L. G. Telser
University of Chicago

A Theory of Self-enforcing Agreements*

. . . a prudent ruler ought not to keep faith when by so doing it would be against his interest, and when the reasons which made him bind himself no longer exist. If men were all good, this precept would not be a good one; but as they are bad, and would not observe their faith with you, so you are not bound to keep faith with them. Nor have legitimate grounds ever failed a prince who wished to show colourable excuse for the nonfulfillment of his promise. [NICCOLO MACHIAVELLI, *The Prince*]

In a self-enforcing agreement each party decides unilaterally whether he is better off continuing or stopping his relation with the other parties. He stops if and only if the current gain from stopping exceeds the expected present value of his gains from continuing. No outside party intervenes to enforce the agreement, to determine whether there has been violations, to assess damages, and to impose penalties. The theory gives a solution of the Prisoners' Dilemma. Application of the theory to transactions between a buyer and a seller gives limits on the sequence of prices that induces repeated transactions. Applied to a group of sellers, the theory describes the conditions under which competition is more profitable than collusion.

I. Introduction

A self-enforcing agreement between two parties remains in force as long as each party believes himself to be better off by continuing the agreement than he would be by ending it. It is left to the judgment of the parties concerned to determine whether or not there has been a violation of the agreement. If one party violates the terms then the only recourse of the other party is to terminate the agreement after he discovers the violation. No third party intervenes to determine whether a violation has taken place or to estimate the damages that result from such violation. No

* I am grateful to Jack Gould, Sam Peltzman, and the members of the Seminar on Applied Price Theory, University of Chicago, for their helpful comments and criticism. All responsibility for any errors that may be present in this paper is mine.

(*Journal of Business*, 1980, vol. 53, no. 1)

0021-9398/80/5301-0003$01.51

third party decides whether a violation has been "willful" or "accidental." A party to a self-enforcing agreement calculates whether his gain from violating the agreement is greater or less than the loss of future net benefits that he would incur as a result of detection of his violation and the consequent termination of the agreement by the other party. If the violator gains more than he loses from the violation, then he will violate the agreement. Hence both parties continue to adhere to an agreement if and only if each gains more from adherence to, than from violations of, its terms.

Many economic transactions illustrate self-enforcing agreements. Since it is costly to rely on the intervention of third parties such as the courts to enforce agreements and to assess damages when they are violated, the parties to an agreement devise its terms to make it self-enforcing, if this can be done cheaply enough. Thus, one of the strongest incentives for honesty of a seller is his desire to obtain the continued patronage of his customer. In cases where a similar transaction between the two parties is unlikely to be repeated in the future so that a loss of future business is not an effective penalty, substitutes for self-enforcing agreements will appear. These substitutes often can avoid the use of third parties as a means of enforcing agreements. A prudent person avoids making a transaction with someone he suspects will be unreliable. Therefore, people seek information about the reliability of those with whom they deal. Reliability, however, is not an inherent personality trait. A person is reliable only if it is more advantageous to him than being unreliable. Therefore, the information about someone's return to being reliable is pertinent in judging the likelihood of his being reliable. For example, an itinerant is less likely to be reliable if it is more costly to impose penalties on him.[1]

1. Information about the reliability of purchasers may be one of the reasons for the success of credit cards. The argument is as follows. Most firms do not offer a lower price to customers who pay in cash or by check than to those who use credit cards, although a firm retains from 95% to 97% of its receipts from a credit card transaction and pays the remainder as a fee to the bank issuing the credit card. It would appear that the net return to the seller is lower on a credit card sale than on a sale paid by check or in cash, yet it charges all of its customers the same price regardless of their mode of payment. (It may accept payment by check from its well-known customers.) Those who use credit cards can pay by check to the bank issuing the credit card, and they do not incur finance charges except for payments deferred for more than 1 month. What services does a credit card give? The answer is that the bank keeps a record of the purchases and payments of the credit card user that enables it to judge the financial reliability of the user. Hence a person who has a valid credit card is certified as reliable to a firm who knows nothing else about him. Therefore, the credit card enables a stranger to buy something and pay for it without using cash as a means of payment. Even so, the customer may offer to pay by check after showing the seller his credit card in order to establish his reliability if the firm will charge him a lower price for his purchase. One objection to this practice is that it would not assure the seller that the buyer's check would clear. A stronger objection is that such a practice would create a free-rider problem. The bank would be certifying the reliability of its credit card holders and would be unable to be remunerated for doing so. In order to solve this free-rider problem, the

A basic hypothesis of this approach is that someone is honest only if honesty, or the appearance of honesty, pays more than dishonesty. Hence, if someone thinks he can gain by dishonesty with impunity then he will be dishonest. This hypothesis leads to important conclusions. For example, if two parties are in a sequence of transactions such that both know for sure which transaction is the last, then each also knows that violating the terms of the agreement on the last transaction cannot evoke a sacrifice of the net gains thereafter. Consequently, each party would be under no compulsion to abide by the terms at the last transaction. Since the same argument applies to the next to the last transaction and so on to the first, we would be driven to the conclusion that no finite sequence of transactions can be self-enforcing if both parties know for sure which transaction is the last one. Therefore, assuming that every sequence of transactions must be finite, a sequence of self-enforcing transactions must have no last one. How can this be? A sequence has no last term if there is always a positive probability of continuing. As long as this is true, anyone who violates the terms at one time incurs the risk of losses in the future. Therefore, there is no certainty of gain from a violation of the agreement on any transaction because there is always a positive probability of continuing to another transaction. Owing to this, there is a positive probability that current violations are punished later.[2]

