lenders and borrowers cannot exceed the sum of the number of borrowers (the number of transactions that would occur with perfect matching) plus the number of lenders (the minimum number of carryovers from one lender to another caused by mismatches). An intermediary would need to make a transaction with each lender and each borrower. Therefore, the number of transactions required with the intermediary equals the number of transactions under the worst mismatching of lenders with borrowers. Financial intermediation, in this case, cannot produce any savings for the economy with respect to transactions costs.[2]

---

2. Obviously, intermediation would exist if the combination of the intermediary-lender or intermediary-borrower transactions had advantages relative to direct borrower-lender transactions. This justification does not help us to understand the existence of intermediaries. It is merely an alternative way of asserting it.

1.4 Uncertainty and Intermediation

Uncertainty about the outcomes of transactions made over time can enter the economy in many ways. The process that determines the output produced by given quantities of inputs can be random. Social uncertainty exists when the uncertainties in individual processes are combined in a way that causes uncertainty with respect to overall output. Individuals can still face private uncertainty in production even when the underlying uncertainties offset each other for the economy as a whole.[3] Both social and private uncertainty are assumed to be inherent in the production processes in the economy and cannot be eliminated through the application of additional resources.

The stochastic nature of the production process is not the only source of uncertainty in the economy. Uncertainty can also arise from i) distinguishing good investments and ii) from the dependence of the outcomes of investments on the resources directed towards monitoring and enforcement. The first type of uncertainty occurs because of the costs involved in identifying different types of investment prior to undertaking them. The possibility that an investor's return can be increased and his risk reduced by assessing different investments means that information costs are an important determinant of the investor's level of uncertainty. The second type of uncertainty arises because it is costly to monitor and enforce agreements with respect to both the supply of productive inputs and the sharing of outputs.[4] The degree of this type of uncertainty depends on the costs of assuring conformity to prior agreements. The purchaser of inputs would use additional resources in enforcing agreements only if the benefits (in the form of higher return and lower risk) exceed the additional costs.

---

3. This distinction between social and private uncertainty corresponds to the distinction made in finance theory between systematic and unsystematic risk or between undiversifiable and diversifiable risk.

4. It can be argued that this type of uncertainty corresponds to the concept of "uncertainty" that Frank Knight distinguished from "risk". Risks are insurable because the probabilities of different outcomes are independent of the presence of insurance. In contrast, incentives to monitor and enforce are altered by the presence of insurance, making the uncertainties endogenous.

Each type of uncertainty has different implications for the arrangements that can be expected to develop between the lender and the borrower. As a result, different sources of uncertainty are treated separately in the following discussion.

The economist's standard device for examining uncertainty is the two-period model in which households choose between present consumption and future consumption. Investments with uncertain returns permit households to transfer purchasing power between the two periods. For the present, it is assumed that investment outcome can be determined without any cost in the final period.

1.4.1 Private risk

a) Investment uncertainty

The simplest case of uncertainty arises when the outcomes of individual investments are uncertain but the outcome of the investments in aggregate is known in advance. Suppose there are λ individuals with X to invest and that each of the possible investment projects in the simplest case costs X . Each investment project gives a payment in the second period of either $X(1+r+\varepsilon)$ with probability 0.5 or $X(1+r-\varepsilon)$ with equal probability. A household holding only one investment has an expected gross return of r from the investment with a standard deviation of ε. In addition, a fixed cost, c, is incurred in acquiring each asset. The net return expected by this household from acquiring one asset equals r-c.

If individuals were risk-neutral, they would choose to hold only one asset. Even though the outcome would be uncertain, the probability, p, of gaining ε above the expected return would offset exactly the cost to investors of the prospect of losing ε. While investors would not realize any advantage from holding more than one investment, they would incur additional costs.

Risk-averse investors value an uncertain income stream less than a known income stream. They can avoid uncertainty by holding more than one investment but in so doing incur additional costs. Risk-averse investors

would continue to divide their resources among additional investments until the additional costs of adding further investments offset the additional benefits at the margin. Each investor could achieve this degree of diversification at a cost of θc, where θ is the number of securities at which the benefits of diversification are maximized and c, as before, is the fixed cost of acquiring each asset. The λ identical individuals in the economy would carry out diversification at a cost of $\lambda\theta c$.

An intermediary could reduce the costs to individuals of holding diversified portfolios. Consider the case where the intermediary holds a portfolio that maximizes the benefits from diversification. In assembling this portfolio, the intermediary would incur a cost of θc, the fixed cost, c, for each of θ assets. In addition, the intermediary must negotiate with each lender at a cost of λc. The total cost of this degree of diversification would be $(\lambda+\theta)c$. Investors would find indirect finance a more efficient means of diversification whenever $\frac{\lambda + \theta}{\lambda\theta}$ is less than one. This condition shows that intermediation is more likely the larger the number of investors over which the intermediary can spread its costs.

It can easily be shown that the intermediary would acquire more investments than typical investors would hold on their own. The benefit to the investor of an additional security in the intermediary's portfolio remains the same as that of an additional security in his own portfolio because he holds $1/\lambda$ of a portfolio λ times the size of his own. The cost to the intermediary of acquiring that additional security is c and the marginal cost perceived by the investor is c/λ. As a result, the intermediary would respond to its customers' wishes by holding a larger number of different securities than each customer would have held.

Acquisition costs prevent the intermediary from acquiring sufficient investments to become completely diversified. To achieve such a degree of diversification, the intermediary would have to acquire shares of all the investments available in the economy. It would do so only if the sum of the benefits to all its customers from eliminating the risk exceeded the costs involved.

A definite form for the intermediary is implied by this explanation.[5] A brokerage intermediary would emerge only if the costs involved in achieving diversification were search costs and did not arise from the direct holding of securities. With only search costs, the broker could direct his customers towards appropriate securities so that they could gain their desired level of diversification.[6] The mutual fund intermediary would be more likely to emerge if the costs involved were directly related to the holding or acquisition of securities. Such costs could not be avoided by using a broker and would require the intermediary to hold securities. This explanation, based on private risk, does not appear to provide any justification for a deposit-taking intermediary.

This simple diversification model forms the basis of explanations of indirect finance advanced by Klein (1973) and Kane and Buser (1979). Klein argues that imperfect security diversifiability is a necessary condition for the emergence of intermediaries to perform this function (p. 930) and notes that "an economic incentive is provided for the pooling of funds and financial intermediaries are the logical instrument for such pooling arrangements." Klein does not specify explicitly the nature of the institutions that would emerge. Kane and Buser (1979) study the degree of diversification among security issues held by U.S. commercial banks. The costs of diversification that are identified include "differences between odd-lot and round-lot trading fees, asset indivisibilities ... administrative costs associated with selecting, evaluating, managing and continually keeping track of a large number of securities ... and information risk" (p. 23). Starting from the Evans and Archer (1968) observation that the majority of diversification benefits can be gained by holding a relatively small number of securities, the authors attempt to explain the number of issues of government securities held by commercial banks. The authors note the argument by Fama, that the number of securities required for diversification may increase

---

5. This explanation for indirect finance also has implications for the organization of the firm. The benefits from diversification could also be gained by horizontal integration.

6. The broker's ability to protect the value of his information is considered later.

substantially in the absence of normally distributed security returns. Kane and Buser emphasize information risk arising from the imprecision of numerical techniques for estimating risks as a source of diversification beyond the degree suggested by Evans and Archer.[7]

Kane and Buser's model of simple diversification can explain only a mutual fund intermediary. This might be the consequence of its restrictive assumptions: i) the only source of risk is in the outcome of investments, ii) investors are homogeneous with respect to their willingness to accept risk, and iii) investors are concerned solely with the outcome after one period. The relaxation of assumptions (i) and (iii) can be usefully considered now.

b) Individual consumption risk

The model of diversification developed above limits uncertainties to the outcomes of investments. Diamond and Dybvig (1983), hereafter referred to as DD, develop an alternative model in which the uncertainty occurs in consumers' demands, extending the model beyond the standard two-period presentation. In this model, investors can be of two types. Type I must consume in period 1 whereas Type II can delay consumption to period 2. In DD's initial model, the proportion of each type is known over the economy, but individuals do not know which type they will be before period one.

Each individual invests a sum of 1 in period 0. The nature of the available technology is such that interruption of production is costly. If production is interrupted in period 1, the individual can consume only C_1 but if the investment continues to period 2 he can consume C_2 ($C_2 > C_1$). The possibility of intermediation arises because, ex ante, the investor would prefer some distribution that gives him consumption $C_1 + u$ in period 1 if he turns out to be Type I and consumption $C_2 - (1+r)u$, where r is the rate of interest, in period 2 if he turns out

---

7. While this explanation of intermediation predicts the emergence of mutual funds, Kane and Buser study commercial banks. Such an emphasis is appropriate because of their attention to the shareholder's concern with diversification. Implicitly, they treat the shareholder in the bank as an investor in a diversified mutual fund.

to be Type II. According to DD, financial institutions emerge as a device through which investors can obtain consumption streams across different contingencies that are preferable to those possible for the investor in isolation. Financial institutions perform this role by pooling the resources of all investors and then by making payments of $C_1 + u$ to Type I investors in period 1 and payments of $C_2 - (1+r)u$ to Type II investors in period 2.

A parallel should be noted between this explanation and the model of simple diversification. In each case the investor can improve on the possibilities available to him in isolation. In one case he avoids the risk inherent in his investment opportunities whereas in the other he avoids the risk inherent in his consumption needs.

Diversification in the DD framework serves a function similar to diversification of investment risks. Investors can avoid the risks specific to individual securities through investment diversification. Similarly, diversification of investors could, at least conceptually, permit them to avoid the risks specific to their consumption demands. The analogy is not perfect, however. Pooling investments to reduce risk does not require any identification of the experience of individual investors. All investors share the same experience. Pooling risks with respect to consumption needs requires identification of individuals according to type so as to determine the appropriate payment. Thus, intermediation of consumption risks involves monitoring problems. It should be noted, however, that private intermediation of consumption risks has developed in areas where monitoring proves to be inexpensive. Life insurance companies provide protection of loss of income as a result of death while disability insurance covers the risk of income loss arising from inability to earn a normal income. In the former case at least, the monitoring appears to be quite easy.

It might appear that DD have created an example in which intermediation arises in the absence of transactions costs. Such an impression is misleading. In the absence of any transactions costs each individual could make a contingent contract with every other individual that would achieve exactly the same outcome as using the DD intermediary.

This contract would specify a payment to be made between individuals in the circumstance that either became a Type I investor. This payment would permit a Type I investor to obtain consumption equal to $C_1 + u$. Under this arrangement a Type II investor has an incentive to make a claim to be Type I because this declaration entitles him to a payment in period 1 and protects his own resources from claim. Private contracts could leave unscrupulous investors with a payment from others and with their own resources intact. The pooling of assets with an intermediary turns the investor's decision into an either/or choice. Declaration as Type I leads to a payment of $C_1 + u$ from the intermediary together with the forfeit of any claim to payment in period 2. As long as the present value of the payment in period 2 exceeds that of the payment in period 1, Type II investors have an incentive to be truthful.

It should be noted that the either/or choice is strictly a result of the assumption in the DD model that Type I and Type II investors consume in different periods. It fails to overcome the problem of designing an intermediary to diversify consumption risks in the standard two-period model. In that model, any consumption risk must reflect different consumption needs in the same period so that the either/or choice is between a large or small payment at the same time and thus an incentive always exists to choose the larger payment. When the payments are separated over time as in the DD model, the size of the payments can be arranged to create self-policing of investors' declarations.[8]

The DD model appears to explain the emergence of a deposit-taking institution when individual risks cancel over the population of investors. The payments over each period are known, as are the number of claimants.[9] Still, the significance of this result should not be overestimated. When DD extend their model to the case where the proportion of Type I and Type II investors is stochastic, the resources that can be

---

8. DD discuss the problem of multiple equilibrium even when the proportions of each type of investor are known. With a correct price of the claims across periods, it is unclear why any other equilibrium than the one described above would ever emerge.

9. DD do not discuss how the proportions of Type I and Type II investors can be known to an intermediary in advance. It would seem to require that (i) the proportions be known for the entire economy and (ii) a single intermediary does business with everyone in the economy.

distributed by the intermediary depend on the number of Type I investors. Any payment to Type I investors that is fixed in advance makes the payment to Type II investors contingent on the realized proportion of Type I investors. In the DD analysis the existence of an intermediary is justified by its ability to provide a desired pattern of consumption across the possibilities for Type I or Type II investors. Consistency with this objective would require a payment to Type I investors contingent on the realized proportions of these investors relative to those in the Type II category. Thus, when the DD model is extended to incorporate a stochastic element, it appears to explain an institution that offers obligations contingent on outcomes rather than a deposit-taking institution that offers fixed-money claims.

1.4.2 Social risk

Up to this point, the only uncertainty analyzed has been uncertainty with respect to individual investments (i.e., over the economy as a whole, the proportion of poor investments is known with certainty). This analysis can now be extended to consider uncertainty with respect to the outcome for the economy as a whole. As a simplification, all investments are assumed to have equal returns given the outcome for the economy so that the only source of uncertainty is aggregate uncertainty.

Homogeneous investors

There would not be any role for intermediaries if all investments were subject to the same risks and investors were homogeneous with respect to their attitudes towards risk. Investors would hold only one security because there are neither any possible gains from diversification nor any benefits from trading risk at the margin.

Heterogeneous investors

If investors were not homogeneous, it would be possible to shift risk among them according to their willingness to accept risk. The assumption

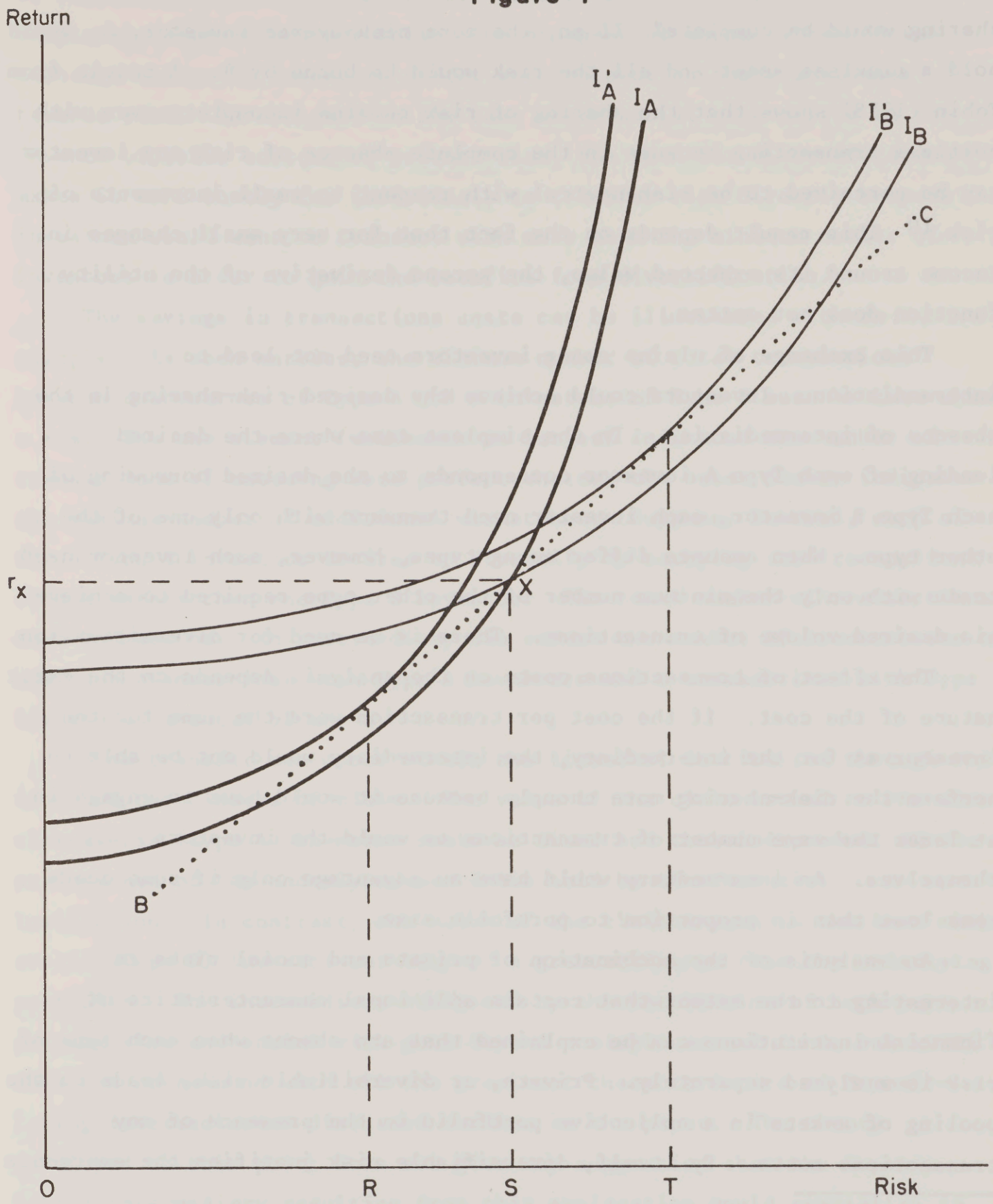
of heterogeneity in attitudes towards risk has been criticized by Klein (1983):

> One should be hesitant to accept risk aversion explanations for contractual terms because these explanations are logically equivalent to relying on tastes to explain behavior (p. 370)... .

