Communication Equilibrium in Games

Joseph Farrell

Number 382

June 1985
Communication Equilibrium in Games

Joseph Farrell

Number 382

June 1985
Communication Equilibrium in Games

Joseph Farrell

GTE Laboratories and MIT; June 1985
1. INTRODUCTION

In a simultaneous-move game of incomplete information, the standard solution concept is Bayesian equilibrium. This is the natural generalization (to incomplete-information games) of the notion of Nash equilibrium: the modification is that, instead of choosing a probability distribution over actions, each player chooses such a probability distribution for each of his possible "types." The different players may be thought of as making their choices in separate rooms: there is no communication among them.

Myerson (1983) has argued for a different solution concept, allowing for the possibility of centralized communication. In a correlated equilibrium, each player secretly reveals his private information to a disinterested trustworthy mediator who is committed in advance to a known behavioral rule. The mediator collates this information, and sends each player secretly a recommendation for his move. The requirement for equilibrium is that no player be tempted to lie and/or to cheat, given that he believes that all others report honestly and act obediently.

Correlated equilibrium allows for communication among players, but in a somewhat artificial way. Often there is no disinterested mediator available. Agents about to make strategic choices often talk directly to one another. Accordingly, it is important to ask whether communication without a mediator is the same as communication with one.

In this paper, we introduce the concept of a communication equilibrium: an outcome that is an equilibrium when players can talk to one another in a specified way. We show that every Bayesian equilibrium is a communication equilibrium, but not vice versa, and that every communication equilibrium is a correlated equilibrium, but not vice versa. Thus on the one hand costless direct communication can count; on the other hand, a mediator can have positive value.
We then ask what kinds of information can be revealed by unmediated communication. In the standard case of independently distributed types, we show that there are invariants that are always preserved under unmediated communication, but need not be preserved when a mediator is available. This represents the fact that, without a mediator, it is impossible for each of two players to make his revelation contingent on what the other says.

Possibly because of game theory's beginnings in the two-person zero-sum case (von Neumann and Morgenstern, 1944) and the excessive emphasis given to the Prisoners' Dilemma since the subject turned to the more general case, communication has been badly neglected in game theory. Many theorists informally think as if all strategic relationships were essentially adversarial, and therefore (absent differential costs of sending messages, as in Spence, 1970, etc.) communication can be ignored. It is also something of a surprise to realize how important payoff-irrelevant moves may be for the outcome of a game.

Roughly, payoff-irrelevant moves matter to the extent that there are opportunities for coordination in the game. There are no such opportunities in two-person zero-sum games, but in general each player may be prepared in effect to delegate some control over his move to the other player (the term "his opponent" is misleading). This can be accomplished through message-sending even if messages do not directly affect payoffs.
2. DEFINITIONS

We consider a game $G$ which can be described as follows: Players are indexed by 1, 2, ..., $n$. Player $i$ observes the value of a random variable $t_i$ which is distributed on the finite set $T_i$ according to the probability distribution $\pi_i$; we assume that $\pi_i(t) > 0$ for all $t \in T_i$. Player $i$ also chooses a move $x_i \in X_i$. We write $\underline{t} = (t_1, ..., t_n)$ and $\underline{x} = (x_1, ..., x_n)$. Player $i$'s payoff is given by $u^i(\underline{x}, \underline{t})$. The priors $\pi_i(\cdot)$ and the payoff functions $u^i(\cdot, \cdot)$ are common knowledge.

We write $T = T_1 x ... x T_n$ and $X = X_1 x ... x X_n$. We write $\Delta(A)$ for the set of probability distributions on any set $A$. An outcome of the game is a function $w : T \rightarrow \Delta(X)$.

An outcome $w : T \rightarrow \Delta(X)$ is a Bayesian equilibrium outcome if there exist strategies, functions $s_i : T_i \rightarrow \Delta(X_i)$, such that

1. The strategies $s_i$ constitute a Bayesian equilibrium:

2. The strategies $s_i$ lead to the outcome $w$; that is, for each $t \in T$, the weight assigned by $w(t)$ to the point $x \in X$ is equal to

   \[ \frac{\prod_{i=1}^{n} s_i(t_i)(x_i)}{\sum_{x \in X} \prod_{i=1}^{n} s_i(t_i)(x_i)} \]

   It is a standard result that a Bayesian equilibrium will exist in a finite game such as ours.