A theological argument helps illustrate the importance of a finite uncertain horizon. Suppose that after a sinful life on earth, the punishment after death is the torment of the soul in hell. Assume that a sinner can obtain forgiveness by taking various actions before death such as good works, confession to a priest, and so on. If the time of death is uncertain, then a potential sinner incurs the risk of being unable before he dies to perform the good works or take the other appropriate actions that can avoid punishment in hell. If the date of death were certain, then he could allow some time before his death for those actions that would remove the consequences of the sinning that follows his death.

---

bank issuing the credit card often requires that those sellers who will accept its credit card not charge a higher price to those of its customers who use its credit card as a means of payment.

2. The cartel's dilemma illustrates this argument. Let two firms contemplate collusion in order to increase their return. If one sells the agreed upon quantity while the other, relying on this, sells the quantity giving him the maximal net return, then the cheating firm can get a higher net return than he would under collusion. Assuming both firms are equally intelligent and avaricious, Cournot (1960) concludes that collusion is not a stable equilibrium even with only two firms. For an application to the Prisoners' Dilemma, see Luce and Raiffa (1957). These authors also suggest that an uncertain finite horizon can lead to a resolution of the Dilemma (p. 102 and appendix 8). For a detailed analysis of whether collusion is more profitable than competition using the theory of a finite horizon which is a random variable, see Telser (1972, chap. 3, sec. 5; chap. 5, sec. 5). The relation between the number of competing firms and the profitability of collusion compared to competition is discussed in Telser (1978, chap. 1, sec. 10).

Even a nonbeliever in the existence of hell can accept this argument provided he is willing to concede there is a small positive probability that hell exists and that the disutility of the torments of hell is unboundedly large. Given the uncertainty of the date of death, although the short-term gain of sinning exceeds that from a good life, one should lead a good life all of the time because of the large expected long term penalty of sinning.[3]

We shall study two situations in which self-enforcing agreements may occur. First, there is an exchange between two parties such that one sells a good or a service to the other for which he receives a payment. Second, two or more parties seek an agreement among themselves under which they propose to undertake a common venture. In both situations there is a sequence of transactions over time such that the ending date is unknown and uncertain. The penalty for failure to abide by the terms of the agreement is to stop the sequence of transactions so there would be no future relations among the parties. The problem is to see under what circumstances there are mutually satisfactory terms for a self-enforcing agreement.

II. A Formal Theory

Assume two parties are considering whether or not to begin a sequence of transactions that is equivalent to a self-enforcing agreement. We calculate the expected gain to each party of such a sequence.

Let u_j denote the net gain to a party during time period j if a transaction occurs at time j. The sequence terminates either because one of the parties stops it or because an event occurs which neither party can prevent. This makes a distinction between the stopping that is the result of an action taken by one of the parties and which one of them induces, and the stopping that is fortuitous, and, therefore, autonomous, which is equivalent to the result of a random event. We shall confine our attention at the outset to fortuitous stopping.

Let T denote the time of stopping, which is a random variable for the stopping that is autonomous. Let p_t denote the probability that T equals t so that p_t is the probability of a horizon of duration t. Since the stopping time is autonomous, p_t does not depend on current or past gains. We may assume that the parties know the sequence of stopping

3. This argument may also explain why some organized religions oppose suicide. Suicide makes the date of death more certain and would lower the risk of inadequate preparation for preventing the punishment of the soul after death (I am indebted to J. A. Telser for this point). The argument in the text also explains why the old are more likely to be virtuous than the young even if the taste for sinning is independent of age. The old have less time left for repentance than the young and therefore have a smaller expected return from sinning. It also follows that a position of trust is more likely to be given to an older than a younger person.

probabilities. It is certain that the process comes to a halt sooner or later so that

$$\sum_{0}^{\infty} p_t = 1. \tag{1}$$

Let q_t denote the probability of a horizon lasting for more than t periods.

$$q_t = \sum_{t+1}^{\infty} p_j. \tag{2}$$

The expected duration of the horizon, denoted by $E(T)$, satisfies

$$E(T) = \sum_{0}^{\infty} tp_t = \sum_{0}^{\infty} q_t. \tag{3}$$

Notwithstanding (1), $E(T)$ can be infinite.[4]

Given the possibility of stopping at time t, there is the question of whether the gain at time $j < t$, namely u_j, should also depend on t. It should not. Since the horizon is a random variable, the return at an earlier time cannot depend on the value of a random variable at a later time.

There are other possible complications. One we consider below is whether fortuitous events that cause the process to stop for one or both of them are independent of each other. Another is whether the probability of stopping depends on the age of the agreement, that is, on how long it has gone on. As we shall see, the present formulation is general enough to accommodate these factors.

