Review of Economic Studies (1985) LII, 557-573
0034-6527/85/00390557$02.00

A Theory of Credibility

JOEL SOBEL
University of California, San Diego

This paper presents models in which one agent must decide whether to trust another, whose motives are uncertain. Reliability can only be communicated through actions. In this context, it pays for people to build a reputation based on reliable behaviour; someone becomes credible by consistently providing accurate and valuable information or by performing useful services. The theory provides a justification for long-term arrangements without binding contracts. It also describes those situations where it pays an agent to cash in on his reputation.

1. INTRODUCTION

Many decisions depend on trusting someone. Consumers must decide whether to believe advertisements, banks need to judge the reliability of loan applicants, and employers must decide how much responsibility to delegate to their employees. This paper attempts to model situations that arise when someone is uncertain about whether to trust the people he deals with. In this context, reputations for reliability are valuable. An agent becomes credible by consistently providing accurate, valuable information or by always acting responsibly.

I analyse a series of simple, abstract models of information transmission. In the basic model, there are two agents, the Sender (S) and the Receiver (R) who are to play the game a known, finite number of times. At each stage both players learn the value of a parameter measuring the importance of that period's play of the game (that is, the value of making a correct decision) and S obtains payoff-relevant information; S then sends a signal to R, who makes a decision that affects the welfare of both players. After both players learn the consequences of R's decision, the process repeats. At each stage R must decide what action to take. In order to do that he must assess the credibility of S. The difficulty arises because R is uncertain about S's preferences when play begins. I model this by assuming that with positive probability S has identical preferences to R (S is a friend) and with positive probability S has completely opposed preferences to R (S is an enemy). This uncertainty creates a trade-off for enemy Senders. At each stage, providing honest information enhances S's reputation, but only at the expense of foregoing an opportunity for immediate gain by duping R. The possibility that S is a friend lends credibility to what he says, but this credibility may also be exploited by an enemy S.

The principal equilibrium has the following form. In the basic model, where S has no control over the importance of a given play, S typically conveys accurate information for the first several periods. R raises the probability he places on S being friendly after receiving accurate information if he believes that an enemy would have lied to him with positive probability. The more important the information, the more likely an enemy will attempt to deceive R. An enemy will eventually take advantage of R by misleading him. When this happens, the Sender loses all opportunities for deception in the future. I describe the basic model in Section 2. In Section 3, I analyse the one-stage version of the model. Section 4 describes equilibrium behaviour when the game lasts for more than one period.

In Sections 5 and 6, I modify the basic game. In Section 5, S may reduce the importance of the game in any period. This leads to a significant change from the earlier results: the specification of equilibrium now includes a range of information levels in which R trusts no one. That is, in order to influence R, S cannot offer information in a certain range. If S happens to receive information in this range, then he must reduce its importance in order to maintain credibility in future periods.

R controls the importance of the game in Section 6. In the principal equilibrium of this section, R rewards accurate information by increasing the importance of future plays.

The seventh section is a conclusion.

Recent papers by Kreps and Wilson (1982*b*), Kreps, Milgrom and Roberts, and Wilson (1982), and Milgrom and Roberts (1982) were the first to present models in which players, like the unfriendly sender in this paper, exploit uncertainty and take short-term losses in order build a reputation and make long-term gains. Wilson (1985) reviews these papers and other applications of the type of reputation effects that I analyse. My paper owes much to this work.

2. THE BASIC MODEL: THE PRINCIPAL-DOUBLE-AGENT PROBLEM

There is a finite sequence of dates indexed (backwards) by $t = T, T-1, \dots, 1$. Thus, at date t there are t periods remaining. There are two players, S and R. There is a sequence of independent, identically-distributed random variables $\{A_t\}$, each with differentiable probability distribution function $G(A)$ and density $g(A)$ supported on $[\underline{A}, \bar{A}]$, where $\underline{A} < \bar{A}$. A measures the importance of a date's play. Thus, a correct decision becomes more valuable, and an incorrect decision becomes more costly, when A increases. In addition, at the start of period t, S observes the value of a binary random variable M_t, which is independent, identically distributed, and equally likely to be equal to -1 or 1. At date t, S and R play the following game. After observing M_t, S sends a message to R by selecting an element from a set of possible signals, N. R processes the information in S's signal and chooses a real-valued action, y, which determines players' payoffs for that period. After R learns the value of M_t, and both players learn their payoffs, the process repeats.

The game can be interpreted as follows. In each period, S (the "spy") comes to R (the "representative" of a government) with a description of a situation. Both S and R can evaluate the importance of the situation. R is responsible for making a decision, while S has relevant information. S sends a message to R that can be interpreted as an estimate of M. R must then decide how to use this information, taking into account his uncertainty about S's motives. In particular, because he believes that it is possible that the spy is a "double agent", R will be reluctant to trust S fully, particularly on important matters.

Players maximize the undiscounted sum of their single-period payoffs. Continuously differentiable von Neumann–Morgenstern utility functions, $U^E(\cdot)$ and $U^R(\cdot)$, determine the single-period payoffs. R is uncertain about S's preferences. He believes that there is a prior probability $p_T \in (0, 1)$ that S is a friend (F), that is, $U^R(\cdot) = U^F(\cdot)$ describes S's preferences, and probability $1 - p_T$ that S is an enemy (E), that is, $U^E(\cdot)$ describes S's preferences. Thus, S and R have identical preferences when S is a friend, while their preferences differ, in a sense made precise below, when S is an enemy. All aspects of the game other than whether S is a friend or an enemy and the value of M_t in a given period are common knowledge.

I make the following assumptions on the utility function $U^i(\cdot)$, $i = R$ and E.

*Assumption A*1. $U^i(\cdot)$ depends only on $|y-M|$ and A.

Assumption A1 has two consequences. First, it says that the single-period payoffs do not depend on t. Second, it is a symmetry assumption. It says that utility depends only on the distance between y and M. It allows a simplification of the analysis, in which S's two possible lies (communicating 1 when $M=-1$ and communicating -1 when $M=1$) play analogous roles and the probability of lying therefore need not depend on M.

*Assumption A*2. $U^i(\cdot)$ is strictly concave in y.

*Assumption A*3. The solution to the problem:

$$\max_y U^i(y-M, A)$$

is $y=M$ if $i=R$ and $y=-M$ if $i=E$.

Assumptions A2 and A3 guarantee that R has a unique best response to S's signals and makes it unnecessary for R to use mixed strategies in equilibrium. Suppose that i believes that the probability that $M=1$ is r. Then his best action, call it $y^i(r, A)$, is the solution to the problem:

$$\max_y rU^i(y-1, A)-(1-r)U^i(y+1, A).$$

Assumptions A2 and A3 guarantee that $y^i(r, A)$ is well defined for all $r\in[0, 1]$, $A\in[\underline{A}, \bar{A}]$, and $i=R$ and E. Assumptions A1, A2, and A3, combine to imply that $y^R(0, A)\equiv y^E(1, A)\equiv -1$, $y^R(1/2, A)\equiv y^E(1/2, A)\equiv 0$, $y^R(1, A)\equiv y^E(0, A)\equiv 1$, $y^i(r, A)+y^i(1-r, A)\equiv 0$ and that the $y^i(\cdot)$ are strictly monotonic in r. R prefers higher actions in $[-1, 1]$ to lower if $M=1$, and lower actions in $[-1, 1]$ to higher if $M=-1$. E's preferences are the reverse. Thus, the interests of E and R are totally opposed in the sense that if M is known and E weakly prefers an action u to another action v, then R weakly prefers v to u.