An outcome $w : T \rightarrow \Delta(X)$ is a correlated equilibrium outcome (see e.g., Myerson, 1983) if, when a trusted mediator asks for reports of $t$, performs any necessary randomization, and then instructs players secretly as to the value of $x_i$ they should choose, there are no incentives for a player to misreport his private information or to disobey his instructions, assuming that other players are reporting truthfully and acting obediently.
We say that a game $G'$ is a communication version of $G$ if $G'$ consists of three phases:

1. Player $i$ learns $t_i \in T_i$ (just as in $G$)

2. There are opportunities for players to announce messages, i.e., to speak. (We assume that all players hear all announcements; this is unrestricted if $n = 2$.) These messages do not directly enter payoff functions, but may convey information and influence others' final choices.

3. The players simultaneously choose $x_i \in X_i$ (just as in $G$).

We emphasize that payoffs are given by $u^i(x, t)$ just as in $G$: given $t$ and $x$, there is no need to know what messages were announced in order to calculate payoffs. However, as we will see below, sometimes the messages will affect $x$.

Every outcome $w'$ of $G'$ determines an outcome $w$ of $G$ in a natural way (simply forget the information in $w'$ concerning the messages announced). If $w'$ is the outcome corresponding to some perfect Bayesian equilibrium in $G'$, then we say that $w$ is a communication equilibrium outcome of $G$.

For instance, if we make the communication opportunities vacuous, we see that any Bayesian equilibrium outcome $w$ of $G$ is also a communication equilibrium outcome. We will show below that there can be others.
3. RELATIONSHIPS OF THE THREE EQUILIBRIUM CONCEPTS.

Theorem 1:

1. Every Bayesian equilibrium outcome is a communication equilibrium outcome, but not vice versa.

2. Every communication equilibrium outcome is a correlated equilibrium outcome, but not vice versa.

Proof:

1. First, we can always set $G' = G$. Hence every Bayesian equilibrium outcome is a communication equilibrium outcome. To see that the converse fails, consider Example 1:

Example 1: This is a symmetric game. For each player, $T = \{a, b\}$, where $0 < a < 1 < b$. The probability that $t_1 = a$ is $p$, and $t_1$ and $t_2$ are independent. For each player, $X = \{y, z\}$. Payoffs are given by:

```
Player 2
     z
    y    |
     (1 - t_1, t - t_2) (-t_1, 0)
Player 1
      z
     y    |
         (0, -t_2) (0, 0)
```

Lemma 1: If $p < a$, then the unique rationalizable (hence, the unique Bayesian equilibrium) outcome of Example 1 is for $(z, z)$ to be chosen for all $t$.

Proof: Let $q$ denote the probability that player 2 will play $y$, and consider player 1's choice. If $t_1 = b$, then $z$ dominates $y$. If $t_1 = a$, then $y$ is the better move if
\[ q(1 - a) + (1 - q)(-a) > 0 \]

i.e., if \( q > a \). However, since \( z \) dominates \( y \) for player 2 whenever \( t_2 = b \), we know \( q \leq p \). Hence if \( p < a \), then player 1 cannot rationally play \( y \). Likewise, nor can player 2. This proves Lemma 1.

(Note: This is an adaptation of a more striking example with infinite \( T \), described in Farrell, 1983.)

**Lemma 2**: The outcome which assigns probability 1 to \((y, y)\) when \( t = (a, a) \), and assigns probability 1 to \((z, z)\) for other values of \( t \), is a communication equilibrium outcome.

**Proof**: Define \( G' \) as follows. In Stage I, each player simultaneously announces "A" or "B". Stage II is \( G \). Consider the following strategy:

If \( t = b \), announce "B", and then play \( z \) whatever the other player announces. If \( t = a \), announce "A"; then play \( y \) if the other player announced "A", and play \( z \) if he announced "B".

It is a perfect Bayesian equilibrium in \( G' \) for both players to use this strategy. If they do so, the outcome is as described above. This proves Lemma 2.

Lemma 2 displays an example of a communication equilibrium outcome that is not a Bayesian equilibrium outcome (in \( G \)); thus the proof of part (1) of Theorem 1 is complete. Costless communication can make a difference.

