Review of Economic Studies (1984) LI, 415-432
0034-6527/84/00290415$02.00

Information Reliability and a Theory of Financial Intermediation

RAM T. S. RAMAKRISHNAN
Massachusetts Institute of Technology

and

ANJAN V. THAKOR
Indiana University and Northwestern University

This paper is an analysis of when it will be beneficial for agents engaged in the production of information to form coalitions. The model is cast in a financial market framework, thus leading to an identification of conditions sufficient for the existence of financial intermediaries. Intermediation is shown to improve welfare if informational asymmetries are present, and the information generated to rectify these asymmetries is potentially unreliable. The usual appeal to transactions costs to explain intermediation is not needed.

INTRODUCTION

Suppose some agents produce and disseminate (costly) information about economic variables whose values are unknown ex ante. Reward for this activity is direct compensation provided by those who have an interest in the information. Reliability of the information cannot be assessed ex post without error, however. Only a "noisy" indicator of this reliability is available. Thus, the compensation of each information producer (i.p.) must be based on this indicator. The objective of this paper is to establish (sufficient) conditions under which i.p.'s will find it profitable to collaborate and operate jointly.

This tendency to centralize information production can be viewed as an explanation for the emergence of the traditional pure broker as described in the literature on financial intermediation. However, the explanation here is rooted exclusively in informational asymmetries[1] and the companion issue of information reliability. No assumptions are made about scale economies in information production or in processing financial claims, nor does the model depend on transactions costs.

The financial intermediaries we seek to explain are diversified information brokers. In our model, a coalition of i.p.'s constitutes a diversified information broker. Examples are Standard and Poor's Value Line, credit bureaus, financial newspapers, Moody's, check guarantee services, econometric modellers, consultants, investment bankers and accounting firms. Our intermediary acquires and processes information and its compensation depends—albeit in an imperfect way—on how well it performs its task. This corresponds closely with the functioning of the intermediaries above. As in our model, their output is information, an intangible service whose reliability is not directly observable. Hence, some performance surrogate must be used. The surrogate is often reputation-based; the intermediary's performance is assessed on the basis of how well its predictions did relative to observed events subsequent to the predictions. Future demand for the intermediary's services—and hence its future payoff—is then made dependent on this assessment. The stochastic evolution of reputation creates a natural incentive for risk averse i.p.'s to form coalitions. There are two similarities between such a reputation-based surrogate and our indicator. First, both are noisy measures of information quality. Second, both determine the intermediary's payoff.[2]

Niehans (1980) describes the brokerage function of financial intermediaries as follows.

> Similarly, financial intermediaries may be brokers, middlemen, or dealers in assets, bringing borrowers and lenders together at lower costs than if the parties had to get together directly. The basis of their existence, from this point of view—is the cost of evaluating credit risks. In a competitive system—the interest they receive—simply reflects their own marginal transactions costs. As dealers, financial intermediaries do not transform the claims they help to exchange. Their assets and liabilities relate to funds of the same type; in particular, they have the same liquidity or "moneyness". An impressive example of a highly developed intermediary system in which the brokerage function predominates is the Eurodollar market (p. 167).

Thus, the key to a theory of financial intermediaries as brokers is an explanation of *why* intermediation reduces the cost of exchanging capital. A major component of this cost is the cost of information production. The model we develop provides an economic rationale for the emergence of financial intermediaries based on their ability to lower information production costs. A firm (or a borrower) may need to raise capital from a number of investors (or lenders). Without an intermediating information broker, there would be enormous duplication in information production as each investor attempts to screen the firm. This can be avoided by appointing an i.p. to certify the firm's economic worth (or the borrower's likelihood of default). The transition from the i.p. acting as an individual information broker to a coalition of i.p.'s operating as a diversified information broker (an intermediary) will occur if the latter can further reduce information production costs. This reduction is achieved in our model because incentive costs—arising from the moral hazard created by each i.p.'s propensity to generate unreliable information—decline with group size. The larger the number of i.p.'s constituting the intermediary, the smaller is the (expected) cost of screening each firm (borrower).

The basic model is presented in Section 1, where i.p.'s are assumed to operate as individuals. Intermediation with costless internal monitoring by group members is considered in Section 2. The impact of the absence of internal monitoring is analysed in Section 3. Section 4 examines the welfare implications of intermediary size. Two alternatives to costless internal monitoring are examined in Section 5. One is costly internal monitoring. The other alternative requires *no* monitoring. Section 6 concludes. (For expositional continuity, all proofs are contained in an appendix.)

1. INFORMATION PRODUCTION

Consider a market with many participants, each endowed with a distinctive attribute known to itself but unknown to others. As in Akerlof (1970), the market is assumed to know the cross-sectional distribution of attributes, so that each participant is valued as if its attribute was the cross-sectional average. For concreteness, we call the participants "firms" and refer to the unknown attribute as the "value" of the firm.

It is assumed that the market provides frictionless trading opportunities and that firms are publicly owned. Each firm seeks to maximize its shareholders' wealth and all firms are in the process of selling *new* shares to the public. Further, equilibrium share prices are determined in a *tatonnement* process; no trades take place out of equilibrium, and traders can deduce each other's information by observing trading behaviour. Thus, if a trader were to acquire information about the true value of a firm's shares, he would be unable to exploit this information. And, if information is costly, no trader will want to invest in the activity (see Grossman (1976) and Grossman and Stiglitz (1980)).[3]

Firms with values exceeding the cross-sectional average will, however, have an incentive to reveal their true values because doing so increases the proceeds from the sale of new shares. One alternative, proposed by Spence (1974), is for firms to self-select by emitting signals with attribute-related costs. If agents are risk averse, or if the ex post outcome is not costlessly observable, signals are likely to involve deadweight losses—as in Bhattacharya (1979) and Leland and Pyle (1977)—rather than involving costless contingent contracts, as in Bhattacharya (1980) and Ross (1977).

Another possibility is that firms could have themselves screened through an external certification process, as suggested by Stiglitz (1975). That is, a firm could approach an "outside" i.p. to certify its value. Provided that expectations are rational and expeditiously revised, it follows that all firms except those at the lower end of the value spectrum will opt to be screened.[4]

We assume that there is no viable signalling mechanism that dominates the screening alternative. This rules out costless contingent contract signalling equilibria (see Bhattacharya (1980)) and assumes that costly signalling entails deadweight losses greater than screening costs. Thus, the "need" to rectify informational asymmetries pertaining to firm values should lead to the formation of a market for information. We examine the structure of this market and demonstrate that welfare is improved if the sellers of information form coalitions rather than operate individually.

It is assumed to be privately costly for an i.p. to produce information. Thus, subsequent to receiving a contract to screen a firm, the i.p. would like to avoid investing in information production if it can do so undetected. To ensure the reliability of information and the consequent feasibility of the market for information, some ex post indicator of information quality should be available. The incentive compatibility problem, arising from an i.p.'s tendency to deceive, could then be resolved by making each i.p.'s compensation contingent on this indicator.

