Discussion Paper No. 992

On Modeling Cheap Talk in Bayesian Games*

by

STEVEN A. MATTHEWS
Northwestern University

and

ANDREW POSTLEWAITE
University of Pennsylvania

August 25, 1992

* This paper has been prepared for the Stanley Reiter Festschrift, *The Economics of Informational Decentralization: Complexity, Efficiency, and Stability*, ed. J. Ledyard, Kluwer Academic Publishers, Amsterdam. We each thank George Mailath for valuable comments, and the National Science Foundation for support.

The "cheap talk games" studied in this chapter are games in which players with private information exchange payoff-irrelevant messages. Crawford and Sobel (1982) and Green and Stokey (1980) introduced the simplest cheap talk games, those in which a "sender" with private information sends a message to a "receiver" who then takes an action. Equilibrium refinement criteria for sender-receiver games have been studied extensively,[1] and variations of sender-receiver games have been used to model a variety of phenomena.[2] Little work, however, has been done on more general cheap talk games. A few studies have considered multiple informed parties, but with only one round of pre-play communication.[3] Even fewer studies have considered games with multiple rounds of communication, and then with only one informed party.[4]

We examine issues that arise in the modeling of face-to-face conversation among multiple informed parties. The modeling approach we follow represents a conversation as a finite, or even infinite, sequence of rounds of public message exchange that occurs before some underlying game is played. This is a straightforward extension of the predominant approach taken in the literature cited above. However, new is and difficulties arise when this approach is extended to settings with multiple communication rounds and multiple informed parties. Using two examples, we show that the number of communication rounds can interact consequentially and arguably artificially with the number of informed players. The among of information revealed in plausible equilibria will be shown to depend in unexpected ways on the number of communication rounds.

The basic framework, consisting of an underlying Bayesian game to which a finite

---

[1] Farrell (1985, forthcoming) first formulated a refinement criterion for sender-receiver games. Subsequent criteria are developed in, e.g., Blume (1992a, 1992b), Blume and Sobel (1991), Matthews *et. al.* (1991), Myerson (1989), and Rabin (1990).

[2] E.g., Forges (1990), Matthews (1989), Stein (1990).

[3] E.g., Austen-Smith (1990), Farrell and Gibbons (1989), Matthews and Postlewaite (1989), Palfrey and Srivastava (1991).

[4] Forges (1990) and Blume and Sobel (1991).

number of communication rounds is appended, is described in Section 1. Also introduced there is a preliminary refinement criterion, one which extends the "strong announcement-proof" criterion developed in Matthews *et. al.* (1991) for sender-receiver games. The criterion is illustrated in Section 2 by our first example. In this example, an equilibrium in which information is effectively revealed does not even exist if there is only one round of message exchange. But if there is more than one round, the unique announcement-proof equilibrium outcome entails consequential information revelation.

In order to eliminate possible end-round effects, a formulation of games with an infinite number of communication rounds is presented in Section 3. Desirable properties for an "announcement-proof" refinement criterion for such games are discussed there, although a complete definition is not given and must be deferred. Section 4 contains an example to illustrate end-round effects, and the consequent desirability of the infinite-round model. In this example, if the number of communication rounds is finite, all information is revealed in the only announcement-proof outcome. But if the number of rounds is infinite, equilibria in which no information is revealed are also announcement proof — the addition of infinite opportunities to communicate may actually restrict communication.

1. Communication Games

The underlying Bayesian game, to which a communication structure will be appended, has n players. The private information of player i is reflected in his *type*, $t_i \in T_i$, where T_i is a finite set. Types are independently distributed. Player i knows his type, but the others view it as distributed according to a probability distribution $\pi_i \in \Delta(T_i)$. The support of π_i is $S_i(\pi_i) \equiv \{t_i \in T_i \mid \pi_i(t_i) > 0\}$. The belief profile is $\pi \equiv (\pi_1,\ldots,\pi_n)$, with support $S(\pi) \equiv \prod_{i=1,n} S_i(\pi_i)$. The set of actions available to player i is A_i, also assumed finite. The payoff to player i from action profile a and type profile t is $u_i(a, t)$. Actions are taken simultaneously. This defines the underlying game, $G^0(\pi)$.

Our concern is with the effect of communication on the play of this game. We focus on face-to-face communication, which occurs when people talk freely in a group, such as an informal committee meeting, before choosing payoff-relevant actions.

Focusing on face-to-face communication has implications for how the communication stage should be modeled. For example, face-to-face messages are not influenced by correlation devices or filtered through mediators (e.g., Myerson, 1982, 1989). Also, face-to-face messages should be public, i.e., each message one player sends should be received, and commonly known to be received, by all the other players. Accordingly, we model the communication stage as K rounds of public, unmediated message exchange before the underlying game is played. In this section we assume $K < \infty$, postponing a treatment of games with an infinite number of communication rounds until Section 3.

In each round, player i sends a message from a finite set M_i. A *message profile* is an n-tuple of messages, $m \in M \equiv M_1 \times \cdots \times M_n$. A *talking strategy* for player i in round k is a mapping τ_i^k of his type and the sequence of past message profiles to a probability distribution on messages. Thus, $\tau_i^k(m_i \mid t_i, m^1, \ldots, m^{k-1})$ is the probability with which player i sends message $m_i \in M_i$ in round k when his type is t_i and the past message profiles are $m^1, \ldots, m^{k-1}$.

