The Economic Theory of Agency: The Principal's Problem

By Stephen A. Ross*

The relationship of agency is one of the oldest and commonest codified modes of social interaction. We will say that an agency relationship has arisen between two (or more) parties when one, designated as the agent, acts for, on behalf of, or as representative for the other, designated the principal, in a particular domain of decision problems. Examples of agency are universal. Essentially all contractural arrangements, as between employer and employee or the state and the governed, for example, contain important elements of agency. In addition, without explicitly studying the agency relationship, much of the economic literature on problems of moral hazard (see K. J. Arrow) is concerned with problems raised by agency. In a general equilibrium context the study of information flows (see J. Marschak and R. Radner) or of financial intermediaries in monetary models is also an example of agency theory.

The canonical agency problem can be posed as follows. Assume that both the agent and the principal possess state independent von Neumann-Morgenstern utility functions, $G(\cdot)$ and $U(\cdot)$ respectively, and that they act so as to maximize their expected utility. The problems of agency are really most interesting when seen as involving choice under uncertainty and this is the view we will adopt. The agent may choose an act, $a \in A$, a feasible action space, and the random payoff from

this act, $w(a, \theta)$, will depend on the random state of nature $\theta(\epsilon\Omega$ the state space set), unknown to the agent when a is chosen. By assumption the agent and the principal have agreed upon a fee schedule f to be paid to the agent for his services. The fee, f, is generally a function of both the state of the world, θ, and the action, a, but we will assume that the action can influence the parties and, hence, the fee only through its impact on the payoff. This permits us to write,

$$(1) \qquad f = f(w(a, \theta); \theta).$$

Two points deserve mention. Obviously the choice of a fee schedule is the outcome of a bargaining problem or, in large games, of a market process. Much of what we have to say is relevant for this view but we will not treat the bargaining problem explicitly. Second, while it is possible to conceive of the fee as being directly functionally dependent on the act, the theory loses much of its interest, since without further conditions, such a fee can always be chosen as a Dirac δ-function forcing a particular act (see S. Ross). In some sense, then, we are assuming that only the payoff is operational and we will take this point up below. Now, the agent will choose an act, a, so as to

$$(2) \qquad \max_{a} \, E_{\theta}\{G[f(w(a, \theta); \theta)]\},$$

where the agent takes the expectation over his subjectively held probability distribution. The solution to the agent's problem involves the choice of an optimal act, a^o, conditional on the particular fee schedule, i.e., $a^o = a(\langle f \rangle)$, where $a(\cdot)$ is a

mapping from the space of fee schedules into A.

If the principal has complete information about the fee to act mapping, $a(\langle f \rangle)$, he will now choose a fee so as to

$$
(3) \quad \max_{\langle f \rangle} E_{\theta}\big\{ U[w(a(\langle f \rangle)), \theta) \\ - f(w(a(\langle f \rangle), \theta); \theta)] \big\},
$$

where the expectation is taken over the principal's subjective probability distribution over states of nature. If the principal is not fully informed about $a(\cdot)$, then $a(\cdot)$ will be a random function from his point of view. Formally, at least, by appropriately augmenting the state space the criterion (3) could still be made to apply. In general some side constraints on $\langle f \rangle$ would also have to be imposed to insure that the problem possesses a solution (see Ross). A market-imposed minimum expected fee or expected utility of fee by the agent would be one economically sensible constraint:

$$
(4) \quad E_{\theta}\big\{ G[f(w(a, \theta); \theta)] \big\} \geq k.
$$

Since utility functions are assumed to be independent of states, θ, one of the important reasons for a fee to depend directly on θ would be if individual subjective probability distributions differed. In what follows we will assume that both the agent and the principal share the same subjective beliefs about the occurrence of θ and write the fee as a function of the payoff only,

$$
(5) \quad f = f(w(a, \theta)).
$$

Notice that this interpretation would not in general be permissible if the principal lacked perfect knowledge of $a(\cdot)$. More importantly, though, surely aside from simple comparative advantage, for some questions the *raison d'être* for an agency relationship is that the agent (or the principal) may possess different (better or finer) information about the states of

the world than the principal (agent). If we abstract from this possibility we will have to show that we are not throwing out the baby with the bath water.