Any information that does not lead to a correct valuation of the firm is viewed as being unreliable. The production of information is assumed to require "effort" and the i.p. can expend any effort, $\alpha \in [0, 1]$. To correctly identify a firm's value, $\alpha = 1$ must be chosen; any $\alpha < 1$ yields unreliable estimates of value. We assume α is unobservable and an ex post indicator, β, is used to measure it:

$$\beta: [0, 1] \times \boldsymbol{\theta} \rightarrow \{0, 1\},$$

where $\boldsymbol{\theta}$ is the state space of the random variable θ, which represents an exogenous source of noise in the ex post detection of the i.p.'s choice of α. The probability mass function of β is given by Prob $(\beta = 1 | \alpha = 1) = r$, Prob $(\beta = 0 | \alpha = 1) = 1 - r$, Prob $(\beta = 1 | \alpha < 1) = q \ \forall \alpha \in [0, 1)$, and Prob $(\beta = 0 | \alpha < 1) = 1 - q \ \forall \alpha \in [0, 1)$, where $1 > r > q > 0$.

All i.p.'s are identical, each with preferences described by the utility function

$$V(m, \alpha) = U(m) - \alpha A,$$

where m denotes monetary wealth, A is a positive real valued scalar, and $U(\cdot)$ is twice continuously differentiable and bounded, with $U'(\cdot) > 0$ and $U''(\cdot) < 0$. It will be useful to define the inverse function

$$\psi(U) \equiv U^{-1}(\cdot).$$

Since $U'(\cdot)>0, U$ is invertible. Also, $\psi(\cdot)$ is strictly increasing, convex and twice continuously differentiable everywhere. We use primes to denote partial derivatives.

The functioning of the market can be summarized as follows. All firms expecting to realize a net gain from being correctly identified will retain i.p.'s. The contracts negotiated with the i.p.'s are public knowledge, precluding undetected side payments. This ensures that with known preferences the expected utility maximizing actions of i.p.'s can be computed by all ex ante. Thus, if the offered contracts are incentive compatible (induce a choice of $\alpha=1$), the credibility of i.p.'s will not be suspect.[5] The information production market is perfectly competitive, and each i.p. must be offered a contract that yields him at least his reservation expected utility, R, if he chooses $\alpha=1$. Firms will, therefore, search for a contract which is incentive compatible, minimizes their expected cost, and guarantees each i.p. no less than his reservation wage. That is, we are assuming that firms are risk neutral toward the fees to be paid to the i.p.'s. Competitive behaviour in the information production market implies that the i.p.'s, on their part, will seek organizational forms that reduce information production costs.

As long as side payments from firms to i.p.'s are ruled out, no i.p. choosing $\alpha=1$ will have an inducement to disseminate anything but the truth. Of course, this assumes that the probability distribution of the ex post indicator depends only on the effort expended.

Assume initially that i.p.'s operate independently. The ramifications of group formations will be explored later. Since an i.p. will produce information regarding one firm at a time, each firm will separately contract with only one i.p. A contract can be defined as

$$\phi(\beta)=\begin{cases} W & \text{if } \beta=1 \\ X & \text{if } \beta=0 \end{cases}$$

where W and X are the fees paid to the i.p. by the firm. Let $U(W)=w$ and $U(X)=x$.

Let $EU(\alpha=j)$ be the i.p.'s expected utility if he chooses $\alpha=j$. Then,

$$EU(\alpha=1)=rw+(1-r)x-A \tag{1}$$

$$EU(\alpha<1)=qw+(1-q)x-A. \tag{2}$$

In order to motivate the i.p. to select $\alpha=1$ and ensure incentive compatibility (IC), we must have $EU(\alpha=1)\geqq EU(\alpha<1)$.[6] In order to ensure that an individually rational (IR) i.p. accepts the contract, we need $EU(\alpha=1)\geqq R$. These constraints are

$$\begin{aligned} &\text{IR:} \quad rw+(1-r)x-A\geqq R \\ &\text{IC:} \quad (r-q)(w-x)\geqq A(1-\alpha) \quad \forall\alpha\in[0,1]. \end{aligned} \tag{3}$$

Since the IC constraint will be most binding for $\alpha=0$, we can rewrite it as

$$\text{IC:} \quad (r-q)(w-x)\geqq A. \tag{4}$$

The firm pays W if $\beta=1$ and X if $\beta=0$. Thus the expected fee, given that the i.p. chooses $\alpha=1$, is $rW+(1-r)X$. Hence, the firm's problem is to

$$\text{minimize}_{w,x}\ r\psi(w)+(1-r)\psi(x) \tag{5}$$

subject to (3) and (4).

We now have the following result.

Theorem 1. *If each firm negotiates separately with one i.p., the optimal contract offered to each i.p. will be*

$$\bar{W} = \psi[R + A(1-q)(r-q)^{-1}], \qquad \bar{X} = \psi[R - Aq(r-q)^{-1}],$$

where bars denote optimal values. Moreover, the expected cost to the firm in this case is greater than it would be if the actions of i.p.'s were observable without error ex post.

The intuition is transparent. In absence of noise in the ex post verification of actions, the terminal payoff uncertainty faced by an i.p. disappears. This risk dissipation benefits the risk averse i.p. who is now willing to accept a reduced payoff in expected value terms. While this result itself is hardly surprising, the explicit characterization of the optimal second best contract serves as a benchmark with which to assess the desirability of intermediary formation.

2. INTERMEDIARY FORMATION WITH INTERNAL MONITORING

Since i.p.'s are risk averse and their payoffs uncertain, it would appear that they should benefit by collaborating and thereby reducing the risk faced by each. If the indicators used by the different firms are imperfectly correlated, one would ordinarily expect diversification benefits from the pooling of payoffs. To accentuate the possible benefits, we assume that each indicator is stochastically independent of every other. Throughout *this* section we shall assume that when i.p.'s merge, they can avail of additional *internal* monitoring of each other's actions. In the next section we examine the case without internal monitoring.

Internal monitoring of each other's actions by the i.p.'s (who constitute an intermediary) can take one of two forms. A noisy indicator that provides information in addition to that conveyed by the external indicator may be employed. Alternatively, internal monitoring may be perfect. To keep things simple, we shall assume the latter. Internal monitoring is also assumed to be costless initially, but this assumption is relaxed later. We consider the case in which $N(>1)$ identical i.p.'s form an intermediary that contracts with N distinct firms. Given perfect internal verification of actions, the i.p.'s *collectively* determine a production strategy. This strategy defines the number of firms, $n(<N)$, about which reliable information will be generated. The *total* effort is then shared *equally* by all members. Thus, each member's input is n/N and the intermediary's total payoff in every state is also shared equally among the N i.p.'s. This equal sharing of work and compensation is stipulated because all i.p.'s are identical.[7] The effect of such a mechanism is the complete elimination of moral hazard internally. The contracting firms, however, must still contend with moral hazard in their attempt to construct incentive compatible compensation structures for the i.p.'s they transact with.

In this setting, internal monitoring is the ability of the i.p.'s within the intermediary to allocate work among themselves and to verify that each has performed the assigned task. The internal review committees of many intermediaries perform this role.

We seek to establish that mergers of screening agents will be beneficial in this setting even if there are no scale or scope economies in information production. We will initially consider two i.p.'s merging to form an intermediary. The implications of larger groupings are examined in Section 4.

Coalescing has obvious risk sharing benefits for the i.p.'s. But will the firms contracting with these i.p.'s also gain from the merger? If the firms continue to use the same (second best) optimal contract they used before the i.p.'s coalesced, the answer is: not always. A proof of this is outlined below.