After K rounds of message exchange, actions are taken. An *action strategy* for player i is a mapping α_i from his type and the sequence of message profiles to a probability distribution on his action set. Thus, $\alpha_i(a_i \mid t_i, m^1, \ldots, m^K)$ is the probability that player i takes action $a_i \in A_i$ if his type is t_i and the sent message profiles are $m^1, \ldots, m^K$.

This completes the definition of $G^K(\pi)$, the game with K rounds of message exchange and prior beliefs π.[5]

To discuss sequential equilibrium, belief formation must be made explicit. Accordingly, define a *belief function* for player i in round k to be a mapping from his type and past messages to a set of probability distributions over the types of the other players. The beliefs of player i about the type of player j at the end of round k cannot depend directly on i's type, nor on the messages sent by players other than j in round k; also, all players $i \neq j$ must have the same beliefs about j's type.[6] We write $\beta_j^k(t_j \mid m^1, \ldots, m^{k-1}, m_j^k)$

[5] A special case is a sender-receiver game. Assuming player 1 is the sender and 2 the receiver, it takes the form of a one-round game, $G^1(\pi)$, with A_1 and T_2 both singletons.

[6] These claims follow from the independence of the players' types, the simultaneity of their messages in each round and their actions afterwards, and the "consistent

to denote the probability held by each player $i \neq j$ at the end of round k that the type of player j is t_j, given that the message sent in round k by player j was m_j^k and the message profiles previously sent were $m^1, \ldots, m^{k-1}$. The profile of beliefs is $ß^k = (ß_1^k, \ldots, ß_n^k)$.

A (*sequential*) *equilibrium* of $G^K(\pi)$ is a profile of talking strategies, action strategies, and belief functions which satisfy Bayes consistency and the best response property — we refer the reader to Kreps and Wilson (1982) for a full definition. Let $E^K(\pi)$ be the set of equilibria. Each equilibrium $e \in E^K(\pi)$ gives rise to an (*equilibrium*) *outcome* $o(e)$, which we define to be a mapping from type profiles into probability distributions on actions. Thus, $o(a \mid t, e)$ is the probability according to e that actions a are taken when the types are t. An outcome captures all payoff-relevant consequences of an equilibrium, and it depends only on the equilibrium's talking and action strategies. The payoff to type t_i of player i from equilibrium e is,

$$u_i(e \mid t_i, \pi_{-i}) \equiv \sum_{t_{-i} \in T_{-i}} \sum_{a \in A} \pi_{-i}(t_{-i})\, o(a \mid t, e)\, u_i(a, t).^7 \qquad (1)$$

This definition of an outcome also has the advantage that outcomes of games with different numbers of rounds can be directly compared, as they are all maps from $T \equiv T_1 \times \cdots \times T_n$ to $\Delta(A)$. Still, our attention will not be entirely on outcomes, since they abstract from some noteworthy features of an equilibrium, such as the sequencing and timing of information revelation.

Every outcome of $G^K(\pi)$ is an outcome of $G^{K+1}(\pi)$, i.e., cheap talk expands the set of outcomes.[8] Referring to outcomes obtainable only with cheap talk as "communication outcomes" is problematic, however, since messages can be used merely to achieve correlated equilibria rather than to communicate relevant information about types.[9] We

---

assessment" property of a sequential equilibrium (Kreps and Wilson, 1982, p. 872).

[7] Notice that even if no other player believes that player i might be type t_i, i.e. even if $\pi_i(t_i) = 0$, the equilibrium payoff $u_i(e \mid t_i, \pi_{-i})$ is well defined by (1).

[8] An equilibrium of $G^K(\pi)$ defines an equivalent equilibrium of $G^{K+1}(\pi)$: append to the equilibrium talking strategies of $G^K(\pi)$ a pooling talking strategy for each player in round K+1, and belief functions that ignore all messages sent in round K+1.

[9] For example, suppose players 1 and 2 play the Battle of the Sexes (BOS), and player 3

wish to focus instead on outcomes which use communication to alter players' beliefs about types, and thereby to alter their best-reply correspondences in the underlying game. Thus, since the correlated equilibria of $G^0(\pi)$ which are achievable by public messages must be mixtures of its Nash equilibria, we say that an outcome of $G^K(\pi)$ is a *communication outcome* if it is not a convex combination of outcomes of $G^0(\pi)$.

Some cheap talk outcomes are implausible. This has been noted in sender-receiver games, for which a variety of equilibrium refinement criteria have been proposed (see footnote 1). We discuss and extend one of them here, the announcement-proof criterion of Matthews *et. al.* (1991). However, our feeling is that most of the points we shall make could also be made using extensions of other cheap talk refinement criteria.

The following is a rough description of the announcement-proof criterion, as applied to a sender-receiver game.[10] The criterion is used to test, via an indirect argument, the proposition that a "putative equilibrium" is an accepted way of playing the game. It cannot be an accepted way of playing if there is a surprise "announcement" that the sender can make, instead of sending his equilibrium message, which will induce the receiver to take an action which the sender prefers to the equilibrium outcome. The announcement is a message which has a literal meaning, and that literal meaning is a claim that the sender is one of an identified set of types which are deviating from the putative equilibrium. The announcement is "believed" if after hearing it, the receiver takes the action that is optimal for herself when the sender's type is as claimed in the announcement. The announcement, as well as other, "associated" announcements sent according to the same "announcement strategy," are deemed "credible" if precisely the types which these announcements declare are deviating want these announcements to be

---

is a strategic dummy. Let the type of player 3 be M or W with equal probability. Then the correlated equilibrium of the BOS in which its two pure strategy equilibria are played with equal probability is achieved as a Nash equilibrium of $G^1(\pi)$ in which player 3 reveals his type in the communication round, and then the favorite equilibrium of player 1 (player 2) is played if player 3 has revealed his type to be M (W).

