Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out or Peer Monitoring?

By RICHARD ARNOTT AND JOSEPH E. STIGLITZ*

We examine a situation in which insurance is characterized by moral hazard. When market insurance is provided, supplementary mutual assistance between family and friends (unobservable to market insurers) will occur. When nonmarket insurers have no better information than market insurers, the mutual assistance not only crowds out market insurance but is also harmful and therefore dysfunctional. Alternatively, when nonmarket insurers can observe each other's effort perfectly, mutual assistance is beneficial. These results point to the potential importance of peer-monitoring mechanisms in mitigating moral hazard. (*JEL* 026)

The economics literature over the past 15 years has directed attention to the ubiquity of moral-hazard and incentives problems. One way the market responds to moral hazard is to provide only partial insurance, since then individuals still have *some* incentive to take care to avoid the accident. However, they must then bear more risk than they would like. A principal function of many nonmarket institutions, meanwhile, is to help those who have suffered some misfortune, which entails the provision of insurance: the marriage vows formalize and sanctify the mutual insurance aspects of the family; the acid test of a friend is his willingness to help in times of need; charity is regarded as meritorious and is subsidized by the government; and many government social assistance programs, such as unemployment insurance and workmen's compensation, have a strong insurance component. The importance of nonmarket insurance is illustrated by what happens if an individual catches pneumonia as a result of going on a hiking trip with inadequate rain gear. His employer gives him compensated sick leave; part or all of his medical expenses are reimbursed by his insurance policy or the state; uncovered medical expenses may be partially deductible from his income tax; and family and friends rally round to provide other forms of support. Such extensive support, while directly helpful, deleteriously affects individuals' care to avoid accidents. In terms of the example, had the individual borne all the costs of catching pneumonia himself, he might have taken the trouble to carry adequate rain gear. Thus, it is not obvious that the insurance provided by nonmarket institutions is always beneficial or, more specifically, whether nonmarket insurance institutions, when they supplement market insurance, improve the economy's ability to handle the moral-hazard trade-off between risk-bearing and incentives.

We address this issue by inquiring whether the reciprocal provision of insurance within families and between friends, which we term nonmarket insurance,[1] is welfare-improving when it supplements market insurance. We

*Arnott: Department of Economics, Boston College, Chestnut Hill, MA 02167; Stiglitz: Department of Economics, Stanford University, Stanford, CA 94305-6072. Financial support from the National Science Foundation, the Olin Foundation, and the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged. We thank the referees and the seminar participants at Stanford University, in particular Ken Judd and Suzanne Scotchmer, for very helpful comments.

[1]The term "social insurance" is perhaps more appropriate but is used in some countries to refer to social security. The term "informal insurance" is appropriate for the example, but the phenomenon we identify arises in formal, nonmarket institutions as well.

assume that a market insurer can observe his clients' market insurance purchases but not the nonmarket insurance they obtain through informal arrangements. We first show that nonmarket insurance will always be provided. Moral hazard causes fully insured individuals to expend too little effort. In response, the competitively determined market insurance contract rations the amount of insurance that can be obtained at the equilibrium price. The contract achieves this by specifying both a price and a quantity and by stipulating that it will pay in the event of accident only if the insured has no additional market insurance. Insured individuals would like to obtain additional insurance at the market price. Since they take the market insurance contract as given, they perceive that they can effectively do so by entering into informal insurance arrangements. They neglect that when everyone enters into such arrangements, the accident frequency will change, as will the market insurance contract.

After showing that nonmarket insurance will always be provided, we ask whether the level of expected utility is higher with such insurance than without it. The moral-hazard problem arises because of the inability of the insurance firm to monitor the actions of the insured. There are often other individuals, such as the members of the insured's family, who are in a better position to monitor the insured's action than the insurance firm. The welfare consequences of nonmarket insurance turn out to depend on whether these monitoring capabilities can be effectively harnessed to reduce the moral-hazard problem. If the members of the family (the providers of nonmarket insurance) do not monitor each other, we show that such nonmarket insurance always lowers welfare. The nonmarket insurance leads individuals to take less care; market insurance firms respond by providing less insurance; thus, in the new equilibrium, nonmarket insurance displaces market insurance. Since nonmarket insurance involves less risk-pooling, welfare is reduced; the less-effective insurance crowds out the more-effective insurance. The nonmarket insurance is dysfunctional.

If, alternatively, the members of the family do monitor each other, they will take greater care than they would without monitoring, which mitigates the moral-hazard problem. Thus, there appear to be two offsetting effects, the risk-pooling advantages of market insurance versus the monitoring advantages of nonmarket insurance. However, it turns out that, with perfect monitoring within the family, the latter effect dominates, and the nonmarket insurance is welfare-enhancing.

This is an example of what we call *peer monitoring*. Peer monitoring is an important mechanism for controlling moral hazard. It arises in credit markets. In less-developed countries, loans are often made to groups of individuals; the members of a group then have an incentive to monitor each other. In developed countries, one of the functions of co-signers on loans is to provide additional monitoring. Partnership arrangements also encourage monitoring. Peer monitoring is also important in labor markets; workers are often in a better position to monitor their co-workers than are employers, which may be one of the advantages of team production.