Next, we prove part (2) of Theorem 1. Again, the positive statement is easy to prove: whatever the process in \( G' \) that yields an outcome \( w \), a mediator can commit himself to imitating that process internally. The fact that \( w \) was supported as an equilibrium in \( G' \) implies that it will be an equilibrium for players to report truthfully and to obey instructions.
To see that not every correlated equilibrium outcome is a communication equilibrium outcome, consider Example 2.

Example 2. This too is a symmetric two-player game, with $T_i = \{c, d\}$ and $X_i = \{z, w\}$. To represent the payoffs, we use four bi-matrices, one for each possible $t \in T = \{c, d\} \times \{c, d\}$.

\[ t = (c, c): \]

\[
\begin{array}{c|cc}
\text{z} & \text{w} \\
\hline
\text{z} & (1, 1) & (0, 0) \\
\text{w} & (0, 0) & (0, 0) \\
\end{array}
\]

\[ t = (c, d): \]

\[
\begin{array}{c|cc}
\text{z} & \text{w} \\
\hline
\text{z} & (0, 0) & (-2, 0) \\
\text{w} & (1, -2) & (-2, -3) \\
\end{array}
\]
\[ t = (d, c): \]

\[
\begin{array}{c|cc}
 & z & w \\
\hline
z & (0, 0) & (-2, 1) \\
\hline
w & (0, -2) & (-3, -2) \\
\end{array}
\]

\[ t = (d, d): \]

\[
\begin{array}{c|cc}
 & z & w \\
\hline
z & (0, 0) & (0, 0) \\
\hline
w & (0, 0) & (1, 1) \\
\end{array}
\]

The types are independently distributed, with \( 1 > \pi_1(c) > 1/2 \).

**Lemma 3:** In Example 2, the following outcome is a correlated equilibrium outcome:

- if \( t = (d, d) \), then \( x = (w, w) \);
- Otherwise, \( x = (z, z) \).

**Proof of Lemma 3:** Suppose a mediator announces that he will instruct each player to play \( z \), unless both players report their types as \( d \), in which case he will instruct both to play \( w \). If this induces honest and obedient behavior, then we have shown that the outcome given is a correlated equilibrium.

Consider player 1's incentives, assuming that player 2 is honest and obedient. Suppose first that \( t_1 = c \). If player 1 reports truthfully, he will be asked to play \( z \) whatever player 2's report, and he knows that player 2 will play \( z \). Would he then prefer to play \( w \)? Since \( \pi_2(c) > 1/2 \), he would not. Thus, if 1 reports truthfully when \( t_1 = c \), he will then obey the mediator's instruction. His expected payoff is \( \pi_2(c) \).

Would player 1 benefit from lying when \( t_1 = c \)? If he falsely reports that \( t_1 = d \), then what happens depends on \( t_2 \):
With probability $\pi_2(c)$, player 2 truthfully reports $t_2 = c$. The mediator then asks each player to play $z$. Player 1 now knows that $t_2 = c$ and that $x_2 = z$. Accordingly, consulting the first of our payoff tables, player 1 sees that he should play $z$, and gets payoff 1.

With probability $1 - \pi_2(c)$, player 2 truthfully reports $t_2 = d$. The mediator then asks each player to play $w$. Player 1 now knows that $t_2 = d$ and $x_2 = w$. Consulting the second payoff table, second column, 1 sees that whatever he does he gets payoff (-2).

Hence the expected payoff to a player of type $c$ who lies is $\pi_2(c) + (1 - \pi_2(c))(-2) < \pi_2(c)$. A player of type $c$ will not choose to lie.

Next, consider a player 1 of type $d$. If he tells the truth, one of two things may happen.

With probability $\pi_2(c)$, player 2 truthfully reports $t_2 = c$. Then the mediator asks both players to play $z$. Player 1 thus consults the first column of the third payoff table, and sees that he might as well obey. This gives him payoff 0.

With probability $[1 - \pi_2(c)]$, 2 reports $t_2 = d$. The mediator asks both players to play $w$. Player 1 now knows $t_2 = d$, $x_2 = w$. Consulting the fourth table, he sees he should obey, and gets payoff 1.

Thus if $t_1 = d$, player 1 can get $1 - \pi_2(c)$ by telling the truth; and if he tells the truth he will obey instructions.