However, Gordon (1974) has shown that differences in apparent tastes for risk can arise from differences in resource endowments -- even with identical preferences. In his explanation for implicit labour contracts, he assumes that workers are more risk-averse than employers because their human capital is not as easily marketable as other capital and hence they are less diversified. Investors would want to exchange securities if, with their existing holdings, some investors had different valuations for risk at the margin than others.

This process is illustrated in Figure 1. The risk-return opportunities in the economy are represented by a point X in the risk-return space because the outcomes of all investments are equal given the state of the economy. Implicit in this assumption is the absence of any risk-free asset. At point X, investors A and B are on indifference curves I_A and I_B respectively. As can be seen, their marginal rates of substitution are not equal at X. An opportunity exists for B to issue risk-free securities to A at an interest rate lower than r_x. Whether B could issue risk-free securities would depend on the size of his wealth relative to the worst possible outcome for the risky asset. If B's wealth is inadequate to cover all contingencies, he could issue only low-risk securities to A. As the volume of lending grows, A's exposure to risk decreases whereas B's increases. At the margin, B would require more compensation than before to take on risk whereas A would be less willing to pay to escape risk. An equilibrium is established where A holds a portfolio that consists of both risky assets and the risk-free securities issued by B, whereas B holds risky securities in an amount that exceeds his total wealth and finances them by borrowing from A.

Figure 1



OR- A's holdings of risky assets
RS A's holding of safe assets
OT- B's holdings of risky assets
ST- B's issue of safe assets
BC- Opportunity line with lending and borrowing

How far will this risk-sharing go? Is it possible that the risk-sharing would be complete? If so, the more risk-averse investor, A, would hold a riskless asset and all the risk would be borne by B. A result from Tobin (1958) shows that the sharing of risk remains incomplete even with costless transacting because in the complete absence of risk any investor can be perceived to be risk-neutral with respect to small increments of risk.[10] This result depends on the fact that for very small changes in income around its expected value, the second derivative of the utility function does not matter.

This exchange of claims among investors need not lead to intermediation. Investors could achieve the desired risk-sharing in the absence of intermediaries. In the simplest case where the desired lending of each Type A investor corresponds to the desired borrowing of each Type B investor, each investor need transact with only one of the other type. When amounts differ among types, however, each investor need trade with only the minimum number of the other type required to achieve his desired volume of transactions. There is no need for diversification.

The effect of transactions costs on the analysis depends on the exact nature of the cost. If the cost per transaction were the same for the investor as for the intermediary, the intermediary would not be able to perform the risk-sharing more cheaply because it would have to engage in at least the same number of transactions as would the investors themselves. An intermediary would have an advantage only if some costs rose less than in proportion to portfolio size.

An analysis of the combination of private and social risks is interesting to the extent that certain additional characteristics of financial institutions can be explained that are absent when each type of risk is analyzed separately. Private, or diversifiable risk, leads to the pooling of assets in a collective portfolio in the presence of any transactions costs. By itself, diversifiable risk justifies the emergence

---

10. This same principle has been used by Grossman and Hart (1981) to show that risk-sharing in labour contracts must always be incomplete and by Arrow and Lind (1970) to suggest that the government should behave as if it were risk-neutral with respect to investment projects.

of only mutual fund intermediaries. But some risks remain even with a completely diversified portfolio. The prospect of mutually beneficial exchange arises if investors differ with respect to their attitudes towards risk. In contrast to the case where social risk exists in the absence of other sources of private risk, the need for diversification makes it more costly for individuals to transfer risk directly because investors would want to transact with more than the minimum number of investors in order to gain the benefits from diversification.

The savings in transactions costs can be illustrated with an example. As seen earlier, the minimum number of total transactions between n investors of Type A and m investors of Type B must be less than n + m. The one element (whichever is larger) indicates the number of transactions if matching were perfect whereas the other number indicates the maximum number of additional mismatches. Suppose, instead, each of the n Type A investors desires to diversify by carrying out transactions with m rather than one Type B investor. The total number of transactions then becomes nm. An intermediary can reduce the number of transactions by making n transactions with Type A investors and m transactions with Type B investors, a total of m + n.

Risk transfer can occur within the financial institution through the issue of two classes of claim. The value of the claim issued to the more risk-averse (Type A) investor could be fixed and thus independent of the performance of the underlying portfolio held by the financial institution. In contrast, the value of the claim issued to the less risk-averse (Type B) investor can vary to a greater degree than the underlying portfolio. The intermediary explained by the combination of both pure private risk and social risk goes beyond a simple mutual fund, but still does not correspond with all the prominent characteristics of deposit-taking intermediaries discussed earlier in the paper. The theory developed to this point does not give any suggestion that the deposit-taking intermediary resulting from this explanation would specialize in the holding of non-marketable securities.

1.5 Information Costs

The models considered so far have the common assumptions that the outcome of investments can be ascertained in advance and that all contracts can be enforced, in each case with no resource costs. Moreover, none of these models can explain the phenomenon of deposit-taking institutions which issue liabilities with a fixed-money value and which tend to specialize in the holding of non-marketable securities. This failure is a definite shortcoming because such institutions do exist and, indeed, are a prominent part of our financial system. In this section the theory is extended to include information costs to determine if they contribute to a theory of intermediation that better explains the types of financial institutions that actually exist.

Two aspects of imperfect information are distinguished in the following analysis. First, identification costs: the costs of search and verification arising from imperfect information. In this context, information costs require the identification of investment opportunities that, once identified, give an assured return. The second aspect of imperfect information concerns the outcome to the investor from the time of the investment decision onwards. This aspect forces the investor to use resources to monitor and enforce investment performance to protect or enhance his expected return.

1.5.1 Identification costs

To this point, it has been assumed that investors identify the quality of investment projects without incurring costs. The model can be extended by adding search and verification costs. Search costs are incurred in seeking out investment opportunities. Verification costs are the costs of assessing the eventual productivity of an investment once that investment has been identified.

Search costs differ from verification costs in terms of the transferability of the information gained. Information from the search can be easily transferred because once an investment is found its location can be revealed to other investors. In contrast, verification provides

only private benefits because it is assumed that an individual cannot transfer the results of a verification to another person in a convincing way. The other individual must verify the quality of the investment for himself.

Chan (1983), following Leland and Pyle (1977), builds a model of financial intermediation in which the advantage of indirect finance arises from spreading the search costs among many investors. In isolation each investor would incur the cost of the search for investment opportunities until he finds a productive investment. The intermediary can search among investments and, once a productive investment is found, offer a share of it to other investors. A crucial assumption in this analysis is that investment projects can absorb more resources than are available from the typical investor.[11]

Two difficulties arise from Chan's model. First, the model explains only a limited form of intermediary -- the broker -- because it lacks any reason for the intermediary to take the investment into his own portfolio and to offer claims upon himself to ultimate investors. The intermediary performs only an identification function and once this function has been fulfilled, its role is complete. Second, consideration of the difficulty inherent in keeping information private suggests that Chan is unsuccessful in explaining the persistence of intermediation. The broker's client is in exactly the same position as the broker after he gains the information through search. Under such perfect information, the broker would be unable to realize any return from his information.

Leland and Pyle (1977) suggest that the creation of a true intermediary provides a solution to overcome this appropriability problem.[12] By holding the investment in its own portfolio, the true

---

11. Given the certainty of the investment projects once identified, Chan cannot invoke a diversification motive in his model.

12. Leland and Pyle's solution is only one among several ways of overcoming the problem of appropriability. Other forms of contracts between the broker and the investor that overcome the problem could be considered. This difficulty would not arise if it were possible for the intermediary to pre-empt the investment by committing itself to supply all the funds required for the investment. Such pre-emption means that some of the returns from search can be appropriated by the intermediary. The appropriability problem could also be overcome if the broker's customers were conscious of a continuing relationship with the broker. The awareness of the need to provide a return to the broker might restrain short-run opportunism if the number of customers were quite small. This prospect diminishes as the number of customers increases.

intermediary need not disclose the nature of the investment and can protect its returns from search. As Leland and Pyle recognize, by itself this extension fails to explain completely the emergence of a true intermediary. The customer must be satisfied that the intermediary holds productive investments. Thus, although solving the appropriability problem, the creation of an intermediary generates in its place a credibility problem. Unless investors can be assured that the intermediary holds productive investments, they will not hold the intermediary's liabilities. Simple disclosure of the assets in the intermediary's portfolio to satisfy investors does not solve the problem; it just reintroduces the information appropriability problem that the intermediary is designed to overcome.

The presence of verification costs, as distinct from search costs, alters the analysis considerably. Verification costs could take either of two forms: i) a lump sum charge that is independent of the accuracy of the measurement or ii) a charge that depends to some degree on the fineness of the measurement. The benefits of verification, unlike those of search, cannot easily be shared among investors.

Verification costs explain the existence of financial intermediation to the extent that the intermediary is capable of spreading some of its verification costs among its customers. If the verification costs are independent of the accuracy of measurement, the intermediary cannot offer any advantage over direct finance. To establish the value of the assets he is acquiring by investing in the intermediary, the client would have to duplicate on his own the intermediary's verification process. Indirect finance would just add another level to the verification costs. Indirect finance may, however, benefit investors when verification costs depend on accuracy. If investment through the intermediary permits investors to verify at a lower degree of confidence than they would need if they invested directly, then it would be possible for the total cost of investing through intermediaries to be less than the costs of direct finance.

How can the presence of an intermediary reduce the need for investors to verify the quality of the investment? The intermediary could establish

some device to permit the investors to reduce their effort. One such device would be the commitment by the owner of the intermediary of some of his wealth to the intermediary's portfolio together with the obligation that the investors have a prior claim on the assets. As a consequence, investors would be able to reduce the accuracy of their verification because the value of the assets held by intermediary would exceed the amount necessary to cover their claim. For example, if the owner's wealth subscribed to the intermediary equalled the sum of claims to investors, any investor would need verify only that the value of the assets is not less than 50 per cent of the value reported by the intermediary. The existence of an intermediary would depend on the costs saved through less intensive verification relative to the opportunity cost to the owner of using wealth as a substitute for verification.

The economies arising from savings in verification costs would appear to explain either a mutual fund or a deposit-taking intermediary but not a broker. Use of a broker would not eliminate the need for an investor to verify the investment. On the other hand, a subscription of part of the intermediary owner's wealth into a mutual fund could reduce an investor's need for verification. This subscription would not offer the investor greater protection but, according to Leland and Pyle (1977), it would be an indication to the investor that the owner believed that the investment was productive. Alternatively, the owner could subscribe his own capital to an intermediary and provide investors with a guaranteed claim to overcome the credibility problem.[13]

In summary, the introduction of verification costs does not solve the problem of the appropriability of information even though it does provide an explanation of intermediation. An intermediary that performed search services would still need to establish some device in order to derive the full benefit from the greater efficiency of collective search. Verification, in contrast, is an activity that yields purely private benefits. The need for intermediation arises because some forms

---

13. A further issue, not discussed here, arises from the ability of the entrepreneur to withdraw his capital at the expense of the depositors.

of guarantee can serve as a substitute for verification and, as a consequence, lead to real resource savings.

1.5.2 Monitoring and enforcement[14]

The analysis of search and verification deals with only one dimension of the problems that arise when information has a cost. In this first stage of analysis, projects are sought out and examined with respect to their productivity. But once an investment is verified, other problems arise. Can the investor be assured the funds will be directed towards the agreed-upon purpose? Moreover, once the funds are committed, can the outcome of the investment be assured? In the real world, investment outcome is not independent of the resources used by the suppliers of funds to supervise the investment. Monitoring costs are incurred throughout the life of the investment to ensure that the funds are used for their intended purpose and that any commitment with respect to the use of the complementary inputs is honoured. Enforcement costs must be paid to ensure that borrowers fulfill repayment obligations once the outcome of the investment is known and repayment becomes due. For present purposes, the distinction between these costs does not matter -- the important difference is between those costs that occur after the funds are committed and the search and verification costs that are incurred prior to commitment.

Individual lenders are quite able to monitor and enforce on their own the contracts they make with borrowers. Two conditions must be present before a role for intermediation can emerge. First, the costs of monitoring and enforcing the behaviour of any borrower must rise less than in proportion to the scale of funds. Second, the funds needed by any one borrower must exceed the resources committed by any one lender. The fulfillment of the first condition establishes the presence of economies of scale whereas the second condition makes possible the realization of

---

14. A more elegant version of this approach to the explanation of financial intermediation can be found in Diamond (1984).

the scale economies. The second condition can be achieved either by i) the assumption that the scale of investment projects exceeds the resources held by potential lenders or ii) the assumption that lenders are risk-averse and avoid risk by holding a diversified portfolio.

Consider the alternative ways in which the lender can overcome the problems of monitoring and enforcement. Direct finance may under some circumstances be more expensive than some form of delegated monitoring and enforcement that consolidates this function for a number of lenders under one agent. Still, the consolidation of these functions does not eliminate the problem. Rather, the lenders must be concerned with the behaviour of the agent delegated to monitor and enforce. The delegation of these tasks to either a broker or a mutual fund intermediary is an unsatisfactory solution because the outcome of the investment is borne solely by the lender, despite the actions of the agent. Incentive systems could be devised to overcome these problems but in any system the payment to the agent would have to be contingent in some way upon performance. The deposit-taking institution appears to provide a more efficient solution to monitoring and enforcement problems. The agent, who is delegated with the responsibility for monitoring and enforcement, becomes the residual claimant to income. Thus his returns are directly dependent upon his performance. By this interpretation the deposit-taking institution serves as a device to overcome the problems present in any delegation of monitoring and enforcement to agents.

The question still remains of the credibility of the fixed-value commitment incorporated in the deposit claim. The agent, if he performs his monitoring and enforcement duties, gains all the benefits from greater activity and could be expected to maintain a level of activity at which the additional gains from further effort exactly match the additional costs. The establishment of a deposit-taking institution just shifts the monitoring and enforcement problem away from the ultimate lenders to the level of the agent himself. How do lenders assure themselves that their agent fulfills his commitment to a degree sufficient to assure the fixed value of their claims? The solution is identical to that described earlier to enforce appropriate behaviour on the agent who verifies. The

agent must subscribe sufficient wealth to the institution so that either: i) the added wealth is adequate to protect depositors whatever the degree of supervision that he chooses to perform or ii) that the losses to him from failure to perform his function exceed the benefits from doing so. In effect, the intermediary must post a performance bond.