This raises an intriguing issue related to mechanism design with a principal and many agents when moral hazard is present. The principal can monitor his agents himself (direct monitoring), hire a supervisor to monitor his agents (supervision), or set up a mechanism to induce agents to monitor each other (indirect monitoring). While the literature has considered direct monitoring (e.g., Steven Shavell, 1979; Bengt Holmstrom, 1979) and supervision (e.g., James Mirrlees, 1976; Stiglitz, 1975; Jean Triole, 1988), it has largely ignored the design of indirect monitoring systems.[2] An indirect monitoring system will encourage peer monitoring through the creation of interdependence: the dependence of one agent's utility on others' effort. A good example is a university department. A faculty member's utility depends not only on his salary, perfor-

[2]One exception is H. Lorne Carmichael (1988). He considers peer review in the university setting, which is a form of indirect monitoring system, and argues that the institution of tenure is needed to make peer review incentive-compatible.

mance, reputation, and working conditions, but also directly on his department's and university's reputation. Furthermore, his salary, performance, reputation, and working conditions all depend to some extent on the department's quality. For these reasons, faculty members have a strong incentive to monitor one another's performance, which they do through peer review, teaching evaluations, and so on. In other contexts, production may be physically organized to facilitate peer monitoring; the open office, the assembly line, and team production are examples. In this paper, we treat the peer-monitoring system as exogenous and examine only the extreme cases in which there is either perfect monitoring or no monitoring.

I. The Basic Model without Nonmarket Insurance

Moral hazard is an asymmetric-information phenomenon, and its defining characteristic is hidden action. In the context of insurance, the probability distribution of observable outcomes depends on the insured's unobservable actions. The insurer would like to write insurance contingent on the insured's actions but, since these are unobservable, must base his insurance on observable outcomes. Because of moral hazard, there is a trade-off between risk-bearing and incentives in the provision of insurance. At one extreme, if full insurance is provided,[3] the insured is equally well off whatever the outcome and therefore has no incentive to take care. At the other extreme, if no insurance is provided, the individual faces the appropriate incentives but is fully exposed to risk.

We first describe the canonical moral-hazard model without nonmarket insurance (Arnott and Stiglitz, 1988a). There is a single, fixed-damage accident. The probability of its occurrence, p, depends on the individual's effort at accident avoidance, e. The probability-of-accident function is strictly convex: $p' < 0$, $p'' > 0$. The individual's wealth is w, and d is the damage caused by the accident.

If an accident occurs, the individual receives a (net of premium) payout of α, and his consumption is

$$\text{(1a)} \qquad y_1 = w - d + \alpha.$$

If an accident does not occur, he pays the premium β, and his consumption is

$$\text{(1b)} \qquad y_0 = w - \beta.$$

For simplicity, we assume a separable, event-independent utility function.[4] Expected utility is then

$$\text{(2)} \quad \mathrm{EU} = u(y_0)(1-p) + u(y_1)p - e$$

with $u' > 0$, $u'' < 0$.

The individual chooses effort so as to maximize expected utility, taking α and β as given:[5]

$$\text{(3)} \quad e(\alpha,\beta) = \operatorname*{argmax}_{\{e\}} \mathrm{EU}(\alpha,\beta,e).$$

It is straightforward to show that, with the form of utility function assumed, $\partial e/\partial\alpha < 0$ and $\partial e/\partial\beta < 0$; as more insurance is provided, the individual reduces effort. Substitution of (3) into (2) gives $V(\alpha,\beta) \equiv \mathrm{EU}(\alpha,\beta,e(\alpha,\beta))$, from which an individual's indifference curves in $\alpha-\beta$ space can be plotted (see Fig. 1). The slope of an indifference curve is

$$\text{(4)} \qquad \left.\frac{d\beta}{d\alpha}\right|_{\bar{V}} = \frac{u_1'p}{u_0'(1-p)}$$

where $u_0 \equiv u(y_0)$ and $u_1 \equiv u(y_1)$. The locus of (α,β) for which there are zero profits, the zero-profit locus (ZPL), is $(1-p)\beta - p\alpha = 0$. Substitution of (3) into this equation

[3]Full insurance—equalization of the marginal utilities of income in all events—implies no incentive to take care when the utility function is event-independent, which we assume, but not generally otherwise.

[4]The general form of the utility function with event i is $U_i(y_i,e)$. Separability implies $U_i(y_i,e) = u_i(y_i) - e$. Event-independence implies that the utility-from-consumption function $u_i(y)$ is independent of the event; the accident causes neither pain nor pleasure and does not alter tastes.

[5]Throughout the paper, we ignore the complications that arise from the possibility that $e = 0$.