[10] We refer the reader to Farrell (1985, forthcoming) and Matthews *et. al.* (1991) for discussions of the roles played by a putative equilibrium, a rich language with literal meanings, and the sense in which announcements are surprises.

believed, i.e., prefer the outcomes which arise if these announcements are believed to the putative equilibrium outcome. A credible announcement upsets the putative equilibrium, which therefore cannot be the received way of playing the game. The test of whether an equilibrium is "announcement-proof" is whether a credible announcement exists.

We now formalize and extend this rationale to the finite-round games $G^K(\pi)$. For simplicity, we adapt only the simplest criterion in Matthews *et. al.* (1991), the "strong announcement-proof" criterion. We also simplify by restricting attention to pure strategy announcements.

The definition is inductive on the number of rounds. The set of announcement-proof equilibria of $G^K(\pi)$ will be denoted $A^K(\pi)$. Define $A^0(\pi) \equiv E^0(\pi)$, so that all equilibria of $G^0(\pi)$ are announcement-proof. Then assume that the sets $A^k(\pi) \subseteq E^k(\pi)$ have been defined for all priors π and numbers of rounds $k < K$. We define $A^K(\pi)$ by stating the conditions for an equilibrium $e \in E^K(\pi)$ to be announcement proof.

A first condition for e to be announcement proof is that it should induce announcement-proof equilibria on the continuation games starting after round 1. Only then will no credible, equilibrium-destroying announcement exist in subsequent rounds. Recall that $\beta^1(m) = (\beta_1^1(m_1),\ldots, \beta_n^1(m_n))$ is the equilibrium belief profile at the beginning of round 2 when m is the message profile sent in round 1. Let $e(m) \in E^{K-1}(\beta^1(m))$ be the continuation equilibrium induced by e when m is sent in round 1. Then, this first condition is:

(AP1) $\quad e(m) \in A^{K-1}(\beta^1(m))$ for all $m \in M$.

It remains to determine whether an equilibrium-destroying announcement exists in round 1 itself.

A (pure) *announcement strategy* for player i is a nonempty collection $\mathcal{D}$ of nonempty, disjoint subsets of T_i. For each $D \in \mathcal{D}$, the pair $\langle D, \mathcal{D}\rangle$ is an *announcement* associated with the announcement strategy $\mathcal{D}$. The set of *deviating types* is $T(\mathcal{D}) = \{t_i \in T_i \mid t_i \in D \text{ for some } D \in \mathcal{D}\}$. The remaining types in T_i are *nondeviating*. The interpretation of player i announcing $\langle D, \mathcal{D}\rangle$ instead of playing his round 1 talking strategy is that he is making a speech along the following lines:

"My type is in D. If my type had been in another set $\hat{D} \in \mathcal{D}$, I would have announced that. If my type had not been in any set in $\mathcal{D}$, I would have played my equilibrium talking strategy instead of making this announcement."

The first credibility condition for an announcement strategy is that the other players should believe that each supposedly deviating type of player i is possible (recall that $S_i(\pi_i)$ is the support of π_i):

(C0) $T(\mathcal{D}) \subseteq S_i(\pi_i)$.

An announcement $\langle D, \mathcal{D} \rangle$ by player i which satisfies C0 is *believed* if it causes the other players to adopt beliefs $\hat{\beta}_i(D)$, where $\hat{\beta}_i(D)$ is defined by $\hat{\beta}_i(t_i \mid D) = 0$ if $t_i \notin D$, and

$$\hat{\beta}_i(t_i \mid D) = \frac{\pi_i(t_i)}{\sum_{s_i \in D} \pi_i(s_i).} \quad \text{if } t_i \in D. \tag{2}$$

If player i makes an announcement $\langle D, \mathcal{D} \rangle$ that is believed, and the others play their equilibrium strategies, the beliefs entering round 2 are given by

$$\hat{\beta}(D, m_{-i}) = (\hat{\beta}_i(D), (\beta_j^1(m_j))_{j \neq i}), \tag{3}$$

when m_{-i} is the message profile sent by the other players in round 1. Player i then should predict the play of an announcement-proof equilibrium of the K–1 round game with this prior, i.e. an equilibrium in $A^{K-1}(\hat{\beta}(D, m_{-i}))$. If he is type t_i and pessimistically expects the worst equilibrium in this set, he expects the contribution to his expected payoff from deviating to be,

$$\underline{u}_i(D \mid t_i, m_{-i}) \equiv \underset{\hat{e} \in A^{K-1}(\hat{\beta}(D, m_{-i}))}{\text{minimum}} \sum_{t_{-i} \in T_{-i}} \sum_{a \in A} \pi_{-i}(t_{-i})\, \tau_{-i}(m_{-i} \mid t_{-i})\, o(a \mid t, \hat{e})\, u_i(a, t). \tag{4}$$

The expected payoff of this pessimist from deviating is,

$$\underline{u}_i(D \mid t_i) \equiv \sum_{m_{-i} \in M_{-i}} \underline{u}_i(D \mid t_i, m_{-i}). \tag{5}$$

If, on the other hand, type t_i expects the best equilibrium in $A^{K-1}(\hat{\beta}(D, m_{-i}))$ to be played, he will expect a payoff of $\overline{u}_i(D \mid t_i)$ to result from the announcement $\langle D, \mathcal{D}\rangle$, where $\overline{u}_i$ is defined analogously to $\underline{u}_i$ (taking a maximum instead of a minimum in (4)).