If two i.p.'s merge and each is compensated with the second best contract of Theorem 1, *each* i.p. gets $\psi(\bar{w})$ if $\beta=1$ is observed for both firms, $[\psi(\bar{w})+\psi(\bar{x})]/2 \equiv \hat{K}$ (say) if $\beta=1$ is observed for one firm and $\beta=0$ is observed for the other, and $\psi(\bar{x})$ if $\beta=0$ is observed for both. Let $U(\hat{K})=\hat{k}$. By Jensen's inequality, we have $\hat{k}>[\bar{w}+\bar{x}]/2$. Define $[\bar{w}+\bar{x}]/2 \equiv \bar{k}$.

If both i.p.'s choose $\alpha=1$, the probability is r^2 that $\beta=1$ for both firms, $2r(1-r)$ that $\beta=1$ for only one firm, and $(1-r)^2$ that $\beta=0$ for both firms. If both i.p.'s choose $\alpha=0$, the probabilities are $q^2, 2q(1-q)$ and $(1-q)^2$ respectively. The expected utility of *each* i.p. is

$$r^2\bar{w}+2r(1-r)\hat{k}+(1-r)^2\bar{x}-A=R+2(\hat{k}-\bar{k})r(1-r)$$

if $\alpha=1$ is chosen by both, and is

$$q^2\bar{w}+2q(1-q)\hat{k}+(1-q)^2\bar{x}=R+2(\hat{k}-\bar{k})q(1-q)$$

if $\alpha=0$ is chosen by both. As long as $r(1-r)<q(1-q)$, the intermediary's optimal strategy will be to expend zero effort for each firm.

The problem here is that the merging of i.p.'s improves risk sharing but distorts incentives. Incentive compatibility can be restored by increasing the difference between W and X, thereby increasing the reward for demonstrated diligence. But the risk aversion of the i.p.'s implies that this will increase expected costs. Thus, the risk sharing property of mergers reduces expected screening costs, but the accompanying incentive distortion increases these costs. We will show, however, that, internal monitoring always results in the first effect dominating the second, so that firms will prefer to transact with merged i.p.'s rather than independents.

In dealing with a two-i.p. intermediary, the most general contract makes the intermediary's payoff from *each* firm equal to

$$\begin{array}{l} W \text{ if } \beta=1 \text{ for both firms,} \\ K_1 \text{ if } \beta=1 \text{ for firm one and } \beta=0 \text{ for firm two,} \\ K_2 \text{ if } \beta=0 \text{ for firm one and } \beta=1 \text{ for firm two,} \\ X \text{ if } \beta=0 \text{ for both firms.} \end{array} \tag{6}$$

Since in equilibrium each firm will choose a contract that motivates the intermediary to pick $\alpha=1$ for both firms, the probability of the intermediary receiving K_1 is the same as the probability of its receiving K_2. We can, therefore, set $K_1=K_2=K$. Such a contract will be referred to as an "Internal Monitoring Joint Contract" (IMJC).

Joint contracts are similar to reputation-based contracts. Even if the intermediary's assessed performance with a particular firm is not good, the firm is willing to pay a high fee if the intermediary did well with all the other firms. The reputation index is the intermediary's overall performance—or the number of firms for which the intermediary "scores" $\beta=1$—which is estimated ex post.

A special case of an IMJC is a "separate" contract that entails each firm paying the intermediary W if $\beta=1$ for that firm and X if $\beta=0$ for that firm. Since the intermediary members share payoffs equally, such a contract is a special case of the IMJC with $K=[W+X]/2$. It can be proved (the proof is omitted for space considerations) that, except for the case of logarithmic i.p. preferences, the IMJC always strictly dominates the separate contract.[8] With logarithmic preferences, the two are equivalent. Thus, in our analysis we restrict our attention to joint contracts.

Those familiar with Holmstrom's (1979) result on monitoring may find this surprising. The performance of the intermediary with respect to one firm conveys *no* information about the intermediary's effort in screening any other firm. In Holmstrom's (1979, 1982) terminology, the indicator for one firm is "noninformative" in its ability to reveal the intermediary's effort choice for any other firm. Hence, it *may* appear that the intermediary's payoff from one firm should depend only on what is revealed by the indicator for that firm. This thinking is reinforced by Baiman and Demski (1980) who find that when the firm contracts directly with the i.p.'s, joint contracts—which reward each i.p. on the basis of the outputs of both i.p.'s—are inferior to contracts that make each i.p.'s payoff contingent only on his own performance. This result is an immediate consequence of Holmstrom's (1979) noninformative monitoring proposition.

Our observation differs radically from Baiman and Demski's (1980) application of Holmstrom's proposition because we are concerned with intermediaries and they are not. Thus, Baiman and Demski ignore explicit intermediary formation and the concomitant availability of internal monitoring and risk sharing. In our model two different firms contract with *one* intermediary which, in turn, utilizes internal monitoring and determines the sharing of stochastic terminal wealth and the allocation of work. This reduces the adverse impact of the exogenous uncertainty on each i.p. and thus slashes incentive costs. The reason for this is that joint contracts reduce the "incentive spread". Because the i.p.'s are risk averse, a larger "incentive spread" (the difference between the payoff if $\beta=1$ and the payoff if $\beta=0$) is undesirable. With joint contracts, the intermediary's motivation to choose $\alpha=1$ for the first firm arises not only from the fact that it increases the probability of getting $\beta=1$ for the first firm, but also because it improves the likelihood of getting a high compensation from the other firm. This means that a lower incentive spread can be used (without disturbing incentive compatibility). The same is true for the second firm. The mutual benefit created by a joint incentive spread reduction decreases incentive costs.

Using the IMJC we now provide a rationale for intermediation.

Theorem 2. *With internal monitoring, the expected cost of information production per firm for an intermediary is strictly lower than the expected (second best) cost if each i.p. functions independently.*

The intuition behind this theorem is clear. The formation of an intermediary offers diversification benefits to its members. With perfect internal monitoring, there is no need to be concerned about these benefits being nullified by the increased internal moral hazard that accompanies group formation. Thus, as long as the intermediary as a whole can be motivated to work, the reservation utility constraint of each i.p. can be satisfied at lower cost than would be possible without diversification. This theorem indicates that financial intermediaries may exist under a much weaker set of conditions than those customarily advanced.

3. INTERMEDIARY FORMATION WITHOUT INTERNAL MONITORING

To examine whether intermediation will be worthwhile in the absence of monitoring, we again assume that two i.p.'s form an intermediary to screen two firms. They are unable to observe each other's effort and ex post verification of effort beyond the publicly observable indicator is impossible. Since the i.p.'s are identical and cannot monitor each other, the intermediary's total earnings will be shared equally.

The two i.p.'s unite solely to internalize the benefits of risk sharing. For an i.p. to raise its effort from $\alpha < 1$ to $\alpha = 1$, its incremental expected utility from the resulting stochastic increase in its share of the payoff must at least match the decline in its welfare due to the higher effort. However, improvements in the total payoff are shared equally by the i.p.'s. Thus, in contrast to the internal monitoring case, the intermediary is now beset with incentive distortion and a free rider problem. Additional wealth is shared, but the higher effort supply that creates it is not. Hence, in order to motivate each i.p. to take the desired action, contracts need to make greater payoff distinctions between good and bad indicator outcomes. This *could* increase expected screening costs.

Since the implemented contracts will motivate both i.p. members to opt for $\alpha = 1$ in equilibrium, we can focus exclusively on the Nash equilibrium strategy of both i.p.'s choosing $\alpha = 1$. Each firm will use joint contracts and pay the intermediary W if $\beta = 1$ for both firms, K if $\beta = 1$ with only one firm, and X if $\beta = 0$ for both firms. In a Nash equilibrium, this contract should motivate an i.p. member to choose $\alpha = 1$ if the other member is expected to choose $\alpha = 1$. Theorem 3 establishes that without internal monitoring the joint contracting solution will be worse than the second best solution obtained with independent i.p.'s. The situation with "no internal monitoring joint contracts" will be referred to as the NIMJC case.