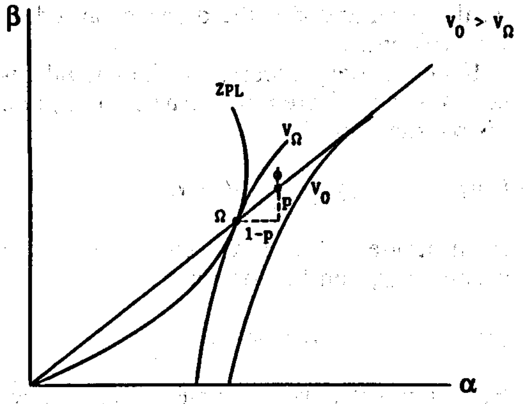

FIGURE 1. THE INDIVIDUAL PERCEIVES THAT HE CAN OBTAIN ADDITIONAL INSURANCE AT THE MARKET PRICE BY ENTERING INTO A MUTUAL INSURANCE PACT

gives the ZPL as a function of α and β, which is plotted in Figure 1. Its slope is

$$(5)\quad \left.\frac{d\beta}{d\alpha}\right|_{\mathrm{ZPL}} = \frac{p+(\alpha+\beta)p'(\partial e/\partial\alpha)}{(1-p)-(\alpha+\beta)p'(\partial e/\partial\beta)}.$$

In Figure 1, Ω is the point of optimal insurance, contingent on the unobservability of effort;[6] it occurs at the point of maximum utility on the zero-profit locus.

We now investigate competitive equilibrium for this model with insurance purchases observable (Arnott and Stiglitz, 1988a, b). Define $q \equiv \beta/\alpha$ to be the price of insurance. Since Ω is on the ZPL, $q_\Omega \equiv (\beta/\alpha)_\Omega = [p/(1-p)]_\Omega$. The precise shapes of the indifference curves and the ZPL are inessential for our analysis.[7] What is essential, however, is that at Ω the slope of the ZPL exceeds the price of insurance. Since indifference curves are positively sloped and since Ω is at a point of tangency of an indifference curve and the ZPL, the ZPL must be positively sloped at Ω. Now consider increasing α by one unit. To maintain zero profits, β must be increased by more than the price of insurance, $q_\Omega = [p/(1-p)]_\Omega$. An increase in β of $[p/(1-p)]_\Omega$ causes zero profit to be made on the *marginal* unit of insurance; but the increase in α raises the probability of accident, and so β has to be increased further to offset losses on *inframarginal* units of insurance. Hence,

$$(6)\quad \left(\left.\frac{d\beta}{d\alpha}\right|_{\bar V}\right)_\Omega = \left(\left.\frac{d\beta}{d\alpha}\right|_{\mathrm{ZPL}}\right)_\Omega > q_\Omega.$$

Since at Ω the slope of the indifference curve exceeds the price of insurance, individuals would like to acquire more insurance at the price q_Ω (see Pauly, 1974). Thus, competitive decentralization of Ω requires rationing of insurance at the price q_Ω.[8] The intuition for this result is that, because of moral hazard, at any price of insurance, the individual expends too little effort. Rationing induces him to increase his effort. The easiest[9] way for such rationing to be

[6] Ω is the optimal *deterministic* contract. The circumstances in which randomization is Pareto-improving are investigated in Arnott and Stiglitz (1988c).

[7] Indifference curves are positively sloped and can be nonconvex

$$\left(\left.\frac{d^2\beta}{d\alpha^2}\right|_{\bar V} > 0\right)$$

as can be seen from (4). As the amount of insurance provided increases, holding utility fixed, u_1'/u_0' falls, but since more insurance causes the probability of accidents to rise, $p/(1-p)$ increases. The zero-profit locus is positively sloped for small amounts of insurance but may bend backwards.

[8] The point Ω can be decentralized by the insurer offering any locus of (α,β) such that Ω is the point of maximum utility on the locus and insisting that he be the sole insurer. One such possibility entails the insurer offering the zero-profit locus. This particular form of nonlinear pricing in this context was considered by Elhanan Helpman and Jean-Jacques Laffont (1975). The analysis of the paper carries through whatever form of pricing is employed to decentralize Ω.

[9] Insurers could alternatively agree to sell a client who has purchased a total amount of insurance from other insurers of (α',β') an insurance policy $(\hat\alpha,\hat\beta) = (\alpha_\Omega - \alpha', \beta_\Omega - \beta')$, which would bring his total insurance up to $(\alpha_\Omega, \beta_\Omega)$. However, this is unnecessarily complicated. In fact, almost all insurance policies contain exclusivity provisions (life and air-flight accident insurance are the exceptions, but moral hazard is unimportant for both types of insurance).

accomplished is for each insurer to offer the contract $(\alpha_{\Omega}, \beta_{\Omega})$, conditional on its clients purchasing no additional insurance. This condition, which we term the exclusivity provision, is enforceable, since it is assumed that insurers are able to observe all insurance purchases.

To sum up, with moral hazard, when nonmarket insurance is absent and insurance purchases are observable, the social optimum conditional on the unobservability of effort is decentralizable. Competitive equilibrium entails each individual being restricted to purchase all his insurance from a single insurer and being rationed in the amount of insurance he can purchase at the equilibrium price.