Putting aside these considerations for the moment, we now calculate the expected gain to one of the parties. Define s_t as follows:

$$s_t = \sum_{j=0}^{t} u_j, \tag{4}$$

so that s_t gives the sum of the net gains to a party for a sequence of transactions lasting for $t + 1$ periods. Observe that this allows for discounting, since u_j has a time subscript, j, and the present value of a given u_j can decrease as j increases. The probability of s_t is p_{t+1} so that $p_{t+1} s_t$ is the contribution to the expected value of a horizon of length t. Summing over horizons of all possible durations gives the expected value of the gain, denoted by $E(u)$, as follows:

$$E(u) = \sum_{0}^{\infty} p_{t+1} s_t. \tag{5}$$

Another way of writing $E(u)$ which uses (2) and (4) is given by

$$E(u) = \sum_{0}^{\infty} q_t u_t. \tag{6}$$

4. See Feller (1962, vol. 1, chap. 11, sec. 1, theorem 2).

We now consider some complications. A self-enforcing agreement refers to a sequence of transactions between at least two parties, and fortuitous events can affect at least one of them bringing their relation to an end. A relation can continue between two parties if and only if no fortuitous event stops the process for either one of them. The length of the horizon, t, satisfies

$$t = \min < t_1, t_2 >, \tag{7}$$

where t_i is the time when some event would occur that would prevent the continued participation of the ith party. Therefore, p_t is the probability of $T = t$ with t as defined by (7). Consequently, it does not matter in this formulation whether the fortuitous events stopping a relation among the parties are independent or dependent.

The age of the relation between the two parties does not affect the probability of ending their agreement if and only if the conditional probability of stopping is the same at all points in time. In particular, the conditional probability of stopping at time t must be the same as the probability of never starting, which is p_0. This holds if and only if

$$p_t = (1 - p_0)^t p_0 \tag{8}$$

(Feller 1962, vol. 1, chap. 13, sec. 9). We note for future reference that in this case

$$q_t = (1 - p_0)^{t+1}. \tag{9}$$

Also, from (3) and (9), we obtain

$$E(T) = (1 - p_0)/p_0. \tag{10}$$

The sequence of transactions can stop because one of the parties believes this would make him better off. Let δu_t denote the increment in the net gain to a party who stops the sequence at time t. We now refer to induced stopping and not to fortuitous stopping that is beyond the control of either party. Write the sequence δu^t so that $\delta u^t = < 0, \ldots, 0, \delta u_t, -u_{t+1}, -u_{t+2}, \ldots >$. Using this notation, the sequence of net gains to a party who takes an action bringing the transactions to a halt from period $t + 1$ onward is given by $u + \delta u^t$ (the superscript shows when the violation occurs). Therefore,

$$E(u + \delta u^t) = E(u) + q_t \delta u_t - \sum_{t+1}^{\infty} q_j u_j. \tag{11}$$

We may conclude from (11) that the gain from continuing the sequence of transactions exceeds the gain from stopping at time t if and only if for all t,

$$E(u + \delta u^t) - E(u) = q_t \delta u_t - \sum_{t+1}^{\infty} q_j u_j \leq 0. \tag{12}$$

In applications, the net gain u_t depends on the terms of the transaction at time t and δu_t depends on how much the party can gain from a violation of these terms. Therefore, the problem of finding terms that can give a self-enforcing agreement reduces to determining whether there is a sequence $\{u_t\}$ that can satisfy (12) for the implied δu_t so that $E(u)$ would give the maximum.

It is convenient to pause at this point and consider whether this theory captures the important aspects of the problem under study. If one party takes an action giving him a net gain of δu_t at time t which is an unfavorable surprise to the other party and constitutes a violation of the agreement, then the theory assumes that the victim responds by terminating the agreement. This gives the maximum penalty that the victim can impose on the violator of the self-enforcing agreement. If the maximum penalty cannot deter a violation because $E(u + \delta u^t) - E(u) > 0$, then no smaller one suffices. Hence (12) is both necessary and sufficient for a self-enforcing agreement.

Second, the victim may not immediately discover the violation. If he makes the discovery after violations have been going on for k periods, the gains to the violator would be $\delta u_t, \delta u_{t+1}, \ldots, \delta u_{t+k-1}$. Slowness in the detection of violations raises the gains to the violator. In order to simplify the algebra, the preceding formulation assumes discovery with a minimum delay ($k = 1$). This does not change the validity of the analysis in any important way.

Third, the reader may object that the theory fails to allow for deviations from the expected gains and seems to require rigid and continuously perfect adherence to expectations. Here, too, nothing essential is lost with this approach. To see why, consider how we would take deviations into account. Assume there is a band around the expected gains such that actual gains falling within the band are admissible. Gains outside these bands imply a violation of the agreement. The deviation between the actual and the expected gain should behave like a sequence of independent and identically distributed random variables if the parties are in compliance with the agreement. There is a violation either when the actual gains go outside the limits or remain too close to one of them for too long. It follows that the parties would pay careful attention to those gains which are close to the limits. Consequently, the limits themselves have a role in the analysis allowing for deviations just like u_t in the present theory. Therefore, without loss of generality we may proceed with our analysis.

Since δu_t is the increment of gain to a party who stops the sequence at time t, $\delta u_t > 0$ is a weak necessary condition for stopping.