Under this assumption the problem is considerably simplified but much of interest does remain. Suppose, first, that we are simply interested in the properties of Pareto-efficient arrangements that the agent and the principal will strike. Notice that the optimal fee schedule as seen by the principal is found by solving (3) and is dependent on the desire to motivate the agent. In general, then, we would expect such an arrangement to be Pareto-inefficient, but we will return to this point below. The family of Pareto-efficient fee schedules can be characterized by assuming that the principal and the agent cooperate to choose a schedule that maximizes a weighted sum of utilities

$$
(6) \quad \max_{\langle f \rangle} E\big\{ U[w - f] + \lambda G[f] \big\},
$$

where λ is a relative weighting factor (and where strategies have been randomized to insure convexity). K. Borch recognized that the solution to (6) is obtained by maximizing the function internal to the expectation which requires setting

$$
(P.E.) \quad U'[w - f] = \lambda G'[f]
$$

when U and G are monotone and concave. (See H. Raiffa for a good exposition.) The P.E. condition defines the fee schedule, $f(\cdot)$, as a function of the payoff w (and the weight, λ). (See R. Wilson (1968) or Ross for a fuller discussion of this derivation and the functional aspect of the fee schedule.)

An alternative approach to finding optimal fee schedules was first proposed by Wilson in the theory of syndicates and studied by Wilson (1968, 1969) and Ross. This is the similarity condition that solves for the fee schedule by setting

(S) $$U[w - f] = aG[f] + b$$

for constants $a > 0$, b. If $\langle f \rangle$ satisfies S then, given the fee schedule, it should be clear that the agent and the principal have identical attitudes towards risky payoffs and, consequently, the agent will always choose the act that the principal most desires. Ross was able to completely characterize the class of utility functions that satisfied both P.E. and S (for a range of λ) and show that in such situations the fee schedule is (affine) linear, L, in the payoff. (The class is simply that of pairs $\langle U, G \rangle$ with linear risk tolerance,

$$-\frac{U'}{U''} = cw + d \quad \text{and} \quad -\frac{G'}{G''} = cw + e,$$

where c, d and e are constants.) In fact, it can be shown that any two of S, P.E., or L imply the third.

A question of interest that naturally arises is that of the relation that S and P.E. bear to the exact solution to the principal's problem. (A comparable "agent's problem" can also be posed but we will not be concerned with that here. Some observations on such a problem are contained in Ross.) The solution to the principal's problem (3) subject to the constraint (4) and to the constraint imposed by the condition that the agent chooses the optimal act from his problem (2) can, under some circumstances, be posed as a classical variational problem. To do so we will assume that the payoff function is (twice) differentiable and that the agent chooses an optimal act, given a fee schedule, by the first order condition

(7) $$\underset{\theta}{E}\{G'[f(w)]f'(w)w_a\} = 0,$$

where a subscript indicates partial differentiation. The principal's problem is now to

(8) $$\max_{\langle f \rangle} \underset{\theta}{E}\{H\} \equiv \max_{\langle f \rangle} \underset{\theta}{E}\{U[w - f] + \Psi G'f'w_a + \lambda G\}$$

where Ψ and λ are Lagrange multipliers associated with the constraints (7) and (4) respectively. Changing variables to $V(\theta) \equiv f(w(a, \theta))$ where we have suppressed the impact of a on V and assuming, without loss of generality, that θ is uniformly distributed on $[0, 1]$ permits us to solve (8) by the Euler-Lagrange equation. Thus, at an optimum

(9)
$$\frac{d}{d\theta}\left\{\frac{\partial H}{\partial V'}\right\} - \frac{\partial H}{\partial V}$$
$$= U' + \Psi G'\frac{d}{d\theta}\left[\frac{w_a}{w_\theta}\right] - \lambda G' = 0;$$

or the marginal rate of substitution,

(10) $$\frac{U'}{G'} = \lambda - \Psi\frac{d}{d\theta}\left[\frac{w_a}{w_\theta}\right].$$

This is an intuitively appealing result; the marginal rate of substitution is set equal to a constant as in the P.E. condition plus an additional term which captures the constraint (7) imposed on the principal by the need to motivate the agent. To determine the optimal act, a, we differentiate (8) with respect to a which yields

(11) $$\underset{\theta}{E}\{U'[1 - f']w_a + \Psi G''(f'w_a)^2 + \Psi G'f''(w_a)^2 + \Psi G'f'w_{aa}\} = 0,$$

where we have made use of (7). Substituting the boundary conditions permits us to solve for the multipliers Ψ and λ.