This explanation of intermediation in terms of monitoring and supervision now explains too much. Indeed all the devices available to intermediaries to reassure ultimate creditors can also be used by ultimate debtors in borrowing directly from ultimate lenders. Thus, monitoring and enforcement costs alone serve to explain only appropriate behaviour of ultimate lenders but not the presence of deposit-taking intermediaries. As will be seen in the subsequent analysis, monitoring and enforcement costs do play an important role in the composite theories that serve to explain deposit intermediaries.

1.5.3 Intermediation and the marketability of securities

The discussion of theories presented above distinguishes among brokers, mutual funds and deposit-taking intermediaries according to their form of organization. As discussed earlier, observation suggests that these institutions differ not only in form but also in terms of the types of business they perform. Brokers and mutual funds appear to concentrate on transactions in marketable securities whereas deposit intermediaries concentrate a large part of their holdings in non-marketable securities.[15] Any explanation of the various forms of intermediaries should be able to explain the apparent specialization of the various intermediaries into different types of activity.

So far no use has been made of the distinction among the types of business associated with each kind of intermediary. These differences suggest that a single theory of intermediation would not be suitable to explain all forms of intermediation and that the differences among types

---

15. This distinction is not entirely clearcut. The most notable form of broker deals in stocks and bonds but some brokers do deal in mortgages. Most mutual funds specialize in holding stocks and bonds but some hold mortgages and, now, some hold real estate.

of intermediaries are differences of substance. In particular, brokers and mutual funds may be found to perform functions different from those performed by deposit-taking intermediaries.

The distinction between marketable and non-marketable securities may provide some useful clues for evaluating various explanations for the existence of financial institutions. It is useful at this point to discuss this distinction, at least in a preliminary way. In a literal sense, the term "marketable" is applied to securities that have gained the approval of regulators for distribution to the general public. In that sense they can be transferred without any legal restrictions with respect to holders.

In economic terms, the distinction between marketable and non-marketable securities is a question of degree. In a trivial sense, any security can always be marketed at some price. The legal sense of the word captures some of the essence of the economic usage. Securities legislation requires borrowers to make detailed information available with respect to financial condition, identity of principals, their commitments and many other factors. The accuracy of this information may be attested so that all lenders do not have to repeat entirely the exercise of determining this information on their own. Still, this public provision of information need not eliminate the need for individual lenders to seek out information. Despite the public release of information, potential holders of marketable securities may investigate many dimensions of the security and its issuer. In contrast, the lender takes on the responsibility of collecting information for non-marketable securities. Some of this information may be the same as required under securities legislation. In some cases, this information may be gained through a continuing business relation with the borrower built up over a long time. The intangible quality of this subjective information may restrict the ability of lenders to transfer it to other parties in any convincing way.

While the idea of the legal responsibility for supplying information is an element of marketability, more than this appears to be involved in the concept. Differentiation among securities would exist even in the absence of securities legislation. The distinction between marketable and

non-marketable securities appears to correspond with the degree to which information required to verify and monitor the value of the investment is publicly supplied by the borrower. While the matter is one of relative emphasis, marketable securities are identified with those for which the borrower supplies the bulk of information required by investors whereas with non-marketable securities, the lender gathers more of the information.[16]

What types of borrowers are likely to convey information to the general public themselves? The size of the borrowing and the quality of the information appear to be important. If the size of a borrower's demand for funds is such that it requires access to the funds of many lenders, the borrower would be more likely to supply much of the information on his own and thus alleviate some of the need for each of the prospective lenders to acquire verification of information. Similarly, only those firms that have a "track record" that permits them to supply tangible information would be able to issue marketable securities.[17]

The issuers of marketable and non-marketable securities may also differ with respect to their need for monitoring. With marketable issues, the distribution among many lenders diffuses the incentive for any one lender to monitor and enforce the performance of the borrower with respect to the terms of the contract. Any lender who chooses to supervise the investment would be able to appropriate only a share of the benefits for himself. The remainder would flow to all investors at large. In this circumstance, lenders must be assured that an effective mechanism exists to enforce performance by the borrowers. As Fama and Jensen (1983a,b) note, this monitoring mechanism may take a variety of forms.

Doubts in the minds of investors with respect to the monitoring and enforcement mechanism may limit the ability of a borrower to issue marketable securities. Instead, the borrower may be constrained to issue

---

16. A parallel can be drawn here with the "sorting" and "signalling" literature on the labour market. Signalling corresponds to the actions of issuers of market securities whereas sorting corresponds to the behaviour of lenders who acquire non-marketable securities.

17. A promoter or underwriter may be able to substitute his reputation for that of an ultimate borrower in some instances.

securities only to lenders who intend to monitor and enforce the terms of the contract on their own. Indeed, in some cases the monitoring and enforcement role required of the lender may be quite minimal. For example, the borrower may have established his reliability with one lender but may be unable to convey this reliability to a wider group of lenders at reasonable cost.

The fact that deposit-taking intermediaries are identified with the holding of non-marketable securities suggests that these intermediaries participate in the monitoring and enforcement function to a greater degree than other intermediaries. Moreover, the fixed money value of deposit liabilities appears to be consistent with the need to create appropriate incentives for agents to carry out effective monitoring and enforcement. In the earlier discussion of these functions, it was noted that each of these activities could be provided for in a contract between ultimate borrowers and ultimate lenders. Such a contract would incorporate both a fixed payment by the ultimate borrower, which would be independent of the investment outcome, and a commitment of wealth by the borrower such that the lender would be assured that any commitment would be met. By itself, the monitoring and enforcement activity explains the form of the contract between lenders and borrowers. While it is conceivable that the ultimate lenders and borrowers could reach the same contractual arrangements on their own as they could through an intermediary, there are other factors that may lead to a preference for an intermediary.

The ability of the ultimate borrower to commit his wealth to the project as a substitute for monitoring and enforcement effort on behalf of lenders depends on the borrower's wealth position and on his willingness to risk his wealth.[18] If the borrower's wealth position is insufficient to assure lenders that they need not spend resources to monitor and enforce the contract, then a role may exist for deposit-taking intermediaries. For example, if some part of the costs of monitoring and

---

18. Any unwillingness of the borrower to commit his own wealth would depend on the presence of risk. Following the approach of Gordon (1974) in the implicit-contract literature, borrowers can be assumed to be risk-averse at the margin because they are undiversified as a consequence of the large proportion of their wealth tied up in their investment project.

enforcement is independent of the number of investment projects that are supervised, intermediaries that supervise a range of investments would have a cost advantage over individual investors.

The introduction of systematic, or undiversifiable, risk together with monitoring and enforcement costs also increases the probability of the emergence of a pure deposit-taking intermediary. Without the need for monitoring and enforcement, the risk-transfer function by itself explains only an institution that passively accepts the return and risk of marketable securities supplied by the market. When the investment requires monitoring and enforcement, the incomplete risk-sharing claim may also require supervision on the part of its holder, especially to the extent that another claim-holder earns a residual claim. In this case, any misspecification of the return on investment by the residual claimant would distribute wealth away from the claim-holder to the residual claimant. The offering of a fixed-claim deposit by the residual claimant may be mutually beneficial to both the depositor and the residual claimant in that the depositor can save the resources required to monitor and enforce the outcome of the investment. This arrangement requires the residual claimant to bear more risk than would be optimal in the absence of transactions costs. The savings by the depositor from the reduced need for supervision together with the value of not bearing any risk must be sufficient to compensate the residual claim-holder for the additional risk he must bear.

The introduction of diversifiable risk also makes the monitoring and enforcement functions more likely to be performed by intermediaries. In order to diversify or reduce their risk, ultimate investors want to supply funds to a variety of borrowers. In isolation, they would incur multiple costs of supervision because they invest in more than the minimum number of projects. An intermediary can, as a consequence, limit the supervision expenses by investing on behalf of a number of investors in multiple projects and then committing its own wealth to a degree that is sufficient to allow the investors to reduce the resources they use for supervision.

1.5.4 Implications of alternative models of intermediation

The literature on the theory of financial intermediation contains more complete development of theories designed to explain the portfolio behaviour of intermediaries than theories explaining the existence of and functions performed by intermediaries. This imbalance poses the danger that theories of portfolio choice developed for intermediaries do not adequately reflect the functions performed by those intermediaries. In some cases the separation of theories explaining behaviour from theories explaining the existence of particular features of the economy may not matter in that the behaviour stands independently from the source of the feature. Tobin (1980), for example, argues that the consequences of changing the money supply can be derived even with incomplete understanding of the reasons for the existence of money. In other cases, the separation of the explanation of behaviour from the explanation of function may come at some cost.

The behaviour of deposit-taking institutions in response to market changes is modeled typically on the basis of the Tobin-Markowitz model of portfolio selection. As Santomero (1984) observes

> ... The Tobin-Markowitz asset portfolio models are used for quantity choice in a perfect market. The results follow directly from the finance literature and add little more. Their insights, and they are many, come from the realization that the bank asset problem is a special case of the standard portfolio choice model. (p. 590)

The choice of this model depends implicitly on the assumption that intermediaries serve mainly the risk-transfer function between investors on the basis of differences in their degree of risk aversion. Even in this context, these models fail to incorporate the transactions costs required to explain the existence of intermediaries. More fundamentally, this modelling fails to capture the essential elements of deposit-taking intermediaries to the extent that their existence depends on advantages in monitoring and enforcement as well as in risk transfer.

Whether the Tobin-Markowitz model is suitable for analyzing the behaviour of intermediaries depends on the degree to which it gives appropriate predictions. The monitoring and enforcement approach suggests that these functions serve to explain both the existence of deposit-taking intermediaries and their specialization relative to others in the holding of non-marketable securities. What difference does the presence of a monitoring and enforcement role make in explaining the behaviour of an intermediary? There are a number of possibilities.

First, much of the literature that tries to explain bank behaviour has little relevance to an institution whose existence purports to be explained by transfer of risk alone. The customer-relationship literature (Hodgman, 1963) suggests that it is mutually advantageous for banks and their customers to make decisions taking into account their continuing relationship. Similarly, literature on credit rationing stresses the informational problem faced by banks and their customers. Neither of these issues appears relevant to a deposit-taking institution that derives its rationale solely from portfolio diversification or risk-transfer motives.

Second, the monitoring and enforcement explanation gives a different interpretation to the problem of bank capital than do arguments based on diversification or risk-shifting. In these latter theories the bank-capital problem can be viewed as part of the optimal portfolio choice made by bank shareholders on the basis of expected returns, risks and their attitudes towards risk. In contrast, the monitoring and enforcement explanations of deposit-taking institutions suggest that bank capital creates confidence among depositors so that they are willing to place their deposits with the institution, permitting it to perform its specialized functions. The difference between the two approaches can be illustrated by the predictions of the consequence of an exogenous shock that increases the riskiness of shareholders' capital in a bank. The portfolio model suggests that shareholders might rearrange their portfolios so as to hold less of the risky bank capital and more of the safe asset. Theories emphasizing monitoring and enforcement would suggest that the bank's owners have to subscribe more capital to assure depositors that their claims could be met.

Finally, the monitoring and enforcement explanation leads to predictions in areas where portfolio explanations would not. For example, financial intermediaries would be expected to be more important relative to total financial market activity in developing, as opposed to developed, economies.[19] In developed economies, more enterprises could be expected to have reached a stage at which they can effectively convey information directly to potential lenders.

1.6 Payments Intermediaries

So far the analysis has neglected one prominent feature of financial institutions that channel funds from ultimate lenders to ultimate borrowers. Many of these institutions also perform the additional and economically important task of supplying the major portion of the payment media in the economy.

This raises a number of questions about the preceding analysis. Does the failure to explain the medium of payment limit the usefulness of the preceding explanations? Or, is the function of supplying the medium of payment an additional function totally independent from an intermediary's investment functions? These questions are explored in this section. The analysis begins with the "storehouse" payments intermediary which serves as a repository for individuals' payments balances and then proceeds to analyze the "transfer" intermediary, the liabilities of which serve as a medium of exchange in themselves.

The storehouse function can be illustrated by reference to the analysis of Orr (1970) who derives a demand for money of the form:

$$M = \frac{4}{3}\left(\frac{3kv^2}{4r}\right)^{1/3}$$

---

19. Financial institutions are only one of the devices that can be used to economize on the costs of information. Horizontal and vertical integration among firms also serve the same function. The above argument also suggests a large role for vertically integrated firms in developing economies.

where M = the firm's average cash balance,
k = the cost of each transfer to and from the investment portfolio,
r = the interest rate or opportunity cost of holding cash balances, and
v^2 = the variance of changes in daily cash.

The economies of scale derived from consolidating the cash balances of several individuals or firms within an intermediary can be shown through an examination of the relationship between v^2 and M. Cash flow variance, v^2, depends on both the size and number of transactions. Orr (1970) discusses a doubling in the size of each transaction and a doubling of the number of transactions:

> ... the model can accommodate a doubling of sales by doubling each separate receipt and expenditure invoice.... The model then would predict ... that the optimal balance should rise by a factor of $2^{2/3}$. As another possibility, a doubling of sales may come about because transactions occur twice as often; that is, t [the number of transactions in a day] is doubled in value, with no change in average invoice. The "law of large numbers" is strongly operative in this event, since there will be additional opportunities for offsetting changes, and the desired balance increases by a factor of only $2^{1/3}$. (p. 65)

The latter case is more relevant for present purposes and suggests strong economies of scale from the consolidation of the money holdings of several individuals at an intermediary. The ability to hold proportionately smaller money balances relative to the volume of transactions permits the intermediary to offer a larger interest return from holding other assets.

The explanation of a "storehouse" payment intermediary is still incomplete in that the costs of using the storehouse have not been specified. Traditionally, the advantage of holding cash balances directly has been expressed in terms of a convenience yield: the avoidance of the costs of switching between securities and money. The use of a storehouse intermediary in turn involves costs in that the individual must go to the intermediary either to withdraw money to make a payment or to deposit money upon gaining a receipt. The presence of these costs means that the individual must again decide on the level of cash balances outside the storehouse relative to those held there. Unless the costs per deposit and

withdrawal at the storehouse are below the costs of transactions between cash and earning assets, the use of the storehouse would be dominated by the direct holding of cash and bonds.[20]

This explanation of the storehouse intermediary undoubtedly accounts for some of the payments functions of banks in a modern economy but it does not account for all. Bank liabilities are also used as a medium of exchange, alleviating the need for depositors to withdraw and deposit funds while present physically at the bank.

A transfer intermediary is a natural progression beyond the storehouse intermediary. The latter permits the realization of economies with respect to cash by combining and consolidating the holdings of individuals that would have been held in isolation. The savings from the consolidation come at the cost that individuals do not have immediate and costless access to their banknotes. The transfer intermediary permits consolidation of individual payment balances at the same time as providing immediate access. Individuals transfer claims to banknotes rather than the notes themselves. Hence, the transfer intermediary saves individuals the costs of gaining access to money balances which would be incurred if only storehouse intermediaries existed.