Theorem 3. *With no internal monitoring, the expected cost of information production for each firm approaching an intermediary is strictly higher than the expected (second best) cost if the i.p.'s function independently.*

This theorem contradicts the frequently encountered claim that the emergence of financial intermediaries is a manifestation of the desire of risk averse individuals to join together for risk-sharing purposes. It also points out that the intermediary's access to internal monitoring plays a central role in Theorem 2. Payoff pooling is inefficient here because diversification has perverse *incentive* effects. Without internal monitoring, pooling tends to blur the distinction between i.p. effort inputs. The marginal gain to an i.p. from incremental effort diminishes, because diversification smooths rewards. Hence, the ex post sharing of payoffs creates an information externality. To counter it and to preserve incentive compatibility, it is necessary to increase each i.p. member's marginal inducement to select $\alpha = 1$. This creates an augmented screening cost for each firm. Thus, the benefits of diversification are more than offset by the costs of increased moral hazard.

4. SIZE BENEFITS IN INTERMEDIATION

We now consider the effects of expanding the size of the intermediary beyond two members. Since internal monitoring appears to be crucial to improving welfare, we will assume that costless internal monitoring is available. The cases of costly internal monitoring and alternatives to internal monitoring are addressed in Section 5. The advantages of size are formalized as follows.

Theorem 4. *As the number of i.p.'s constituting the intermediary goes to infinity, the expected cost of information production for each firm converges to the cost that would be incurred if the action of each i.p. were observable by "outsiders" without error ex post.*

Thus, this theorem indicates that information production can be undertaken most efficiently by a natural monopoly. Note that this result does not depend on economies of scale in information production. Moreover, the informativeness of the ex post external indicator of an i.p.'s action is not enhanced as more indicators are employed. In other words, the noise present in a specific indicator is unaffected by the number of indicators used because an indicator of one i.p.'s action conveys no information about any other i.p.'s action. Moreover, the results of the internal monitoring undertaken by i.p.'s are not revealed to outsiders.

The intuition is that the availability of internal monitoring permits the intermediary to write "forcing contracts" for its members, giving each member a sure payoff, contingent on the member choosing $\alpha = 1$. Such an arrangement will be possible only if the intermediary's aggregate payoff is nonstochastic. This is because the budget constraint necessitates that the sum of individual compensations equal the intermediary's total earnings. Hence, each member's reward cannot be expunged of exogenous risk unless the intermediary's total reward is also nonrandom. The intermediary can achieve this in a probabilistic sense—despite receiving a stochastic payment from each firm—because it is infinitely large and thus capable of diversifying away the risk associated with independently distributed random variables. It might appear that one hitch in this is that the absence of payoff-related noise, even in the limit, could create moral hazard since the intermediary may decide to choose $\alpha = 0$ for some firms. This possibility is ruled out, however, since the (probabilistic) convergence argument requires *every* i.p. to select $\alpha = 1$. Realizing this, the intermediary finds it in its interest to enforce $\alpha = 1$ for every firm.

In Theorem 4, the intermediary's total payoff depends on the number of i.p.'s for which $\beta = 1$ is observed. In the context of the reputation-based surrogate discussed in the Introduction, the intermediary can be thought of as receiving N information production contracts at time zero, with the fee on each contract fixed. At time one, the market makes an (imperfect) assessment of the quality of the intermediary's output on each contract and concludes that on $n(\leqq N)$ contracts the quality was good. This now becomes the reputation index that determines how many new contracts the intermediary receives at time one. The higher the n, the greater is the demand for the intermediary's service and consequently, the greater is its payoff at time one. Thus, although explicit indicators of the type modeled here may not be widely observed, market forces that reward demonstrated quality operate analogously.

5. INTERMEDIATION WITHOUT COSTLESS INTERNAL MONITORING

Our rationale for financial intermediation has up to this point assigned a pivotal role to costless internal monitoring. This is admittedly restrictive. It is, therefore, important to know if weaker conditions will suffice. Two less stringent alternatives are examined in this section.

A. *Costly internal monitoring*

In the preceeding sections a noisy ex post indicator with zero direct usage cost was the only monitoring device assumed to be available to "outsiders". Assume now that in addition a monitoring technology capable of producing *perfect* information ex post about the actions of any independent i.p. is also available at a direct cost of $\pi_1 \in (0, \infty)$ to all.[9] We shall keep our discussion brief because Diamond (1984), in a related paper, provides a detailed characterization of the conditions for intermediation using the assumption of perfect and costly monitoring.

Let $\delta_s \equiv r\psi(\bar{w}) + (1-r)\psi(\bar{x})$ be the second best contracting cost for any firm using the noisy indicator to monitor an individual i.p. Let $\delta_f \equiv \psi(A+R)$ denote the first best contracting cost, incurred by the firm if it uses the perfect monitor. The firm's total screening cost in this case will be $\pi_1 + \delta_f$. The perfect monitor should, therefore, be strictly preferred to the noisy monitor only if $\delta_s > \pi_1 + \delta_f$.

Now suppose two i.p.'s merge to form an intermediary, and are able to internally monitor each other perfectly at a direct cost of π_2 each. Let each firm's expected contracting cost with the IMJC arrangement (with the firm using the noisy indicator for the intermediary) be

$$\delta_{im}^0 \equiv r^2\psi(w_{im}^0) + 2r(1-r)\psi(k_{im}^0) + (1-r)^2\psi(x_{im}^0).$$

Hence, each firm transacting with the intermediary incurs a total screening cost of $\pi_2 + \delta_{im}^0$, because the intermediary must be allowed to recoup its investment in costly internal monitoring. From Theorem 2 we know that $\delta_{im}^0 < \delta_s$, and from Theorem 4 we know that $\lim_{N\to\infty} \delta_{im}^0 = \delta_f$, where N is the size of the intermediary. Thus, we can write the IMJC cost for each firm as a function of intermediary size, i.e.

$$\delta_{im}^0 = \delta_{im}^0(N) \quad \text{with } \partial\delta_{im}^0(N)/\partial N < 0.$$

Our explicit characterization of contracting costs in a world in which internal monitoring is costly now gives us a condition which explains when intermediary formation will be superior to direct contracts between firms and independent i.p.'s. The condition is

$$\pi_2 + \delta_{im}^0(N) \leqq \min\{\delta_s, \pi_1 + \delta_f\}. \tag{7}$$

That is, if N^* is the smallest N for which (7) holds, then all intermediaries with less than N^* i.p.'s are infeasible, while any intermediary containing more than N^* i.p.'s strictly improves welfare.[10]

B. *Functional divisibilities and horizontal effort allocation*

We have thus far assumed a specific work allocation scheme within the intermediary. When two i.p.'s merge, each member is assigned responsibility for screening one firm. This is a sensible arrangement because all i.p.'s are identical and incentive compatible contractual mechanisms guarantee that in equilibrium each i.p. will choose $\alpha = 1$. There is, however, another possibility. The intermediary could require *each* of its two members to supply $\alpha = 1/2$ for *each* firm. Whether such client assignment is feasible is an open question since it requires divisibility of tasks. With functional specialization within the intermediary, we visualize work allocations being made so that the total effort of each member is equal; i.e. each i.p.'s total effort is $\alpha = 1$. Such work assignment is defined as Horizontal Effort Allocation (HEA). In contrast to Section 2, it will be assumed that there is *no* internal monitoring. Hence, *no* member i.p. can verify ex post whether its partner actually performed the assigned task. The question is whether intermediation is worthwhile in this setting. In Theorem 5 we show that, if screening is functionally divisible, there is at least one equilibrium in which HEA is a perfect substitute for costless internal monitoring.