If $t_1 = d$ and player 1 lies, he will not learn $t_2$, and will know that $x_2 = z$. Whatever $t_2$ may be, $x_1 = z$ is at least as good a move as $x_1 = w$; and it gives 1 a payoff of 0. Since $0 < 1 - \pi_2(c)$, lying is unprofitable for 1 if $t_1 = d$. This concludes the proof of Lemma 3.
We complete our proof that a correlated equilibrium mechanism need not be a communication equilibrium mechanism with Lemma 4:

**Lemma 4:** The outcome described in Lemma 3 is not a communication equilibrium outcome.

**Proof of Lemma 4:** Lemma 4 is a corollary of Theorem 5 below. However, that apparatus is required only in order to allow for mixed Bayesian strategies in $G'$. It is very straightforward to show that neither player in Example 2 is willing to reveal his type before knowing the other's.

To see this, notice that, if player 2 knows or believes that $t_1 = d$, then he will set $x_2 = w$. This is desirable for player 1 if $t = (d, d)$, but undesirable otherwise. If $t_1 = d$, revealing the fact would give 1 a payoff of $\pi_2(c)(-2) + (1 - \pi_2(c))(1) = 1 - 3\pi_2(c)$, while concealing it gives 1 a payoff of 0 (since 2 will also not reveal $t_2$, and then each player will want to choose $x_2 = z$). Since $1 - 3\pi_2(c) < 0$, player 1 will conceal.

Lemmas 3 and 4 together complete the proof of Theorem 1. Theorem 1 shows that decentralized costless communication can matter, and that it cannot necessarily be properly modeled by assuming a mediator. We next examine a benchmark case in which communication cannot matter: if there is so much conflict that players can never trust one another.

**Theorem 2:** In a two-person zero-sum game, every correlated equilibrium outcome gives the same payoffs in each state $t$ as Bayesian equilibrium.

**Remark:** Another way to say this is that only "inessential" communication that does not affect payoffs is possible in equilibrium in such a game. Intuitively, the reason is clear. By choosing a message, each player (say, 1) may affect the action rule $x_2: T_2 \rightarrow \Delta(X_2)$ used by the other. Incentive-compatibility in reporting for player 1 means that he prefers the rule associated with a report $t_1$ when his true type is $t_1$. But since the game is zero-sum, that means that, if 2 is asked to follow one of the rules used in equilibrium,
he would do better to follow any other. This proves Theorem 2 for the case where the rule can be inferred from the mediator's message. In order to allow for the more general case, we give the following proof.

Proof of Theorem 2: Although in general a player in a zero-sum game whose opponent pays attention to his claims can profit by lying, it is difficult to write down a successful lying strategy for the general case, so we consider instead a deviation from a proposed correlated-equilibrium mechanism to randomized uninformative messages.

Consider a correlated-equilibrium, in which each player sends messages to the mediator revealing his private information, and then the mediator suggests appropriate moves. Suppose that player 1 begins randomizing his messages so that they convey no information about his type, and playing a Bayesian equilibrium strategy independent of what the mediator suggests for him. If player 2 knew this was happening, his best response would be to play his Bayesian strategy, and both would get Bayesian equilibrium payoffs. If player 2 did not know about 1's deviation, on the other hand, then player 2 would not use his best response, and would therefore do no better than get his Bayesian equilibrium payoffs. Hence player 1 would do at least as well as his Bayesian payoffs.

Thus, in any correlated equilibrium, each type of player 1 must get at least his Bayesian equilibrium expected payoff. This implies that player 2 can do no better than his. The same argument, reversing players 1 and 2, gives the bounds the other way, thus proving Theorem 2.

Remark: We cannot quite say that communication will not occur. For example, consider the game which gives payoffs of zero to both players whatever happens. Many correlated equilibria will exist in which the players communicate. Theorem 2 expresses a sense in which such uninteresting communication is the only kind that can occur in equilibrium in a two-person zero-sum game.
Corollary to Theorem 2: Every communication equilibrium outcome gives the same payoffs (to each type) as Bayesian equilibrium.