In order to describe the properties of a sequence $\{u_t\}$ capable of sustaining a self-enforcing agreement, it is necessary to make assumptions about the relation between δu_t and u_t. Assume first that δu_t and u_t

vary inversely. Thus, the simplest relation of this kind would be linear so that

$$\delta u_t = a_0 - a_1 u_t, \qquad a_0, a_1 > 0.$$

But then, $u_t \geq a_0/a_1$ would make δu_t negative. Hence if u_t exceeds this lower bound, the sequence would be self-enforcing. In general, a sequence would be self-enforcing whenever u_t is large enough if δu_t and u_t vary inversely.

Alternatively, assume that δu_t is proportional to u_t. Thus, let

$$\delta u_t = \beta u_t, \quad \beta > 0. \tag{13}$$

By virtue of (12), the necessary and sufficient condition for continuing becomes

$$\beta q_t u_t \leq \sum_{t+1}^{\infty} q_j u_j \tag{14}$$

Before giving a general result for this case, it is helpful to consider two examples. First, assume that $q_t u_t = r^t$ with $0 < r < 1$. Condition (14) is equivalent to $\beta \leq r/(1 - r)$. If $\beta = 1$, then $r > \frac{1}{2}$. If $\beta = 2$, then $r > \frac{2}{3}$. In the second example, let $q_t u_t = t^{-\alpha}$ with $\alpha > 1$. The series on the right side of (14) is approximately $[1/(\alpha - 1)][1/(t + 1)^{\alpha-1}]$. In order to satisfy (14), conditions on β and α are required. Write $\alpha = 1 + \sigma$. Then (14) holds if and only if

$$\beta(t + 1)^{\sigma}/t^{1+\sigma} \leq 1/\sigma \text{ for all } t \geq 1. \tag{15}$$

The function on the left decreases monotonically with t and has a maximum at $t = 1$. Therefore, to satisfy (15) it is necessary and sufficient that

$$\beta 2^{\sigma} \leq 1/\sigma. \tag{16}$$

For $\beta = 1$, it can be shown that (16) holds for all $\sigma < 0.64115$. More generally, (16) gives the implication of an inverse relation between β and σ. Both examples share the property that the series $\Sigma q_j u_j$ must not converge too rapidly in order to satisfy (14). This means that an agreement is self-enforcing if and only if the expected horizon is long enough.

Treating time as a continuous parameter gives convenient expressions for the necessary and sufficient condition of a self-enforcing agreement. Let $f(t)$ denote a probability density function so that

$$F(t) = \int_0^t f(s)\, ds \to 1 \text{ as } t \to \infty.$$

The probability of going on longer than t is $1 - F(t)$. Let $u(s)$ denote the gain at time s. Then

$$E(u) = \int_0^\infty u(s)[1 - F(s)]ds \tag{17}$$

corresponds to (6). The necessary and sufficient condition for continuing is

$$\delta u(t)[1 - F(t)] \leq \int_t^\infty u(s)[1 - F(s)]ds, \tag{18}$$

where $\delta u(t)$ is the gain from cheating. Define the function $H(t)$ as follows:

$$H(t) = \int_t^\infty u(s)[1 - F(s)]ds \tag{19}$$

so that $H'(t) = -u(t)[1 - F(t)] < 0$. If $\delta u(t)$ is proportional to $u(t)$, $\delta u(t) = \beta u(t)$, $\beta > 0$, then (18) becomes

$$0 < -\beta H'(t) \leq H(t) \text{ for all } t \geq 0. \tag{20}$$

Rearranging terms in (20) gives an interesting condition on H(t). It follows that $0 < -\beta dH/H \leq dt \rightarrow -\beta \log H \leq t + c$, where c is a constant. Consequently,

$$H(t) \geq e^{-(t+c)/\beta} \tag{21}$$

gives a lower bound on H.

In this theory each party compares his current gain from cheating the other party to his expected gain from continuing his relation honestly with the other party. The probability of continuing does not depend on the past record of transactions the parties have had with each other though it may vary with time. Some may wish to argue that previous favorable experience raises the probability of continuing while previous unfavorable experience lowers this probability. Equivalently, assume that a party to an agreement accumulates a stock of goodwill toward the other party that depends on his past experience with the other party. Favorable past experience raises and unfavorable past experience lowers this stock of goodwill. He terminates his relation with the other party when he has no remaining stock of goodwill.

The argument postulating a stock of goodwill based on past experience faces a fatal objection because it is inconsistent with rational behavior. To see why, suppose that a buyer accumulates goodwill toward a seller based on the excess of his favorable over his unfavorable experiences with that seller. Each past favorable experience raises and each past unfavorable experience lowers the buyer's goodwill. A seller who knows that the buyer behaves in this way has various tempting ways of cheating the buyer. Such a seller dealing with a new

customer may deliberately behave honestly toward him at first to gain his confidence so that he can cheat him more profitably later. Moreover, he need only maintain the stock of goodwill of his old customers at a level just high enough to obtain their continued patronage. In the process he can cheat them, but not too often. The accumulation of a fund of goodwill of a buyer toward a seller that depends on past experience stands as a temptation to the seller to cheat the buyers and convert their goodwill into ready cash. It is the prospect of the loss of future gain that deters and the existence of past goodwill that invites cheating. Therefore, rational behavior by the parties to an agreement requires that the probability of continuing their relation does not depend on their past experience with each other. We recognize this condition as equivalent to an efficient market with rational traders.