So far, payments intermediaries have been explained in terms of economies in the consolidation of cash balances and, in the case of transfer intermediaries, the savings of the costs of physical exchange. A prior question, however, must be answered. Under what conditions would an individual give up possession of money holdings in return for a promise either to repay these money holdings in the future on demand or to transfer these money balances to other parties when requested? The depositor, clearly, must believe that these promises will be kept with a sufficiently high degree of probability that the expected losses from failure to fulfill these promises are less than the costs from foregoing

---

20. This explanation of the storehouse may appear to be a violation of the attempt to explain intermediaries under the condition that costs be the same for intermediaries as for individuals. Intermediaries can be assumed to pay the same transactions costs as individuals for transfers between bonds and money. New transactions, the deposit or withdrawal of cash at the storehouse, have been added. For the storehouse to exist, these transactions must be cheaper than transfers between cash and bonds.

the services of payments intermediaries. The confidence problem becomes even greater with the transfer intermediary because, in addition to the transfer intermediary and its customer, third parties are involved. The individual entitled to the payment may not be a customer of the intermediary but must have confidence in the intermediary to accept the transfer of one of its claims in discharge for a payment receivable. To the extent that an intermediary fails to gain the confidence of individuals beyond its range of customers, it must remain a storehouse intermediary and fail to become a transfer intermediary.

How can a payments intermediary gain the confidence of both its customers and third parties with which it deals? Initially it must hold a portfolio of assets that gives assurance to its customers and to third parties that the payments will be made. Here the storehouse differs from the transfer intermediary. The storehouse intermediary offers its customers savings from reductions of their cash balances. These savings are realized only if some of the consolidated cash balances are invested in assets that yield a return. Thus the storehouse intermediary must invest in earning assets. The need to invest is less urgent for the transfer intermediary because the customer benefits from avoiding the cost of the more frequent physical transfer of cash required with the storehouse intermediary. Nevertheless, the transfer intermediary offers the same benefits from consolidation of cash balances as does the storehouse intermediary, so it too can reduce the costs to its customers from cash management by investing some portion of the consolidated cash in earning assets and passing the benefits on to its customers.

This last step brings the storehouse and transfer intermediaries into the scope of the meaning of the term intermediary as it is used in the remainder of this paper. Once these institutions substitute claims for cash on the asset side of their balance sheets, they are serving as intermediaries by lending to ultimate borrowers and borrowing from ultimate lenders.

The need for payments intermediaries to maintain the confidence of their customers and third parties constrains their choice of portfolios. The portfolio held by the intermediary must be such that customers and

third parties can be assured their claims will be met. This assurance can come in either of two ways. Owners can subscribe their own wealth to the intermediary so that the margin between the intermediary's total assets and its outstanding liabilities is substantial. Similar results can be obtained through the composition of the intermediary's portfolio. If it can be demonstrated that the portfolio is held in stable valued securities, then the costs to customers and others in verifying the security of their claims would be reduced.[21] Unless the customers and third parties can assure themselves of the security of their claims at little cost, the savings in transaction costs from payments intermediaries could be totally offset.

The payments explanation differs from those offered earlier in this paper. The source of advantage for the payments intermediary lies not in its ability to bear or manage the risks of assets better than the ultimate lenders could on their own. It need not have any advantages with respect to investing in or dealing with the ultimate borrowers. Instead, the advantage arises from an ability to economize on cash by consolidating the holdings of many individuals, thus avoiding or minimizing the fixed cost in making transactions. In contrast to intermediaries that gain their advantage through asset management, the payments intermediary can be explained by the nature of simple transactions costs alone.

---

21. It is not suprising that the earliest payments intermediaries evolved from goldsmiths. Their trade required them to hold a valuable inventory. Moreover, by its characteristics, the value of gold is relatively easy to ascertain.

Chapter 2
A RATIONALE FOR REGULATION OF FINANCIAL INSTITUTIONS

I now want to examine the rationale for regulation arising from each of the theories of intermediation examined previously. In each case the simplest conditions are postulated for the existence of each type of intermediary considered. The question is then asked whether under these conditions a case can be made for the regulation of financial institutions.

In a sense, the arguments here provide only limited guidance for regulation of financial institutions in the real world. For example, the question of whether mutual funds should be regulated is not addressed. Rather, I ask whether the minimal conditions required to produce mutual funds imply a need for regulation. In reality, mutual funds operate under more complex conditions than the basic hypothetical premises necessary for their existence, and the case for regulation depends on these actual conditions. While the approach here may appear artificial, it does have a strong justification. It identifies the minimal conditions that cause a need for regulation.

2.1 The Efficient Regulation Approach

The approach taken in this chapter can be best described as an "efficient regulation" model. Problems that must be overcome by unregulated intermediaries and their customers are delineated and examined. I then discuss the general forms of regulation that would be mutually beneficial to the institutions and their customers. In Chapter 3 the types of regulation most appropriate for dealing with these apparent problems are reviewed.

Several dangers in this approach should be made explicit. First, it asks only if the regulations in question can be justified in assisting intermediaries and their customers to reach more efficient contracts than they could reach in the absence of regulation. A negative answer may be

quite decisive, assuming that the relevant dimensions of the contract problem have been considered. Regulations that cannot be justified must find their explanation in alternative theories of regulation.[22] More important, however, is the need to recognize the problems arising from a positive answer. This shows only the consistency of the regulation with the contracts that appear mutually beneficial to the intermediaries and their customers. It does not rule out explanation by other approaches.

The second danger, which may be more significant, is that either the nature of the problems to be overcome by lenders and borrowers or the way in which a regulation overcomes a problem have been misunderstood through a lack of insight during this investigation. The solution for this shortcoming must be vigilance on the part of the reader to avoid too easy acceptance of the arguments that follow.

2.2 Regulation and the Mutual Fund Intermediary

Minimal conditions for the existence of mutual fund intermediaries require the presence of purely private risk together with transactions costs involved in the acquisition of claims issued by borrowers. Neither social risk nor information costs are needed to explain the mutual fund intermediary. The transactions costs that prevent individual investors from acquiring a diversified portfolio could arise from the record keeping surrounding the existence of the debt.

In the simplest model explaining the existence of mutual fund intermediaries, there is no need for regulation. Investors would be able to choose the mutual fund whose holdings correspond most closely to their tastes without incurring costs. Investors could effectively monitor and enforce the investment activities of the mutual fund without cost. Finally, the return from the the intermediary's portfolio could be ascertained easily so that investors would be assured of appropriate payment, given the fund's investment performance.

---

22. Other theories that might then be used include the captive theory of Stigler (1971) and the taxation by regulation theory of Posner (1971).

2.3 Regulation of Risk-Transfer Intermediaries

Contingent payments intermediaries can arise in two circumstances -- each with some form of transactions cost: i) heterogeneous investors with uncertain investment outcome at the individual level or ii) uncertainty of investments at the aggregate level. In either case, the intermediaries provide an opportunity for mutually beneficial exchange among investors by issuing two types of liabilities. One, which corresponds closely to shareholders' capital in banks and which will be called equity, offers a return related to, but more volatile than, the underlying portfolio held by the intermediary. The other, deposits, corresponds with the deposit liabilities of intermediaries. The existence of these two classes of assets serves as a device whereby risk-averse investors can transfer risk from themselves to less risk-averse investors. The most risk-averse investors hold the safe deposits whereas the less risk-averse hold bank equity.

The conditions of the model that lead to risk transfer among customers of the financial institution fail to provide an argument for regulation of the institution. In the absence of any information costs, depositors can determine in advance the nature of the risks to which they will be subject by examining the portfolio held by the intermediary. Similarly, at the end of the investment period, the outcome of the intermediary's investments is known and the appropriate payment can be enforced without cost.

It might be argued that the depositor should be protected in this instance from the intermediary choosing to hold an excessively risky portfolio. Such a possibility is not relevant initially because with the minimal conditions required to explain this intermediary the investor can be fully informed without any cost of the portfolio held by the intermediary. Such a problem arises only if the depositor is locked into the intermediary over a period in which it is possible for the intermediary to alter its behaviour. This case captures the spirit of the subsequent analysis where enforcement of contracts becomes relevant.

2.4 Regulation of Deposit-Taking Institutions

Information, monitoring and enforcement costs are required to explain the existence of the type of deposit-taking institutions found in a typical developed economy. The deposit-taking intermediary exists because the costs of collecting information, monitoring portfolios, and enforcing payments are less when a centralized institution -- the intermediary -- is used. The intermediary does not, however, eliminate all the problems faced by the ultimate lenders, problems that arise from information, monitoring, and enforcement costs. Rather, ultimate lenders must be concerned with these problems with respect to the intermediary in its role as their agent.

2.5 The Unregulated Intermediary

We can start analyzing the role of regulation by examining the way in which financial intermediaries might operate in the absence of regulation. Two questions are relevant. What problems would financial intermediaries and their customers have to overcome? What institutional forms would permit them to overcome these problems?

Financial intermediaries, like any firm or other party to contracts that extend over time, must establish confidence in the minds of suppliers of funds that repayment will be made according to agreed-upon terms. The ultimate lender tries to avoid the dangers of misappropriation of funds on the part of the borrower.

Misappropriation can arise from two separate sources: the disposal of assets and the diversion of cash flow. The opportunity for misappropriation through disposal of assets depends largely on the characteristics of the assets involved. The more specific the asset is to the activity in question, the less opportunity there will be for misappropriation by the agent. Specific capital has a narrower market than general capital and, as a consequence, the opportunity for disposal of the asset would be more limited. The opportunity for misappropriation of cash flow depends primarily on the time span over which the

diversion of funds occurs because the probability of detection can be expected to increase with time. Therefore, misappropriation could be expected to be a greater problem for activities involving a large cash flow per unit of time.

Non-financial firms borrow from the general public, and from financial institutions to finance both working capital and fixed capital. Although such loans are, in a sense, backed by "potential earnings", the existence of real capital instills some confidence in lenders that the loan will be repaid or at least gives some indication of the realizable value should the loan go into default. The capital of the owner of the firm serves two purposes. First, it indicates how much he stands to lose through mismanagement before the loan goes into default. Second, it gives the ultimate lender a margin of valuation of capital above the amount loaned. In other words, the lender needs to spend fewer resources in order to assure the value of his capital.

The pledge of property by any firm protects its creditors against the consequences of both mismanagement and misappropriation, but the degree of protection will depend on the circumstances. The value of the capital in its alternative uses outside the firm protects against mismanagement in that if the firm defaults on promised payments the assets can be sold by the creditors to recover part of their claims. The extent of recovery depends on the degree of specificity of the capital. If the capital is general in its uses, its value outside the firm will be close to its value within the firm, and for a given cost of the capital the creditor will have substantial protection. If the capital is specific to uses in the firm, its value outside will be below its value within the firm and the degree of protection will accordingly be less. The capital also provides protection against misappropriation to the extent that the assets can be observed and easily measured.

Financial firms face the same problems as non-financial firms but with a number of important differences. A non-financial firm borrows funds to acquire assets in order to carry out some real activity. Similarly, a financial institution could be characterized as borrowing funds to carry out the activity of lending. The parallel breaks down

however, because the business of lending, unlike real activity, does not generally require the holding of real assets. A financial institution holds a set of claims against others.

The lender to an intermediary faces a different problem than a lender to a non-financial firm. The assets held by the intermediary are themselves claims and any borrower would have to evaluate the underlying assets on which the financial institution lends funds. This alternative, however, is self-defeating. Financial institutions are explained in the presence of information, monitoring, and enforcement costs by precisely the savings of resources that arise from the delegation of these functions to an agent.

The opportunities for misappropriation are greater for a financial institution than for a non-financial firm because of the difference between the non-financial firm's holdings of real capital and the financial firms' holdings of financial claims. Holdings of real capital limit the options available to the principals of the non-financial firm to misappropriate the lender's funds. One opportunity available to the owner of the non-financial firm is the sale of the real capital to a third party. Yet this possibility could be ruled out if the claim of the lender restricted the sale. Indeed, property law has evolved in such a way that the buyer of capital goods or property might be unable to escape the lenders' claims. In this case, the seller could only realize his equity position in the asset. In contrast, a financial institution holds a portfolio of claims on others and serves as an agent for ultimate lenders by holding a portfolio on their behalf in order to economize on information, monitoring, and enforcement costs. Yet it is this delegation of portfolio-holding to an agent that creates the problem of credibility. To be assured of the security of his funds, the lender would have to monitor the initial use of the funds together with the use of any funds repaid. Moreover, the monitoring of claims on capital may be more expensive than the monitoring of the capital itself.

Klein (1974) and subsequently Klein and Leffler (1981) suggest a solution for the problem of establishing confidence in a financial firm that is similar to that postulated for the non-financial firm.

The financial institution can acquire real capital to create credibility in its undertaking to repay its borrowings. This explanation gains some credibility from historical experience. Early bankers generally were established in other lines of commerce and used the capital accumulated for their other activities as security for their lending business.

A bank starting de novo faces a different problem than one that starts in association with an established enterprise. The amount of capital needed directly as a consequence of the mechanical processes involved in intermediation is itself likely to be small. Klein and Leffler argue that capital required to establish confidence in the intermediary would have to take a form such as elaborate buildings that would have little use for any function other than intermediation. The amount of capital required in such a form to establish credibility reduces the resources the intermediary has to carry out its functions.

The established intermediary may not face the same problem. Its capital can take the form of its value as an established and trusted financial institution. As long as the present value of the gains from misappropriation fall short of the capitalized future returns from its activity of intermediation, the intermediary will have an incentive to honour its claims. Thus the role of capital in an established financial institution is very different from that in a non-financial firm. Much of the value of an established financial institution consists of the trust built up over time through its function as a financial institution.

2.6 Alternative Forms of Regulation

The problems facing intermediaries and their customers in reaching efficient contracts that maximize the potential gains from trade provide the starting point for a framework for the analysis of regulation. "Efficient" regulation could take a variety of forms and would enable intermediaries and their customers to reach a wider range of transactions than they could in its absence. A public regulator could substitute its own confirmation services for the decentralized collection of information by customers of intermediaries. It could also establish a set of standardized contracts so as to reduce the costs to customers in

determining the set of rights they have obtained through their transactions with an intermediary.[23] Finally, efficient regulation might provide a more effective monitoring mechanism than would be possible through the uncoordinated efforts of individual customers.

The earlier analysis suggests that deposit-taking institutions manage risk by monitoring and enforcing the behaviour of ultimate borrowers. From this perspective, this type of financial institution is little different than other firms in that they offer their customers a service. While it might be argued that the service offered by financial institutions involves a contract over time, financial institutions are not unique in this respect -- any warranty is a contract over time. Moreover, all financing of economic activity requires contracts over time. It is apparent, however, that the activities of financial institutions appear to be regulated in a very different manner than those of many other corporations that offer customers a service.

Economists generally accept that a role exists for government to establish a legal framework for the establishment and enforcement of contracts. For example, in discussing the role of government in monetary arrangements, Milton Friedman (1960) states:

> What is involved is essentially the enforcement of contracts, if the failure of an issuer to fulfill his promise is in good faith, or the prevention of fraud, essentially of counterfeiting, if it is not. Both are functions that most liberals would wish the state to undertake. It so happens that the contracts in question are peculiarly difficult to enforce and fraud peculiarly difficult to prevent. (p.6)

The actual framework governing financial institutions differs from that for many other businesses in one essential way. Much of the other regulation of contracts and corporate behaviour is <u>remedial</u> in that it specifies the forms of recourse that are available to one party after the other party to the contract fails to meet its terms. In contrast, the

---

23. These contracts could economize on transactions costs if, for example, they were overriding contracts in the sense that they would govern all transactions of a particular type. Parties to such transactions would need only to establish that the transaction fits the type and then they could disregard the "small print".

legal framework surrounding financial institutions can be characterized as preventive in that it tries to prevent the failure of the other party to meet the terms of the contract either by limiting the range of permissible activities for any class of financial institution or by specifying prohibited activities.[24,25]

The choice between preventive regulation and remedial regulation must be faced whenever governments establish a framework to regulate any form of economic activity. Moreover, the government's choice is not the same across different forms of economic activity. What factors serve to shape the government's choice among alternative forms of regulation? For the present, we will assume that the government chooses the form of contract that is efficient for the market participants. In other words, the contract would have a form that minimizes the combined costs of doing business for both parties to the transaction. These costs include the expense of establishing the contract, of monitoring the degree of its observance, and of enforcing it, and also any costs that result from inability to constrain behaviour fully to the terms of the contract.