For it to be deemed credible, the first-round announcement strategy $\mathcal{D}$ of player i must satisfy three conditions. First, it must be true that deviating types should prefer using the announcement strategy to the equilibrium, even if they are pessimistic:

(C1) $\underline{u}_i(D \mid t_i) \geq u_i(e \mid t_i, \pi_{-i})$ for all $D \in \mathcal{D}$ and $t_i \in D$, with at least one of these inequalities strict.

Second, nondeviating types who are still believed possible should prefer the equilibrium to any announcement associated with $\mathcal{D}$, even if they are optimistic:

(C2) $\overline{u}_i(D \mid t_i) \leq u_i(e \mid t_i, \pi_{-i})$ for all $t_i \in S_i(\pi_i) \backslash T(\mathcal{D})$.

Third, the announcement strategy should be internally consistent: deviating type $t_i \in D$ should prefer announcement $\langle D, \mathcal{D}\rangle$ to any other announcement associated with $\mathcal{D}$. Again, we state this condition conservatively in terms of optimists and pessimists:

(C3) $\underline{u}_i(D \mid t_i) \geq \overline{u}_i(\hat{D} \mid t_i)$ for all $D \in \mathcal{D}$, $t_i \in D$, and $\hat{D} \in \mathcal{D}$.

DEFINITION 1: For player i, an announcement strategy $\mathcal{D}$ and the associated announcements $\langle D, \mathcal{D}\rangle$ are *credible* in round 1 of $G^K(\pi)$ relative to equilibrium e if $\mathcal{D}$ satisfies C0 – C3. Equilibrium e is *announcement proof* in $G^K(\pi)$ if e satisfies AP1 above and,

(AP2) no player in round 1 has a credible announcement strategy relative to e.

Now we define $A^K(\pi)$ to be the equilibria in $E^K(\pi)$ which are announcement proof by this definition. This extends the strong-announcement-proof criterion in Matthews *et. al.* (1991) from sender-receiver games to all finite round games $G^K(\pi)$.

We have formulated this criterion as simply as possible to capture some of the more compelling rationales for a cheap talk refinement. Alternative formulations of a credible announcement strategy might, for example, require consistency with other possible announcement strategies, or take into account the so-called "Stiglitz critique," or consider mixed announcement strategies (see Matthews *et. al.,* 1991). Further, the pessimism and optimism assumptions in C0 – C3 might be weakened, or announcements might more generally give rise to a range of possible beliefs when they are "believed," or announcements might include proposals of new equilibria as well as claims about types. We hope that future research will resolve these and other issues raised in this paper.

2. The Effect of Adding Communication Rounds (Example 1)

This example shows how adding rounds of communication can increase the amount of information transmitted, even with a large message space. In the example, the game with one round has no equilibrium in which players effectively reveal information, but the game with two rounds has an announcement-proof equilibrium with consequential revelation.[11] An important lesson of the example is that care must be taken in determining whether a Bayesian game is *communication-impervious,* i.e., robust to the players achieving, through pre-play communication, outcomes unforeseen by the economist.[12] The logic of the example indicates that generally, for any number *k*, effective pre-play communication may require at least *k* rounds of communication.

The intuition for the example is the following. The game has two players with private information. Player 3 is uniformed, and only he can take an action. Player 3 cares only about the information of player 2. Player 1 knows whether it is "safe" for player 2 to disclose that information to player 3; player 2 has a separating equilibrium strategy only if he has first learned from player 1 that the situation is "safe." Player 2 will not disclose if he has learned nothing from player 1, or if he has learned from player 1 that the situation is "unsafe." Consequently, information is effectively released only if

---

[11] Forges (1990) also shows that adding rounds can introduce outcomes. Her example has only one informed player; adding rounds is consequential only because it allows the uninformed player to participate in a "jointly controlled lottery" by sending messages according to a mixed strategy simultaneously with the other player (see also Matthews and Postlewaite, 1989). The phenomena we explore here are different, arising even when the players use pure message strategies.

[12] Communication-imperviousness is discussed in Matthews *et. al.* (1991, fn. 7), and Palfrey and Srivastava (1991) and Seidman (1990) analyze the communication-imperviousness of various games. See also Blume and Sobel (1991).

the number of communication rounds is at least two, so that player 2 has a chance to receive a message from player 1 before deciding whether to disclose his information.

The following table gives the payoffs and information structure.

Types of Players 1 and 2		Actions of Player 3: A	B	C
	$1a, 2a$	6, 6, 6	0, 0, 0	4, 4, 4
	$1a, 2b$	5, 5, 0	6, 6, 6	4, 4, 4
	$1b, 2a$	6, 6, 6	0, 0, 0	4, 4, 4
	$1b, 2b$	6, 6, 0	0, 0, 6	4, 4, 4

Player 1 is type $1a$ or type $1b$, where $1a$ represents the "safe" situation and $1b$ the "unsafe" situation. Player 2 is type $2a$ or $2b$. The types are privately known and equally likely. Thus, each row has equal probability of being the true row of payoffs; player 1 knows whether the true row is one of the top two or one of the bottom two, and player 2 knows whether it is one of the first and third or the second and fourth rows. After messages are sent, player 3 takes an action, A, B, or C. Players 1 and 2 have the same payoff functions. Player 3 cares only about the type of player 2; he takes action A if he knows it is $2a$, action B if he knows it is $2b$. Both types of player 2 would like to reveal themselves truthfully to player 3 if the type of player 1 is $1a$. But both types of player 2 would like to convince player 3 that they are type $2a$ if player 1 is type $1b$.