*Assumption A*4. $U^i(y^R(0, A)-M, A)\equiv 0$ for all M and A.

*Assumption A*5. If $U^E(\cdot)\neq 0$ then $|U^E(\cdot)|$ is strictly increasing in A, for all M and y.

Assumption A4 is a normalization that guarantees that a totally uninformative signal, which induces R to take the action $y^R(1/2, A)$, yields zero utility to E and F. Assumption A5 makes increases in A increase the importance of the game to E.

I make these restrictive assumptions mostly for convenience. A representative utility function that satisfies Assumptions A1-A5 is $U^R(y-M, A)\equiv A(1-(y-M)^2)$.

Further, except that the principal equilibria require complicated mixed strategies, the qualitative properties of equilibria do not change if the utility functions are linear. For example, the stage game could have the following payoff matrix.

$R \backslash E$	$M=-1$	$M=1$
$y=-1$	$A, -A$	$-A, A$
$y=1$	$-A, A$	$A, -A$

That is, R has two pure strategies available ($y=-1$ and $y=1$) and always disagrees with E on which is best. Very little is lost by replacing the general $U^i(\cdot)$ by linear or quadratic utility functions.

3. THE ONE-STAGE GAME

I begin the analysis with a discussion of the one-stage game. For a fixed prior p_1, an equilibrium consists of the Sender's signalling rules, $q^i(n|M, A)$, and the Receiver's action rule, $y(n, A)$, which satisfy for $i=E$ and F, $M=-1$ and 1, and all $n \in N$:

*Condition E*1. (a) $q^i(\cdot) \geqq 0$ and $\sum_{n \in N} q^i(n|M, A)=1$.
(b) If $q^i(n_0|M, A)>0$ for some $n_0 \in N$, then n_0 solves: $\max_{n \in N} U^i(y(n, A)-M, A)$.

*Condition E*2. $y(n, A)=y^R(r(n, A), A)$ where $r(n, A)$ is the probability that $M=1$ given n and A so that,

$$r(n|A)=\frac{p_1 q^F(n|1, A)+(1-p_1)q^E(n|1, A)}{p_1(q^F(n|1, A)+q^F(n|0, A))+(1-p_1)(q^E(n|1, A)+q^E(n|0, A))}$$

if

$$p_1(q^F(n|1, A)+q^F(n|0, A))+(1-p_1)(q^E(n|1, A)+q^E(n|0, A))>0.$$

Conditions E1 and E2 describe a Nash Equilibrium. The first condition says that S's signalling rule is a probability distribution[1] and that any signal S sends in equilibrium is a best response to R's action rule. Condition E2 uses Bayes' Rule to compute R's posterior distribution on M given his information: the value of A and the signal n; $r(n, A)$ is the probability that $M=1$ given A and n. If S uses a signal that is unexpected, that is, if S sends a signal n such that

$$p_1(q^F(n|1, A)+q^F(n|0, A))+(1-p_1)(q^E(n|1, A)+q^E(n|0, A))=0,$$

then Bayes' Rule cannot be used to compute $r(n, A)$. If R does not expect n_0 in equilibrium, it must be the case that n_0 is not the unique maximizer of $U^i(y(n, A)-M, A)$. Since $r(n, A)$ determines $y(n, A)$, the specification of $r(\cdot)$ cannot be made arbitrarily in equilibrium. Finally, Condition E2 guarantees that R's action is a best response to the signal.

Notice that the equilibrium conditions hold pointwise in A. To ease notational clutter, I suppress reference to A for the remainder of this section.

By the assumptions on preferences and Condition E1, it follows that at most two actions are induced (signalled for with positive probability) in equilibrium. If $Y=\{y: y=y(n)$ for some $n \in N\}$, then $Y \subset [-1, 1]$ and the maximization problem in Condition E1 is equivalent to picking $y \in Y$ to maximize $U^i(y-M)$. This problem has a solution if and only if Y has a least and greatest element; call these $\underline{y}$ and $\bar{y}$ respectively. $\underline{y}$ and $\bar{y}$ typically depend on the data of the problem, A and p_1. In order to satisfy Condition E1(b), S sends a signal that induces R to take action $\underline{y}$ if $S=F$ and $M=-1$ or $S=E$ and $M=1$, and $\bar{y}$ if $S=F$ and $M=1$ or if $S=E$ and $M=-1$. Without loss of generality, I restrict attention to signalling spaces that have two elements, say $N=\{-1, 1\}$, and assume that F weakly prefers to make signal i if $M=i$. That is, I assume that $y(-1)=\underline{y}$ and $y(1)=\bar{y}$.

These comments make a complete characterization of equilibria in the one-stage model routine.

Theorem 1. *If $T=1$, then there is an equilibrium in which $y(-1)=y(1)=y^R(1/2)$. This is the only equilibrium when $p_1 \leqq \frac{1}{2}$. If $p_1 > \frac{1}{2}$, then there is exactly one other equilibrium; in it, $y(-1)=y^R(1-p_1)$ and $y(1)=y^R(p_1)$.*

The proof of Theorem 1 is in the Appendix. The no-transmission equilibrium, in which R takes the same action regardless of S's signal, always exists. If R chooses to ignore S's signal, then it is a best response for S to make non-informative signals. However, if S is sufficiently reliable, then there is an equilibrium in which R uses the signal in order to determine his action. This equilibrium appears to be more sensible than the first for two reasons. First, the no-transmission equilibrium leads to an expected payoff of zero for both types of S and for R, which is lower for all players (both types of S and R) than the payoffs available in the other equilibrium. Second, one expects F's signals to be informative in equilibrium. In what follows, I restrict attention to equilibria in which F sends different signals for different values of M. These equilibria involve non-trivial communication wherever it is possible. Assuming that signals are labeled so that $y(1) \geqq y(-1)$, this means that $q^F(n|M)=1$ if and only if $n=M$. This type of signal allows F to communicate the maximum amount of information to R. I call F *honest* if he uses this type of signalling rule; I intend to concentrate on honest equilibria: those equilibria in which F is honest.

Corollary 1. *If $T=1$, then there is a unique honest equilibrium in which $y(-1)=y(1)=y^R(\frac{1}{2})$ if $p_1 \leqq \frac{1}{2}$ and $y(-1)=y^R(1-p_1)$ and $y(1)=y^R(p_1)$ if $p_1 > \frac{1}{2}$.*

The proof of Corollary 1 is in the Appendix.