Form of Capital The choice between preventive and remedial regulation will, for the protection of creditors, be influenced by the form of the capital held by the parties to the contract. This capital provides the wronged parties with an opportunity for recourse by permitting them to take possession of the capital and realize its value on the market as a way of assuring compensation according to the terms of the contract. The form of capital is crucial to the wronged party in terms of its ability to serve as compensation.[26]

---

24. It should be noted that recourse provisions have also been used in the regulation of Canadian financial institutions. At one time all shareholders of a bank were liable for double indemnity. In the case of the bank's failure, they would lose the value of their shares and be further liable to bank creditors to the extent of the par value of their shares.

25. An argument similar to the one used in this section has been developed by Shavell (1984a,b) with respect to the use of tort liability and regulation for controlling activities that create risks of harm to others.

26. Shavell (1984a) discusses the role of the quantity of wealth available to meet a liability claim. The emphasis on form of capital here reflects the different problem under consideration. Shavell discusses liabilities in which the harm may not have any relationship to a prior transaction between the parties. For financial intermediation, initially at least, the assets of the intermediary must exceed its claims to the extent there is any owners' equity. The form of the assets, as a consequence, becomes important.

The importance of the difference among types of capital can be illustrated by two extreme examples. Consider first a type of capital that is general in its usage in that, even though its current use may be marginally superior to any other use, its value in alternate uses is very close to that in its current use. At the other extreme, consider a unit of capital that is specific in that it does not have any value outside its current use.[27] Recourse may be an appropriate way to protect the creditors of firms holding general capital. They can be assured with some certainty that they will recover at least the value of the capital regardless of the success of the firm. In contrast, the creditors of firms holding specific capital cannot depend upon the firm's assets as a remedy for failure.

What characteristics determine the value of a firm's capital outside its current use? In general, the more specific a capital good is to its current use, the less will be its value in other uses. The concept of specificity in itself does little to advance our understanding of the problem. Rather, we must consider the sources of specificity. One obvious meaning of specificity refers to physical adaptability to other tasks. A screwdriver may be valuable in many tasks other than its current use. In contrast, a sophisticated bottling machine may be useful only in bottling a particular type of beer. Although it may be adaptable to bottling soft drinks or other types of beer, many of its special attributes would lose their value in these other uses.

A less obvious meaning of specificity refers to capital that is strongly identified with a particular individual who may have invested time and effort in determining the idiosyncrasies of the particular capital good. Operation of the capital good by this individual gives a higher stream of income than operation by someone else. Thus, a change in the identity of the user of a capital good may alter its value substantially.

---

27. It should be noted that this difference in the value of the capital occurs only when the firm is in failure. When the firm continues as a going concern, the value of the capital is the same in both cases. Benjamin (1978) argues that the value of the collateral to the creditor may be immaterial if the loss of the collateral imposes costs on the debtor. In the case of the failure of financial institutions, the owners' equity has already been eliminated and limited liability prevents any further losses.

Transactions Costs The institutional framework for any economic activity will also be influenced by the scale of transactions costs that would be incurred under the remedial approach relative to those incurred under preventive regulation. Two types of transactions costs can be distinguished that are relevant to the choice between remedial and preventive regulation. The first are costs incurred in establishing the contractual relationship between the parties to the transaction. The second are those costs that arise from the failure of one of the parties to the contract to satisfy its terms. These two sources of transactions costs are sufficiently different that each is considered separately.

Costs of Contracting The costs of reaching a contract can arise from a wide variety of sources such as the costs of finding appropriate matches to the two sides of the transaction, the costs of establishing agreement, and the costs of documenting the terms of the transaction in a way that is acceptable to both parties. If there is reliance on liability the importance of these costs will depend on the nature of the relationship between the parties. These costs will be unimportant where no contractual relationship exists between the parties (for example, in the case of an automobile accident) but will be more substantial where the prospect of harm is perceived to arise within a contractual relation.

Similiar costs arise with regulation. Shavell (1984) notes that

> the administrative costs of regulation include the public expense of maintaining the regulatory establishment and the private costs of compliance. (p. 364)

The transactions costs of government regulation relative to the costs of reliance on liability depend on the number of parties to the contract and their turnover. Both large numbers of parties and turnover in these parties can be accommodated at relatively low expense in some circumstances. Where the transaction poses few complications, where the needs of the contracting parties are relatively homogeneous, and where the contracting parties are well informed about the performance of the other parties to the contract, standard contracts in the context of common law may be relatively efficient. For example, companies that publish magazines manage contracts with large numbers of subscribers. Few problems appear to arise from the turnover among subscribers. With

homogeneous subscribers, there is little need to alter the product. New subscribers are likely to be well informed about the nature of the magazine.

Turnover of parties to a contract would appear to be costly if the needs of different parties were heterogeneous. Each time one party to a contract is replaced by another, the contract would have to be redefined to reflect the particular interests of the new party. When new parties were poorly informed with respect to the terms of the contract and the qualities of the parties on the other side, they would have to invest resources to gain enough information to make them willing to enter into the contract. Some of these costs could be alleviated with use of public regulation in place of reliance on common law. For example, a prescribed set of standards for these contracts could alleviate the costs of reaching these contracts.

Costs of Remedy The costs of using the tort system once harm has occurred are described by Shavell (1984) to be

> broadly defined to include the time, effort, and legal expenses borne by both parties in the course of litigation or in coming to settlements, as well as the public expenses of conducting trials, employing judges, empaneling juries, and the like. (p.364)

Some of the factors affecting the costs of remedy under common law are similar to those influencing the cost of reaching a contract. For a large number of creditors the costs of coordinating the establishment of a need for remedy and the implementation of a remedy would be high. The form of government regulation may also be influenced by the scale of transactions costs likely to be incurred under the remedial approach.

Use of remedy is less likely when there are many creditors, each with a small stake. A large number of creditors means that the costs of coordinating the establishment of need for a remedy and the implementation of the remedy would be high. Moreover, when each creditor has only a small stake in the overall value of the remedy, he will commit only a small amount of resources to assure its realization. In effect, the combination of many creditors with small amounts at stake turns the remedial process into a public good. In these circumstances, the threat

of a remedy being invoked and the diligence with which it is pursued could be expected to diminish as the number of creditors rises and their average stake falls.

The advantage of the preventive approach with respect to transactions costs would tend to depend directly on the number of creditors even when the state acts as an agent for the creditors. The preventive approach requires the state to supervise debtors to assure their conformity to a range of permitted activity. In contrast, under remedial regulation the state would be required to invoke the remedies with respect to the debtor and then to evaluate the legitimacy of the creditors' claims. This latter aspect of the remedial approach obviously becomes more costly as the number of creditors increases.

2.7 Regulation of the Financial Industry

How do the characteristics of the financial industry correspond to the qualities appropriate to either of these forms of regulation? First, the financial industry has highly specific capital in two regards. The ability of a financial institution to attract customers depends on its ability to establish trust or confidence that its claims will be repaid. Klein (1974) argues that this confidence can be achieved by the establishment of "brand name" capital. An intermediary will be discouraged from dissipating its "brand name" capital for the purpose of short-run gains if the capitalized return exceeds the benefits from short-run opportunism. Brand-name capital represents the epitome of specific capital: once a financial institution fails, its brand name becomes worthless. Thus, brand name capital is an extreme form of specific capital.

The theory of deposit-taking institutions also suggests a further respect in which their capital would be specific. Deposit-taking institutions have been identified as monitors and enforcers of loan contracts on behalf of ultimate lenders. This function of managing risk is performed by acquiring non-marketable securities for which the institution takes the responsibility for screening information about the borrower. The value of these assets is specific to the deposit-taking institution who has verified the expected returns from the project, who

has gained the information required to monitor the projects and who understands the problems with respect to enforcement. These dimensions of the customer relationship have to be built up with experience over time. The value of these claims would be less for an outside party who has not gained the knowledge embodied in the customer relationship.

In addition to the specific nature of the capital used in intermediation, another aspect of intermediation also appears to favour the use of preventive regulation. Intermediaries typically have many customers, each with rather small amounts on deposit relative to the total. Under these conditions, the transactions costs in coordinating recourse for customers appear to be substantial and the incentive for any one depositor to commit resources to increase the probability of a remedy is also slight.

A final determinant of the regulatory approach arises from the costs of the information required by customers. The remedial approach requires the creditor to identify the conditions under which invocation of the remedy may have legal cause. The monitoring and enforcement aspects make it difficult to determine this condition without the collection of various types of specialized information, the delegation of which to an agent provides one source of benefit from intermediation. On the other hand, preventive regulation may be less costly -- certain types of activity, once identified, can provide clear evidence of breach of contract without the need to determine bankruptcy.

Summary Two dimensions of economic activity -- specificity of capital and transactions costs -- are identified as influencing the choice between a preventive approach and a remedial approach to the regulation of business activity. Aspects of these factors should serve to explain the approach taken to any type of economic activity. Three, at least, appear instructive for understanding the particular approach used for financial institutions. Intermediation uses forms of capital specific to that activity, it involves a large number of creditors relative to debtors and it faces frequent turnover of both debtors and creditors. These aspects suggest that preventive regulation would be more appropriate for financial institutions.

2.8 Regulation of Payments Intermediaries

The final type of intermediary examined consists of payments intermediaries that exist by reason of their ability to allow customers to economize on their money holdings. Payments intermediaries invest in income-earning assets and provide a return to their customers that they would not have earned had they held the money balances themselves. Many of the arguments justifying regulation of financial institutions explicitly incorporate the payments functions of intermediaries. One group of arguments stresses the possibility of so-called "contagion" effects. Even though contagion effects may not be confined to payments intermediaries, the emphasis in the literature ties the two together. This convention will be observed in the following discussion.

So far the analysis of regulation has focused on the relationship between financial institutions and their customers. Many arguments for regulation go beyond this and stress interdependencies among customers of financial institutions or among the financial institutions themselves. These interdependencies are supposedly the basis of contagion effects. The literature suggests that these contagion effects provide a major justification for the regulations under which banks operate.

Contagion is defined in the Oxford English Dictionary as:

> The communication of disease from body to body by contact direct or indirect

or alternatively, as

> the contagious or "catching" influence or operation of example, sympathy and the like.

Its use in economics appears to be consistent with these meanings. Contagion arises when it is revealed that the quality of the liabilities of a financial institution, or more particularly a bank, falls below the expectations held by its customers. This triggers a response that causes depositors to reassess, and presumably reduce, their opinion

of other financial institutions with which they deal. Contagion occurs when this lower valuation causes them to withdraw their funds from the other financial institutions. This possibility of contagion is an oft-cited reason for the regulation of financial institutions. In this section, financial institutions are compared with automobile manufacturers to determine whether the prospect of contagion provides a qualitatively different type of argument for the regulation of financial institutions than for the regulation of automobile manufacturers.

For the financial institutions, the event causing contagion is a reduced prospect that a claim on the institution will be repaid at full value in the future. For the automobile manufacturer, the comparable event is the increased prospect that the repair expense for the manufacturer's product will be greater at future dates than previously expected. These events are roughly parallel. In the financial institution, a deterioration occurs in the services provided by bank deposits in the future. The depositor could remedy the problem by discontinuing the purchase of any services of bank deposits. Similarly, a deterioration occurs in the future services provided by an automobile to the consumer. The customer can remedy this problem by not buying the car.

This section proceeds by examining a variety of possible meanings of contagion to determine whether some types of contagion are unique to, or at least different by a scale of magnitude for, financial institutions. For simplicity, the arguments are kept separate and can be classified into the following themes:

i) interdependence of demands for services of different producers,
ii) interdependence of demands of customers of the same producer,
iii) role in payments system.

2.8.1 Interdependence of demands for services of different producers

How do the effects of changed expectations of quality differ for a financial institution compared to a manufacturer of consumer durables? More specifically does the demand for the product of one supplier of financial services affect the demand for the product of another supplier of financial services in a way different than for other products.

A decrease in the quality of some product or service with its price unchanged can be considered to be identical to an increase in price. Given an unchanged quality and price for alternative products, customers can be expected to switch from the more expensive producer to other producers. Economic theory provides little guidance with respect to the range of substitutions that consumers make. Other financial institutions might benefit from the demand displaced from a given financial institution but they may not be the only beneficiaries. Customers might also switch to holding currency, direct financial instruments such as stocks or bonds, or even real assets. Similarly, the decreased demand for a particular brand of automobile will also be reflected in a variety of ways. The most obvious is an increased demand for other brands of automobiles but other possibilities include an increased demand for bicycles, trucks, or public transportation.

The effects just discussed are no more than the usual substitution effects in response to a change in price. If this interdependence among consumers reflects contagion, then it is a broader but less significant phenomenon than generally claimed. It is broader in that every event that alters demand produces some contagion. It is less significant in that recognition of this interdependence is well known but has never been suggested as a rationale for the regulation of markets.[28]

Demands for related products may be dependent in a further way. Consumers may classify together the qualities of the output of different producers of similar commodities. For example, the discovery of a deterioration of quality of one financial institution or one make of automobile may cause consumers to alter their judgement of the quality of the deposits of other financial institutions and the quality of other makes of automobiles.[29] This interdependence of customer demands also appears to fit the dictionary meaning of contagion.

---

28. This statement may not be wholly accurate. Disasters such as floods or famines which create excess demand at current prices often lead to a call for non-price rationing.

29. Certainly in the past consumers appear to have perceived a poorer quality of North American cars relative to Japanese and German cars and only to a lesser degree, the poorer quality of particular makes.

An argument for regulation based on this concept of contagion leads to a number of problems. Automobiles or any other goods appear to be no different than deposits at financial institutions with respect to this argument. If the argument does produce a case for regulation, the question arises as to the form the regulation should take. What standards of quality should be enforced for financial institutions? Should the prospect of the failure of a financial institution to honour its outstanding claims be less than 10 per cent, 1 per cent, or any other arbitrary figure? Or alternatively, should commodities be placed into groups in which homogeneous quality is enforced?

A more sophisticated variant of this argument about interdependence among producers suggests that the business of intermediation depends on confidence and that confidence in the financial industry as a whole is a public good. This argument, however, is just a variant of the preceding one. The public good aspect of confidence means that the actions of any one producer of financial services affect the confidence that consumers place in other producers. Similar public good aspects prevail with respect to automobile manufacturers. On theoretical grounds there does not appear to be any difference in kind with respect to public good characteristics. Any difference would have to arise through the empirical importance of the public good effect.

The implications of regulation based on this argument would be quite profound. It suggests that any case where the demands for different goods are interdependent, this interdependence forms the basis for an argument for regulation that would limit the ability of producers to redefine the characteristics of their product. Nevertheless, redefinition of products may be beneficial for both producers and consumers. Disposable syringes, for example, might initially have been regarded as a deterioration in quality from the glass syringe. But the accompanying decrease in expense apparently more than compensated for this deterioration.

To say that this argument for regulation is not particularly strong, is not to suggest such regulation does not occur. Minimum standards of quality are enforced for a wide variety of products and services. Most notably, it is impossible in Canada to obtain the services of a doctor who

has not graduated from medical school and served a prescribed period as an intern.

Despite the similarity in public confidence effects for bankers and automobile manufacturers, there may be differences in the consequences of changes in confidence or perceived quality in the two cases. In particular, are there differences in the relationship between producer and consumer that alter the significance of the public-confidence argument? When an automobile company decreases the quality of its cars, prospective buyers of new cars have the opportunity to find out that the quality of these cars has decreased. Those that find out would be less willing to purchase the cars, but if they do so they will be aware of the new level of quality. In contrast, customers of the financial institution face a different situation. As in the case of the automobile manufacturers, new customers can avoid the product with reduced quality or, at least, be aware of the reduced quality. The major difference arises with respect to "old" customers.