In G^0, the game without communication,[13] player 3 takes an action to maximize his expected payoff according to his prior beliefs. The unique equilibrium action is C.

Now consider G^1, the game obtained by appending one round of communication. In this game player 2 is unable to wait until he learns the type of player 1 before deciding whether to disclose his type. As a result, in no equilibrium does he reveal his type, and the unique equilibrium outcome is still C. To show this, note that because player 2 receives no information about the type of player 1 at the time he sends a message, and because player 3 cares only about the type of player 2, the game is viewed by players 2

[13] In discussing the examples, we write G^k instead of $G^k(\pi)$ when the prior π is understood to be uniform.

and 3 as equivalent to the sender-receiver game between them alone that is obtained by averaging rows one and three and rows two and four, respectively:

	A	*B*	*C*
2*a*	6, 6	0, 0	4, 4
2*b*	5.5, 0	3, 6	4, 4

It is straightforward to show that *C* is played with probability one in any equilibrium of this sender-receiver game; no (partially) separating equilibrium exists because type 2*b* would then pretend to be type 2*a*. Consequently, in G^1 the unique equilibrium outcome is still *C*, and player 2 does not effectively reveal information.

We note in passing that this pooling outcome of G^1 is announcement-proof.[14] The only interesting announcement to consider that might upset a pooling equilibrium is for type 2*a* of player 2 to announce his type.[15] If this announcement is believed, player 3 takes action *A*, and type 2*a* receives a payoff of 6 instead of his equilibrium payoff of 4. But then type 2*b* would obtain an expected payoff of 5.5 from making the same announcement, and he would prefer to tell this lie than to receive his equilibrium payoff of 4. Thus, the announcement by 2 that his type is 2*a* is not credible, and the equilibrium is announcement proof.

Now we turn to G^2, the game with two communication rounds. This game has the following equilibrium in which player 2 effectively reveals information. In the first round, player 1 plays a separating strategy and so reveals his type, but player 2 stays silent (i.e. plays a pooling strategy). In the second round, player 2 plays a separating strategy and so reveals his type if he learned in the first round that player 1 is type 1*a*. But if he learned that player 1 is type 1*b*, player 2 stays silent in the second round. Subsequently, player 3 takes action *A* if the types are (1*a*,2*a*), *B* if they are (1*a*,2*b*), and *C* if they are (1*b*,2*a*) or (1*b*,2*b*). These strategies constitute an equilibrium.

---

[14] Its uniqueness does not imply that it is announcement proof, since some games do not have announcement-proof equilibria (Matthews *et. al.*, 1991).

[15] Formally, this announcement is $\langle D, \mathcal{D}\rangle$, with $D = \{2a\}$ and $\mathcal{D} = \{\{2a\}\}$.

Refer to an equilibrium of G^2 with these strategies as e. The two rounds are necessary for information to be effectively revealed; in particular, players 1 and 2 speaking simultaneously and informatively is not an equilibrium, as the analysis of G^1 indicated ($2b$ would mimic $2a$). We now show that e is announcement proof.

We first show that AP1 holds, i.e., that any announcement in the first round is not credible relative to e. The only announcement player 1 could make that might overturn the equilibrium, since his equilibrium strategy is revealing, is to refuse to reveal his type.[16] If this announcement is believed, beliefs are not updated and the continuation game is G^1. As we have seen, C is taken in any (announcement-proof) equilibrium of G^1. Thus, player 1's announcement (if believed) gives a payoff of 4 to type $1a$, which is less than the payoff of 6 that he gets from e. So C1 is not satisfied, and the announcement is not credible.

The only possible announcements of player 2 in the first round that could upset equilibrium e must reveal his type, since his equilibrium strategy is concealing. Formally, we have three announcement strategies to consider, $\mathcal{D}_1 = \{\{2a\}\}$, $\mathcal{D}_2 = \{\{2b\}\}$, and $\mathcal{D}_3 = \{\{2a\}, \{2b\}\}$. Neither $\mathcal{D}_2$ nor $\mathcal{D}_3$ are credible relative to e, since it is not credible for type $2b$ to announce his type: if $2b$ were to be believed if he announced his type, B would be taken and his expected payoff would be 3, which is less than his equilibrium payoff of 5 (so C1 fails). Thus, the only announcement strategy that could be crdible is $\mathcal{D}_1$, i.e., a revealing announcement by type $2a$ only. If announcement $\{2a\}$ is believed, it induces action A, and type $2b$ would receive a payoff of 5.5 from untruthfully making it. As 5.5 is greater than $2b$'s equilibrium payoff of 5, credibility condition C2 fails. Hence, $\mathcal{D}_1$ is also not credible relative to e. This completes the demonstration that e satisfies AP1.