III. Self-enforcing Agreements between a Buyer and Seller

Before giving the theory it is helpful to give an example. Consider a firm as a buyer and a worker as a seller to it. Assume that the worker initially accepts a lower wage rate in order to acquire skills that will be useful as long as he remains with that employer. The sacrifice in his current earnings during the training period is equivalent to an investment by the worker in firm specific human capital upon which he expects to receive a return subsequently. This return comes in the form of a higher wage rate later on in compensation for the lower wage during the earlier period of employment. If the firm does not pay the higher wage rate subsequently then it violates the agreement and the worker can quit. The firm also incurs part of the cost of investment in firm specific human capital if it pays the worker a wage rate during the training period above the value of his marginal product at that time. The firm expects to obtain a return on its investment subsequently if it pays the worker a wage rate below the then current value of the worker's marginal product. If the worker quits, then the firm stands to lose the investment that it has made in the worker. Therefore, these circumstances raise the problem of seeing whether it is possible to have a self-enforcing agreement.[5]

A formal statement of the situation is as follows. The buyer expects a benefit in period t of b_t^* for which he expects to pay x_t^*. The seller expects to incur a cost of a_t^* in period t for which he expects to receive x_t^*. The gain the buyer expects is $u_t^* = b_t^* - x_t^*$. The gain the seller expects is $v_t^* = x_t^* - a_t^*$. The expected benefit of the buyer, b_t^*, depends on the expected cost incurred by the seller, a_t^*. The lower is a_t^*, the

5. For an empirical analysis of firm specific human capital, see Telser (1972, chap. 8).

lower is b_t^*. The buyer is said to violate the agreement if he gets b_t^* and pays $x_t < x_t^*$. Hence

$$u_t = b_t^* - x_t > u_t^* = b_t^* - x_t^*, \tag{22}$$

and

$$\delta u_t = u_t - u_t^* = x_t^* - x_t \leqslant x_t^*. \tag{23}$$

The maximal gain of the buyer occurs when $x_t = 0$. The seller is said to violate the agreement if he incurs the cost $a_t < a_t^*$ and gets x_t^* from the seller. Hence the seller obtains

$$v_t = x_t^* - a_t > x_t^* - a_t^*, \tag{24}$$

and the gain to the seller is δv_t where

$$\delta v_t = v_t - v_t^* = a_t^* - a_t \leqslant a_t^*. \tag{25}$$

If the buyer violates the agreement in period t, then the seller detects the violation afterward and, in a self-enforcing agreement, imposes the penalty of stopping the transactions with the buyer. Therefore, the buyer would gain δu_t and would sacrifice the expected gains that he would obtain by faithful adherence to the expectations. Similarly, if the seller violates the agreement by furnishing a commodity that proves to be less beneficial to the buyer than the buyer expected, then the penalty that the buyer imposes in a self-enforcing agreement is termination of future transactions with the seller.

Condition (12) applies in the present situation. The buyer obtains a maximal gain by continuing the sequence of transactions if and only if for all t, it is true that $E(u + \delta u) \leqslant E(u)$. Therefore, taking the upper bound for δu_t given by (23), we obtain a sufficient condition for the buyer to continue the sequence of transactions which is given as follows:

$$q_t x_t^* \leqslant \sum_{t+1}^{\infty} q_j u_j^* = \sum_{t+1}^{\infty} q_j (b_j^* - x_j^*). \tag{26}$$

We may conclude that the buyer is willing to continue the sequence of transactions if there is a sequence $\{x_t^*\}$ such that

$$\sum_{t}^{\infty} q_j x_j^* \leqslant \sum_{t+1}^{\infty} q_j b_j^*. \tag{27}$$

The equivalent sufficient condition for the seller is that

$$q_t x_t^* \leqslant \sum_{t+1}^{\infty} q_j v_j^* = \sum_{t+1}^{\infty} q_j (x_j^* - a_j^*). \tag{28}$$

Rearranging terms in this expression gives

$$\sum_{t+1}^{\infty} q_j a_j^* \leq \sum_{t+1}^{\infty} q_j x_j^* - q_t x_t^*. \tag{29}$$

Hence there is a self-enforcing agreement if there is an $\{x_t^*\}$ that can satisfy (27) and (29).

To illustrate these conditions, consider the important special case where $a_t^* = a, b_t^* = b$, and $p_t = (1 - p_0)^t p_0$ so that $q_t = (1 - p_0)^{t+1}$ (cf.[8]). Then (27) is equivalent to $x(1 - p_0)^{t+1} \leq b(1 - p_0)^{t+2}$, which reduces to

$$x \leq b(1 - p_0) < b. \tag{30}$$

Condition (29) is equivalent to $a(1 - p_0)^{t+2}/p_0 \leq x(1 - p_0)^{t+2}/p_0 - x(1 - p_0)^{t+1}$. This becomes

$$a(1 - p_0) \leq x[(1 - p_0) - p_0]. \tag{31}$$

The inequality in (31) can be satisfied only if $p_0 < \frac{1}{2}$. This means that the expected horizon must exceed one period (cf. [10]). Let $\theta = p_0/(1 - p_0) = 1/E(T)$. Then we can rewrite (31) in the shape as follows:

$$a/(1 - \theta) \leq x. \tag{32}$$

If $p_0 < \frac{1}{2}$ then $\theta < 1$. Hence we have $a < a/(1 - \theta)$. Combining (30) and (32), we may conclude that there is a self-enforcing agreement if

$$a < a/(1 - \theta) < x < b(1 - p_0) < b. \tag{33}$$

We must necessarily have

$$a/b < (1 - p_0)/(1 - \theta) = 1 - 2p_0, \tag{34}$$

or (33) cannot hold. Notice that the presence of uncertainty about continuing the sequence of transactions narrows the range in which an admissible x can fall.[6]