Discussion: The proof of Theorem 2 suggest the following heuristic analysis of when communication will be important. When there is essential communication in equilibrium, each player is allowing the other to choose his move to some extent, since it is common knowledge that one message will induce one action rule, another another. Such delegation is wise precisely when payoff functions are sufficiently "positively correlated." A game of pure coordination ($u^1 = u^2$) is the polar case, opposite to that of the zero-sum game ($u^1 = -u^2$). A general game has some aspects of each type. Often they are inextricably mixed; but in some games they are separable. An example is the problem of a group of risk-lovers wishing to meet to play poker. If they coordinate successfully on the meeting-place, they will play an enjoyable game of pure conflict; otherwise they will drink alone. We expect them to communicate honestly concerning where to meet, but not concerning their playing strategies, knowledge of odds, etc.
4. ANNOUNCEABLE INFORMATION

We now turn to a limitation of decentralized communication that has nothing to do with incentives to tell the truth. As in Example 2 above, it is not possible both for 1's revelation of $t_1$ to depend on the value of $t_2$, and for 2's revelation of $t_2$ to depend on the value of $t_1$. Such a mutual dependence is possible with a mediator, as in Example 2. In this section, we define the concepts of public information structure and announceable information structure. We show that a mediator can (if information is verifiable ex-post) generate any public information structure having the right "mean"; but that in the absence of a mediator various other constraints must be satisfied. We apply this result (Theorem 5) to complete the proof of Theorem 1 and to discuss bargaining with two-sided uncertainty.

In this section, we are concerned with constraints on what it is logically possible to reveal. Accordingly, we suppress for now the role of incentives, by assuming that private information is verifiable ex-post (and adequate punishment is available for liars). If private actions cannot be controlled (e.g., if there is a lot of noise in the observation of actions), it may still be desirable to have some, but less than complete, revelation of private information. Just what revelations can be achieved without a mediator is the subject of this section.

We assume (with some loss of generality if $n > 2$) that every message spoken is publicly heard by all. Therefore, at each stage of a communication equilibrium, each player's information consists of his private information together with the information inherent in what has been announced (which may, of course, reduplicate some or all of the player's private information). This latter information we call the public information at that stage. As communication happens, the public information changes (in a way that depends on the true state $t$) until, just before moves in $G$ are actually selected, each player sees the final public information together with his own private information.
We can represent the final public information as a posterior probability distribution $p$ on $T$. Each player $i$'s information is then the conditional distribution generated by restricting $p$ to $T_1 \times \ldots \times T_{i-1} \times \{t_i\} \times T_{i+1} \times \ldots \times T_n$.

A **public information structure** is a function $\sigma: T \rightarrow \Delta(\Delta(T))$, where $\sigma(t)(p)$ describes how likely the posterior $p$ on $T$ is to become the final public information, given that $t \in T$ is the true state. We will suppose for simplicity that each $\sigma(t)$ has finite support in $\Delta(T)$, so that $\sigma$ maps $T$ into $\Delta_f(\Delta(T))$, where $\Delta_f(\ast)$ represents the set of distributions with finite support.

Write $\{p_1, \ldots, p_N\}$ for the union of the supports of $\sigma(t)$ as $t$ varies over $T$. That is, $p_1, \ldots, p_N$ are the distributions on $T$ which can arise under $\sigma$ as final public information. Thus, for each $i$, $\sigma(t)(p_i) > 0$ for some $t \in T$, while if $p \in \Delta(T)$ is not one of the $p_i$ then $\sigma(t)(p) = 0$ for all $t \in T$.

The **mean** of $\sigma$ is the distribution $\nu(\sigma)$ on $T$ given by

$$\nu(\sigma) = \sum_{t \in T} \sum_{i=1}^{N} \frac{\sigma(t)(p_i)}{\Delta(T)} p\, dp$$

(4.1)

$$= \sum_{t \in T} \sum_{i=1}^{N} \pi(t, c(t)(p_i)) p_i$$

(4.2)

so that, for any $t' \in T$,

$$\nu(\sigma)(t') = \sum_{t \in T} \sum_{i=1}^{N} \pi(t, c(t)(p_i)) p_i(t').$$

(4.3)

**Theorem 3:** Every public information structure has mean $\pi$.

**Proof:** Let $m_1, \ldots, m_M$ be all possible message histories. Suppose that state $t$ leads to message-history $m_i$ with probability $p(m_i | t)$. Then, after $m_i$, the final public information is $q(\ast | m_i)$, where by Bayes' rule

$$q(\ast | m_i) \in \sum_{t \in T} p(m_i | t) \frac{\pi(t)}{p(t)} = p(m_i | t) \pi(t).$$