To complete the argument that e is announcement proof, we must show that AP2 holds, i.e., that the continuation equilibria that it induces in the one-round games starting after the first round are announcement proof. Depending on whether player 1 has revealed his type to be $1a$ or $1b$, there are two equilibrium continuation games starting

---

[16] Formally, this is $\langle D, \mathcal{D}\rangle$, with $D = \{1a, 1b\}$ and $\mathcal{D} = \{\{1a, 1b\}\}$.

with the second round. Both are essentially sender-receiver games between 2 and 3. We handle them simultaneously by letting p denote the updated probability of type $1a$, so that the payoffs of the sender-receiver continuation game between 2 and 3 are the following:

	A	B	C
$2a$	6, 6	0, 0	4, 4
$2b$	$6-p$, 0	$6p$, 6	4, 4

Denote this sender-receiver game, with a uniform prior on player 2's type, as $G^1(p)$. In equilibrium e, the game starting in the second round is $G^1(1)$ if player 1 is type $1a$, and it is $G^1(0)$ if he is type $1b$. By Proposition 6.1 in Matthews *et. al.* (1991), the only announcement-proof outcome of $G^1(1)$ is the separating one (since this outcome gives both types of player 2 their maximum possible payoffs). It is also easy to see that the unique announcement-proof outcome of $G^1(0)$ is the pooling one. These are in fact the continuation equilibrium outcomes induced by e. Thus, AP2 holds, and equilibrium e is announcement proof in G^2. (A similar argument shows that for any integer $k \geq 2$, the outcome of e is the only announcement-proof outcome of G^k.)

3. Infinite Rounds of Communication

Now we turn to phenomena which arise if there is not a final communication round, so that messages sent in any round can receive a response in the next. In this section we introduce a formal analysis of the game with a countably infinite number of communication rounds, $G^\infty(\pi)$. The players in this game take their payoff-relevant actions a_i only after publicly exchanging an infinite sequence of messages, $\vec{m} = (m^1, m^2, \ldots)$.

Measure-theoretic problems prevent the definition of an outcome, and hence of payoffs, if the players can play arbitrary profiles of talking and action strategies.[17] We

---

[17] Given a talking strategy sequence, $\tau_i = (\tau_i^1, \tau_i^2, \ldots)$ for each i, a probability measure $P(\cdot \mid t)$ on $(M^\infty, \mathcal{F}^\infty)$ exists for each type profile t. But the action functions, $\alpha_i(a_i \mid t_i, \vec{m})$, are choice variables and may not be measurable with respect to this endogenous measure. Thus the outcome, $o(a \mid t, e) = \int \alpha(a \mid t, \vec{m})\, P(d\vec{m} \mid t)$, may be undefined.

accordingly restrict the feasible talking strategies, in way we think is realistic. Allowing a player to talk meaningfully in an infinite number of rounds does not seem reasonable, particularly when the amount of information to convey is finite. Furthermore, our focus is on the elimination of end-round effects, not the consequences of unending conversation. So, a sequence of talking strategies will be assumed feasible only if all but a finite number of them can be interpreted as "staying quiet," i.e., pooling.

To facilitate the formulation, we now let each message set M_i contain a designated message q, which shall be interpreted as *staying quiet*. With this, our feasibility requirement is the following:

(F) For each i and sequence of talking strategies $\tau_i = (\tau_i^1, \tau_i^2, \ldots)$, τ_i is *feasible* only if $k_i < \infty$ exists such that for all $k > k_i$,
$\tau_i^k(q \mid t_i, m^1, \ldots, m^{k-1}) = 1$ for all $t_i \in T_i$ and all $(m^1, \ldots, m^{k-1}) \in M^{k-1}$.

Thus, a feasible sequence of talking strategies for player i requires all his types to stay quiet in all but a finite number of rounds. This definition accomplishes our purpose of eliminating end-round effects. But it does not limit the number of rounds in which a player can meaningfully speak; each player, in reply to the talking strategies of the others, can choose to speak meaningfully after the others have become silent.

A profile of feasible talking strategies induces positive probabilities on only a finite number of sequences $\vec{m}$ (as each M_i is finite). These probabilities define the expectation with respect to $\vec{m}$ of the product of any profile of action strategies $\alpha_i(a_i \mid t_i, \vec{m})$. In this way any profile of action strategies and feasible sequences of talking strategies gives rise to a well-defined outcome, from which payoffs can be defined as in (1). A (sequential) equilibrium e, a profile of sequences of talking strategies, sequences of belief functions, and action strategies, is then defined as before, with each player's optimization constrained to feasible talking strategies. Let $E^\infty(\pi)$ be the set of equilibria.[18]

---

[18] Equilibria exist. For each equilibrium of $G^0(\pi)$, a corresponding equilibrium of $G^\infty(\pi)$ is that in which all types of all players are quiet in every round, all players refuse to update their beliefs regardless of the messages received, and each player plays the given equilibrium action strategy regardless of the messages received.

Defining announcement proofness for the infinite round game also presents problems. For the finite round games we proceeded by backwards induction. But without a last round, conditions like AP1 become self-referential in defining a set $A^{\infty}(\pi)$. Another approach would be to let $A^{\infty}(\pi)$ be the limit of the sets $A^{k}(\pi)$ as $k \to \infty$. In addition to convergence problems, this approach is flawed by not eliminating end-round effects: any equilibrium in the limit of the $A^{k}(\pi)$ sets is in all but a finite number of them, so that it is announcement proof in all the finite-round games with a sufficiently large number of rounds.