When $p_1 > \frac{1}{2}$, E always lies to R; for either value of M, E and F always make different signals. Although R knows that this is E's strategy, he is sufficiently confident that S is a friend to take different actions for different signals. When $p_1 \geqq \frac{1}{2}$, there is no equilibrium in which F is honest and E always lies. Given signalling rules of this form, R would choose to take a higher action when he receives a lower signal, but then S's signalling strategy would not be optimal. Instead, when F is honest and $p_1 < \frac{1}{2}$, E must randomize to cover up the information in F's signal. The strategy presented in the proof of Corollary 1 makes $r(-1)=r(1)=\frac{1}{2}$.

The one-stage game is similar to the information transmission models of Green and Stokey (1981*a*) and Crawford and Sobel (1982). These models have more general uncertainty about the state of nature (M), but no uncertainty about preferences. Green and Stokey (1981*b*) and Holmström (1980) analyze similar one-stage models, but allow the uninformed player to commit himself to an action rule.

4. THE MULTI-STAGE GAME

When there is more than one stage, E sometimes wants to supply useful information to R at first in order to take advantage of his reputation later. Now I build on the analysis of the one-stage game to describe equilibria of the finitely-repeated game. As in Section 3, I restrict attention to signalling spaces that have two elements, $N=\{-1, 1\}$, and I identify the signal $n=-1$ with the lower action induced. I characterize only the honest equilibria. Corollary 1 provides a justification for this in the one-period game; a similar result, Corollary 2, holds for the finitely-repeated game. Also, I assume that the state $M=-1$ and $M=1$ lead to symmetric behaviour so that, in particular, the probability that E and F send the same signal does not depend on M. This assumption actually is a straightforward consequence of the symmetry assumptions on utility functions (for a proof, see Sobel, 1983).

With these simplifications, an honest equilibrium consists of signalling rules for S, denoted by $q_t^i(A, p)$ for $i=E$ and F, which represent the probability that i tells the truth (reports $n=M$) given A with t periods remaining and p is the probability that R believes $S=F$, and action rules for R, denoted by $y_t(n, A, p)$. Given the symmetry assumptions, an honest signal induces an action that differs in absolute value from M by $z_t(h, A, p) \equiv 1-y_t(1, A, p) \equiv 1+y_t(0, A, p)$, while a dishonest signal induces an action that differs in absolute value from M by $z_t(d, A, p) \equiv 1+y_t(1, A, p) \equiv 1-y_t(0, A, p)$. In an honest equilibrium, $q_t^i(\cdot)$, $y_t(\cdot)$, and $z_t(\cdot)$ must satisfy:

Condition $E0'$.

$$p_{t-1}(l, A, p)=\begin{cases}0 & \text{if } l=d \text{ or } p=0\\ p/[p+(1-p)q_t^E(A, p)] & \text{if } l=h \text{ and } p=0.\end{cases}$$

Condition $E1'$. (a) $0 \leqq q_t^i(\cdot) \leqq 1$
(b) $q_t^F(\cdot)=1$
(c) If $q_t^i(A, p)>0$, then $W_t^i(h, A, p) \geqq W_t^i(d, A, p)$, and if $1>q_t^i(A, p)>0$, then $W_t^i(d, A, p) \geqq W_t^i(h, A, p)$ where $V_0^i(p, A) \equiv 0$, $\bar{V}_t^i(p)$ is the expected value of $V_t^i(p, A)$, $W_t^i(l, A, p) \equiv U^i(z_t(l, A, p), A)+\bar{V}_{t-1}^i(p_{t-1}(l, A, p))$.

Condition $E2'$. $y_t(n, A, p)=y^R(r_t(n, A, p), A)$ where $r_t(n, A, p)$ is the probability that a signal is honest given A so that $r_t(n A, p)=p+(1-p)q_t^E(A, p)$.

Given p_T, Condition E0′ defines a sequence of probabilities, $p_T, p_{T-1}, \ldots, p_1$, where if A_t is the realized value of A in period t, and $l_t=d$ or h depending on whether S is dishonest or honest in period t, then $p_{t-1}=p_{t-1}(1_t, A_t, p_t)$.

The function $p_{t-1}(\cdot)$ defined in Condition E0′ is an expression for the probability that $S=F$ given S's signalling rule. When $l_t=d$, $p_{t-1}(\cdot)=0$ since in an honest equilibrium R believes that F always is honest. Further, once R discovers a lie he remains certain that S is an enemy regardless of what happens in the remaining periods. The second line in Condition E0′ follows from Bayes' Rule since $p+(1-p)q_t^E(A, p)$ is the probability of an honest signal. The probability p_t summarizes all of the information that R needs to know about S when t periods remain. In general, strategies may depend on more detailed descriptions of the history than the sequence $p_T, p_{T-1}, \ldots, p_1$. However, this information is not relevant to R in an honest equilibrium. To see why, I argue inductively. By Corollary 1, there is a unique honest equilibrium when $T=1$; p_1 completely determines equilibrium strategies. Therefore, all equilibrium strategies for the final period do not depend on past information other than p_1. Similarly, Theorem 2 shows that an honest equilibrium for any t-period game is determined uniquely by p_t. Conditioning on histories may be used to select from a number of equilibrium outcomes for the remainder of the game[2], but since honesty is enough to predict definite outcomes, information not included in p_t plays no role in the t-period equilibrium. Following Corollary 2, I discuss an example of an equilibrium in which strategies depend on more general descriptions of history than the posterior probability that S is F.

Condition E1′ describes S's signalling rule. Condition E1′ (a) says that S's signalling rule is a probability distribution over signals, Condition E1′(b) says that F is honest, and Condition E1′(c) guarantees that S best responds to R's action rule. Condition E2′ uses Bayes' Rule to compute R's posterior distribution on M given his information: the value of A; the signal n; and the probability that $S = F$, p. Bayes' Rule determines the probability that S is telling the truth, $r_t(\cdot)$. Notice that Condition E2′ implies that R's period-t payoff determines his action in period t. This is because his payoff with t periods remaining depends on $y_t(\cdot)$ only in period t. Thus, in order to best respond to S's signal, R must behave myopically.

I must define equilibrium strategies for all p, not simply for the equilibrium realizations $p_T, p_{T-1}, \ldots, p_1$. This is because only by defining these strategies can the players compute payoffs associated with non-equilibrium behaviour. The definition of equilibrium guarantees that continuations of honest equilibria are also honest equilibria. Moreover, once S lies the relationship between R and S stops in that S's signals are not informative and R's actions do not depend on S's signal.

Let $V_t^R(p, A)$ be the expected value function for R when t periods remain and R believes that $S = F$ with probability p. The corresponding value functions, $V_t^i(p, A)$ for $i = E$ and F, have been defined in Condition E2′.

Theorem 2 characterizes the equilibrium.

Theorem 2. *There is a unique honest equilibrium. In it, for $i = E$, F, and R, $\bar{V}_t^i(p) = 0$ for $p \leqq 2^{-t}$ and is strictly positive and increasing in p and t for $p > 2^t$. There is a function $H_t(p)$, increasing in p and t, such that E never lies in period t if $A < H_t(p)$. The probability that S's signal is honest and the degree to which R believes S increases in p and t.*

Before discussing Theorem 2, I present a justification for concentrating on honest equilibria.