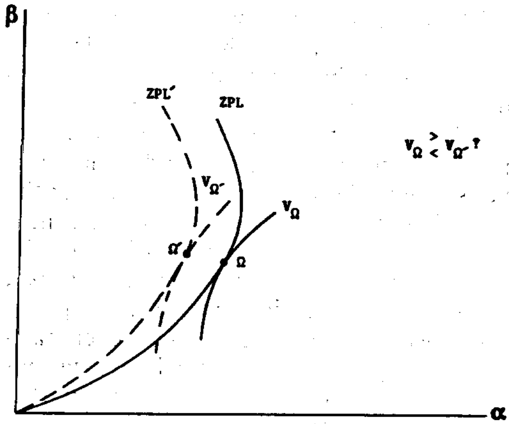

FIGURE 2. THE EFFECT OF NONMARKET INSURANCE WHEN IT REDUCES EFFORT

II. Effort Unobservable by the Nonmarket Insurer

We now consider the simplest possible extension of this model that allows for the simultaneous provision of market and nonmarket insurance. We assume that, although an insurer can observe his clients' insurance *purchases*, he cannot observe the nonmarket insurance they acquire.[10]

As described earlier, nonmarket insurance may be provided through many institutions. To simplify, we treat only nonmarket insurance provided reciprocally by pairs of symmetrical individuals. We label the two partners in a pair, H (husband) and W (wife). The two partners have the same tastes and probability-of-accident functions, and their accident probabilities are statistically independent.[11] The form of the nonmarket insurance is as follows: H and W agree that if one spouse has an accident and the other does not, the latter will transfer δ to the former.

The issues to be addressed can be posed in terms of Figure 2. The solid-lined ZPL and V_{Ω} are the zero-profit locus and equilibrium-indifference curve in the absence of nonmarket insurance (the same as in Fig. 1); the dashed-lined ZPL and $V_{\Omega'}$ are the corresponding curves with nonmarket insurance. Assume, for the sake of argument, that the provision of nonmarket insurance reduces effort. There are then two offsetting effects on equilibrium utility. On one hand, with α and β held fixed, the provision of nonmarket insurance increases utility; it shifts indifference curves to the left. On the other hand, since by assumption the provision of nonmarket insurance increases the probability of accident, the insurance firm must lower the payout for any level of the premium to continue making zero profits; as a result, the ZPL also shifts to the left. The analysis that follows examines which effect is dominant.

In this section, we characterize the equilibrium for the case in which an individual's accident-avoidance effort is observable by neither his partner nor market insurers. Equilibrium may entail a combination of market and nonmarket insurance. Subsequently, we shall investigate the efficiency properties of the equilibrium.

[10] If an insurer were able to observe the nonmarket insurance that his clients acquire, he could write the contract contingent on their nonmarket insurance. The resulting equilibrium would be efficient conditional on the information available (the efficiency result depends on there being only one commodity and one type of accident; see Arnott and Stiglitz, 1986, 1989).

[11] This assumption simplifies the analysis and does not affect the qualitative results. Depending on context, the accident probabilities of the partners may be positively correlated.

There are four events: 1) neither the individual nor his partner has an accident; 2) the individual has an accident, but his partner does not; 3) only the partner has an accident; 4) both have accidents. Let e denote the individual's effort and $\tilde{e}$ his partner's. Then, the probability that neither the individual nor his partner has an accident is $[1-p(e)][1-p(\tilde{e})]$ and similarly for the other events. We assume that the market insurer sells individual rather than group insurance policies[12] (i.e., he sells each partner a policy (α,β); if neither individual has an accident, each has to pay the market insurer the premium β, etc).

Thus, an individual's expected utility is

$$(7)\quad \begin{aligned} \mathrm{EU} = {} & u(w-\beta)[1-p(e)][1-p(\tilde{e})] \\ & + u(w-d+\alpha)p(e)p(\tilde{e}) \\ & + u(w-\beta-\delta)[1-p(e)]p(\tilde{e}) \\ & + u(w-d+\alpha+\delta)p(e) \\ & \times[1-p(\tilde{e})]-e \end{aligned}$$

which may be written more succinctly as

$$(7')\quad \begin{aligned} \mathrm{EU} = {} & u_0(1-p)(1-\tilde{p})+u_1 p\tilde{p} \\ & + u_2(1-p)\tilde{p}+u_3 p(1-\tilde{p})-e \end{aligned}$$

where $u_0 \equiv u(w-\beta)$, $u_1 \equiv u(w-d+\alpha)$, $u_2 \equiv u(w-\beta-\delta)$, $u_3 \equiv u(w-d+\alpha+\delta)$, and $\tilde{p} \equiv p(\tilde{e})$.

We assume that H and W are smart and take into account how the other will adjust effort in response to a change in δ.[13] Both assume that the market contract will be unaffected by their actions, which is perfectly reasonable in the atomistic environment we envisage.

We adopt the Nash assumption. W, in deciding on her level of effort, takes α, β, and δ as well as H's effort as fixed. W believes that H is rational and selfish and will accordingly choose the level of effort that maximizes his expected utility, given that W is acting similarly.[14] Then, from (7), the equation characterizing her level of precaution is

$$(8)\quad \begin{aligned} & [-u_0(1-\tilde{p})+u_1\tilde{p}-u_2\tilde{p} \\ & \qquad + u_3(1-\tilde{p})]p'-1=0 \end{aligned}$$

which gives

$$(9a)\qquad e=\hat{e}(\alpha,\beta,\delta,\tilde{p})$$

and by symmetry

$$(9b)\qquad \tilde{e}=\hat{e}(\alpha,\beta,\delta,p).$$

Combining (9a) and (9b) yields

$$(10a)\qquad e=e(\alpha,\beta,\delta)$$

$$(10b)\qquad \tilde{e}=e(\alpha,\beta,\delta).$$

From (7), the individual and his partner perceive expected utility to be related to δ in the following way:

$$(11)\quad \begin{aligned} \frac{\partial \mathrm{EU}}{\partial\delta} = {} & [-u_2'(1-p)\tilde{p}+u_3'p(1-\tilde{p})] \\ & +\{[-u_0(1-\tilde{p})+u_1\tilde{p}-u_2\tilde{p} \\ & \quad +u_3(1-\tilde{p})]p'-1\}\frac{\partial e}{\partial\delta} \\ & +[-u_0(1-p)+u_1 p \\ & \quad +u_2(1-p)-u_3 p]\tilde{p}'\frac{\partial\tilde{e}}{\partial\delta}. \end{aligned}$$