6. The argument in the text does not require either $u_t^* = b_t^* - x_t^* > 0$ or $v_t^* = x_t^* - a_t^* > 0$. Hence the upper bounds for δu_t given by (23) and for δv_t given by (25) are valid even if either u_t^* or v_t^* is negative. We obtain different results by assuming $x_t^* \leq b_t^*$ so that $\delta u_t \leq b_t^*$ in place of (23). A sufficient condition for the buyer to continue now becomes as follows:

$$q_t b_t^* \leq \sum_{t+1}^{\infty} q_j u_j^* = \sum_{t+1}^{\infty} q_j (b_j^* - x_j^*)$$

which reduces to

$$\sum_{t+1}^{\infty} q_j x_j^* \leq \sum_{t+1}^{\infty} q_j b_j^* - q_t b_t^*. \tag{a}$$

For the seller, a sufficient condition for continuing is given by

$$\sum_{t}^{\infty} q_j a_j^* \leq \sum_{t+1}^{\infty} q_j x_j^*. \tag{b}$$

Together these give sufficient conditions for a self-enforcing agreement between the buyer and the seller. We can see the difference between this situation and the one in the

The analytic formulation in Section 2 is useful in giving sufficient conditions for continuing the series of transactions between the buyer and the seller. Corresponding to (26), there is

$$x^*(t)[1 - F(t)] \leq \int_0^\infty [b^*(s) - x^*(s)][1 - F(s)]ds \tag{35}$$

which becomes

$$x^*(t)[1 - F(t)] + \int_t^\infty x^*(s)[1 - F(s)]ds \leq \int_t^\infty b^*(s)[1 - F(s)]ds. \tag{36}$$

For the seller, corresponding to (28), there is

$$a^*(t)[1 - F(t)] \leq \int_t^\infty [x^*(s) - a^*(s)][1 - F(s)]ds$$

which becomes

$$a^*(t)[1 - F(t)] + \int_t^\infty a^*(s)[1 - F(s)]ds \leq \int_t^\infty x^*(s)[1 - F(s)]ds. \tag{37}$$

Define

$$A(t) = \int_t^\infty a^*(s)[1 - F(s)]ds$$

so that $A'(t) = -a(t)\,[1 - F(t)]$. Using a similar notation for $B(t)$ and $X(t)$ we can obtain a sufficient condition for a self-enforcing agreement. Thus, (36) becomes

$$-X'(t) + X(t) \leq B(t) \tag{38}$$

and (37) becomes

$$-A'(t) + A(t) \leq X(t). \tag{39}$$

A necessary condition for (38) and (39) is that

$$A(t) - A'(t) \leq B(t) + X'(t). \tag{40}$$

Consequently, (40) imposes conditions on $X'(t)$ in terms of the given functions $A(.)$ and $B(.)$ that are necessary for (38) and (39). Any function $X(.)$ that satisfies (38) and (39) implies a self-enforcing agreement between the buyer and the seller.

---

text most easily for the special case where $a_j^* = a, b_j^* = b, x_j^* = x$ and $q_j = (1 - p_0)^{j+1}$. For this case (a) and (b) reduce to

$$a \leq x(1 - p_0) \leq b[(1 - p_0) - p_0]. \tag{c}$$

Necessarily, $p_0/(1 - p_0) < 1$ which holds if and only if $p_0 < \frac{1}{2}$. The limits in (c) are wider than the limits in (33) as is readily verified. However, it remains true that the bounds for x sufficient for a self-enforcing agreement are narrowed because $a < a/(1 - p_0)$ and $b > 1 - \theta$.

IV. Self-enforcing Cooperative Agreements

The preceding section analyzes a sequence of transactions between two parties who trade goods for money. Another important application is to a situation where the two parties can cooperate in some venture which does not involve a direct exchange between them. A leading example would be collusion between two firms, say the two sellers of mineral water in Cournot's theory, who can choose the same policy as a monopoly that obtains a maximal net return. The alternative to cooperation is the noncooperative equilibrium that results if each firm acting independently chooses the policy giving it the maximal net return for a given policy of the other firm. Cournot's theory states there is never collusion because one party to a collusive agreement can always gain more by violating it than by adhering to it, although together the firms gain more from collusion than competition. In the present formulation, Cournot asserts that a self-enforcing collusive agreement is impossible even if there are only two firms. This conclusion is correct if the firms operate over a finite certain horizon so that punishment of departures from the agreement is not possible after the last period. However, the presence of uncertainty about the length of the horizon giving the implication that there is no last period, implies different results. Such uncertainty may enable a self-enforcing cooperative agreement to be an equilibrium.