Corollary 2. *If the honest equilibrium is played when $s < t$, then R and both types of S prefer F to be honest in period t than any other equilibrium behaviour.*

I prove Theorem 2 and Collorary 2 in the Appendix.

Corollary 2 justifies the honest equilibrium inductively. If all of the players believe that the honest equilibrium will be played for s-stage games, then they cannot do better than play honestly in an $(s+1)$-stage game. As Corollary 1 establishes the prominence of the honest equilibrium for the one-stage game, players expect it to be played. Thus, they have no reason not to play it in longer games. On the other hand, Corollary 2 does not prove that the honest equilibrium Pareto dominates all other equilibria of the game. This is not true when $T > 1$. For example, let $T = 2$ and suppose that F uses a signalling rule when $t = 1$ that depends on R's action when $t = 2$. In particular, suppose that F uses an honest signalling rule when $t = 1$ if and only if $y_2(n, A, p) = n$ for $n = 0$ and 1, and otherwise F conveys no information so that $q_1^F(n|M) = \frac{1}{2}$ for $n = 0$ and 1 and $M = -1$ and 1. If F chooses to convey no information, then the only equilibrium strategy for R when $t = 1$ is $y_1(n) = y^R(\frac{1}{2})$ for $n = 0$ and 1. Therefore, unless R takes F's most preferred action when $t = 2$, the no-transmission equilibrium is selected when $t = 1$. If p_2 is close to 1, then R prefers to set $y_2(n, A, p) = n$ rather than sacrifice a positive payoff when $t = 1$. With appropriate signalling rules for E, these strategies then comprise a sequential equilibrium (see Kreps and Wilson (1982*a*)). F prefers this equilibrium to the honest equilibrium because he induces his most preferred outcome in both periods. Provided that E lies with positive probability in the honest equilibrium, he too does better since a lie causes R to take E's most preferred action. Notice that in this equilibrium R does not make his myopic best response to the signal when $t=2$, because E will lie to him with positive probability. This is possible because F conditions his $t=1$ signal on R's action. I could restrict attention to histories summarized by the probabilities p_t when I characterized honest equilibria because p_t uniquely determines equilibrium strategies. In general, however, conditioning on earlier strategies can be used to make selections from multiple equilibria. This is what happens in the example. I have chosen to restrict attention to honest equilibria because Corollaries 1 and 2 suggest that F's threat to revert to the no-transmission equilibrium is not plausible.

In the honest equilibrium, E does one of three things, depending on the value of A. If A is small relative to the number of periods remaining and E's current reputation ($A<H_t(p)$), E always tells the truth because he expects to have a more attractive opportunity for deception in the future. Therefore, when A is sufficiently low, R is willing to believe anything that S says, but for this reason honesty does not improve S's reputation. E always lies with positive probability if his current reputation has no future value.

For larger values of A, E lies with positive probability. If p and A are sufficiently large, then E lies with probability one. This type of behaviour only occurs if $p\geqq\frac{1}{2}$ since otherwise R would not trust the signal. For some values of A, E uses a mixed strategy. If S does tell the truth in these situations, then the fact that R thought that E might lie leads R to feel more confident about S in the future. Without further assumptions about preferences, I cannot guarantee that increases in A increase the equilibrium probability of lying. This is because an increase in A may change $y^R(p, A)$ so as to make lying less costly to E. However, if $y^R(\cdot)$ does not depend on A, increases in A always make lying more attractive to E, and the equilibrium probability of lying increases as a result.

Theorem 2 sheds some light on how the relationship between S and R depends on the reputation of S, p, and the number of periods remaining, t. Increasing p is beneficial to R because it increases his chance of dealing with a friend. Increasing p is beneficial to S because R receives more informative signals (honesty is more likely) so the payoffs associated with honesty (for F) and dishonesty (for E) increase. When more periods remain or when p is higher, the value of pretending to be a friend is high to an Enemy Sender.

The variable importance level A provides an incentive for Enemy Senders to supply honest information. If $\underline{A}=\bar{A}$, so that the importance of each stage game does not vary over time, then E signals honestly with positive probability in some periods. However, E is never honest with probability one. Here is why. E tells the truth if he expects to have an opportunity to gain more by lying in the future. This can happen for two reasons. Either there is a chance that future opportunities are more valuable (higher A), or that R will be more trusting in the future (higher reputation). The first cannot happen if $\underline{A}=\bar{A}$ and the second cannot happen unless R has some reason to believe that S is more likely to be a friend, which in turn only happens if R expects E to lie with positive probability and then receives an honest signal. Thus, the role of a variable A in the model is to provide situations in which E always tells the truth.

R prefers to deal with a single Sender for T periods rather than T Senders separately. This is because repeated plays allow him to better evaluate the usefulness of S's information. Long-term arrangements help to moderate the inefficiencies caused by incomplete information.

Theorem 3. *In the honest equilibrium,* $\bar{V}_t^R(p)\geqq t\bar{V}_1^R(p)$ *for all* t *and* p.

Proof. The proof requires two observations. The first step is to observe that if $0<p$ and $\hat{p}<1$, then

$$pV_1^R(\hat{p}, A) \geqq V_1^R(p\hat{p}, A). \tag{1}$$

When $p\hat{p} \leqq \frac{1}{2}$, (1) follows since $V_1^R(p\hat{p}, A)=0$. Otherwise,

$$\begin{aligned} pV_1^R(\hat{p}, A) &= p\hat{p}U^R(y^R(\hat{p}, A)-1, A)+p(1-\hat{p})U^R(-y^R(\hat{p}, A), A) \\ &> p\hat{p}U^R(y^R(p\hat{p}, A)-1, A)+p(1-\hat{p})U^R(-y^R(p\hat{p}, A), A) \\ &> p\hat{p}U^R(y^R(p\hat{p}, A)-1, A)+(1-p\hat{p})U^R(-y^R(p\hat{p}, A), A) \\ &= V_1^R(p\hat{p}, A). \end{aligned}$$

The first inequality follows from the definition of $y^R(\cdot)$, the second inequality follows because $U^R(-y^R(p\hat{p}), A)<0$ whenever $p\hat{p}>\frac{1}{2}$, and the equations follow from the definition of $V_1^R(\cdot)$ and because in the honest, one-period equilibrium if the probability of S being E is less than one half, E is always dishonest.

The second step is to note that for some $r \in [p, 1]$,

$$\bar{V}_t^R(p, A) \geqq V_1^R(p, A)+r\bar{V}_{t-1}^R(p/r). \tag{2}$$

To obtain (2), note that

$$V_t^R(p, A)=rU^R(y^R(r, A)-1, A)+(1-r)U^R(-y^R(r, A), A)+r\bar{V}_{t-1}^R(p/r),$$

where r is the probability that R receives an honest signal in period t. Thus, since R receives an honest signal more often in the first of t periods than in a single period (when E is always dishonest or no information is transmitted), (2) is established.