(4.4)
Now in our previous notation $p_1, \ldots, p_N$ are the distributions $q_i(\ast|m_i)(i=1, \ldots, M)$, while $\sigma(t)(q_i(\ast|m_i)) = p(m_i|t)$. Hence translating the definition of $\mu(\sigma)$ into our present notation gives us

$$\mu(\sigma)(t) = \sum_{\tau \in T} \sum_{i=1}^M \pi(\tau) p(m_i|\tau) q(t|m_i)$$

(4.5)

$$= \sum_{i=1}^M q(t|m_i) \sum_{\tau \in T} p(m_i|\tau) \pi(\tau) \quad \text{rearranging} \quad (4.6)$$

$$= \sum_{i=1}^M p(m_i|t) \pi(t) \quad \text{by (4.4)} \quad (4.7)$$

$$= \pi(t) \text{ since } \sum p(m_i|t) = 1.$$

This proves Theorem 3. Next we show that $\mu(\sigma) = \pi$ is also sufficient for centralized announceability.

**Theorem 4:** Let $\sigma: T \rightarrow \Delta_f(\Delta(T))$ be any function.

If $\mu(\sigma) = \pi$, then a mediator who knows $t$ can induce public information structure $\sigma$ by public announcement.

**Proof:** As before, let $p_1, \ldots, p_N$ be the distributions on $T$ given positive weight in some $\sigma(t)$. The function $\sigma$ is thus described by $\sigma(t^i)(p_j), i = 1, 2, \ldots, |T|; j = 1, 2, \ldots, N$. (Here we simply number the elements of $T$ and ignore the Cartesian product structure.) Suppose that our mediator announces message $m_j$ with probability $p_j(m_j|t^i)$ when the state is $t^i$. Then the posterior distribution $q(\ast|m_j)$ on $T$ is given by

$$q(t^i|m_j) = \frac{p(m_j|t^i) \pi(t^i)}{\sum_{\tau \in T} p(m_j|\tau) \pi(\tau)} \quad (4.8)$$
Simple algebra shows that this posterior distribution will be $p_j$ if we set

$$p(m_j | t^i) = \pi(t^i)^{-1} \left[ \sum_{\tau \in T} \pi(\tau) \sigma(\tau)(p_j) \right] \quad (4.9)$$

It remains to check that, with this specification, $\sum_j p(m_j | t^i) = 1$. To see this, we notice that

$$\sum_j \sum_{\tau \in T} \pi(\tau) \sigma(\tau)(p_j) \cdot p_j(t^i) = \pi(t^i) \quad (4.10)$$

and by assumption this is equal to $\pi(t^i)$. This proves Theorem 4.

Thus we have seen that every function $\sigma: T \rightarrow \Delta_f(\Delta(T))$ with mean $\pi$ can be realized through (randomized) central announcements. This means that a mediator can publicly reveal exactly as much or as little as is desirable, subject only to the condition on the mean. Without a mediator, controlled revelation is harder:

**Theorem 5:** Suppose that $t_i \in T_i$ is independent of $t_j \in T_j$, $i \neq j$. Let $t'_i$, $t''_i$ be elements of $T_i$, $i = 1, ..., n$. For each $i$, let $t'_i$, $t''_i$ be $t'_i$, $t''_i$ but not necessarily respectively. Then if the distribution $p(t)$ can occur as public information in a decentralized revelation system, we have

$$p(t') p(t'') = p(t') p(t''). \quad (4.11)$$

**Proof of Theorem 5:** First, note that (4.11) is true of the prior distribution $p = \pi$ because $t_i$ is independent of $t_j$, $i \neq j$, so that each side of (4.11) is (in the obvious notation) $\pi(t'_1) \pi(t''_1) \pi(t'_2) \pi(t''_2) \cdots \pi(t'_{n})$. Next, we show that (4.11) is preserved by any revelation by a player. Suppose 1 makes a revelation, beginning at a point $p$ of public information satisfying (4.11). Let $p'$ be a possible public posterior after 1's revelation.

Suppose $t'_1 = t''_1, t'_1 = t''_1$. Then since $t'$ and $t''$ are indistinguishable to player 1, no new information on the likelihood ratio between them can emerge from 1's revelation:
\[
\frac{p'(t')}{p'(\tau')} = \frac{p(t')}{p(\tau')}
\]

Likewise, since 1 cannot distinguish between \( t'' \) and \( \tau'' \),
\[
\frac{p'(t'')}{p'(\tau'')} = \frac{p(t'')}{p(\tau'')}
\]

Since (4.11) can be rewritten as
\[
\frac{p(t')}{p(\tau')} = \left[\frac{p(t'')}{p(\tau'')}\right]^{-1}
\]

equations (4.12) and (4.13) assure us that, if (4.11) holds for the public information \( p \), then it also holds for \( p' \). This concludes the proof of Theorem 5.
5. APPLICATION 1: PROOF OF THEOREM 1.