As above, let u_j denote the expected gain to a firm in period j if it cooperates with the other firm.

$$E(u) = \sum_{0}^{\infty} q_j u_j \tag{41}$$

gives the expected net return if there is cooperation. Let v_j denote the gain in period j if the firms do not cooperate. Hence $v_j < u_j$ and

$$E(v) = \sum_{0}^{\infty} q_j v_j \tag{42}$$

gives the expected net return to a firm in the noncooperative equilibrium. Assume there is cooperation until period $t - 1$ and that the firm, counting on the adherence of the other to the cooperative agreement, violates the agreement in period t so that its gain in that period is $u_t + \delta u_t > u_t$. The punishment is a cessation of the cooperative agreement thereafter, which means a sequence of net returns given by v_{t+1}, v_{t+2}, ... Denote the expected return from a violation of the cooperative agreement by $E(u + \delta u^t)$, and we have

$$E(u + \delta u^t) - E(u) = E(v)_{t+1} - E(u)_{t+1} + q_t \delta u_t, \tag{43}$$

where $E(u)_{t+1}$ and $E(v)_{t+1}$ are defined as follows:

$$E(v)_{t+1} = \sum_{t+1}^{\infty} q_j v_j \quad \text{and} \quad E(u)_{t+1} = \sum_{t+1}^{\infty} q_j u_j.$$

A self-enforcing cooperative agreement gives the maximal expected net return if and only if for all t, $E(u + \delta u^t) - E(u) \leq 0$, which by virtue of (43), is equivalent to

$$0 < q_t \delta u_t \leq E(u - v)_{t+1}, \tag{44}$$

for all t. Take the special case where the probability of a horizon longer than t periods is given by $q_t = (1 - p_0)^{t+1}$. Let $\delta u_t = \delta_2$, a constant, and let $u_t - v_t = \delta_1 > 0$, a constant. The necessary and sufficient condition for a self-enforcing cooperative agreement given by (44) implies that $0 < (1 - p_0)^{t+1}\delta_2 \leq \delta_1(1 - p_0)^{t+2}/p_0$, which simplifies to

$$\delta_2 \leq \delta_1(1 - p_0)/p_0 = \delta_1 E(T). \tag{45}$$

Thus, given δ_2 and δ_1, the longer is the expected horizon, the greater is the dominance of the self-enforcing cooperative agreement over the alternative, noncooperation. The values of the parameters δ_1 and δ_2 depend on the underlying cost and demand conditions.

This theory also has interesting implications for multiproduct firms. Let u_{itk} denote the net return to firm i in period t from product k if it colludes with the other firm on this product. Assume the products are independent so that autonomous stopping for any product is a random event independent of autonomous stopping for any of the other products. These assumptions raise the question of whether linkage among such independent products by the two firms can result in a self-enforcing cooperative agreement on all of the products. The point is this. It may be that collusion on one product does not satisfy (44) so that noncooperation gives the highest expected return. On another independent product (44) may be satisfied so that cooperation gives the highest expected return on this product. By linking the two products together, it may be that cooperation on the two may satisfy (44). Thus, the higher expected return on the one product, where collusion would be more profitable in its own right, is a hostage to the other product, where competition would be more profitable in its own right. In place of (44) holding product by product, linkage gives a necessary and sufficient condition for cooperation on all n products as follows:

$$\sum_{k=1}^{n} q_{tk} \delta u_{t,k} \leq \sum_{k=1}^{n} E(u - v)_{t+1,k}. \tag{46}$$

Plainly, (46) can hold although some of the components do not satisfy (44). This argument assumes that if there is a violation of the cooperative agreement on one product by one firm, then punishment takes the

form of noncooperation on all of the products. Otherwise, of course, there would be no linkage.

For substitutable products, linkage is no luxury but a necessity so that there can be a successful self-enforcing agreement on them. Dependence refers to two aspects; relations among the net gains of the products and relations among the autonomous events that cause stopping for a product. Relations among the net gains means that there may be a direct effect on the net gains of one product as a result of actions taken on the other products. Thus, the net return to collusion on one product depends on whether there is cooperation on the related products. Relations among stopping probabilities means that if the sequence for one product terminates, this affects the probability of termination of the other products. Consequently, in contrast to the situation with independent products, it may not be possible to sustain an agreement on one product unless there is agreement on related products. A rigorous analysis would require a more elaborate description of how the products are related.[7]

V. Alternatives to a Self-enforcing Agreement

A self-enforcing agreement is possible if and only if the expected future gains from adherence to it exceeds the current gain from a violation of the agreement. Therefore, given a sequence of gains $\{u_t\}$, a self-enforcing agreement is not possible when the expected horizon is too short or when δu_t is too large. In these cases a third party may intervene to enforce an agreement so that the agreement is not self-