The proof now follows from induction. The theorem is true when $t=1$, and if it holds for $s-1$ periods, then

$$\begin{aligned} V_s^R(p, A) &\geqq V_1^R(p, A)+r\bar{V}_{s-1}^R(p/r) \\ &\geqq V_1^R(p, A)+(s-1)r\bar{V}_1^R(p/r) \geqq V_1^R(p, A)+(s-1)\bar{V}_1^R(p), \end{aligned} \tag{3}$$

where the first inequality follows from (2), the second from the induction hypothesis, and the third from (1). Taking expected values of (3) establishes the theorem. ||

While Theorem 3 indicates the advantages of dealing with a single S, it is a relatively weak result. It would be of interest to show that

$$\bar{V}_t^R(p) \geqq \bar{V}_{t-s}^R(p)+\bar{V}_s^R(p). \tag{4}$$

However, (4) will not hold without further assumptions. First, thus far I have assumed that E and R have totally opposed preferences in any single period, but that they may differ in their assessments of a situation's importance. If E and R do differ in their ordering of A, then there may be situations in which (4) does not hold. To see this, suppose that increasing A increases $|U^E(\cdot)|$, but that increasing A first decreases, then increases, $|U^R(\cdot)|$, so that R loses very little when he is duped for intermediate values of A. As the length of the horizon increases, it is more likely that E will wait for a high value of A and cause more damage to R when he is dishonest. In this case, ex ante, R would prefer to shorten the maximum feasible length of the partnership and use two informants for short periods rather than one for a long period. This problem can be ruled out by assuming that E and R have opposing interests in an extended sense, for example, if $U^E(y-M, A) \equiv -U^R(y-M, A)$ so that E and R always agree on the importance of a decision and disagree on the appropriate action.

Assuming that $U^E(y-M, A) \equiv -U^R(y-M, A)$ does not appear to be enough to prove (4), however. Since R is risk averse, he is willing to pay a premium to discover if S is an enemy. However, because commitment is not possible, he cannot do this by being more credulous in early periods and thereby inducing an unfriendly Sender to be dishonest early. On the other hand, a shorter horizon allows fewer opportunities for deception and R may prefer to limit the length of attachments so that the first periods will provide stronger tests of S's motives. When $U^i(\cdot)$ is linear, (4) is satisfied, indicating that risk aversion may play a role in discouraging long-term associations if commitment is not possible. When $\underline{A} = \bar{A}$, (4) is also satisfied. This suggests that part of the problem with long-term arrangements is the increased possibility that E could deceive R on a very important matter.

Many of the assumptions made in this section are not really necessary. I made the symmetry assumption A3 for exposition convenience. Aside from the fact that the probability of being honest will depend on M, no results will be altered by relaxing the assumption. The distribution of A and the utility functions can depend on t; in particular, introducing discounting presents no problem. More general types of preferences for E may be allowed. If there are only two states of the world, preferences will either be totally opposed or in agreement. If it is common knowledge that E's preferences are identical to R's periodically, then the results do not change; there is complete trust in those periods. On the other hand, if the nature of the uncertainty about M is more complicated, then there are more possibilities for diversity of preferences. In such a variation it might be possible to analyse the effect that increasing the degree by which preferences differ has on the characteristics of equilibrium.

The results remain basically unchanged if R observes A only after he makes a decision. There typically is an interval of values of A such that E always tells the truth, followed by an interval in which mixed strategies are used, and then by an interval in which E always lies. Changing the payoff structure could make more or less randomization necessary. If $U^R(\cdot)$ were not strictly concave, then R would need to randomize. If, for example, R observed the value of A with noise (or, for some reason, was uncertain about the value of A that entered S's utility function), randomization by S would probably not be necessary. What is needed is a way by which R places non-degenerate probability on E telling the truth. This can be done in a variety of ways.

Another modification is worth noting. If R can commit himself to an action rule in a given period, but is unable to commit himself to ignoring information in the future, the equilibrium changes somewhat. R would like to take more drastic (be more credulous) actions in the early periods. This increases the probability that E is dishonest, and thus allows R to be more confident that information provided in future periods is reliable. Formally R's payoff can be divided into two parts, the future expected utility and the utility he expects from the current period's decision. Since he is unable to make commitments, R must use the current period's information optimally. This maximizes the contribution of the current period to expected utility. However, the future consideration argues for R to increase the amount of dishonest behaviour in the present period in order to be more confident, after he receives reliable information, that he is dealing with a friend. If R could make commitments, this effect would cause him to be more credulous in the current period in order to be more confident in the future. However, it will typically not be optimal for E to lie with probability one in period T, and the analysis of this section captures the qualitative properties of the equilibria with commitment.

In the model an enemy has only one opportunity to mislead R. Afterwards, his reputation falls and R believes nothing that he says. While he is providing useful information, S's reputation increases through time. Several simple modifications of the model would alter these results. If S did not receive perfect information, then R could not be certain that a misleading signal was deliberate or accidental. In this case, mistakes would cause F's reputation to fall, and E may be able to deceive R at regular intervals. It is worth adding that if R had the option of replacing S with another player whose reputation was the same as S's original reputation, he would do so in the present model only if S is an enemy who has just lied. If S can make mistakes, then there is a positive probability that a friend will be replaced.

5. REDUCING AGENTS

Theorem 2 showed that if A_t is sufficiently low E will send the same signal as F with probability one. This is costly to E; he makes short-term losses because R is able to make a well-informed decision, but does not improve his reputation. It is only the expectation that future values will be greater than the current one that makes his strategy optimal. However, if E was able to manipulate A_t, he might choose to provide less valuable information in these situations. Continuing the analogy with spying, a double agent might convey accurate, but less than comprehensive, information in order to lower his losses when deception is not justified. To allow for this possibility, in this section it is assumed that S observes the value of A_t, again drawn from the distribution described by $G(\cdot)$, and then selects a reduced value $a_t \in [\underline{A}, A_t]$ to report to R, thereby lowering the potential importance of his services to R. Next he sends a message to R about M as before, and R chooses an action. R learns the reduced value a_t when S reports it; this is the value that enters his utility function. R never learns the true value of A_t. I continue to make all of the other assumptions of Section 2.

This section presents an informal description of the equilibrium for this model. The details are in Sobel (1983). For the reducing model, the strategies of the players include signalling rules for S and decision rules for R, as in Section 2. In addition, S must have a rule that tells them what level of importance a to report given t, p, and A_t. R evaluates S's reputation on the basis of S's report, a, so that an Enemy Sender cannot systematically reduce the importance of the information that he provides without lowering R's trust in him.

There always exists an honest equilibrium with reducing. It involves both types of S reducing A to $\underline{A}$ in every period. This equilibrium can be supported if R refuses to trust anyone who offers information $A > \underline{A}$; consequently, no one offers it. The reduced equilibrium with importance $\underline{A}$ in each period can be analysed as in Section 4. The equilibrium is extreme because R never is able to gain access to any important information. It might be hoped that there are other equilibria that both types of S and R prefer to this one. In particular, just as it was reasonable to expect F to put as much information as possible about M into his signal (that is, to be honest), it seems reasonable to expect F never to reduce the importance of this information. It turns out that this is not possible in general.

