

THE BEST IS THE ENEMY OF THE GOOD*

by

Roy Radner
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey 07974

A quarter of an hour is clearly not enough time for a serious and comprehensive survey of the impact of Herbert Simon's work on economics. I have chosen, instead, to illustrate this impact with an example, in which two different threads of Simon's work eventually combined in a surprising, and I think fruitful way. The choice of this example is perhaps idiosyncratic, but it has a significant personal basis for me, and I hope it is therefore appropriate to this occasion.

My first contact with Herbert Simon's work occurred when I was a graduate student (in statistics) at the University of Chicago. Simon had written a paper for the Cowles Commission, on "A Formal Theory of the Employment Relation", and it was the Cowles Commission's practice to have its papers "refereed" before they were circulated externally. For this ostensible purpose, I was asked to read Simon's paper. As I remember it, I was being considered for a research assistantship at the Commission, and I suspect that the paper was given to me more as a test of my own abilities than anything else. This paper was eventually published in Econometrica (Volume 19, July, 1951), and reprinted in his book, Models of Man (Wiley,

* Talk presented at luncheon in honor of Herbert Simon, AEA Meetings, Atlanta, December 29, 1979.

1957). What I intend to do today is trace the subsequent history of the ideas in that paper.

First, I shall give a brief account of the substance of the paper. In his characteristic way, he set out to explain an empirical observation, in this case the institutional fact that "... services are obtained by buyers in our society, sometimes by a sales contract, (and) sometimes by an employment contract. For example, if I want a new concrete sidewalk, I may contract for the sidewalk or I may employ a worker to construct it for me." Simon elaborated a model of employment contracts in which there was an agreement on a wage and on a set of alternative acceptable behavior patterns of the worker; the employer then observed some aspects of his environment (modelled as exogenous random variables), and on the basis of the observation decided which of the acceptable behaviors the worker should adopt. The set of acceptable behaviors constituted the authority that the worker relinquished to the employer, in exchange for the agreed-upon wage. The consequent utility to each of the parties to the contract depended upon the chosen behavior, the environment, and the wage (differently for the two parties, of course!). A sales contract was modelled as a special case of an employment contract, in which the set of acceptable behaviors is narrowed down to a single behavior, specified in advance of the employer's knowledge of the environment.

In the framework of this model, Simon examined the following two conjectures:

1. The worker "will be willing to enter an employment contract only if it does not matter to him 'very much' which behavior (within the agreed-upon area of acceptance) the employer will choose, or if the worker is compensated in some way for the possibility that the employer will choose a behavior that is not desired by the worker (i.e., that the employer will ask the worker to perform some unpleasant task)."

2. The employer will be willing to pay for the privilege of postponing the selection of the worker's behavior until he has observed the relevant aspects of the environment, and will be willing to pay more, the greater the (prior) uncertainty about the environment.

With the aid of some more precise assumptions about his formal model, Simon was able to confirm the correctness of his conjectures, and to characterize more precisely the conditions under which some (nondegenerate) employment contract would be preferred to a sales contract. He concluded his paper with a discussion of possible extensions of the model to explain the allocation of "authority" and decision-making among the members of an organization, and of the application of these ideas to the theory of planning under uncertainty.

It is important for my subsequent remarks to mention another point. Simon noted that, in general, an employment

contract would not result in a Pareto-optimal arrangement for the employer and worker jointly (and in particular, not in the presence of risk aversion). As Simon said, "... the difficulty lies in the fact that, once agreement has been reached about (the wage and the acceptable set of behaviors), there is no way for the worker to enforce any (informal) understanding that the employer will" ... subsequently choose a behavior that optimizes some joint utility rather than the employer's own utility (within the acceptable set, of course). In other words, "... the worker has no assurance that the employer will consider anything but his own profit in deciding what he will ask the worker to do."

Simon went on to note that "... we might expect the employer to maximize (their joint utility) only if he thought that by doing so he could persuade the worker, in subsequent renewals of the employment contract, to accept a wage sufficiently smaller to compensate him for this. Otherwise, the employer would rationally maximize (his own short-run profit). We might say that the latter behavior represents "short-run" rationality, whereas the former represents "long-run" rationality when a relationship of confidence between employer and worker can be attained. The fact that the former rule leads to solutions that are preferable to those of the latter shows that it "pays" the employer to establish this relationship."

Finally, in this paper, Simon did not use any explicit principles or mechanisms of "bounded rationality". Nevertheless, I might point out that the paper does not consider the problem

of optimizing among the set of all possible contracts, but restricts its attention to a limited set of contracts, which are intended to model an important aspect of the institution of "authority".