Theorem 5. *Suppose screening is perfectly divisible. Then with HEA, intermediation improves welfare even if no internal monitoring is feasible. This is because there is one Nash equilibrium in which HEA is a perfect substitute for costless internal monitoring.*

Since HEA and internal monitoring are equivalent, the result derived in Section 2 should apply with HEA too. The assumption that screening is a divisible activity is not critical. Imperfect divisibility will suffice for intermediaries of finite size; what is needed is that there be some divisibility. Of course, an infinitely large intermediary (à la Theorem 4) will be unattainable if divisibility is not perfect.

Note that the probability of $\beta = 1$ jumps from q to r as α goes from a number in $[0, 1)$ to 1 for any firm. This kind of "impulse" probability production function makes sense when only completely done work "matters". One "weak" link is enough to result in "failure" (a decline in the "success" probability from r to q). Such discreteness means that each i.p. is pivotal; its failure to take $\alpha = 1$ results in failure of the *whole* operation. Therefore, moral hazard is resolved without internal monitoring because no i.p. can get a "free ride" by not taking the predetermined action. In other words, the incremental utility of wealth for any i.p. from working outweighs the incremental disutility of effort. The Pareto efficiency of intermediary formation can be sustained with "less discreteness" in the indicator distribution. For instance, intermediation will help even if the probability of $\beta = 1$ is r with $\alpha = 1$, p with $\alpha = \frac{1}{2}$ and q with $\alpha = 0$, as long as $q < p < r$, and p is close to q. In the limiting case in which the probability of $\beta = 1$ varies in a continuum as α goes from 0 to 1, there should be sufficient convexity in this probability distribution. Higher levels of convexity will facilitate intermediary emergence.[11]

7. CONCLUDING REMARKS

We have explained why economic agents who produce information under conditions of moral hazard may benefit from forming coalitions even in the absence of economies of scale in information production.

Perhaps the best example of the payoff structure we have assumed for the intermediary (see Theorem 4) is the manner in which investment bankers are compensated in firm commitment underwritten offerings. The investment banker guarantees the firm fixed proceeds from the issue and bears the risk of the actual proceeds being less. The investment banker's fee is the spread between the actual proceeds and the amount promised the issuer. A successful issue is one that is fully subscribed at the price at which the investment banker decides to float it. The investment banker's compensation is higher if the issue succeeds than it is if it fails. To ensure the success of the issue, the investment banker should know what price to float the issue at. This will call for information production. If the investment banker does produce the necessary information about the value of the firm's securities (i.e. if it takes $\alpha = 1$), the probability is high that the issue will be "correctly" priced and thus successful. If it does not produce the necessary information (i.e. if it takes $\alpha = 0$), the probability is high that the issue will be "mispriced" and thus unsuccessful. This means that the investment banker's payoff depends on the result of a noisy, market-based indicator, which in this case is the market's assessment of whether the information conveyed by the flotation price (set by the investment banker) is reliable. Thus, an investment banking house that is simultaneously involved in many issues faces a payoff structure similar to the one faced by the N-i.p. intermediary of Theorem 4.

Finally, we turn to the question of which firms get screened and which do not. Any firm that is not screened will be valued at the cross-sectional average. And as more and more firms get "peeled off", the cross-sectional average will decline. Consequently, the only firms that will not be screened are those whose values lie in the interval $[V^-, \hat{V}]$, where V^- is the value of the lowest valued firm in the market and $\hat{V}$ is the value of the highest valued firm that does *not* get screened. If $E(\phi)$ is the expected screening cost per firm, then

$$\hat{V} - \int_{V^-}^{\hat{V}} Vg(V)dV < E(\phi),$$

where $g(V)$ is the density function of V over $[V^-, \hat{V}]$. Thus, this inequality says that the screening process will stop when the gain to even the firm with the highest value among the unscreened firms is less than the expected screening cost. (See Stiglitz (1975)). In our earlier discussion (see footnote 4) we assumed, for simplicity, that the distribution of firm values was discrete and that the values of firms were spaced sufficiently far apart so that all firms except those with values equal to V^- would opt to be screened. Such an assumption is not necessary, however. It is also apparent that the more efficient the screening process—through, for example, an increase in intermediary size as in Theorem 4—the smaller will be $E(\phi)$ and the greater will be the aggregate demand for screening as more firms will find it optimal to be screened.

APPENDIX

Proof of Theorem 1. Let $\bar{w}$ and $\bar{x}$ denote the i.p.'s expected utilities for monetary wealth under the optimal contract if the indicator reveals $\beta = 1$ and $\beta = 0$ respectively. With Lagrange multipliers μ and λ appended to (3) and (4) respectively, the first-order optimality conditions corresponding to the problem in (3)–(5) yield

$$\psi'(\bar{w}) = \mu + \lambda(r-q)r^{-1}, \tag{A.1}$$

and

$$\psi'(\bar{x}) = \mu - \lambda(r-q)(1-r)^{-1}. \tag{A.2}$$

Now, since $r > q$ and $A > 0$, (4) implies that $\bar{w} > \bar{x}$. Because $\psi(\cdot)$ is increasing and convex, we have $\psi'(\bar{w}) > \psi'(\bar{x})$. Further, since $q < r < 1$, it follows that $\lambda > 0$. Thus, (A.2) implies that $\mu > 0$. Consequently, both (3) and (4) are binding equalities. Solving (3) and (4) as simultaneous equations gives us

$$\bar{W} = R + A(1-q)(r-q)^{-1} \quad \text{and} \quad \bar{x} = R - Aq(r-q)^{-1}. \tag{A.3}$$

To prove the second part of the theorem, note that if α is observable without error ex post, the optimal contract to offer each i.p. is

$$\phi(\alpha) = \begin{cases} \psi(A+R) & \text{if } \alpha = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The desired fact, $\psi(A+R) < r\psi(\bar{w}) + (1-r)\psi(\bar{x})$, now follows immediately from Jensen's inequality. (Note $r\bar{w} + (1-r)\bar{x} = A + R$.) ||

Proof of Theorem 2. The intermediary has three options—choose $\alpha = 0$ for both firms, or choose $\alpha = 1$ for both firms, or choose $\alpha = 0$ for one (arbitrarily picked) firm and $\alpha = 1$ for the other. In the last case, each i.p. will supply $\frac{1}{2}$ unit of effort. An incentive compatible contract is one that causes the intermediary to select $\alpha = 1$ for *both* firms. For simplicity, we take $K_1 = K_2 = K = [W + X]/2$ in (6). As we have already indicated, such a contract is generally suboptimal. But if intermediation is beneficial with even this suboptimal contract, it must be of value with more efficient contracts.

Thus, to obtain the optimal contract to offer *each* i.p., one solves the following problem.

$$\text{minimize}_{w,x}\ r^2\psi(w)+(1-r)^2\psi(x) \tag{A.4}$$

where $\psi(w)=W$ and $\psi(x)=X$.