Theorem 4. *If $U^E(y-M, A) \equiv U^F(y-M+1, A)$, $U^E(y-M, A) \equiv 0$, and $T>1$, then in every honest equilibrium there exists a nondegenerate interval of importance levels A that neither type of S provides in some period. Moreover, the reputation of S cannot be monotonically increasing in the level of A provided.*

For a proof of Theorem 4 see Sobel (1983).

The message of Theorem 4 is that the ability to manipulate the importance of a situation adds a new inefficiency to the problem of contracting with an agent who has uncertain motives. In order to avoid placing confidence in an agent who provides accurate, but unimportant, information, R must be wary of an S who reports low values of A. Theorem 4 states that this is not consistent with accepting certain values of A. A typical equilibrium involves S reducing to A those values of A that would cause E to be honest with probability one. F must also reduce these values, for otherwise an enemy S could obtain a good reputation at low cost by imitating F. In order to persuade F not to reduce A, R must conjecture that only unreliable agents provide certain levels of A. This means that reputation cannot generally increase with the level of information provided.

The assumption that $U^E(y-M, A) \equiv U^F(y-M+1, A)$ guarantees that increasing A increases importance for both R and S. The assumption that $U^E(y-M, A) \equiv 0$ guarantees that there are some levels of importance so low that E would prefer to always tell the truth in the hope of making a big gain tomorrow, than lying with positive probability.

This type of model applies to economic situations. For example, let the informed player, S, be in charge of market research for R. S learns two things about R's product. He learns the appeal of the item, measured by A. S also learns whether the product works (or can be produced profitably), measured by M. R wants to market the product if $M=1$, but does not want to market it if $M=-1$. In this context, assume that a friendly Sender's preferences agree with R's, but an Enemy Sender, never wants R to market the product.[3] The results of this section apply if I assume that R learns the value of M after he makes his marketing decision and that S is able to reduce that size of A by claiming that only $a<A$ people might use the product. In this case, R observes a but not A. For this model, the results indicate that E provides reliable information for a period while waiting for an opportunity to convince R not to produce when A is large and $M=1$. Moreover, the larger is A, the more costly it is for E to recommend production. Consequently, the ability to reduce the size of projects is attractive to E. Theorem 4 states that in some situations even F must reduce A in order to maintain credibility. A particular form of equilibrium behaviour involves R marketing only those products that attract a large enough group of people. R does this not because of any economic of scale in production, but because if R believes reports of low values of A then he allows Enemy Senders to deceive him at low cost.

The model of this section has the property that the level of information (magnitude of A) and not just its precision (relationship to M) convey information to R about S's motives. When S has control over the level of information, R must be wary of Senders who do not provide important information; otherwise, E will take advantage of him by supplying important information only when he intends to lie.

6. A LOAN MODEL

The model of this section does not involve direct information transmission, but shares with earlier models the idea of uncertainty about motives. I will call two players the Borrower (B) and the Lender (L). The game repeats a fixed, finite number of times. At any stage, L decides on an amount $A \in [0, 1]$ to lend to B. B, taking M as given, decides whether to invest (I) or to default (D). There are two type of B, E and F, and L assigns the initial probability p_T that $B=F$. The payoffs for a single play of the game are given by the utility functions $U^i(A, j)$. In this section, I assume that

$$U^F(A, D) = U^L(A, D) = U^E(A, I) = -A$$

and

$$U^F(A, I) = U^L(A, I) = U^E(A, D) = A.$$

The payoffs suggest that E cannot produce anything with the loan, but is able to take the money. F cannot costlessly default, but he is able to invest the money and make a profit, which is shared with L. The two types of borrowers differ only in their ability to produce income from a loan. For example, if a friendly borrower can always triple the value of the loan, while an enemy borrower cannot increase the loan's value, then preferences similar to the ones used arise provided that the lender requires that a non-defaulting borrower returns \$2 for each \$1 borrowed.[4]

Informally an equilibrium consists of loans, $A_t(p)$, probabilities that $B = F$, p_t, and probabilities of defaulting, $d_t^i(p, A)$ for $i = E$ and F, such that reputations are updated by Bayes' Rule when possible, L best responds and B puts positive probability only on strategies that maximize utility. In this definition, L is able to commit himself to a value A_t in each period, but cannot make commitments about future loans. It turns out that when $U^i(\cdot)$ is linear, this restriction does not influence L's payoffs.

Theorem 5. *The equilibrium to the loan model is*

$$A_t(p) = \begin{cases} 2^{-k} & 2^{-k} \geqq p \geqq 2^{-k-1}, \quad k = 0, \ldots, t-1, \\ 0 & 2^{-t} > p \geqq 0, \end{cases}$$

$$d_t^F(p, A) \equiv 0,$$

$$d_t^E = \begin{cases} 1 & \text{if } A > A_t(p), \text{ or if } A = A_t(p) \text{ and } p \geqq \frac{1}{2} \\ p/(1-p) & \text{if } A = A_t(p) \text{ and } p < \frac{1}{2} \\ 0 & \text{if } A < A_t(p). \end{cases}$$

The associated value functions are

$$V_t^E(p) = \begin{cases} 2^{-k} & 2^{-k} \geqq p \geqq 2^{-k-1}, \quad k = 0, \ldots, t-1, \\ 0 & 2^{-t} > p \geqq 0, \end{cases}$$

$$V_t^F(p) = \begin{cases} t - \sum_{i=1}^{-k} 2^{-i} & 2^{-k} \geqq p \geqq 2^{k-1}, \quad k = 0, \ldots, t-1, \\ 0 & 2^{-t} \geqq p \geqq 0 \end{cases}$$

and $\hat{V}_t^L(p) \equiv pV_t^F(p) - (1-p)V_t^E(p)$.

The proof is a routine verification, I omit it.

In equilibrium, L is unwilling to loan the full amount to B unless he is confident that B is a friend. The level of confidence does not depend on the number of periods remaining; until full confidence is reached ($p \geqq \frac{1}{2}$), the amount of loan increases with each successful loan. Also, the value of the game is increasing in t and p for all players and if $p > 0$, then $\lim_{T \to \infty} V_T^i(p) > 0$, so that if there is some chance that B is a friend, there are T's large enough so that it is worthwhile for L to make loans. Notice that F never defaults, and E is indifferent between defaulting or investing in the first period. When L can commit himself to A_t in period t, $d_t^E(p, A_t)$ is not uniquely determined although it cannot be equal to one (the largest feasible amount) unless L can commit himself to the entire future sequence of loans. On the other hand, if $d_t^E(p, A_t(p)) = p/(1-p)$ when $p \leqq \frac{1}{2}$, the L's prior on the probability of default is $\frac{1}{2}$ and he is indifferent about amounts that he lends at t; in particular, he is willing to lend $A_t(p)$ and he does not need to be able to commit himself to do so. As a final remark about the equilibrium, note that L would prefer to have T opportunities to lend to a given B, rather than to make T individual loans to different Borrowers (each with probability p_T of being friendly). This is consistent with Theorem 3 and underlines a basic idea of the analysis: Long-term arrangements are of value when there is uncertainty about preferences because past transactions provide relevant information to agents.