[12] Where the insurance company cannot identify the nonmarket insurance partners (friends, for example), this is the natural assumption. However, in other contexts, notably the family, the market typically provides group policies. We have run through the analysis when the market provides group rather than individual policies. Group policies are Pareto-superior, since they contain an extra policy parameter: an insurance policy specifies the group premium payable if neither has an accident, the net payout if only one has an accident, and the net payout if both have an accident. In other respects, the qualitative results are the same when the market provides group policies as when it provides individual policies.

[13] We would obtain the same qualitative results if we assumed instead that H and W ignore that the other will adjust effort in response to a change in δ.

[14] This is not an infinitely repeated game. If it were, the cooperative outcome might be obtained. We comment on this later.

In so doing, they neglect that, since other couples too behave in this way, insurance companies adjust α and β in response to a change in δ. Combining (8), (10a), (10b), and (11), and noting that the equilibrium is symmetric, gives

$$(12)\quad \frac{\partial \mathrm{EU}}{\partial \delta}=(-u_2'+u_3')(1-p)p + [1+(u_2-u_3)p']\frac{\partial e}{\partial \delta}.$$

Furthermore, from (8),

$$(13)\quad \frac{\partial e}{\partial \delta} = -\frac{[u_2'p+u_3'(1-p)]p'}{(p''/p')+(p')^2(u_0+u_1-u_2-u_3)} < 0.$$

At Ω (the competitive equilibrium in the absence of nonmarket insurance), $\delta=0$, $1+(u_2-u_3)p'=0$ [from (8) since $u_0=u_2$ and $u_1=u_3$], and $-u_2'+u_3'>0$ (incomplete insurance). Hence, from (12),

$$(14)\quad \left.\frac{\partial \mathrm{EU}}{\partial \delta}\right|_{\Omega}=(-u_2'+u_3')(1-p)p>0.$$

Thus, at the competitive equilibrium in the absence of nonmarket insurance, the partners perceive a mutual insurance pact to be beneficial and would therefore provide one another with nonmarket insurance to supplement their market insurance. The intuition for this result is as follows. At Ω, the partners are rationed in the amount of insurance they can purchase at the price q_{Ω}. Each perceives that, by entering into a mutual insurance pact, he can acquire additional insurance at this price, which pays out when he, but not his partner, suffers an accident. More specifically, at Ω, since

$$\frac{\partial \mathrm{EU}}{\partial \alpha}=pu_3' \quad \text{and} \quad \frac{\partial \mathrm{EU}}{\partial \beta}=-(1-p)u_2'$$

from (12),

$$(15)\quad \frac{\partial \mathrm{EU}}{\partial \delta}=(1-p)\frac{\partial \mathrm{EU}}{\partial \alpha}+p\frac{\partial \mathrm{EU}}{\partial \beta}.$$

An individual regards a unit increase in δ as equivalent to a unit increase in α with probability $1-p$ (the probability that his partner is not sick when he is) combined with a unit increase in β with probability p (the probability that his partner is sick when he is not), or equivalently, as an expected increase of $1-p$ in the amount of insurance obtained at the price q (i.e., movement from Ω to ϕ in Fig. 1). As already noted, in reasoning in this way, individuals neglect that, when everyone enters into such a pact, which reduces effort [eq. (13)] and increases the probability of accident, market insurers are forced to offer a less attractive contract in order to maintain zero profits.

H and W choose δ to maximize their expected utilities, taking α and β as given. From $\partial \mathrm{EU}/\partial\delta=0$, $e=\bar{e}$, and (8), one obtains $\delta=\delta(\alpha,\beta)$. By observing how the probability of accident responds to changes in α and β, market insurers will implicitly take into account that δ responds to α and β according to $\delta=\delta(\alpha,\beta)$. Competition, meanwhile, will continue to result in the equilibrium market contract maximizing expected utility subject to zero profits. Thus, in the presence of nonmarket insurance, the equilibrium market contract maximizes

$$(16)\quad \mathrm{EU}=u(w-\beta)(1-p)^2 + u(w-d+\alpha)p^2 + u(w-\beta-\delta)(1-p)p + u(w-d+\alpha+\delta)p(1-p)-e$$

subject to

$$\text{(i)}\quad \beta(1-p)-\alpha p=0$$

$$\text{(ii)}\quad e=e(\alpha,\beta,\delta(\alpha,\beta))$$

where (ii) is obtained by combining (10a) and $\delta=\delta(\alpha,\beta)$.