Like S or P.E. (10) defines the fee schedule as a function of w. (Notice that we are tacitly assuming that, at least for the optimal act, the payoff is (a.e. locally) state invertible. This allows the fee to take the form of (5).) It follows that (10) will coincide with P.E. if and only if Ψ is zero, or if $\Psi \neq 0$, we must have

$$(12) \qquad \frac{d}{d\theta}\left[\frac{w_a}{w_\theta}\right] = b(a),$$

a function of a alone.

In particular, using these conditions we can ask what class of (pairs of) utility functions $\langle U, G \rangle$ has the property that, for any payoff structure, $w(a, \theta)$, the solution to the principal's problem is Pareto-efficient. Conversely, we can ask what class of payoff structures has the property that the principal's problem yields a Pareto-efficient solution for any pair of utility functions $\langle U, G \rangle$.

A little reflection reveals that the only pairs of $\langle U, G \rangle$ that could possibly belong to the first class must be those which satisfy S and P.E. for a range of schedules (indexed by the λ weight in P.E.). Clearly if (10) is to be equivalent to P.E. for all payoff functions, $w(a, \theta)$, then Ψ must be zero and the motivational constraint (7) must not be binding. For this to be the case, for an interval of values of k (in (4)), the satisfaction of P.E. must imply that the agent chooses the principal's most desired act by (7). For any fee schedule, $\langle f \rangle$, the principal wants the act to be chosen to maximize $E_\theta\{U[w-f]\}$ which implies that

$$(13) \qquad \underset{\theta}{E}\{U'(1 - f')w_a\} = 0.$$

If (13) is to be equivalent to the motivational constraint (7) for all possible payoff structures, then we must have

$$(14) \qquad U'(1 - f') = G'f'$$

which, with P.E. (or (10) with $\Psi = 0$) yields a linear fee schedule in the payoff. But, as shown in Ross, linearity of the fee schedule and P.E. imply the satisfaction of S and the $\langle U, G \rangle$ pair must belong to the linear risk-tolerance class of utility functions described above.

Since the linear risk-tolerance class, while important, is very limited, we turn now to the converse question of what pay-off structures permit a Pareto-efficient solution for all $\langle U, G \rangle$ pairs. If $\Psi = 0$ we must, as before, have that the motivational constraint is not binding for all $\langle U, G \rangle$ or (13) must always imply (7). The implication will always hold if there exists an a^* such that for all a there is some choice of the state domain, I, for which

$$(15) \qquad w(a^*, \theta) \geq w(a, \theta), \qquad \theta \in I.$$

Conversely, from P.E., we must have that for all $G(\cdot)$

$$(16) \qquad \underset{\theta}{E}\{G'[f](1 - f')w_a\} = 0$$

implies (7) where f is determined by P.E. Since $\langle U, G \rangle$ can always be chosen so as to attain any desired weightings of w_a in (7) and (16) the special case of (15) is the only one for which motivation is irrelevant. Given (15) all individuals have a uniquely optimal act irrespective of their attitudes towards risk.

If $\Psi \neq 0$, then to assure Pareto efficiency we must satisfy (12). This is a partial differential equation and its solution is given by

$$(17) \qquad w(a, \theta) = H[\theta B(a) - C(a)],$$

where $H(\cdot)$, $B(\cdot)$ and $C(\cdot)$ are arbitrary functions. (The detailed computations are carried out in an appendix.) This is a rich and interesting class of payoff functions. In particular, (17) is a generalization of the class of functions of the form $l(\theta - a)$, where the object is to pick an act, a, so as to best guess the state θ. It therefore includes, for example, traditional estimation problems, problems with a quadratic payoff function, and all problems with payoff functions of the form $|\theta - a|^\xi h(a)$, and many asymmetric ones as well. It is not, however, difficult to find plausible payoff functions which do not take the form of (17). (The class of the form (15) will generate such functions.)