We do not have a complete definition of announcement proofness for $G^{\infty}(\pi)$. Instead, we specify two properties that a good definition, in our view, should have. These properties will suffice for the analysis of some games, such as those in the next section.

The first property can be shown to be a consequence of the finite-round definition of announcement proof.[19] Suppose the prior π is such that only the type of player i is unknown. Suppose also that a pure strategy equilibrium $e \in E^{K}(\pi)$ exists in which each type of player i receives his maximum possible payoff within the set of equilibrium payoffs of the underlying game as the beliefs in it vary over those which can possibly arise from Bayesian updating of π. Then e is announcement proof: players other than i have known types and hence nothing to announce, and the types of player i cannot gain by announcing because they have no payoffs greater than their equilibrium payoffs. If in addition each t_i is indifferent over the equilibria of the underlying game when the beliefs in it are those which t_i induces in equilibrium e, then e is the unique announcement-proof equilibrium (up to payoff equivalence for player i). For, in this case the talking strategies in e itself can serve as a credible round 1 announcement to upset any other equilibrium.[20]

We formalize this property as an axiom that should be satisfied by any candidate set $A^{\infty}(\pi)$ of announcement-proof equilibria.

---

[19] Proposition 6.1 of Matthews *et. al.* (1991) proves it for sender-receiver games.

[20] The equilibrium e is assumed to be in pure strategies at this point so that it can be converted to a pure strategy announcement of the form $\langle D, \mathcal{D} \rangle$, as defined in Section 2. The indifference of each t_i over the consequences of deviating is required to make nonbinding the conservatism embodied in the use of pessimistic and optimistic expectations in C1 – C3.

AXIOM 1: Suppose prior π and equilibrium $e \in E^{\infty}(\pi)$ satisfy the following:

(*i*) the talking strategies in e are pure strategies;

(*ii*) player i exists such that $\pi_{-i}(t_{-i}) = 1$ for some $t_{-i} \in T_{-i}$;

(*iii*) $u_i(e \mid t_i, \pi_{-i}) = \max_{\hat{e} \in F} u_i(\hat{e} \mid t_i, \pi_{-i})$ for each $t_i \in S_i(\pi_i)$,[21]

where $F \equiv \cup\{E^0(\hat{\pi}) \mid \hat{\pi} \in \Delta(T_1) \times \cdots \times \Delta(T_n),\ S(\hat{\pi}) \subseteq S(\pi)\}$; and

(*iv*) for all $t_i \in S_i(\pi_i)$, if $\vec{m}$ is the equilibrium message sequence when i's type is t_i, then $u_i(\hat{e} \mid t_i, \pi_{-i}) = u_i(e \mid t_i, \pi_{-i})$ for all $\hat{e} \in E^0(\beta(\vec{m}))$, where $\beta(\vec{m})$ is the equilibrium belief profile that $\vec{m}$ induces.[22]

Then, the announcement-proof equilibria consist of e and all other equilibria in $E^{\infty}(\pi)$ which give player i the same payoffs:

$$A^{\infty}(\pi) = \{\hat{e} \in E^{\infty}(\pi) \mid u_i(\hat{e} \mid t_i, \pi_{-i}) = u_i(e \mid t_i, \pi_{-i}) \text{ for all } t_i \in S_i(\pi_i)\}.$$

Axiom 1 provides a compelling definition of an announcement-proof criterion, but only for a small class of games. However, something can be said about other games if they have continuation games to which Axiom 1 applies. Suppose a player has an announcement strategy in some round of play of an equilibrium such that each of the announcements associated with it — if they are believed — generate games for which Axiom 1 defines announcement-proof equilibria. Then, the credibility of the announcement strategy can be determined on the basis of conditions C0 – C3, applied under the assumption that play subsequent to an announcement will be as specified in Axiom 1. This is the gist of Axiom 2 below, for which we must introduce more notation.

Consider an equilibrium $e \in E^{\infty}(\pi)$. Given a $(k-1)$-round history of messages, $h = (m^1, \ldots, m^{k-1})$, let $\hat{\pi} = \beta^{k-1}(h)$ denote the updated equilibrium beliefs after this history, and let $\hat{e} \in E^{\infty}(\hat{\pi})$ be the continuation equilibrium induced by e. Denote by

---

[21] Recall that $S_i(\pi_i) \subseteq T_i$ is the support of π_i, and $S(\pi) \equiv \prod S_i(\pi_i)$.

[22] To check understanding, note that because the other players' types are known and player i's talking strategies are pure, $\beta(\vec{m})$ takes the form $\beta(\vec{m}) = (\beta_i(\vec{m}), \pi_{-i})$, where $\beta_i(\vec{m})$ is a truncation of π_i to a subset containing t_i.