A customer can purchase an automobile and continue to receive services from that automobile with little concern for any continuing relationship with the manufacturer. Financial claims are different. Customers can receive the services of financial claims only if they maintain a continuing relationship with the financial institution. The service, and hence the value of any financial claim, depends on the actions of the financial institution over the entire period in which the customer holds the claim. Thus, a difference arises because of the nature of the continuing relationship required to gain the use of deposit services.

2.8.2 Interdependence among customer demands

A second class of argument with respect to contagion suggests that it arises from an interdependence among demands of customers. At the simplest level, this interdependence could mean that the probability that one customer revises his expectation of quality increases the probability that other customers make similar revisions.

Information can be easily transferred from one person to another and, with few exceptions, individuals have little incentive to keep information

to themselves. Much evidence suggests that individuals form their perceptions of market conditions on the basis of information gained from friends and acquaintances.

As the discussion to this point suggests, interdependence among customers is hardly unique to financial institutions. The fact that one person's choice of financial institution affects the choices of others is not any different from the fact that one person's choice of automobile affects another's. In each case, the consequence of this interdependence of demands is that the perception of changed quality by one individual may "spread" to the demands of others. While such an effect certainly fits the description of contagion, it does not in itself appear to justify any form of regulation. Producers are undoubtedly conscious that poor service to one customer may jeopardize the business of other customers.

It may be suggested that the simple argument about interdependence of demands fails to capture an essential feature of the process of financial intermediation. The demands of customers may affect the actual as well as the perceived quality of the product that can be offered to other customers. In other words, financial institutions are different in that not only are demands of customers interrelated but changes in the demands of one customer alter the quality of the product available to other customers.

This argument depends on the specific factors that determine the ability of a financial institution to offer a specified quality of deposit to any customer. The probability that a customer's deposit will be redeemed on the agreed-upon terms depends on the behaviour of other customers unless the financial institution holds a portfolio of perfectly safe assets. Each depositor has some probability of withdrawing funds from the financial institituion and the probability that the institution can meet its claims can be derived from a cumulation of the individual distributions. Any change by a customer in his judgement of the quality of his claim alters the probability that he may make a withdrawal at any time. Thus, a shift in demand by some customers not only alters the other customers' perceptions of the quality of their claims but also alters the actual quality of their claim.

As before, it is useful to inquire whether this source of interdependence of demand is any different for financial institutions than for other producers. Here some difference in kind appears to be evident. The quality of automobiles that can be offered to one customer appears to be in general independent of the demands from other customers. There are parallels, however. If the general demand for the type of car that someone owns drops sharply, that individual will be concerned about both the continued existence of the producer and also about the future resale value of his car. If the producer goes out of business, the quality of his car may deteriorate because of a lack of availability of spare parts or service facilities. Thus, the argument about the dependence of quality on the state of demand does not appear unique to financial institutions, though it may be more important for them empirically.

2.8.3 Role of the payments system

In any modern economy the major means of payment consist of the transfer of claims held at financial institutions. Consequently, individuals hold their transactions balances as deposits at financial institutions. A totally distinct argument from those considered to this point suggests that banks should be regulated because of their unique role in the payments system. The particular feature of the payments system that forms the basis for this argument is its role as a "network utility" in which the benefits to each customer from the existence of the system depend on the number of other customers who use the system. Just as the sole subscriber to a telephone system would not gain any benefit from the system, the only person to use a payments system does not gain any benefit from exchange. To a point, additional customers in a network utility increase the benefit to other customers. Similarly, if customers withdraw from a network utility, they may impose costs on others as well as on themselves.[30]

---

30. It should not be assumed that the addition of customers to network utilities will always add to the benefits of existing customers. The addition of a computerized canvasser to a telephone system may reduce rather than increase the benefits to other customers from the system. A counterpart of this effect for the payments system might be the addition of a customer whose cheques are returned regularly because of insufficient funds.

The existence of a network utility in itself does not create any apparent rationale for regulation. If the network utility were operated by a single enterprise, all the external effects among customers would remain internal to the supplier of the network services. The operator of this utility could take the dependence among customers' demands into account in determining pricing policy. This dependence would not, however, be captured if the network system were offered in parts by different producers. In such a circumstance any producer of the network services would take into account only the effects on his customers of adding another customer to the system.

In terms of the present analysis of the payments mechanism, any deterioration in the quality of the deposits at any one bank has effects that spread beyond that bank and its customers. It reduces the usefulness of deposits offered by other banks as a medium of exchange. This network-utility argument is not new but is essentially an externality argument in a different guise. Milton Friedman (1960), for example, cited as one reason for government intervention with respect to money:

> the pervasive character of the monetary nexus means that the failure of an issuer to fulfill his promises to pay has important effects on persons other than either the issuer or those who entered into a contract with him in the first instance or those who hold his promises. (p.6)

Similarly, James Tobin (1980) argues:

> Another time-honoured observation of monetary economists is the analogy of money and language. Both are means of communication. The use of a particular language or a particular money by one individual increases its value to other actual or potential users. (p.86)

To summarize to this point, the network-utility argument suggests that individuals who use money benefit from the number of other people who also use money as a medium of exchange. The analogy can be drawn between the payments system and a telephone system. In the extreme, neither money nor a telephone is any use to an individual if that person is the only one using it.

As appealing as it may seem, it may be useful to examine the network-utility argument more carefully. Why, for example, should anyone care who else uses the payments system? To answer this question, first consider an economy in which there is a large number of buyers and sellers and in which market participants are informed costlessly of the array of trading opportunities available to them. Under these circumstances any individual does not care whether his trading partner uses money as long as some individual trading the same goods does use money. Trade with non-money users is more costly to money users than it would be if the money users had been able to trade with another money user. Still, all money users have that opportunity and will trade with a non-money user only as long as the trade is as advantageous as any other trade. The money user can then shift the greater expense of a non-money transaction to the other individual who prefers not to use money. Thus, in these circumstances, it becomes immaterial to the money user whether his trading partner uses money.

The analysis can be extended by adding two types of transactions costs: search costs and trading costs. Search costs refer to the costs of locating a buyer or a seller of a given product, whereas trading costs refer to the costs of carrying out a transaction once a buyer or seller has been located. Each of these costs has been identified as a source of the usefulness of money. Clower (1969), for example, suggests that money permits traders to economize on the costs of search. Alchian (1977), on the other hand, suggests that the homogeneity of money reduces the information costs of carrying out an exchange with another trader.

Consider the effects of search costs that cannot be overcome entirely by the use of money. These costs mean traders must commit resources to seeking out trading partners. Once a trading partner has been found, the trader has the choice of trading with that partner or of seeking another partner through the expenditure of more resources. Under these circumstances traders care whether their trading partners use money. If they do not, the traders must decide whether to trade with them without the use of money or to seek other partners. The expected cost of finding trading partners who use money limits the extent to which money users must bear the higher costs of trading with a non-money users.

Now consider the effects of any trading costs that would remain once a buyer has been located. These costs result from the need to establish the terms of the transaction and may include, among other things, the costs of determining the nature and quality of the objects being exchanged. If these trading costs constitute the major costs of exchange in the economy, then money users should not care whether they trade with other money users. Unless non-money users make the terms of the trade for others as advantageous as the terms that can be gained in trading with other money users, the individual would choose to trade only with other money users. In the absence of search costs, traders will not bear any additional expense from doing so.

The analysis presented here suggests that the basic premise of the network-utility argument for the regulation of financial institutions has some validity, but depends on the nature of the transactions costs in the economy. An individual does care if his trading partners use money when it is costly to search for trading partners. On the other hand, if it is relatively easy to find additional trading partners in that the major trading costs in the economy take the form of establishing the terms of transactions, any individual should be indifferent as to whether his trading partners choose to use money or not.

2.8.4 Contagion effects: a summary

Three possible meanings for the term "contagion" as applied to financial institutions have been identified in this chapter. Each reflects some type of external effect that could exist in the operation of financial markets. The first involves dependence among the demands for services of different producers; the second involves dependence among customers of the same financial institution; and the third involves dependence among producers of payments services. Each possible source of "contagion" has been discussed for the purpose of understanding the rationale for regulation of financial institutions. In each case, the source of contagion is shown to depend on the existence of some form of externality arising from the behaviour of some economic actor. The first two arguments are just versions of externality arguments that may be

equally applicable to many other goods and services. The third argument provides a possible rationale for regulation of a more limited range of products. Still, while it is well known that externalities can provide a rationale for regulation, it is less frequently remembered that just establishing the possible presence of externalities is not a sufficient argument for regulation. Rather, the empirical magnitude of the costs of the externality must be weighed against the costs and benefits of regulation. Traditionally, economists have judged the externalities in the payments system to be large.

Chapter 3
IMPLICATIONS OF EFFICIENT REGULATION FOR THE FORM OF REGULATION

In this chapter I examine the reasons for the different forms of regulation applied to financial institutions. In Chapter 1 the emergence of different types of financial institutions was explained in terms of the problems that must be overcome by direct lenders and borrowers. In Chapter 2 the rationale for the regulation of financial institutions as a whole was analyzed from the perspective that regulation serves as a collective substitute for the contracts that intermediaries and their customers would want to reach privately. In this chapter the analysis is applied to the question: what forms of regulation can be justified by this explanation of regulation as a whole?[31]

The approach taken towards the regulation of financial institutions in this paper is derived directly from an assumption of "efficient regulation." Regulation is seen as a device by which private parties can supplement the range of enforceable contracts they might reach on their own. The purpose of a contract is to broaden the range of exchange between parties beyond those transactions in which both sides of the exchange are carried out simultaneously. As Posner (1973) notes,

> There are many contingencies that may prevent the process of exchange from operating to reallocate resources to higher valued uses, especially when the exchange is carried out over a period of time rather than simultaneously or when the performance of one or both parties involves a complicated undertaking. To minimize breakdowns in the process of exchange is the function of the law of contracts. (pp. 41-42)

By their very nature, financial transactions involve a separation of the obligations of the different parties over time. The scope for financial transactions depends on the effectiveness of the system of contracts that has evolved for these types of transactions. From the

---

31. Financial institutions may be regulated for reasons other than efficient contracting. Monetary policy and taxation are two such reasons. These other considerations are beyond the scope of this study.

standpoint of "efficient regulation", the regulation of financial institutions exists to provide their customers with a broader range of possible contracts.

Efficient regulation thus implies one central question: what aspects of contracts would depositors and financial institutions both wish to have enforced? In Chapter 2, it was established that the preventive approach to contracts may be more appropriate for financial institutions than the remedial approach because of the nature of their business. The type of assets held by financial institutions makes restitution difficult once the terms of the contract with the depositor have been breached. It is necessary, however, to enquire whether the particular problems of transactions in financial markets are reflected in the types of preventive regulation applied to financial institutions.

3.1 Contract Problems of Financial Institutions

What problems must be overcome by depositors in the formation of contracts with respect to their business with financial intermediaries? An important consideration with respect to the regulation of financial institutions is the emphasis that must be placed on global versus marginal considerations.[32] By its very nature the deposit-taking institution is organized to minimize conflict over margins between the institution and its customers. The fixed-money value of the deposit contract reduces the need for depositors to monitor the behaviour of operators at the margin because the operators of the institutions benefit fully from their efforts to maximize the returns from their portfolios. The global incentive replaces the marginal incentive as the focus of attention for depositors. If the financial institutions remain solvent, depositors are assured of repayment on agreed-upon terms. Insolvency prevents fulfillment of the contract.

---

32. The fixed-value deposit was explained in Chapter 1 as a device to reduce the need of depositors to continually monitor and supervise the management of the institution.

From this perspective, regulation can be used to limit the actions of financial institutions where they face incentives incompatible with their depositors' interests. The analysis considers three such possibilities:

i) Financial institutions might wish to hold riskier assets than would be desired by their depositors.
ii) Financial institutions might invest more in equity relative to debt than would be in the interests of their depositors.
iii) The operators of financial institutions might have an incentive to misappropriate resources from the financial institution against the interests of depositors.

3.1.1 Excessive risk

Excessively risky investments are one source of inconsistent incentives between financial institutions and their depositors. This represents a specific instance of the general problem of inconsistent incentives that exists in any contract specifying a fixed-value obligation between debtor and creditor. The debtor prefers the course of action that maximizes his wealth, whereas the creditor prefers the course of action that maximizes the expected value of his obligation from the debtor. This conflict can be illustrated by the example in Table 1, developed by Fama and Miller (1972).

> ... a firm has two mutually exclusive production decisions a and b available at period 1, and either can be carried out without additional expenditures of resources at period 1 ...There are also assumed to be two possible states of the world at period 2. The price p (1) at period 1 of a contingent claim to $1 to be received only if state 1 occurs at period 2 is $0.5; and likewise the price p (2) at period 1 of $1 to be received at period 2 if state 2 occurs is $0.5. At period 1, the firm is assumed to have bonds in its capital structure in the form of a promise to pay $5 at period 2, whichever state occurs.
>
> For each of the production decisions, Table 1 shows the payoffs in the two states at period 2, along with the period 1 market values of the firm, its bonds, and its common stock.

> Thus if production plan a is chosen, the period 1 market value of the firm's bonds is B(1) = 5(0.5) + 5(0.5) = 5, and the value of the shares is S(1) = 2(0.5) + 2(0.5) = 2. On the other hand, if plan b is chosen, the firm is not able to deliver in full on its debt promise if state 1 occurs. Hence with this production plan the period 1 market value of the bonds is B(1) = 1(0.5) + 5(0.5) = 3, and the value of shares is 2.5 (pp. 179-80).

Table 1

Production plan	Payoff at period 2		Market values at period 1		
	State 1	State 2	V(1)	B(1)	S(1)
a	7	7	7	5	2
b	1	10	5.5	3	2.5

Note: p(1) = p(2) = 0.5 Promised payment on debt is $5 at period 2.

Fama and Miller go on to note that although the market value of bonds is higher with plan a, the market value of shares is higher with plan b. Thus a conflict arises between creditors and shareholders over the investment strategy that should be followed by the firm.

Fama and Miller tend to minimize the importance of this possible inconsistency between debtor and creditor interests. They suggest that side payments between debtors and creditors provide a means by which the inconsistency of interests could be resolved:

> the bondholders could give the shareholders a subsidy of $0.5 to induce them to choose plan a; then the shareholders would have as much wealth as if plan b were chosen, and the bondholders would have more (p. 180).

Their arguments are less relevant for indirect lending through financial institutions than for direct lending and borrowing because financial institutions exist for the purpose of delegation of monitoring and enforcement. Such a side payment to financial institutions would be predicated on the lenders' awareness of the opportunities of both the

intermediary and the final borrower. Yet the current interpretation of intermediation is based on the premise that the fixed-value deposit eliminates the need for depositors to monitor and supervise the intermediary very closely.