Given the assumed information technology, it can be shown that the nonmarket insurance is unambiguously harmful and dysfunctional. The line of proof is straightforward: welfare is at least as high if the market insurer chooses α, β, and δ as if he chooses just α and β, with δ being chosen by the nonmarket insurer; and if the market insurer chooses α, β, and δ, he will set $\delta = 0$.

The equilibrium without nonmarket insurance cannot be improved upon, and if it were possible, it would be desirable to outlaw the provision of nonmarket insurance. The intuitive rationale for this result is as follows. The provision of nonmarket insurance does not enhance the risk-sharing capabilities of the economy. Rather, such insurance crowds out market insurance. Not only is it less effective than market insurance since it randomizes an individual's event-contingent consumption and is provided by a risk-averse agent (see John M. Marshall, 1976), but also the simultaneous provision of market and nonmarket insurance violates exclusivity, which typically creates uninternalized externalities (see Arnott and Stiglitz, 1989).

The above analysis was predicated on the assumptions that a market which provides insurance against the accident in question exists and that there are no transaction costs associated with the provision of insurance. If market insurance against a given accident does not in fact exist, voluntary nonmarket insurance is unambiguously beneficial.[15] When transactions costs are present, nonmarket insurance may be beneficial if it is provided at lower transaction cost than market insurance.

III. Effort Observable by the Nonmarket Insurer

This case is more interesting, since there appear to be two offsetting effects. On one hand, because individuals have information on their partner's effort, which an insurance company does not, the provision of nonmarket insurance has the potential of enhancing the risk-sharing capabilities of the economy. On the other hand, the provision of insurance by a risk-neutral agent is typically more efficient than provision by a risk-averse agent, if they have access to the same information. Furthermore, the simultaneous provision of market and nonmarket insurance violates exclusivity. This line of reasoning suggests that the provision of nonmarket insurance in this case may be beneficial in some circumstances and harmful in others.

We continue with the same model. When effort is observable within the family but not to the insurance firm, and when, as we have assumed, individuals are identical, family members will effectively choose the level of precaution to take cooperatively. Each will take α and β to be fixed and choose δ and e to maximize

$$(17)\quad \mathrm{EU} = u_0(1-p)^2 + u_1 p^2 + u_2(1-p)p + u_3(1-p)p - e.$$

This yields the following first-order conditions:

$$(18a)\quad e:\ \left[-2(1-p)u_0 + 2pu_1 + (1-2p)(u_2+u_3)\right]p' = 1$$

$$(18b)\quad \delta:\ (-u_2' + u_3')p(1-p) = 0.$$

Equation (18b) implies that

$$(18b')\quad \delta = \frac{d-\alpha-\beta}{2}.$$

Because the partners can observe each other's effort and treat α and β as fixed, they perceive there to be no moral-hazard problem associated with the insurance they provide and hence provide full insurance (or as full as possible). This stands in contrast to the previous section where, as a result of the inability of each partner to observe the other's effort, only partial nonmarket insurance was provided [see (12)].

The insurance firm effectively chooses α and β to maximize expected utility, subject to (18a), (18b), and the zero-profit constraint. The competitive equilibrium with nonmarket insurance is characterized by the constraints and first-order conditions of this program.

[15] Market insurance is generally unavailable for the multitudinous small (but cumulatively substantial) risks faced in everyday life. How nonmarket institutions handle such risks is an important and interesting question.

We now investigate the welfare properties of the equilibrium. To do this, we assume that the planner chooses α, β, and δ, knowing that individuals choose e according to (18a), which takes account of the fact that δ is chosen with effort observable, and subject to the break-even constraint on market insurance.

Substituting the zero-profit constraint into (17) gives

$$\text{(19)}\quad \mathrm{EU}(\beta,\delta)=u(w-\beta)(1-p)^2 + u\left(w-d+\frac{\beta(1-p)}{p}\right)p^2 + u(w-\beta-\delta)p(1-p) + u\left(w-d+\frac{\beta(1-p)}{p}+\delta\right)\times p(1-p)-e.$$

The corresponding first-order condition for δ [using (18a)] is

$$\text{(20)}\quad \frac{\partial \mathrm{EU}}{\partial\delta}=(-u_2'+u_3')p(1-p) - \left(\beta u_1'p'+\frac{\beta(1-p)}{p}u_3'p'\right)\frac{\partial e}{\partial \delta}=0.$$

Using (18a) again,

$$\text{(21)}\quad \frac{\partial e}{\partial\delta}=\frac{(1-2p)(u_3'-u_2')}{\Delta}$$

where

$$\Delta=\frac{\beta p'}{p^2}[2pu_1'+(1-2p)u_3'] -2p'(u_0+u_1-u_2-u_3)-\frac{p''}{(p')^2}.$$

Substituting (21) into (20) gives

$$\text{(20')}\quad \frac{\partial \mathrm{EU}}{\partial\delta}=\frac{(u_3'-u_2')}{\Delta}\times\left[\beta u_1'p'-2p'p(1-p)\times(u_0+u_1-u_2-u_3)-\frac{p''}{(p')^2}p(1-p)\right].$$

Both Δ and the expression in brackets are unambiguously negative, and hence $\partial\mathrm{EU}/\partial\delta=0$ if and only if $u_3'=u_2'$ [i.e., iff $\delta=\delta^*\equiv(d-\alpha-\beta)/2$]. Furthermore, $u_3'>u_2'$ for $\delta<\delta^*$ and $u_3'<u_2'$ for $\delta>\delta^*$, and so δ^* is the utility-maximizing δ. Thus, when effort is observable by the nonmarket insurer, the equilibrium is constrained efficient.