Some of the ideas in this paper reappeared, in different forms or in different contexts, approximately twenty years later, in papers by Theodore Groves and by Stephen Ross. In his paper, "Incentives in Teams" (Econometrica, 1973), Groves considered the problem of devising incentives (reward functions) that would induce the members of a team to use decision and communication rules that would be optimal from the point of view of the organizer (e.g., society, or an employer). The solutions and techniques he developed there led to, among other things, the well-known Groves-Ledyard mechanism for solving the "free-rider" problem in the allocation of public goods.

Ross formulated the so-called "principle-agent" problem (A.E.R., 1973), in which the principle and his agent that is intended to induce the agent to act on the principal's behalf. Ross showed that the resulting "equilibrium" would not, in general be fully Pareto-optimal for the two parties jointly, although he was able to characterize (within his particular model) those special conditions that would lead to full Pareto-optimality.*

* My available time does not permit me to trace all of the antecedents of these ideas, but, at the usual risk of offending the many left off the list, I should at least mention J. Marschak, Arrow, Borch, Wilson, Hurwicz, and Spence and Zeckhauser.

About the same time (or earlier!) Leonid Hurwicz was developing his theory of decentralization and incentive-compatible mechanisms, which explored, among other things, the nature of the potential conflict between decentralization and Pareto-optimality.

The substantial development of these ideas that occurred during the past decade essentially used the formulation of the theory of games, and in particular the theory of "Cournot-Nash", or noncooperative, equilibrium. In this context, the conflict between equilibrium and Pareto-optimality was not new, of course; just recall the "Prisoners Dilemma Game", the simplest and most dramatic formal example of this conflict.

How are we to reconcile this considerable body of theory with the observation that cooperative arrangements pervade much of economic life, indeed, cooperative arrangements that are not enforced by any formal contracts? (The area of labor relations alone provides numerous illustrations.) A clue to one explanation can be found in the passage from Simon's paper that I quoted above, namely the establishment of long-term relationships.

Long-term relationships permit the use of "self-enforcing" combinations of strategies that sustain cooperative behavior. For example, in the Prisoners' Dilemma there is a unique noncooperative equilibrium pair of actions.* There is

* In fact, a dominant strategy equilibrium.

also a Pareto-optimal pair that yields an outcome that is better for both players; call these the cooperative actions. Suppose that the Prisoners' Dilemma is repeated many times, and that each player uses the criterion of average payoff per period. A promising candidate for a pair of strategies to sustain cooperation would be the following: each player uses his cooperative action until the first time that the other player does not; thereafter he uses his noncooperative (equilibrium) action. Call this the trigger-strategy pair.

Unfortunately, if the number of repetitions of the Prisoners' Dilemma is finite, no matter how large, the trigger-strategy pair cannot sustain cooperation! This is easily seen by working backwards from the end, since one player can obtain a benefit from breaking the cooperative agreement before the other does so. In fact, if one defines equilibrium strategy pairs for the long-run relationship suitably, then all equilibria result in noncooperative actions throughout the relationship. Thus, even equilibrium in a (finite) long-term relationship cannot explain sustained cooperative behavior in a Prisoners'-Dilemma-like situation (which includes the employment relationship we started with).*

* The equilibrium concept used here is that of perfect subgame equilibrium of a supergame. However, if the supergame is infinitely long, and the players use the criterion of long-run average payoff per period, then the trigger-strategy pair is an equilibrium. Analogous results have been proved, in a very general setting, by Aumann and Shapley, and by A. Rubinstein.

Here is where Bounded Rationality comes to the rescur! Notice that, in a long-term relationship, the trigger-strategy pair is almost an equilibrium, because the benefit from breaking the cooperative arrangement is short-lived, and therefore can be advantageous only at or near the end. Furthermore, any benefit from breaking the cooperative agreement before the other player does so will have a small effect on the average of the remaining payoffs if the end of the relationship is sufficiently distant. Hence, if each player is satisfied with a strategy that is almost, but not quite, an optimal response to the other player's strategy, then a pair of behaviors in which each player uses a trigger-strategy until a certain distance from the end will be an equilibrium for the long-term relationship.* This kind of reasoning can be extended to principal-agent situations in which the principal cannot perfectly monitor the agent's actions. Modesty would not prevent me from giving a reference for these results, but I have not yet published them.

So, we have joined two threads from Simon's earlier work, in an application that he may not have anticipated. Bounded rationality may be a glue that holds together cooperation in the face of threats of strict optimization. "The best is the enemy of the good."

* Technically, an epsilon-equilibrium.

I would like to conclude with a prediction. I did not go into sufficient technical detail here to make the point, but I am convinced that there are many more analytical tools (particularly from probability theory) suitable for exploring Herbert Simon's ideas that are readily available to today's economist than was the case thirty years ago. For this reason (and others), I am convinced that we shall soon see - or are already seeing - a new wave of applications and extensions in economic theory of his stimulating insights.