We may conclude, then, that the class of payoff structures that simultaneously solve the principal's problem and lead to Pareto efficiency for all $\langle U, G \rangle$ pairs is quite important and quite likely to arise in practice.

In general, though, it is clear that the solution to the principal's problem will not be Pareto-efficient. This is, however, a somewhat naive view to take. Pareto efficiency as defined above assumes that perfect information is held by the participants. In fact, the optimal solution to the principal's problem implied that the fee-to-act mapping induced by the agent was completely known to the principal. In such a case it might be thought that the principal could simply tell the agent to perform a particular act. The difficulty arises in monitoring the act that the agent chooses. Michael Spence and Richard Zeckhauser have examined this problem in detail in the case of insurance. In addition, if agents are numerous the fee may be the only communication mechanism. While it might in principle be feasible to monitor the agent's actions, it would not be economically viable to do so.

The format of this paper has been such as to allow us to only touch on what is surely the most challenging aspect of agency theory; embedding it in a general equilibrium market context. Much is to be learned from such attempts. One would naturally expect a market to arise in the services of agents. Furthermore, in some sense, such a market serves as a surrogate for a market in the information possessed by agents. To the extent to which this occurs, the study of agency in market contexts should shed some light on the economics of information. To mention one more path of interest—in a world of true uncertainty where adequate contingent markets do not exist, the manager of the firm is essentially an agent of the shareholders. It can, therefore, be expected that an understanding of the agency relationship will aid our understanding of this difficult question.

The results obtained here provide some of the micro foundations for such studies. We have shown that, for an interesting class of utility functions and for a very broad and relevant class of payoff structures, the need to motivate agents does not conflict with the attainment of Pareto efficiency. At the least, a callous observer might view these results as providing some solace to those engaged in econometric activity.

APPENDIX

This appendix solves the partial differential equation (12) in the text.
Integrating (12) over θ yields

$$\frac{\partial w}{\partial a} + [b(a)\theta + c(a)]\frac{\partial w}{\partial \theta} = 0.$$

Along a locus of constant w,

$$\frac{d\theta}{da} = -\frac{\partial w/\partial a}{\partial w/\partial \theta} = b(a)\theta + c(a),$$

is a first order Bernoulli equation that integrates to

$$\theta = e^{\int b(a)}\left[\int e^{-\int b(a)}c(a) + k\right],$$

where k is a constant of integration. It follows that

$$w(a, \theta) = H[\theta B(a) - C(a)],$$

where

$$B(a) \equiv e^{-\int b(a)}$$

and

$$C(a) \equiv \int e^{-\int b(a)}c(a) + k,$$

and $H(\cdot)$ is an arbitrary function.

REFERENCES

K. J. Arrow, *Essays in the Theory of Risk-Bearing,* Chicago 1970.

K. Borch, "Equilibrium in a Reinsurance Market," *Econometrica,* July 1962, *30,* 424–444.

J. Marschak and R. Radner, *The Economic Theory of Teams,* New Haven and London 1972.

H. Raiffa, *Decision Analysis; Introductory Lectures on Choices Under Uncertainty,* Reading, Mass. 1968.

S. Ross, "On the Economic Theory of Agency: The Principle of Similarity," *Proceedings of the NBER-NSF Conference on Decision Making and Uncertainty,* forthcoming.

M. Spence and R. Zeckhauser, "Insurance, Information and Individual Action," *Amer. Econ. Rev. Proc.,* May 1971, *61,* 380–387.

R. Wilson, "On the Theory of Syndicates," *Econometrica,* Jan. 1968, *36,* 119–132.

———, "The Structure of Incentives for Decentralization Under Uncertainty," *La Decision,* Editions Du Centre National De Le Recherche Scientifique, Paris 1969.