The model of this section is much too simple to give useful insights into credit markets. More general preferences can be allowed without destroying the flavour of the results. Yet it seems more important to extend the analysis to a market setting, and allow the Lender to select a repayment schedule (that is, an interest rate).

7. CONCLUSION

If an agent is uncertain about the motives of someone upon whom he must depend, either to provide information or make decisions, then the extent to which he trusts the other will be based on the partner's earlier actions. Thus there is an incentive for an enemy to behave like a friend in order to increase his future opportunities, and for partnerships to last until someone cashes in. I have tried to make this point with two simple models. Technical extensions, such as more general uncertainty, could add further insights.

Finally, it should be emphasized that in my models credibility can only be established through actions that influence past payoffs. The ability to make commitments or to take actions that influence future payoffs would also help to determine an agent's reliability.

APPENDIX

Proof of Corollary 1. If F is honest and

$$q^E(M|M)=\begin{cases}(1-2p_1)(2(1-p_1))^{-1} & \text{if } p_1\leqq\frac{1}{2}\\ 0 & \text{if } p_1>\frac{1}{2}\end{cases}$$

and $q^E(1-M|M)=1-q^E(M|M)$ for $M=-1$ and 1, then S best responds to the action rules given in the statement of the corollary and $r(n)=\max(p_1,\frac{1}{2})$, so that R's action rules best respond to these signals. Further, if F is honest, then $y(-1)<y(1)$ provided $p_1>\frac{1}{2}$, so that the no-transmission equilibrium cannot be an honest equilibrium if $p_1>\frac{1}{2}$. The corollary now follows from Theorem 1. ||

Proof of Theorem 1. If S uses the strategies $q^i(n|M)=\frac{1}{2}$ for $i=E$ and F, $n=-1$ and 1, and $M=-1$ and 1, then $r(n)=\frac{1}{2}$ for $n=-1$, so that Condition E2 is satisfied when $y(-1)=y(1)=y^R(\frac{1}{2})$. In addition, if $y(-1)=y(1)=y^R(\frac{1}{2})$, then these signalling rules satisfy E1. If $y(-1)<y(1)$ in equilibrium, then by E1 the equilibrium signalling rules must be

$$q^F(n|M)=\begin{cases}1 & \text{if } n=M\\ 0 & \text{if } n\neq M\end{cases}$$

and $q^E(n|M)=1-q^F(n|M)$. Therefore, $r(1)=p_1$ and $r(-1)=1-p_1$, so that $y(-1)=y^R(1-p_1)$, and $y(1)=y^R(p_1)$. Thus, an equilibrium with $y(-1)<y(1)$ exists if and only if $y^R(1-p_1)<y^R(p_1)$, which occurs exactly when $p_1>\frac{1}{2}$. The theorem follows because if $y(-1)=y(1)$, then $y(-1)=y(1)=y^R(\frac{1}{2})$. ||

Proof of Theorem 2. Fix p and A and suppose that the theorem holds for all $s = t-1, \ldots, 0$. Setting $V_0^i(p, A) \equiv 0$ for $i = E$, F, and R, it suffices to prove that the theorem holds for t. If q is the probability that E is honest, then y_j, the action induced when $n = j$, r, the probability that $M = n$, v, the probability that $S = F$ given $M = n$, and $z(l)$, the absolute value of the difference between M and the action induced by an honest (if $l = h$) or dishonest (if $l = d$) signal are simple functions of q. Specifically, it must be that

$$r = p + (1-p)q, \tag{A1}$$

$$v = p/[p + (1-p)q], \tag{A2}$$

$$y_{-1} = y^R(1-r, A) \quad \text{and} \quad y_1 = y^R(r, A), \tag{A3}$$

and

$$z(d) = 1 + y_1 \quad \text{and} \quad z(h) = 1 - y_1. \tag{A4}$$

Finally, let $W^i(l, q)$ denote the value to i of an honest (if $l = h$) or dishonest (if $l = d$) signal when q is the probability that E tells the truth. That is, for $i = E$ and F let

$$W^i(h, q) = U^i(z(h), A) + \bar{V}_{t-1}(v)$$

and

$$W^i(d, q) = U^i(z(d), A) + \bar{V}_{t-1}(0) = U^i(z(d), A).$$

The expression for $W^i(d, q)$ follows because the probability $S = F$ is zero after a lie and because $\bar{V}_{t-1}(0) = 0$ by the induction hypothesis. $W^i(\cdot)$ depends on q since q determines $z(\cdot)$ and v through (A1)-(A4).

The probability q is an equilibrium value if and only if

$$W^F(z(h), q) \geqq W^F(z(d), q), \tag{A5}$$

if $q < 1$, then

$$W^E(z(h), q) \leqq W^E(z(d), q), \tag{A6}$$

and if $q > 0$, then

$$W^E(z(h), q) \geqq W^E(z(d), q). \tag{A7}$$

(A5) guarantees that honesty is a best response for F, and (A6) and (A7) guarantee that q is a best response for E.

I now prove that there exists a unique $q \in [0, 1]$ that satisfies (A5), (A6), and (A7). First, note that if (A6) and (A7) are satisfied then $z(h) \leqq z(d)$ and therefore, since $U^F(\cdot)$ is decreasing in z for $z > 0$, (A5) holds whenever both (A6) and (A7) hold. To see this, assume $z(h) > z(d)$. Then $W^E(z(h), q) > W^E(z(d), q)$. This follows from the definition of $W^E(\cdot)$ because $U^E(\cdot)$ is increasing in z for $z \geqq 0$. Thus, (A6) implies that $q = 1$. However, if $q = 1$, then (A1)-(A4) imply that

$$0 = z(h) < z(d) = 2, \tag{A8}$$

which contradicts $z(h) \geqq z(d)$. Therefore, it suffices to show that there exists a unique q that satisfies (A6) and (A7). However, this follows easily since $W^E(z(d), q) - W^E(z(h), q)$ is strictly decreasing in q.

Next, I verify that $\bar{V}_t^i(\cdot)$ has the desired properties. If $p \leqq 2^{-t}$, then when $q = (1-2p)/[2(1-p)]$, (A1)-(A4) imply that $r = \frac{1}{2}$, $v \leqq 2^{-t+1}$, and $z(d) = z(h) = 1$. Thus, since $\bar{V}_{t-1}^i(v) = 0$ for $v \leqq 2^{-t+1}$ by the induction hypothesis and $U^i(1, A) = 0$ by (A3),