Since the intermediary divides its total wealth equally among members, *each* i.p. obtains W if $\beta=1$ is detected for both firms, $K=[W+X]/2$ if $\beta=1$ is detected for only one firm, and X if $\beta=0$ is detected for both firms. The expected utilities of *each* i.p. (with the argument denoting the intermediary's choice of strategy) are

$$EU(\alpha=1 \text{ for both firms}) = r^2w+2r(1-r)k+(1-r)^2x-A \tag{A.5}$$

$$EU(\alpha=1 \text{ for one firm}) \quad = rqw+[r(1-q)+q(1-r)]k+(1-r)(1-q)x-[A/2] \tag{A.6}$$

$$EU(\alpha=0 \text{ for both firms}) = q^2w+2q(1-q)k+(1-q)^2x, \tag{A.7}$$

where $U(K)=k$.

For the intermediary to be motivated to produce information about both firms, the following constraints must be satisfied.

$$\text{IR:}\quad r^2w+2r(1-r)k+(1-r)^2x \geqq A+R \tag{A.8}$$

$$\text{IC:}\quad r(w-k)+(1-r)(k-x) \geqq A[2(r-q)]^{-1} \tag{A.9}$$

$$\text{IC:}\quad (r+q)(w-k)+(2-r-q)(k-x) \geqq A[r-q]^{-1} \tag{A.10}$$

$$\text{IS:}\quad \psi(w)+\psi(x)=2\psi(k) \tag{A.11}$$

where IS stands for Identity Satisfaction.

As $\psi(\cdot)$ is increasing in its argument, we must have $w\geqq k\geqq x$ or $x\geqq k\geqq w$ for (A.11) to be satisfied. But $x\geqq k\geqq w$ violates (A.9). So, $w\geqq k\geqq x$. Further, the strict convexity of $\psi(\cdot)$ implies that

$$w>k>x \quad \text{and} \quad w+x>2k. \tag{A.12}$$

Because $r>q$, (A.12) tells us that (A.10) will hold if (A.9) holds. Thus, (A.10) can be dispensed with and the optimization problem for the firm becomes

$$\text{minimize}_{w,x,k}\ r^2\psi(w)+2r(1-r)\psi(k)+(1-r)^2\psi(x) \tag{A.13}$$

subject to (A.8), (A.9) and (A.11).

Attaching Lagrange multipliers μ, λ and ξ to (A.8), (A.9) and (A.11) respectively, the first order optimality conditions yield

$$\psi'(w)[1-\xi r^{-2}]=\mu+\lambda r^{-1} \tag{A.14}$$

$$\psi'(k)[1+\xi\{r(1-r)\}^{-1}]=\mu-\lambda(2r-1)[2r(1-r)]^{-1} \tag{A.15}$$

$$\psi'(x)[1-\xi(1-r)^{-2}]=\mu-\lambda(1-r)^{-1}. \tag{A.16}$$

Suppose $\lambda\leqq 0$. Then, from (A.14), (A.15) and (A.16), we have

$$\psi'(w)[1-\xi r^{-2}]\leqq\psi'(k)[1+\xi\{r(1-r)\}^{-1}]\leqq\psi'(x)[1-\xi(1-r)^{-2}]. \tag{A.17}$$

Now, if $\xi\leqq 0$, then (A.17) implies that $\psi'(w)\leqq\psi'(k)$ or $k\geqq w$, which violates (A.12). If $\xi>0$, then (A.17) implies that $\psi'(k)\leqq\psi'(x)$ or $x\geqq k$, which is again infeasible because of (A.12). Thus, we must have $\lambda>0$. From (A.15) and (A.16) then, one obtains $\mu>0$.

This means both (A.8) and (A.10) are binding. The optimal solution is obtained by letting (A.8) and (A.10) be equalities and satisfying (A.11). Solving for w and x in terms of k gives us (with stars denoting optimal values and the subscript "im" signifying that the "internal monitoring" case is being considered),

$$w^*_{\text{im}} = [R + A(1+r-2q)\{2(r-q)\}^{-1} - (1-r)k^*_{\text{im}}]r^{-1} \tag{A.18}$$

$$x^*_{\text{im}} = [R + A(r-2q)\{2(r-q)\}^{-1} - rk^*_{\text{im}}](1-r)^{-1}. \tag{A.19}$$

These two equations plus (A.11) specify the optimal solution.

Recall that (A.3) describes the optimal contract with individual (nonintermediary) information production. Suppose we set

$$k^*_{\text{im}} = \bar{k} = (\bar{w} + \bar{x})/2 = R - A(2q-1)\{2(r-q)\}^{-1}. \tag{A.20}$$

Substituting (A.20) in (A.18) and (A.19) results in

$$w^*_{\text{im}} = R + A(1-q)(r-q)^{-1} = \bar{w} \tag{A.21}$$

$$x^*_{\text{im}} = R - Aq(r-q)^{-1} = \bar{x}. \tag{A.22}$$

Therefore, setting $k^*_{\text{im}} = (\bar{w} + \bar{x})/2$ makes the solution in (A.3) coincide with that in (A.18) and (A.19). But, doing so is not allowed by (A.11) which implies

$$k^*_{\text{im}} > (w^*_{\text{im}} + x^*_{\text{im}})/2 = (\bar{w} + \bar{x})/2 = \bar{k}.$$

So, increase k^*_{im} to some number above $\bar{k}$. Both w^*_{im} and x^*_{im}, in (A.18) and (A.19) respectively, will decrease then. Since $\bar{k} = (\bar{w} + \bar{x})/2$ and $\psi(\bar{w}) + \psi(\bar{x}) - 2\psi(\bar{k}) > 0$, by letting $k^*_{\text{im}} > \bar{k}$, we can reduce the quantity $\psi(w^*_{\text{im}}) + \psi(x^*_{\text{im}}) - 2\psi(k^*_{\text{im}})$ to zero, as required by (A.11). We can therefore conclude that the optimal solution is such that $w^*_{\text{im}} < \bar{w}$ and $x^*_{\text{im}} < \bar{x}$.

The expected cost per firm is

$$r^2\psi(w^*_{\text{im}}) + 2r(1-r)\psi(k^*_{\text{im}}) + (1-r)^2\psi(x^*_{\text{im}}) = r\psi(w^*_{\text{im}}) + (1-r)\psi(x^*_{\text{im}}) \quad \text{using (A.11)}$$
$$< r\psi(\bar{w}) + (1-r)\psi(\bar{x}). \quad \|$$

Proof of Theorem 3. For the firm he is entrusted with screening, each i.p. can choose α equal to either one or zero. For a Nash equilibrium, we can assume that each i.p. believes the other will choose $\alpha = 1$. Thus, given that the other i.p. chooses $\alpha = 1$, an i.p.'s expected utilities are (with the argument denoting the i.p.'s own choice of action)

$$EU(\alpha = 0) = rqw + [r(1-q) + (1-r)q]k + (1-r)(1-q)x. \tag{A.23}$$

$$EU(\alpha = 1) = r^2w + 2r(1-r)k + (1-r)^2x - A. \tag{A.24}$$

For incentive compatibility, we need $EU(\alpha = 1) \geqq EU(\alpha = 0)$ and for individual rationality, $EU(\alpha = 1) \geqq R$ is required.