Does the provision of nonmarket insurance in this case stimulate or discourage effort? To answer this, we hold α and β fixed and increase δ from 0 to $(d-\alpha-\beta)/2$. Though the increase in δ unambiguously reduces risk, it has an ambiguous effect on effort. The risk reduction, by itself, encourages a reduction in effort. However, as δ increases, individuals become less selfish in their choice of effort. From (21), the former effect dominates if $p<1/2$, and the latter effect otherwise. In the normal case, $p<1/2$, the trade-off illustrated in Figure 2 is present, but the direct utility-increasing effect of nonmarket insurance unambiguously dominates the effort-reducing effect.

We can draw together the results in the following inequalities:

$$\text{(22)}\quad \mathrm{EU}^1>\mathrm{EU}^{\mathrm{NMO}}>\mathrm{EU}^{\mathrm{M}}>\mathrm{EU}^{\mathrm{NMU}}$$

where EU^1 is expected utility in the first-best case (with effort observable to both market and nonmarket insurers), $\mathrm{EU}^{\mathrm{NMO}}$ is expected utility with nonmarket insurance and with effort observable to the nonmarket insurer but not the market insurer, EU^{M} is expected utility with only market insurance and with effort unobservable to the market insurer, and EU^{NMU} is expected utility with nonmarket insurance and with effort unobservable to both market and nonmarket insurer.

IV. Discussion

The above models were rather stark. Some discussion and interpretation will therefore be useful. The results of the two cases analyzed above lead naturally to the conjecture that, in intermediate situations in which nonmarket insurers observe their partners' effort imperfectly but better than the market insurer, an excessive amount of nonmarket insurance will be provided which may or may not be better than no nonmarket insurance at all. The analysis could be extended to compare the optimal and equilibrium numbers of members in a nonmarket insurance group; in a large group, there is greater diversification of risk but more imperfect observability.

We distinguished between the two cases treated on the basis of the observability of one partner's effort by the other. The essential difference between the two cases was, however, the severity of moral hazard within the partnership, and this depends on more than the observability of effort. Such factors as the duration of the partnership, the discount rate, the frequency of accidents, the severity of punishment for reneging on an agreement, the power of reputation and social pressure, and nonmonetary rewards from cooperation, are also important.[16]

We provided a rather narrow interpretation of our model. Other interpretations are possible. A market insurer and his clients can be replaced by a principal and his agents, and the partnerships (with some elaboration of the model) can be replaced by secondary markets.[17]

In the above analysis, we took the observability of one partner's accident-prevention effort by the other as exogenous and considered only the extreme cases of unobservability and perfect observability. As we noted earlier, the degree of observability depends on the indirect monitoring system—the means by which one partner observes the other's effort, as well as the incentives to do so. Furthermore, in a fuller analysis, the indirect monitoring system would be treated as endogenous. Indeed, how indirect monitoring systems develop in nonmarket insurance institutions is an exciting research topic, as is the more general issue of how principals should design indirect monitoring systems so as to mitigate the incentives problem.

There is widespread belief that when significant market failure occurs, there are strong incentives for nonmarket institutions to develop which go at least part of the way toward remedying the deficiency.[18] This paper has provided a counterexample[19] in

[16]Altruism is also a factor in nonanonymous relationships. In terms of the model, let EU be an individual's expected utility and $\widetilde{EU}$ his partner's. The individual maximizes welfare $W = EU + \lambda\widetilde{EU}$, where $\lambda = 0$ corresponds to the selfish case, $\lambda = 1$ to balanced altruism, and $\lambda = \infty$ to selfless altruism. If both partners are equally altruistic and cooperate (the effort-observable case), the outcome is independent of the degree of altruism. If both partners are equally altruistic and do not cooperate (the effort unobservable case), effort increases with the degree of altruism.

[17]One can develop a typology in terms of the primary and secondary insurance arrangements, each of which may be market or nonmarket. We have considered the case in which the primary insurance arrangement is a market and the secondary arrangement is a partnership. The case primary-market/secondary-market is an insurance market in which exclusivity cannot be enforced.

[18]In anthropology there is a functionalist tradition of long standing which attempts to explain social institutions (political, economic, sociological, cultural, and psychological) as functional adaptations to a society's environment or ecosystem. Functionalist theories differ in their degree of subtlety and sophistication and in their emphasis, but none seems to make a sharp distinction between equilibrium and optimum. In most theories, however, there seems to be a presumption that institutional adaptation to the environment is efficient. See Roger Keesing (1981) for an informative discussion of contemporary traditions in anthropology.