$W^E(z(d), q) = W^F(z(h), q) = 0$. Therefore, if $p \leqq 2^{-t}$, then $\bar{V}_t^i(p) = 0$. Since $\bar{V}_{t-1}^i(p) > 0$ for $p > 2^{-t+1}$, $W^E(z(d), q) - W^E(z(h), q) > 0$ when $q = (1-2p)/[2(1-p)]$. It follows that if $p > 2^{-t+1}$, $q > (1-2p)/[2(1-p)]$ and therefore $r > \frac{1}{2}$ so that $\bar{V}_t^i(p) > 0$ for $i = E$, F, and R. To show that $\bar{V}_t^i(p)$ is strictly increasing when $p > 2^{-t}$, I first show that equilibrium values of r and v increase in p. To see this, observe that if E is indifferent between being honest and being dishonest for a given value of p, an increase in p, keeping $z(d)$ and $z(h)$ constant, causes E to tell the truth with probability one, because $\bar{V}_{t-1}^E(\cdot)$ is increasing by the induction hypothesis; to prevent this, $z(d)$ must increase (and $z(h)$ must decrease), which makes lying more attractive. Hence, to satisfy (A6) and (A7) v increases. That $\bar{V}_t^i(p)$ is strictly increasing in p for $i = E$ and F and $p > 2^{-t}$ now follows directly from the definition of $\bar{V}_t^i(\cdot)$. Finally,

$$V_t^R(p, A) = r\bar{V}_{t-1}^R(v) + rU^R(z,(h), A) + (1-r)U^R(z(h), A)$$

so that $\bar{V}_t^R(\cdot)$ is increasing in r and v. A similar argument shows that if t increases, then so do r and v, and therefore $\bar{V}_r^i(p)$ increases in t for $i = E$, F, and R.

It remains to establish the properties of the function $H_t(p)$. Let $H_t(p)$ be defined to satisfy

$$U^E(2, A) + \bar{V}_{t-1}^E(p) - U^E(0, A) = 0 \tag{A8}$$

or $H_t(p) = \underline{A}$ if the right-hand side of (A8) is not equal to zero for all A. By assumption, $U^E(2, A) - U^E(0, A)$ is strictly decreasing in A, so $H_t(p)$ is well defined. Also, the monotonicity of $U^E(2, A) - U^E(0, A)$ in A, (A8), and the induction hypothesis guarantee that $H_t(p)$ is increasing in t and p. It follows from (A6) that $q = 1$ whenever $A < H_t(p)$. ||

Proof of Corollary 2. The proof is by induction. If the hypotheses of the corollary are true for $s = t-1, \ldots, 0$, then the unique equilibrium is characterized in Theorem 2. Assume that the corollary is true for $s < t$ and, in order to reach a contradiction, assume that for some A, F induces two actions, w^1 and w^2, when $M = 1$ in period t. It follows that, if w^i leads to reputation v^i,

$$U^R(w^1 - 1, A) + \bar{V}_{t-1}^F(v^1) = U^R(w^2 - 1, A) + \bar{V}_{t-1}^F(v^2).$$

If $w^1 > w^2$, then $U^R(w^1 - 1, A) > U^R(w^2 - 1, A)$, and so $\bar{V}_{t-1}^F(v^1) < \bar{V}_{t-1}^F(v^2)$. The induction hypothesis then implies that $v^1 < v^2$ and $\bar{V}_{t-1}^E(v^1) < \bar{V}_{t-1}^E(v^2)$, but then

$$U^E(w^2 - 1, A) + \bar{V}_{t-1}^E(v^2) > U^E(w^1 - 1, A) + \bar{V}_{t-1}^E(v^1),$$

and so E induces w^1 with probability zero. It follows that R can interpret the signal that induces w^1 as being from F whenever $M = 1$, so $v^1 = 1$. This contradicts $v^1 < v^2$. Hence, F induces only one action (sends essentially one signal) when $M = 1$. Similarly, it can be shown that F sends essentially one signal when $M = 0$. If these signals induce different actions in equilibrium, then F is honest; if not, then no information is being transmitted. That all players prefer the honest equilibrium to no transmission follows because the no-transmission equilibrium yields

$$V_{t-1}^i(p, A) = U^i(y^R(\tfrac{1}{2}) - 1, A) + \bar{V}_{t-1}^i(p)$$

since $U^i(y^R(\frac{1}{2}) - 1, A) \equiv 0$, while the honest equilibrium yields $V_t^i(p, A)$ and $V_s^i(\cdot)$ is increasing in s by Theorem 2. ||

First version received August 1983; *final version accepted February* 1985 (*Eds.*).

ACKNOWLEDGEMENT

I am grateful to Sushil Bikhchandani, Beth Hayes, Meg Meyer, Bob Wilson, and referees, who carefully read an earlier version of this paper, spotted errors, and suggested improvements. If the paper has improved, then they deserve the credit. Conversations with Vince Crawford, Luis Guasch, and Barry Nalebuff (who suggested the title of Section 2), as well as the comments of participants at an SSRC Workshop in Economic Theory, a conference on Incentives in Economics and Game Theory, the 1983 Summer Meetings of the Econometric Society, and seminar presentations at UC Irvine and UC San Diego were also of value. I wrote the first version of this paper while I was a Research Fellow at Nuffield College, Oxford. I am grateful to the SSRC for providing the funding through the project "Incentives, Consumer Uncertainty, and Public Policy" and to members of the College for their hospitality. I am also happy to acknowledge funding through NSF grant SES-8204038.

NOTES

1. Condition E1 assumes that N, the set of available signals, is countable. In fact, all of the analyses that follows holds for more general signal spaces; however, only a finite number of distinct signals is needed in order to characterize equilibrium.

2. This type of conditioning is necessary (essentially to make payoffs continuous in actions) when utility functions are linear. There is an example of this in Kreps and Wilson (1982*b*).

3. Thus, E's preferences are not completely opposed to R's preferences. However, the conflict when $M=1$ is enough to derive the results.

4. This specification would require that $U^F(M, D)=M$ instead of $U^F(M, D)=-M$. However, in this case, it will still be of value for F not to default in order to be able to borrow in the future.

REFERENCES

CRAWFORD, V. P. and SOBEL, J. (1982), "Strategic Information Transmission", *Econometrica*, **50**, 1431-1451.

GREEN, J. and STOKEY, N. (1981*a*), "A Two-Person Game of Information Transmission" (H.I.E.R. Discussion Paper 751, Harvard University).

GREEN, J. and STOKEY, N. (1981*b*), "The Value of Information in the Delegation Problem" (H.I.E.R. Discussion Paper 776, Harvard University).

HOLMSTRÖM, B. (1980), "On the Theory of Delegation" (mimeo, Northwestern University).

KREPS, D. M., MILGROM, P., ROBERTS, J. and WILSON, R. (1982), "Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma", *Journal of Econimic Theory*, **27**, 245-252.

KREPS, D. M. and WILSON, R. (1982*a*), "Sequential Equilibria", *Economica*, **50**, 863-894.

KREPS, D. M. and WILSON, R. (1982*b*), "Reputation and Imperfect Information", *Journal of Econometric Theory*, **27**, 253-279.

MILGROM, P. and ROBERTS, J. (1982), "*Predation, Reputation, and Entry Deterrence*", *Journal of Economic Theory*, **27**, 280-312.

SOBEL, J. (1983), "A Theory of Credibility" (U.C.S.D. Discussion Paper 83-21, University of California, San Diego).

WILSON, R. (1985), "Reputations in Games and Markets", in Roth, A. (ed.) *Game Theoretic Models of Bargaining* (Cambridge: Cambridge University Press) forthcoming.