Each firm designs its optimal contract by solving

$$\text{minimize}_{w,x,k}\ r^2\psi(w) + 2r(1-r)\psi(k) + (1-r)^2\psi(x) \tag{A.25}$$

subject to

$$\text{IR:} \quad r^2w + 2r(1-r)k + (1-r)^2x \geqq A + R \tag{A.26}$$

$$\text{IC:} \quad r(w-k) + (1-r)(k-x) \geqq A(r-q)^{-1}. \tag{A.27}$$

Repeating the steps in the proof of Theorem 2 we can prove that both constraints are binding. Hence, the optimal solution (w^0_{nm}, k^0_{nm}, x^0_{nm})—with the subscript "nm" identifying this as the "no monitoring case"—will satisfy (A.26) and (A.27) as equalities. Solving for w^0_{nm} and x^0_{nm} in terms of k^0_{nm} yields

$$x^0_{nm} = \{R - Aq(r-q)^{-1} - rk^0_{nm}\}(1-r)^{-1} = \{\bar{x} - rk^0_{nm}\}(1-r)^{-1}, \tag{A.28}$$

$$w^0_{nm} = \{R + A(1-q)(r-q)^{-1} - (1-r)k^0_{nm}\}r^{-1} = \{\bar{w} - (1-r)k^0_{nm}\}r^{-1}, \tag{A.29}$$

where we have used (A.3).

Therefore, the expected screening cost to each firm with the NIMJC arrangement is

$$\begin{aligned} &r^2\psi(w^0_{nm}) + 2r(1-r)\psi(k^0_{nm}) + (1-r)^2\psi(x^0_{nm}) \\ &\quad = r[r\psi(w^0_{nm}) + (1-r)\psi(k^0_{nm})] + (1-r)[r\psi(k^0_{nm}) + (1-r)\psi(x^0_{nm})] \\ &\quad > r[\psi(rw^0_{im} + (1-r)k^0_{nm})] + (1-r)[\psi(rk^0_{nm} + (1-r)x^0_{nm})] \\ &\quad = r\psi(\bar{w}) + (1-r)\psi(\bar{x}) \text{ (from (A.28) and (A.29)) (by Jensen's inequality).} \\ &\quad = \text{expected screening cost per firm with i.p.'s independent.} \quad \| \end{aligned}$$

Proof of Theorem 4. Consider an intermediary consisting of N i.p.'s, producing information for N firms. Let *each* firm pay the intermediary

$$\phi(i) = \psi(\bar{x} + (\bar{w} - \bar{x})iN^{-1}), \tag{A.30}$$

where the random variable i represents the number of firms for which $\beta = 1$ is observed, and $\bar{x}$ and $\bar{w}$ are defined in (A.3).

Note that (A.30) stipulates that each firm's payment to the intermediary depends only on the overall performance of the intermediary and not on how efficiently the intermediary is observed to screen *that* firm. We will first prove that (A.30) is incentive compatible.

Assume $\alpha = 1$ is chosen by the intermediary for M firms, with $0 < M < N$. Each i.p.'s expected utility is then

$$\textstyle\sum_{i=0}^{N} (\bar{x} + (\bar{w} - \bar{x})iN^{-1}) \text{ Prob } (\beta = 1 \text{ for } i \text{ firms} \mid \alpha = 1 \text{ for } M \text{ firms}) - MAN^{-1} \tag{A.32}$$

where we have imposed equal sharing of the total work input required. Express (A.31) as

$$\begin{aligned} &\bar{x} \textstyle\sum_{i=0}^{N} \text{Prob } (\beta = 1 \text{ for } i \text{ firms} \mid \alpha = 1 \text{ for } M \text{ firms}) \\ &\qquad + N^{-1}(\bar{w} - \bar{x}) \textstyle\sum_{i=0}^{N} i \text{ Prob } (\beta = 1 \text{ for } i \text{ firms} \mid \alpha = 1 \text{ for } M \text{ firms}) - MAN^{-1} \\ &\quad = \bar{x} + N^{-1}(\bar{w} - \bar{x})E(i \mid \alpha = 1 \text{ for } M \text{ firms}) - MAN^{-1} \end{aligned}$$

(where $E(\cdot|\cdot)$ is the conditional expectation operator)

$$\begin{aligned} &\quad = \bar{x} + N^{-1}(\bar{w} - \bar{x})[Mr + (N-M)q] - MAN^{-1} \\ &\quad = \bar{r}w + (1-r)\bar{x} - MAN^{-1} - N^{-1}(N-M)[(r-q)(\bar{w} - \bar{x})] \\ &\quad = R \text{ (by substituting for } \bar{w} \text{ and } \bar{x} \text{ from (8)).} \end{aligned}$$

Thus, no i.p.'s expected utility is affected by the choice of M. So, we can select $M = N$ and incentive compatibility is assured.

The expected cost of information production for each firm is

$$\sum_{i=0}^{N} \psi(\bar{x}+(\bar{w}-\bar{x})iN^{-1}) \text{ Prob } (\beta=1 \text{ for } i \text{ firms} \mid \alpha=1 \text{ for } N \text{ firms}). \qquad \text{(A.32)}$$

Now, i has a binomial distribution with mean rN and variance $r(1-r)N$. Thus, iN^{-1} has a mean of r and a variance of $r(1-r)N^{-1}$. As N tends to infinity, the variance of iN^{-1} goes to zero. In other words, the expression in (A.32) converges almost surely (as $N \to \infty$) to $\psi(\bar{x}+r(\bar{w}-\bar{x}))$, which is equal to $\psi(A+R)$, the screening cost per firm if the actions of the i.p.'s can be observed freely by *all* without error ex post. ||

Proof of Theorem 5. Let firms use "separate contracts" with an intermediary composed of two i.p.'s. Each firm pays the intermediary W if $\beta=1$ for that firm and X if $\beta=0$ for that firm. Suppose the intermediary arranges to split its total terminal wealth fifty-fifty between the two i.p.'s, and asks each i.p. to supply $\alpha=\frac{1}{2}$ for each firm. Let $w=U(W)$, $x=U(X)$, and $k=U([W+X]/2)$.

Each i.p. has three possible courses of action. He can choose $\alpha=\frac{1}{2}$ (his allocated effort) for each firm, or choose $\alpha=\frac{1}{2}$ for one firm and $\alpha=0$ for the other, or choose $\alpha=0$ for both firms. If the total effort expended by the intermediary as a whole in screening a firm is $\alpha=1$, $\beta=1$ will be observed with a probability of r, and the probability of $\beta=1$ is q if $\alpha<1$. In a Nash equilibrium, each i.p. assumes the other i.p. will choose $\alpha=\frac{1}{2}$ for each firm. The expected utilities of each i.p., with the three action choices are (with the argument denoting this i.p.'s action choice)

$$EU(\alpha=\tfrac{1}{2} \text{ for both firms}) = r^2w+2r(1-r)k+(1-r)^2x-A, \qquad \text{(A.33)}$$

$$EU(\alpha=\tfrac{1}{2} \text{ for only one firm}) = rqw+[(1-r)q+q(1-r)]k+(1-r)(1-q)-A/2, \qquad \text{(A.34)}$$

$$EU(\alpha=0 \text{ for both firms}) = q^2w+2q(1-q)k+(1-q)^2x. \qquad \text{(A.35)}$$

These expected utilities are identical to those achievable by the i.p.'s in an intermediary endowed with costless internal monitoring capability (see A.5)–(A.7) in the Proof of Theorem 2). Hence, HEA with separate contracts is equivalent to costless internal monitoring with separate contracts. Similar arguments can be used to prove the equivalence between the HEA and the internal monitoring cases with joint contracts. ||

First version received September 1982; *final version accepted January* 1984 (*Ed.*).