[19]In our example, with effort unobservable to nonmarket insurers, the market by itself is constrained Pareto-efficient. However, there is *perceived* market failure, and the nonmarket institution (the provision of supplementary nonmarket insurance) arises in response to this perceived market failure.

which a nonmarket institution arises spontaneously (through the uncoordinated actions of atomistic agents), which is completely dysfunctional (has effects opposite to those intended).[20] In our stark model, though the market response to imperfect information (the rationing of insurance) did indeed give rise to a nonmarket response (nonmarket insurance), whether the nonmarket response was welfare-enhancing turned out to depend on whether the nonmarket institution was informationally advantaged relative to the market institution. Our example illustrates, in a vivid way, the functionalist fallacy: the fact that an institution (nonmarket insurance) has a clearly identifiable function (to improve risk-sharing by supplementing the rationed insurance provided by the market) does not mean, within a general equilibrium context, that it actually performs that function (indeed, in one case, the nonmarket insurance was completely dysfunctional). We speculate that the possible dysfunctionality of nonmarket institutions is a general phenomenon. This, too, is an interesting topic for future research.

In an expanded version of our model in which there are many kinds of accidents and many commodities, the market is not constrained Pareto-efficient; there is genuine potential market failure (Arnott and Stiglitz, 1989). Our result concerning the possible dysfunctionality of spontaneous nonmarket institutions carries over to this more realistic setting.

[20]In one sense, this result should come as no surprise, since it is by now well recognized that, even in large economies, Nash equilibria are Pareto-efficient only under special circumstances. One of the great achievements of modern economics was to identify a special set of assumptions under which competitive economies are Pareto-efficient.

George Akerlof (1980) has argued that inefficient social customs may persist as Nash equilibria and that there can be an arbitrarily large set of social customs sustainable as Nash equilibria. The point in our paper is related but different. Akerlof considers the possible *persistence* of inefficient institutions but does not investigate how the institutions came into being. We show not only that an inefficient institution can persist, but also that it can arise spontaneously. Furthermore, while in the Akerlof model there are multiple equilibria of which some may be efficient, in our model there is a unique equilibrium.

V. Conclusion

In this paper we developed a simple moral-hazard model in which individuals have an incentive to supplement market insurance with nonmarket insurance that is unobservable to the market insurer. Whether this nonmarket insurance is socially beneficial depends on whether the nonmarket insurance partners can monitor each other's accident-prevention effort better than the market. In the extreme case in which the partners have no more information than the market, the nonmarket insurance is unambiguously harmful. The nonmarket insurance crowds out the market insurance and results in less risk-spreading; it is therefore completely dysfunctional. In the other extreme case, in which the partners can observe each other's accident-prevention effort perfectly, the nonmarket insurance is ameliorative, and the equilibrium is constrained efficient. The peer monitoring by the nonmarket insurers is effectively utilized to mitigate the moral hazard and to improve the risk-sharing capabilities of the economy.

The simple model raises two broad questions which go beyond the context of moral hazard in insurance markets. In what other situations may nonmarket institutions be dysfunctional? And how may the economy utilize peer-monitoring systems to improve efficiency?

REFERENCES

Akerlof, George, "A Theory of Social Custom, of which Unemployment May Be One Consequence," *Quarterly Journal of Economics*, June 1980, *94*, 749–75.

Arnott, Richard and Stiglitz, Joseph E., "Moral Hazard and Optimal Commodity Taxation," *Journal of Public Economics*, February 1986, *29*, 1–24.

______, (1988a) "The Basic Analytics of Moral Hazard," *Scandinavian Journal of Economics*, 1988, *90* (3), 383–413.

______, (1988b) "Equilibrium in Competitive Insurance Markets with Moral Hazard," mimeo, Stanford University, 1988.

______, (1988c) "Randomization with Asymmetric Information," *Rand Journal of Economics*, Autumn 1988, *19*, 344–62.

______, "The Welfare Economics of Moral Hazard," in H. Loubergé, ed., *Information and Insurance: Essays in Memory of Karl H. Borch*, Norwell, MA: Kluwer, 1989.

Carmichael, H. Lorne, "Incentives in Academics: Why is There Tenure?" *Journal of Political Economy*, June 1988, *96*, 453–72.

Helpman, Elhanan and Laffont, Jean-Jacques, "On Moral Hazard in General Equilibrium Theory" *Journal of Economic Theory*, February 1975, *10*, 8–23.

Holmstrom, Bengt, "Moral Hazard and Observability," *Bell Journal of Economics*, Spring 1979, *10*, 74–91.

Keesing, Roger M., *Cultural Anthropology: A Contemporary Perspective*, 2nd Ed., New York: Holt, Rinehart, and Winston, 1981.

Marshall, John M., "Moral Hazard," *American Economic Review*, December 1976, *66*, 880–90.

Mirrlees, James A., "The Optimal Structure of Incentives and Authority within an Organization," *Bell Journal of Economics*, Spring 1976, *7*, 105–31.

Pauly, Mark, "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," *Quarterly Journal of Economics*, February 1974, *88*, 44–54.

Shavell, Steven, "On Moral Hazard and Insurance," *Quarterly Journal of Economics*, November 1979, *93*, 541–62.

Stiglitz, Joseph E., "Incentives, Risk and Information: Notes Towards a Theory of Hierarchy," *Bell Journal of Economics*, Autumn 1975, *6*, 552–72.

Tirole, Jean, "The Multicontract Organization," *Canadian Journal of Economics*, August 1988, *21*, 459–66.