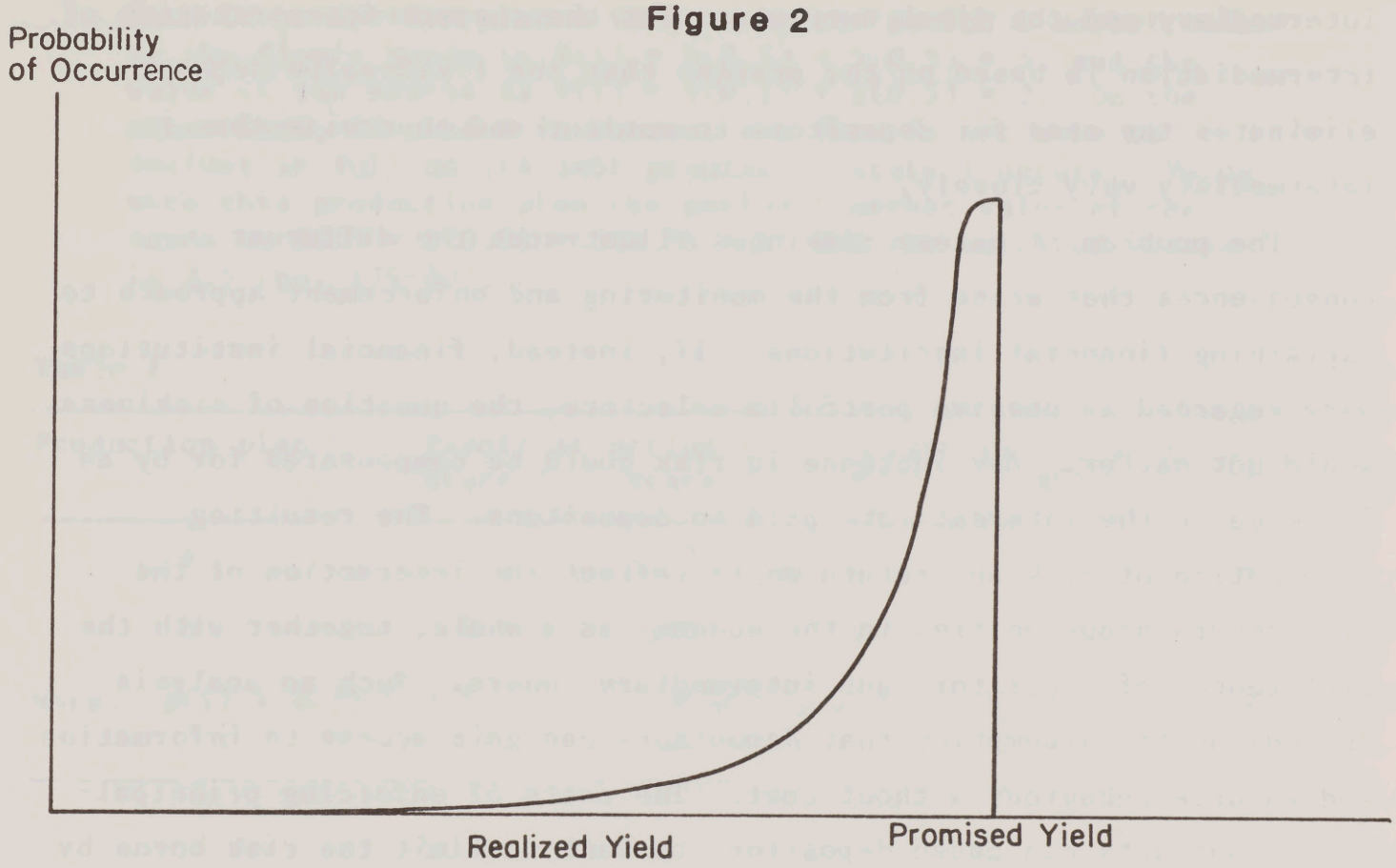
The problem of excess riskiness illustrates the different consequences that arise from the monitoring and enforcement approach to explaining financial institutions. If, instead, financial institutions were regarded as passive portfolio selectors, the question of riskiness would not matter. Any increase in risk could be compensated for by an increase in the interest rate paid to depositors. The resulting combination of risk and return would reflect the interaction of the risk-return opportunities in the economy as a whole, together with the preferences of depositors and intermediary owners. Such an analysis depends on the assumption that depositors can gain access to information and enforce behaviour without cost. The costs of enforcing principal-agent contracts can cause depositors to want to limit the risk borne by intermediaries with whom they do business.

3.1.2 Investment in equity

The argument about the compatibility of the financial institution's incentives with the interests of its customers can be extended to the choice between bonds and equity. One fundamental difference between these two types of securities results from the distribution of payments under different circumstances. The distribution for debt, as can be seen from Figure 2, reaches a maximum at the value of the principal plus the contractual interest payment. Frequently, this value is the mode. Because of concentration at the maximum value, the distribution is asymmetrical about the mean. In contrast, the distribution of payoffs for equity investments is more likely to be symmetrical. More important, however, there need not be any truncation of the distribution at the maximum contractual payment. Small probabilities can exist for very large and very small payments relative to the mean.

What differences do these distributions imply for the functioning of financial institutions? Their significance arises because of the

Figure 2



delegation of monitoring and enforcement from the depositors to their agents, the operators of the intermediaries. The agent has the option of committing resources towards increasing the probability of repayment from the ultimate borrower under different contingencies. In the case of equities, the operator of the intermediary may have an incentive to direct resources towards increasing the prospective payment under circumstances where the expected rate of repayment already exceeds the level that would assure repayment of the intermediary's debt to the depositor. Any increase in expected return accrues entirely to the agent since the return to the depositor has already been assured. Moreover, the increase may be at the expense of a decreased probability of a level of return that would assure repayment of depositors in other states. In the case of debt securities, the efforts of the intermediary are more likely to be compatible with the interests of the depositor. Efforts by the intermediary to increase the probability of its highest return also increase the probability that the depositor's claim will be met.

3.1.3 Misappropriation

To this point, the examples of inconsistency between intermediaries as agents and depositors as principals arise solely because of the agent's stake in the residual remaining after all contractual obligations to the depositors have been met. Another set of principal/agent problems arises in circumstances where the intermediary can benefit in ways other than through its claim on the residual. The most extreme problem arises when the agent uses the intermediary's assets to his own benefit. Misappropriation can occur either directly through the transfer of assets to the agent's own account or indirectly through transactions with enterprises that share some degree of common ownership with the intermediary.

The opportunities for misappropriation are likely to be greater for financial activities than for most other economic activity. Financial activity combines the chance for relatively substantial misappropriation over a short period together with a fairly low probability of detection in the short term compared to other types of activity. The sources of these differences require further elaboration with respect to their significance for financial institutions.

The ability to misappropriate funds over a short period depends on, among other things, the cash flow over that period. Compared to many other activities, the cash flow in financial institutions is substantial. On the asset side, financial institutions receive interest periodically and also have a turnover of their portfolio as existing loans are repaid and new loans are made. Similarly, on the liability side, financial institutions face a turnover of their deposit liabilities. In most other businesses, the cash turnover is smaller relative to assets and, as a result, a smaller proportion of total wealth could be misappropriated over any period.[33]

Misappropriation is difficult to detect in financial institutions because of the nature of the assets held. The stock in trade of any

---

33. Some rough indication of the difference in cash flow can be seen by comparing Stelco, Loblaws, and the chartered banks as a whole. Stelco's revenues in 1983 were roughly equal to assets, Loblaw's revenues were almost six times assets, and cheque clearings through banks alone were 16 times the total assets.

financial institution is a portfolio of non-marketable securities, the value of which depends to a considerable extent on their management by the financial institution itself. It is significant that unlike real capital, which itself may pose problems of verification of ownership and value, the portfolio of a financial institution turns over continually so that it is difficult to assign any element of the portfolio to a particular creditor. Moreover, the creditors of the financial institution would find it difficult to assess the value of any component of this portfolio because they are unlikely to have the specialized knowledge necessary to judge the characteristics of such non-marketable securities.[34] In addition, the value of these securities, unlike that of marketable securities, depends on the skill with which they are managed. Therefore, their value as part of the assets of a going financial concern cannot be measured independently of the fact that they are currently held and administered by the financial institution in question. Although other forms of capital may be equally difficult to value, they may be easier to enumerate. For example, a trucking firm knows the number of trucks it owns. Moreover, neither the number of trucks nor the composition of the fleet are likely to alter substantially over time.

Misappropriation can also result from transactions with associated enterprises. Such transactions could be carried out at unduly favourable terms. Yet such favourable terms pose no problem for the claimant as long as the financial institution remains solvent. The fixed nature of the deposit liabilities means that any benefits from favourable terms to associated enterprises just reduce the profits of the financial institutions. In this case, monitoring the activity of the agent running the intermediary becomes a concern of shareholders, a problem common to any firm. The problems of self-dealing become relevant to depositors only when the scope of the transactions has the potential to jeopardize the solvency of the financial institution.[35]

---

34. The delegation of collecting and interpreting such specialized knowledge is a major part of the monitoring and enforcement explanation for financial institutions.

35. The reason for the different concerns of shareholders and depositors arises because of the types of claims each holds. Depositors hold a fixed claim and need be concerned only with the ability of the financial institution to meet that claim; minority shareholders need to be concerned about such transactions regardless of their scale because any such transactions reduce their wealth.

Problems of self-dealing arise when the agent of the financial intermediary trades on behalf of that intermediary with another enterprise in which the agent also has an interest. This self-dealing can occur in a variety of transactions that are not at arm's length. The intermediary can lend to related enterprises on excessively favourable terms that do not fully reflect the riskiness or other characteristics of the loan. Parallel investments by associates in the same enterprise can create problems in the timing of acquisition or disposal on the receipt of new information. Finally, associated enterprises may exchange services or assets at prices different from those that would be established in an open market. Wealth can be transferred from one element to another in a group of associated enterprises through each of these devices.

3.2 The Consequences of Failure

Bank failures differ in degree, if not in kind, from most other types of business failure largely because of the general form and turnover of bank assets. In many forms of economic enterprise, creditors are able to make claims on particular assets as security for loans made to that enterprise. In many circumstances, unsecured creditors are protected through their ability to realize fairly general capital on the failure of the enterprise without any substantial decrease in its value relative to its value as part of a going concern. These forces provide less protection to the creditor of a deposit institution. Its portfolio turns over continually. Moreover, the nature of much of its deposit business is such that the balances of both its debtors and its creditors fluctuate substantially over time. These factors would make it difficult to assign specific securities to particular creditors. In addition, the value of these securities depends on their management by the deposit intermediary and therefore the assignment of the intermediary's loans to its creditors would not give a substantial degree of protection.

These features -- the costliness of assigning security and the dependence of the portfolio's value on management by the deposit-taking intermediary -- combine to make bank failure more costly to its creditors

than the failure of most other enterprises. These features also create the incentive to be first to withdraw deposits from an intermediary once its ability to meet its claims comes into doubt. Settlements of accounts prior to the declared failure of a financial institution are at the contractual value whereas settlements after the declaration of failure are at a much reduced level because the portfolio's value depends on its being held and managed by the intermediary and all losses are shared pro rata by remaining depositors.

3.3 The Form of Regulation

The application of the "efficient regulation" approach to the regulation of financial institutions can only suggest the types of regulation found in this sector. As Kenneth Arrow (1963) has noted in another context,

> The social demand for guaranteed quality can be met in more than one way, however. At least three attitudes can be taken by the state or other social institutions toward entry into an occupation or toward the production of commodities in general...
>
> 1) The occupation can be licensed, nonqualified entrants simply excluded...
> 2) The state or other agency can certify or label, without compulsory exclusion.
> 3) Nothing at all may be done; consumers make their own choices.
>
> The choice among these alternatives in any given case depends on the degree of difficulty consumers have in making the choice unaided, and on the consequences of errors of judgement. (p. 966-67)

A parallel to these alternatives, which were outlined with reference to medical care, can exist in the regulation of financial institutions. Financial institutions could be left to operate entirely outside the scope of government regulation. Alternatively, the government could use measures of differing degrees of strictness with respect to the regulation of the behaviour of financial institutions. While the efficient

regulation approach can indicate the form of regulation that would appear to be justified logically by the problems of private contracting, by itself it cannot indicate the stringency of the regulation or, for that matter, its necessity.

3.3.1 Incompatibility of incentives

The first problem that must be overcome by intermediaries and their customers concerns the incompatibility of incentives arising from the fact that the intermediary holds either risky or equity investments. This problem could be dealt with by either of two approaches: i) limitations on the holdings of risky assets or ii) constraints upon other elements of the intermediary's balance sheet.

Restrictions on the types of assets that can be held by intermediaries could take a variety of forms. At one extreme are prohibitions on selected classes of assets. Banks, for example, could be prohibited from holding equity securities or, alternatively, the types of security that banks could hold would be limited. This extreme form of regulation would reduce the need for surveillance of the intermediary. Any holding of ineligible assets would in itself be a violation of the terms of the regulation.

Restrictions on holding these classes of assets could also take the form of ceilings. Limited amounts of equity and of loans with limited security could be permitted in the intermediary's portfolio. Compared to the extreme of a prohibition, this form of constraint would permit types of activity that in limited quantities might be mutually beneficial to depositors and the intermediary. While the depositor may either face a higher risk of default or incur greater costs of supervision as a result of these activities, the additional return from the broader portfolio might more than compensate. At the same time, the limitation on the extent of portfolio that could be held in specific assets would still permit economy of supervision for the depositor.

The approaches to incompatibility of incentives discussed so far all relate to the asset side of the intermediary's balance sheet. An

alternative approach involves working through the liability side of the intermediary's balance sheet by requiring the intermediary to maintain a minimum level of shareholders' equity relative to its outstanding deposit liabilities. In some respects a minimum equity requirement is similar in its effects to limitations on the portfolios held by intermediaries. Both reduce the need for depositors to supervise the holdings of the intermediary. Capital requirements achieve this end by reducing the degree of accuracy to which the depositor must assess the portfolio whereas portfolio requirements alleviate the need for monitoring by precluding, or at least reducing, the holdings of securities that require careful monitoring. A capital requirement would appear suitable when definite limits can be attached to the losses that could be realized on the intermediary's portfolio under the least favourable possibilities. Portfolio limitations serve to set the limits to the variability in the portfolio. Even though these policies may be viewed as substitutes, they are also complementary. For example, the level of capital requirements that might be suitable would depend on the range of acceptable assets.

3.3.2 Misappropriation

The prospect of misappropriation creates different problems than those caused by excessive risk and investment in equities. There are two general types of regulation that can be directed towards the problem of misappropriation through self-dealing: structural and behavioural. Structural regulation is designed to prevent misappropriation by setting constraints on the form of organization that may be chosen by financial institutions. In contrast, behavioural regulation attempts to limit the actions of financial institutions so as to reduce the prospects for misappropriation, given their organizational form.[36]

The organizational form of intermediaries and their relationship to associated enterprises can limit the opportunities for misappropriation

---

36. The distinction between structural and behavioural regulation may in fact be one of degree rather than kind. The limiting behavioural regulation may be a prohibition. Generally, the distinction between organizational form and the types of behaviour permitted to an institution organized around any form can be made and will be retained for present purposes.

from self-dealing. To the extent that the entrepreneur behind any financial institution is engaged in other activities, these activities can be organized in association with the intermediary or as independent entities.

Why does the organizational structure of intermediation matter? To answer this question let us consider three different forms of organization with respect to the incentives for misappropriation. The two simplest are i) direct ownership of a financial institution by an enterprise engaged in real activity and ii) direct ownership of an enterprise engaged in real activity by a financial institution. The other alternative to be considered consists of joint ownership of enterprises engaged in both financial and real activities.

What problems would arise if financial firms owned enterprises engaged in real activities as wholly owned subsidiaries? Many of the problems would be the same as those within a financial firm that have been discussed earlier. The relevant question is whether the organization of real and financial activity within the same enterprise leads to different sorts of problems than the organization of financial activity in isolation. In particular, does the organization of the two activities together increase the prospect of misappropriation? Interestingly enough, in this instance it would appear that it does not. The fact that a real firm is owned entirely by the financial institution serves to internalize the costs and benefits of any misappropriation. If, for example, the parent financial institution favours the subsidiary, the decreased net worth of the financial activity of the intermediary would be offset by the increased value of the real subsidiary. Moreover, to the extent any costs were involved, the net worth of the combined enterprises would be reduced. Thus, direct ownership of real enterprises by financial institutions does not appear to add to the problems of misappropriation.[37]

---

37. This result considers only the incentives for misappropriation. It may be argued that separation could be justified in terms of the costs of surveillance and supervision. Separation may also be desirable in practice for reasons of taxation and other reasons not considered here such as equity vis à vis competitors and concern with concentration. Moreover, the argument is independent of the additional risks resulting from equity investment, discussed in section 3.1.