The helpful comments of Stuart Greenbaum, two anonymous referees and Professor Oliver Hart are gratefully acknowledged. Suggestions received at a seminar at the University of Wisconsin-Madison have also helped.

NOTES

1. Although various approaches have been tried to explain financial intermediation (see Benston (1976), and in particular, the review paper by Baltensperger (1980)), the most promising line of attack appears to be the recognition of informational asymmetries. Both Leland and Pyle (1977) and Diamond (1984) use this approach. The similarities and differences between these papers and ours will become apparent as we go along.
2. Because reputation is a multiperiod concept and our model is single period, we do not model reputation directly. But the noisy indicator that determines i.p. compensation in our model serves the same economic function as reputation.
3. In other words, a fully revealing rational expectations equilibrium is assumed to exist in the exchange market for firm shares, so that traders do not produce information.
4. This assumes that the true values of different firms are spaced sufficiently far apart relative to screening costs, so that a full screening equilibrium obtains. See Stiglitz (1975) for illustrations of full screening and partial screening equilibria. Note that our arguments are also valid for a partial screening equilibrium.
5. It seems almost paradoxical that we "punish" the i.p. because our noisy indicator claims he did not produce reliable information, and yet (because an incentive compatible contract is used) we actually have

complete faith that the i.p. has revealed the firm's true value! But that is the nature of incentive compatible compensation policies—to induce "honest" behaviour ex ante, it should be clear to all that the threat of punishment will be executed ex post if aberrant behaviour is detected, even though it is known both ex ante and ex post that the agent has behaved honestly.

6. Note that the two expectations in this inequality are taken with respect to different probability measures. See equations (1) and (2).

7. The assumption that work is shared equally within the intermediary does *not* imply that screening is assumed to be a divisible activity. We are *not* assuming that the intermediary can necessarily divide the total work required to identify a firm's value and assign it to more than one i.p. All that is meant is that if information is produced for $n(<N)$ firms by n i.p.'s, so that $N-n$ i.p.'s are idle, the net welfare is identical. This requires the active i.p.'s to be paid more than the idle i.p.'s by an amount that exactly offsets the effort disutility from being active. The importance of this "flexibility" afforded by perfect internal monitoring will be evident in Section 5 where divisibility of screening is assumed.

8. We wish to emphasize that there is a distinction between *separate* contracts used for an intermediary and individual contracts used for i.p.'s functioning independently (not as members of a group). When a firm negotiates a separate contract, it recognizes that it is dealing with an intermediary; the term "separate"—as opposed to "joint"—is used merely to stress that with separate contracts the intermediary's payoff from each firm is contingent *only* on its assessed proficiency in screening that firm, whereas joint contracts make the intermediary's payoff from each firm depend on its overall performance. Both joint and separate contracts refer to intermediary compensation schemes. By contrast, when i.p.'s function independently, each i.p. deals with only one firm and every firm negotiates a distinct contract with one i.p. In this regard, separate (intermediary) contracts may look similar to individual (nonintermediary) contracts for independent i.p.'s, but the fundamental difference separating them is that in the former case the ability to engage in ex post payoff pooling makes each i.p.'s effective compensation potentially different from that embodied in the relevant separate contract itself, whereas in the latter case the i.p.'s contractual remuneration coincides with his effective compensation. In the subsequent analysis, the only intermediary contracts used are joint contracts. Hence, comparisons of expected screening costs are made between joint contracts and (nonintermediary) individual contracts used to compensate independent i.p.'s.

9. If the noisy indicator is also directly costly, then π_1 can be viewed as the (positive) difference between the direct costs of the perfect and noisy indicators. See Diamond (1984) and Townsend (1979) for other analyses that assume perfect but costly monitoring.

10. This should be compared with Diamond's (1984) condition for intermediation. There are, however, many differences between Diamond's model and ours. Diamond assumes that the manager supplies no effort (unlike the i.p. in our model), and that all managers are risk neutral (unlike our i.p. who has a general concave utility). With this framework, Diamond shows diversification to be of value even for risk neutral agents. The key assumption is that managers face nonpecuniary penalties upon failure. This penalty, which represents a dissipative cost, introduces concavity ("induced" risk aversion) in the objective functions of otherwise risk neutral agents and provides a role for diversification. For another recent paper on this subject, see Boyd and Prescott (1983).

11. Even with perfect divisibility in screening, the HEA case has one bothersome problem relative to perfect and costless internal monitoring. This is the problem of multiple equilibria. The Nash equilibrium in Theorem 5 is one in which every i.p.'s optimal action choice is $\alpha=1$, if he assumes every other i.p. will choose $\alpha=1$. But if each i.p. assumes every other i.p. will choose $\alpha=0$, it will be privately optimal for him to choose $\alpha=0$. Hence, $\alpha=0$ is also a Nash equilibrium. This problem of multiple equilibria does not exist with perfect internal monitoring.

REFERENCES

AKERLOF, G. A., (1970), "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, **84**, 488–500.

BAIMAN, S. and DEMSKI, J. S., (1980), "Economically Optimal Performance Evaluation and Control Systems", *Journal of Accounting Research* (Supplement), **18**, 184–220.

BALTENSPERGER, E. (1980), "Alternative Approaches to the Theory of the Banking Firm", *Journal of Monetary Economics*, **6**, 1–37.

BENSTON, G. (1976), "A Transactions Cost Approach to the Theory of Financial Intermediation", *The Journal of Finance*, **31**, 215–231.

BHATTACHARYA, S. (1979), "Imperfect Information, Dividend Policy, and the 'Bird in the Hand' Fallacy", *The Bell Journal of Economics*, **10**, 259–270.

BHATTACHARYA, S. (1980), "Nondissipative Signalling Structures and Dividend Policy", *Quarterly Journal of Economics*, **95**, 1–24.

BOYD, J. and PRESCOTT, E. (1983), "Financial Intermediaries" (Working Paper, University of Minnesota).

DIAMOND, D. (1984), "Financial Intermediation and Delegated Monitoring", *The Review of Economic Studies* (forthcoming).

GROSSMAN, S. J. (1976), "On the Efficiency of Competitive Stock Markets Where Traders Have Diverse Information", *The Journal of Finance*, **31**, 573–585.

GROSSMAN, S. J. and STIGLITZ, J. E. (1980), "On the Impossibility of Informationally Efficient Markets", *The American Economic Review*, **70**, 393–408.

HOLMSTROM, B. (1979), "Moral Hazard and Observability", *The Bell Journal of Economics*, **10**, 74–91.

HOLMSTROM, B. (1982), "Moral Hazard in Teams", *The Bell Journal of Economics*, **13**, 324–340.

LELAND, H. E. and PYLE, D. H. (1977), "Informational Asymmetries, Financial Structure and Financial Intermediation", *The Journal of Finance*, **32**, 371–387.

NIEHANS, J. (1980) *A Theory of Money* (Johns Hopkins University Press).

ROSS, S. A. (1977), "The Determination of Financial Structure: The Incentive Signalling Approach", *The Bell Journal of Economics*, **8**, 23–40.

SPENCE, M. A. (1974), "Competitive and Optimal Responses to Signals: Analysis of Efficiency and Distribution", *Journal of Economic Theory*, **7**, 296–332.

STIGLITZ, J. E. (1975), "The Theory of 'Screening', Education and the Distribution of Income", *The American Economic Review*, **65**, 283–300.

TOWNSEND, R. M. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification", *Journal of Economic Theory*, **21**, 265–293.