A gap in the proof of Theorem 1 given above was that we did not allow for mixed revelation strategies, i.e., for messages which stochastically convey information. Conceivably it might be possible (e.g.) for 1 to hint very slightly that \( t_1 = x \); then 2 might respond cautiously, and they could proceed with mutual circumspection until they had mutually revealed the appropriate information. Theorem 5 ensures that this cannot happen. For the required information is "whether or not \( t = (x, x) \)." Thus the revelation system would have to have, as one of its final public information states, at least one point with \( p(x, x) = 0 \), and \( p(x, y) > 0 \), \( p(y, x) > 0 \). Hence we would have

\[
p(x, x) p(y, y) \neq p(x, y) p(y, x)
\]

and Theorem 5 guarantees that this cannot be.
6. APPLICATION 2: BARGAINING:

Consider two players bargaining over an object. Each player's value for the object is private information, and the values are independent. The events \( E = \{ \text{1's value exceeds 2's} \} \) and \( F = \{ \text{2's value exceeds 1's} \} \) are both possible. If player 2 owns the object, it is natural for the players to want to know whether or not the event \( E \) has occurred. We ask whether this information could conceivably be revealed by communication between them.

If we write \( u \) for 1's value and \( v \) for 2's value, and each can take on either of two values \((u_1, u_2; v_1, v_2)\), and if

\[
u_1 > v_1 > u_2 > v_2\]  

(4.16)

then we need to reveal whether or not the event \( F = (u_2, v_1) \) has occurred. Of course, full revelation would tell us whether or not \( F \) has occurred, but it might be necessary for subsequent incentive reasons not to reveal (in the course of discovering whether there are gains from trade) just what the values are. Hence we want a system that reveals whether \( E \) or \( F \) has occurred, but without leaking further information if \( E \) has occurred.

With a mediator, such a system is simple. The mediator asks each player for his type, and then announces "E" or "F". If everyone expects that the players will bargain if "E" is announced, and will walk away if "F" is, then truth-telling is an equilibrium.

Without a mediator, the partial revelation cannot be achieved. To see this, apply Theorem 5 to get, for any public information system, \( p \)

\[
p(u_1, v_1) p(u_2, v_2) = p(u_1, v_2) p(u_2, v_1) \]  

(4.17)

Equation (4.17) cannot be satisfied with \( p(u_1, v_1) \), \( p(u_2, v_2) \) and \( p(u_1, v_2) \) strictly positive but \( p(u_2, v_1) \) zero. Hence any system of public communication that establishes whether or not there are gains from trade must also leak
further information. It is not possible, without a mediator, for the players simply to find out whether or not there are gains from trade.
7. CONCLUSIONS

Costless communication in games matters. Outcomes become equilibria which would not be equilibria without communication. This does not happen in two-person zero-sum games, but in general it can be important. Since players very often can talk to one another, our equilibrium notion should reflect the fact.

Mediators matter. Players may be prepared to reveal important information to a trusted mediator (who will pass it on only if it will not be exploited against the player), when they would keep the information secret in the absence of a mediator.

There has been very little work on communication in games. Characterizing the set of communication equilibria may be no easier than characterizing the set of Bayesian equilibria. Theorem 5 gives necessary conditions, but they are far from sufficient. The interaction of incentives with the logical restrictions analyzed in Theorem 5 will be complicated.

An algorithm to determine when a game has enough conflict (relative to its coordination aspects) that communication cannot happen in equilibrium would also be very desirable. Theorem 2 is a first step in this direction. It would be exciting to know how to decompose a general game into "coordination aspects" and "conflict aspects," as hinted on page 12. However, I have made no progress on this.

Some communication structures are more plausible than others. For one direction of work on this, see Farrell (1985). But there is no good reason to focus on the trivial communication structure in which all payoff-irrelevant messages are ignored. That is just one among many, and often not a plausible one. In my opinion, the literature should begin to take more account of communication, which is one of the salient features of the human world.
8. REFERENCES