7. Space constraints do not permit an elaborate description of how this can be done. A sketch of the theory is as follows. Define a commodity in terms of its characteristics so that it has m characteristics given by the coordinates of the m-vector, x. Measure the distance between two products, x and y, by a norm, denoted by $||x - y||$. A norm is a convex function homogeneous of degree one. Assume that the difference in unit value between x and y, denoted by $p(x) - p(y)$, that consumers are willing to pay satisfies

$$|p(x) - p(y)| = ||x - y||. \tag{d}$$

Let $q(x)$ denote the quantity demanded of type x and write the demand function

$$q(x) = G[p(x), s(x)], \tag{e}$$

where $s(x)$ is the number of customers who prefer type x. Assume that $q(x)$ varies inversely with $p(x)$ and directly with $s(x)$. Customers who prefer type x and buy type y instead do so because alternatives to y are more costly. The price they pay to the sellers of y is $p(y)$ and their implicit price is

$$p(x) - p(y) = ||x - y||. \tag{f}$$

The quantity they buy satisfies (e). Customers who prefer type x are indifferent among all other types that are equidistant from x as measured by the norm, which is in money terms. Proceeding in this way leads to a theory of the relations among substitutable products with many useful implications. It turns out that uncertainty and product change are conducive to competition because they tend to shorten the expected horizon and to decrease the punishment for a violation of a potential self-enforcing agreement.

enforcing. There are situations where the parties have an involuntary relation and the intervention of a third party may be necessary to protect the weaker members of the relation. For instance, small children cannot be said to enter into a voluntary agreement with their parents, and the state intervenes to protect them from parental abuse. The domain of the theory of self-enforcing agreement is the cases where there is voluntary consent among the parties. Even so, it may well happen that the probability of a continuing relation is too low to sustain a self-enforcing agreement. This will incline the parties to seek less costly alternatives if there are any. In the absence of such alternatives, no agreement will occur.

Many alternatives to a self-enforcing agreement have a common feature—a sum of money that will be paid or not depending on whether there is fulfillment of the terms. Thus, each party may deposit a sum of money or give an equivalent guarantee of that sum with the understanding that either stands to lose it to the other as a result of failure to abide by the agreement. There is the problem, however, that one party may claim, falsely, that the other has violated the agreement so that it can obtain the sum of money that is subject to forfeiture. Alternatively, there may be deferred payment. The party who would make the payment at the appropriate time may claim, falsely, that the other party has violated the agreement so that it should not receive the deferred payment. In these cases it may be necessary for a third party to intervene in order to settle disputes. A self-enforcing agreement has an object that is equivalent to a deferred payment or bond. This is the expected value of the future gains that is lost by the party who violates the implicit terms of a self-enforcing agreement.[8]

Deferred rebates, deferred wage or salary compensations, stock options, money back guarantees, security deposits in rentals, and posting bonds are among the alternatives to a self-enforcing agreement. Advertising outlays are another important example of an alternative. Such outlays create public awareness of a firm's products and are firm-specific capital that it can lose if the consumers believe that its products are unsatisfactory. Thus, the capital value of the advertising outlays are approximately equivalent to posting a bond as an assurance of satisfactory products.

VI. Conclusions

No one would enter an agreement expecting the other parties to violate it. In a self-enforcing agreement the only penalty that can be imposed on the violator is stopping the agreement. Therefore, aware of this, a

8. Becker and Stigler (1974) discuss alternatives to self-enforcing agreements apparently under the assumption that one of the parties to the agreement is honest and seeks ways to deter the dishonesty of the other party.

potential violator compares the current gain from a violation with the sacrifice of future gains that will result in response to his current violation. These future gains would accrue to him were he to remain faithful to the agreement. He chooses the more profitable alternative. It follows that the parties to a self-enforcing agreement do not expect any violations of it. The terms of the agreement are such that adherence is more advantageous than violation. Were they to expect violations to be more profitable than adherence, they would not embark on the agreement in the first place.

From these premises follow important conclusions. First, self-enforcing agreements are not feasible if the sequence of occasions for transactions has a definite known last element. Although termination is certain to occur sooner or later, when this happens must be uncertain in order to sustain a self-enforcing agreement. Second, for a given sequence of gains, the expected horizon must be long enough or there can be no self-enforcing agreement. Equivalently, the longer is the expected horizon, the greater is the return to the parties from adherence to the terms of the agreement. Third, parties who have a self-enforcing agreement do not expect violations to occur. The theory explains violations of a self-enforcing agreement as the response to unexpected changes in the underlying factors that determine the terms of the agreement. Owing to these unexpected changes, there may be violations if the parties cannot find mutually acceptable terms appropriate for the new conditions. This is to say that highly uncertain conditions are not conducive to self-enforcing agreements.

References

Becker, G. S., and Stigler, G. J. 1974. Law enforcement, malfeasance and compensation of enforcers. *Journal of Legal Studies* 3 (January): 1–18.

Cournot, A. 1960. *Researches into the Mathematical Principles of the Theory of Wealth.* Translated by Nathaniel Bacon from 1838 French ed., introductory essay by Irving Fisher. New York: Kelley.

Feller, W. 1962. *An Introduction to Probability Theory and Its Applications.* 2 vols. Vol. 1, 2d. ed. New York: Wiley.

Luce, R. D., and Raiffa, H. 1957. *Games and Decisions.* New York: Wiley.

Telser, L. G. 1972. *Competition, Collusion and Game Theory.* Chicago: Aldine-Atherton.

Telser, L. G. 1978. *Economic Theory and the Core.* Chicago: University of Chicago Press.