To what extent does the internalization of any favourable treatment persist when the real enterprise is partly rather than wholly owned by the intermediary? In this case, both factors -- the attractiveness of risky activity and the transfer from financial to real enterprises -- become attenuated. The presence of other shareholders in the real enterprise means that any benefits transferred from the financial to the real enterprise must be shared with the other shareholders on a pro rata basis.

The opposite extreme consists of a firm engaged in real activity that has a financial institution as a wholly owned subsidiary. In some ways this example is symmetrical to the ownership of real firms by financial institutions but not completely so. Within a range any favourable treatment of a real-sector parent by the financial-sector subsidiary is internalized. To a point, the increased value of the parent arising from its favourable treatment would be offset by the decreased value of the subsidiary to the parent because the parent is the residual claimant to the income of the subsidiary. Beyond the point at which the prospect of bankruptcy of the financial subsidiary becomes relevant, any redistribution from subsidiary to parent has a net benefit to the parent because it does not share all the costs of bankruptcy. Instead, depositors at the subsidiary deposit-taking institution must bear some of these costs.

If there are minority shareholders in the financial institution, favourable treatment of the parent always has a net beneficial effect on the parent. Prior to the point at which bankruptcy becomes a possibility, the value of equity in the financial institution falls in response to favourable treatment of the parent. In this case, the fall in the value of the equity in the financial institution is shared among the parent and the other shareholders.

The point is really more general. The entrepreneur has an incentive to carry out transactions that increase his wealth. If the entrepreneur holds different proportionate shares in different enterprises, he can increase his wealth by transactions that transfer wealth to those enterprises in which his proportionate stake is the greatest. Inasmuch as intermediaries borrow from ultimate lenders to acquire financial claims

issued by ultimate borrowers, they are, by their very nature, highly levered. Consequently, the incentives tend to direct the misappropriation of wealth away from financial institutions and towards other enterprises held by common owners.

An extreme form of regulatory approach to the avoidance of misappropriation requires the complete separation of ownership between intermediaries and other types of activities. Such an approach would be intended to insulate the ownership of intermediaries from interests that could lead to misappropriation through self-dealing. But the insulation of ownership prevents only one form of misappropriation, that undertaken by a majority owner. It is not a safeguard against other forms of misappropriation because interests other than common ownership may be reflected in the decision-making process of the intermediary.[38]

All the remedies for misappropriation that have been discussed to this point are structural. An alternative to the structural approach is the formulation of rules of behaviour. This approach would permit common ownership of financial institutions and other enterprises but guidelines would be established with respect to acceptable conduct. For example, certain types of transactions might be prohibited among associated companies. The financial institution might be precluded from lending to associated companies or from selling assets to or buying assets from these associated companies. Alternatively, transactions between enterprises that are not at arm's length might be subject to rules requiring valuation by an outside interest. Assets or services sold by an enterprise to its associate might require an outside valuation to establish the terms of the sale.

While the formulation of rules of behaviour might be feasible for the purposes of asset transfer, it is less clear that it would be appropriate as a means of governing the terms of lending between associates. These terms include non-price terms such as the degree of security and other

---

38. The "mutualization" of Canadian banks has been recognized since the introduction of the 10 per cent ceiling to ownership by any one party was introduced in the 1967 Bank Act. An initial discussion of the problems created by mutualization can be found in Chant (1979).

aspects of the loan contract, including interest rates. Any attempt to establish the appropriate terms for these transactions would require the assessor to replicate almost exactly the same review process that would be required by the lender itself. Moreover, the associated enterprise, by the fact of its association, may have an informational advantage over the outside appraiser. In such a case, any discrepancy between the terms deemed appropriate by the outside appraiser and the associated enterprise need not be evidence of misappropriation. It may merely indicate an information advantage for the associated enterprise.

3.4 Conclusions

This paper has examined competing theories of financial intermediation in order to determine their ability to explain the characteristics of actual financial institutions and to assess their implications for the regulation of these institutions. The analysis of the functions of financial institutions has been conducted under the stringent assumption that individuals are able to carry out any given transaction at exactly the same costs as an intermediary. Under this condition, any advantage of an intermediary must be derived from its ability to organize more efficient combinations of transactions than individuals can on their own. This assumption rules out theories of intermediation that rely, for example, on economies arising from specialization in intermediation. No doubt, as a result, some dimensions of existing financial institutions that could be explained by relaxing this assumption remain unexplained under the current approach. Nevertheless, intermediaries would have to exist to be able to gain any benefits from specialization. Any argument for the existence of intermediaries based on specialization would then be uninformative. Gains from specialization would be important, however, in explaining the extent and nature of intermediation once the process becomes established. The seemingly restrictive assumption about the source of cost advantages for intermediaries does not appear to have constrained the analysis unduly. Even with this assumption, explanations are found for each of the

identifiable types of intermediary and, in addition, the explanations appeared to identify each type of intermediary with the activities that correspond to those carried on by their actual counterparts.

The analysis shows that single-factor explanations are inadequate to explain the range of activities of financial intermediaries. The claim by Benston and Smith (1976) that transactions costs are the raison d'être for financial intermediation is incomplete. It fails to specify the source of the transactions costs or how intermediation has a cost advantage over direct transactions between ultimate lenders and borrowers. Answers to these questions require specification of the problems that lenders and borrowers must overcome. Do individuals face uncertainty with respect to the outcome of their investments? Do these uncertainties remain in the aggregate? Can investors alter the outcome of uncertain investments by applying resources to identification, verification, monitoring or enforcement? These questions in turn require specification of the cost of performing each of these functions. Can they be performed collectively by an agent for several investors more efficiently than by investors on their own? Can the agent appropriate the returns from performing the function?

Traditionally, explanations offered for the activity of intermediation emphasize one type of intermediation -- the deposit-taking intermediary. This emphasis may be justified because of their importance and because they perform some of the most heavily regulated activities in the economy. In addition, the deregulation of this sector, or more broadly the reform of regulation, is currently a topic of concern and controversy in a number of countries including Canada. A proper understanding of the role of these institutions would appear to be a prerequisite for any approach to the design of regulation.

The analysis of this paper suggests that deposit-taking intermediaries are more than collectively held portfolios in which risks are diversified and apportioned to investors according to their tastes. Rather, the analysis suggests that financial institutions alter the risks and returns facing investors, not just in the traditional sense of diversification of independent risks, but through the application of real resources to manage the level of risk by monitoring and enforcement. The

analysis should not be interpreted to mean that these activities are all that intermediaries do -- rather it suggests that these are the essential roles of the predominant form of intermediaries found in modern economies.

Acceptance of this view carries with it implications about the interpretation of events and the formation of policy with respect to the financial sector. To illustrate the first, consider the moves that many financial intermediaries have made over the last decade to match the maturities of their assets and liabilities. Previously, many intermediaries made long-term loans and financed them by issuing shorter-term claims. Appropriately, they were described as being involved in maturity intermediation or, following Tobin, in the transformation of assets. From the traditional view, the role of these intermediaries in the transformation of risk was substantially reduced by this change in the structure of their balance sheets. The approach here suggests instead that the primary functions of these intermediaries remained essentially unaltered. Intermediaries, through monitoring and enforcement, still offered their customers substantially different portfolios than they could assemble on their own.

An appreciation of the functions of financial institutions represents a first step towards understanding the role of regulation in the financial sector. Two substantially different arguments for regulation were developed in this paper. The first suggested that regulation could be viewed as a preventive approach to contract enforcement which is required in the case of financial institutions because of the unsuitability of typical creditors' remedies. The specificity of capital together with the large number and turnover of debtors and of creditors tends to favour prevention over remedy. The second argument for regulation applied only to financial institutions that issue liabilities that serve as a medium of payment. Users of the payment system gain benefits to the degree that others choose to use the system. The soundness of suppliers of payment services becomes a matter of general concern.

The relationship between depositors and financial institutions that offer fixed-value deposits leads to three types of contracting problems: incompatibility of incentives, misappropriation, and the discontinuity in

the value of any bank's portfolio when that bank fails. Efficient regulation can serve to supplement private contractual arrangements and has the potential advantage of permitting depositors to reduce their levels of surveillance of financial institutions. The results of this analysis suggest that a number of features common to the regulation of deposit-taking institutions can be explained in terms of efficient regulation: minimum capital ratios, eligibility requirements for investment, prohibitions on the joint ownership of financial institutions and other enterprises, and limitations on the terms and scale of transactions between associated enterprises. These are all consistent with the purpose of an efficient system of regulation designed to minimize monitoring and enforcement costs for depositors. Nevertheless, consistency with efficient regulation does not imply that these regulations in their current, or any other, form are necessarily justified by the efficient regulation approach. This study has considered the regulation of financial institutions from only one of many approaches to the explanation of regulation. It is possible that these forms of regulation may be equally well-explained by other approaches to regulation. Moreover, justification of current regulations would be assured only if it were determined that these regulations provide for more efficient contracts than could be reached by financial institutions and their depositors on their own.

REFERENCES

Alchian, A. (1977). "Why Money?" Economic Forces at Work, Liberty Press, Indianapolis, Ind., pp. 111-123.

Arrow, K.J. (1963). "Uncertainty and the Welfare Economics of Medical Care." American Economic Review, 53, pp. 941-973.

Arrow, K.J. and Lind, R.C. (1970). "Uncertainty and the Evaluation of Public Investment Decisions." American Economic Review, 60, pp. 364-378.

Benjamin, D.K. (1978). "The Use of Collateral to Enforce Debt Contracts." Economic Inquiry, 16, pp. 333-359.

Benston, G.L. and Smith, C.W. (1976). "A Transactions Cost Approach to the Theory of financial Intermediaries." Journal of Finance, 31, pp. 215-231.

Campbell, T.S. and Kracaw, W.A. (1980). "Information Production, Market Signalling, and the Theory of Financial Intermediation." Journal of Finance, 35, pp. 863-882.

Chan, Y. (1983). "On the Positive Role of Financial Intermediation in Allocation of Venture Capital in a Market with Imperfect Information." Journal of Finance, 38, pp. 1543-1568.

Chant, J.F. (1979). "The Banks and the Concentration of Corporate Power." in P.K. Gorecki and W.T. Stanbury. Perspectives on the Royal Commission on Corporate Concentration. Montreal: Butterworth and Co. (Canada) Ltd. for the Institute for Research on Public Policy.

Cheung, S.N.S. (1983). "The Contractual Nature of the Firm." The Journal of Law and Economics, 26, pp. 1-21.

Clower, R.W. (1969). "Foundations of Monetary Theory." In Monetary Theory, R.W. Clower (ed.).

Diamond, D.W. (1984). "Financial Intermediation and Delegated Monitoring." Review of Economic Studies, 51, pp. 393-414.

Diamond, D.W. and Dybvig, P.H. (1983). Bank Runs, Deposit Insurance, and Liquidity." Journal of Political Economy, 91, pp. 401-419.

Evans, J.L. and Archer, S.H. (1968). "Diversification and the Reduction of Dispersion: An Empirical Analysis." Journal of Finance, 23, pp. 761-767.

Fama, E.F. (1965). "Portfolio Analysis in a Stable Paretian Model." Management Science, January, pp. 404-419.

Fama, E.F. and Jensen, M.C. (1983a). "Separation of Ownership and Control." The Journal of Law and Economics, 26, pp. 301-325.

Fama, E.F. and Jensen, M.C. (1983b). "Agency Problems and Residual Claims." The Journal of Law and Economics, 26, pp. 327-349.

Fama, E.F. and Miller, M.H. (1972). The Theory of Finance. New York: Holt, Rinehart and Winston.

Freedman, C. (1977). "Micro Theory of International Financial Intermediation." American Economic Review: Papers and Proceedings, 67, pp. 172-179.

Friedman, M. (1960). A Program for Monetary Stability. New York: Fordham University Press.

Goldberg, V.P. (1976). "Regulation and Administered Contracts." The Bell Journal of Economics, 7, pp. 426-448.

Gordon, D.F. (1974). "A Neo-Classical Theory of Keynesian Unemployment." Economic Inquiry, 12, pp. 431-459.

Grossman, S.J. and Hart, O.D. (1981). "Implicit Contracts Moral Hazard, and Unemployment." American Economic Review: Papers and Proceedings, 71(2), pp. 301-307

Gurley, J.G. and Shaw, E.S. (1960). Money in a Theory of Finance. Washington: Brookings Institution.

Hart, O.D. and Jaffee, D.M. (1974). "On the Application of Portfolio Theory to Depository Financial Intermediaries." Review of Economic Studies, 41, pp. 129-147.

Hodgman, D.R. (1983). Commercial Bank Loan and Investment Policy. Champaign: University of Illinois.

Kane, E.J. and Buser, S.A. (1979). "Portfolio Diversification at Commercial Banks." Journal of Finance, 34, pp. 19-34.

Klein, B. (1974). "The Competitive Supply of Money." Journal of Money, Credit and Banking, 6, pp. 423-453.

Klein, B (1983). "Contracting Costs and Residual Claims: The Separation of Ownership and Control." The Journal of Law and Economics, 26, pp. 367-374.

Klein, B. and Leffler, B. (1981). "The Role of Market Forces in Assuring Contractual Performance." Journal of Political Economy, 89, pp. 615-641.

Klein, M.A. (1973). "The Economics of Security Divisibility and Financial Intermediation." Journal of Finance, 28, pp. 923-931.

Leland, H.E. and Pyle, D.H. (1977). "Informational Asymmetries, Financial Structure, and Financial Intermediation." Journal of Finance, 32, pp. 371-387.

Orr, D. (1970). Cash Management and the Demand for Money. New York: Praeger.

Posner, R.A. (1971). "Taxation by Regulation." Bell Journal of Economics, 1, pp. 22-50.

Posner, R.A. (1973). Economic Analysis of Law. Boston: Little, Brown and Company.

Pyle, D.H. (1971). "On the Theory of Financial Intermediation." Journal of Finance, 26, pp. 737-747.

Pyle, D.H. (1972). "Descriptive Theories of Financial Institutions Under Uncertainty." Journal of Financial and Quantitative Analysis, 7, pp. 2009-2020.

Santomero, A.M. (1984). "Modelling the Banking Firm: A Survey". Journal of Money, Credit and Banking, 16, pp. 576-602.

Sealey, C.W. (1983). "Valuation, Capital Structure, and Shareholder Unanimity for Depository Financial Intermediaries." Journal of Finance, 38, pp. 857-871.

Shavell, S. (1984a). "Liability for Harm Versus Regulation for Safety." Journal of Legal Studies, 13, pp. 357-374.

Shavell, S. (1984b). "A Model of the Optimal Use of Liability and Safety Regulation." Rand Journal of Economics, 15, pp. 271-280.

Stigler, G.J. (1971). "The Theory of Economic Regulation." Bell Journal of Economics, 2, pp. 3-21.

Stigler, G.J. and Becker, G.S. (1977). "De Gustibus Non Est Disputandum." American Economic Review, 67, pp. 76-90.

Stillson, R. (1974). "An Analysis of Information and Transactions Services in Financial Institutions." Journal of Money, Credit and Banking, 6, pp. 517-534.

Tobin, J. (1958). "Liquidity Preference as Behaviour Toward Risk." Review of Economic Studies, 25, pp. 65-86.

Tobin, J. (1980). "Discussion" in J. Kareken and N. Wallace (eds.), Models of Monetary Economics. Minneapolis: Federal Reserve Bank of Minneapolis, pp. 83-90.

Tobin, J. (1982). "A Commercial Banking Firm: A Simple Model." Scandinavian Journal of Economics, 84, pp. 495-530.

Printed in Canada Imprimé au Canada