$\beta(m) = (\beta_1(m_1),\ldots, \beta_n(m_n))$ the equilibrium belief profile at the start of round k+1 if message profile m is sent in round k after history h.[23]

Let $\mathcal{D}$ be an announcement strategy for player i, to be made as a deviation after history h. Assume $\mathcal{D}$ satisfies the support condition C0, with $\hat{\pi}_i$ replacing π_i. Then define $\hat{\beta}(D, m_{-i})$ as in (2) and (3) for each $D \in \mathcal{D}$ and $m_{-i} \in M_{-i}$, with $\hat{\pi}_i$ replacing π_i. Thus, $\hat{\beta}(D, m_{-i})$ is the profile of beliefs at the beginning of round k+1 if i makes announcement $\langle D, \mathcal{D}\rangle$ and the others send messages m_{-i} in round k after history h. Now, most crucially, suppose Axiom 1 applies to the game $G^\infty(\hat{\beta}(D, m_{-i}))$ for each $D \in \mathcal{D}$ and message profile m_{-i} which has positive probability of being sent in round k after history h.[24] For each such (D, m_{-i}), a set $A^\infty(\hat{\beta}(D, m_{-i}))$ of announcement-proof equilibria is defined by Axiom 1. Using this set to replace the $A^{K-1}(\hat{\beta}(D, m_{-i}))$ in (4), one defines the pessimistic and optimistic expected payoffs from deviating, $\underline{u}_i(D \mid t_i)$ and $\overline{u}_i(D \mid t_i)$, for each type of player i. With these defined, one can verify whether C1 – C3 apply to $\mathcal{D}$, again with $\hat{\pi}_i$ replacing π_i.

AXIOM 2: Suppose that for prior π and equilibrium $e \in E^\infty(\pi)$, there exists a history $h = (m^1,\ldots,m^{k-1})$, a player i, and an announcement strategy $\mathcal{D}$ for player i which satisfy the assumptions of the previous paragraph. Then equilibrium e is *not announcement proof* i.e., $e \notin A^\infty(\pi)$, if $\mathcal{D}$ is *credible* in the sense of satisfying C0 – C3.

We go no further here in extending announcement-proofness to the infinite round game. We need only Axioms 1 and 2 in the next section.

4. The Importance of Having the Last Word (Example 2)

This example indicates that the plausibility of an equilibrium can hinge on whether the number of communication rounds is finite or infinite. If it is finite, a known, final opportunity to talk exists; in the last round, a player can reveal information and be sure than no other player can respond by revealing more information later. This is not so with

---

[23] That is, $\beta_j(t_j \mid m_i) = \beta_j^k(t_j \mid m^1,\ldots, m^{k-1}, m_j)$.

[24] That is, for each m_{-i} for which $\tau_{-i}^k(m_{-i} \mid t_{-i}, m^1,\ldots, m^{k-1}) > 0$ for some $t_{-i} \in S_i(\hat{\pi}_{-i})$.

an infinite number of rounds. To the extent that in real conversations no speaker is literally able to have the last word, the example suggests that an infinite-round game is the better model of conversation.

The example exhibits the following specific features. As in Example 1, the game has three players, two with private information and a third who chooses an action. If the number of rounds is finite, the only announcement-proof outcome entails both informed players revealing their types. But if the number of rounds is infinite, then the equilibrium in which both informed players are forever quiet is announcement proof. Thus, communication may be hindered rather than facilitated by the addition of rounds supposedly devoted to communication.

The structure of the game is similar to that of Example 1. Players 1 and 2 each have two types, uniformly and independently distributed. Player 3 takes one of the actions *A*,…., *I*, which result in the following payoffs:

Action of Player 3

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
1*a*, 2*a*	2, 2, 2	3, 0, 3	0, 0, 0	0, 3, 3	0, 0, 0	1, 1, 5	0, 0, 0	0, 0, 0	0, 0, 0
1*a*, 2*b*	2, 2, 2	3, 0, 3	0, 0, 0	0, 0, 0	0, 3, 3	0, 0, 0	1, 1, 5	0, 0, 0	0, 0, 0
1*b*, 2*a*	2, 2, 2	0, 0, 0	3, 0, 3	0, 3, 3	0, 0, 0	0, 0, 0	0, 0, 0	1, 1, 5	0, 0, 0
1*b*, 2*b*	2, 2, 2	0, 0, 0	3, 0, 3	0, 0, 0	0, 3, 3	0, 0, 0	0, 0, 0	0, 0, 0	1, 1, 5

The games G^k with one or more communication rounds, $1 \leq k \leq \infty$, have the following equilibrium outcomes:

(*i*) player 3 takes action *A* (obtained, e.g., by an equilibrium in which players 1 and 2 are quiet in every round);

(*ii*) type 1*a* induces *B* and 1*b* induces *C* (obtained by equilibria in which player 1 uses a revealing talking strategy in some round and player 2 is always quiet);

(*iii*) type 2*a* induces *D* and 2*b* induces *E* (obtained by equilibria in which player 2 uses a revealing talking strategy in some round and player 1 is always quiet); and (*iv*) types (1*a*, 2*a*) induce *F*, types (1*a*,2*b*) induce *G*, types (1*b*, 2*a*) induce *H*, and types (1*b*, 2*b*) induce *I* (obtained by equilibria in which players 1 and 2 both use revealing talking strategies).

If the number of rounds is finite, then only the full communication outcome, (*iv*), is announcement proof. But if $k = \infty$, then both (*i*) and (*iv*), and not (*ii*) and (*iii*), are announcement proof. As space restrictions prevent us from proving all these claims, we focus on showing why the no-communication outcome (*i*) is announcement proof if and only if $k = \infty$.

Consider first the case $k < \infty$. Players 1 and 2 both have credible announcements they can make in the k^{th} round to destroy any equilibrium with outcome (*i*). For example, type 1*a* of player 1 in the k^{th} round of such an equilibrium can announce his type. If this announcement is believed, it results in action *B*, since player 3 would know player 1's type but not player 2's. This gives 1*a* a payoff of 3 instead of the smaller 2 he receives from the equilibrium outcome *A*. Since type 1*b* would achieve a lower payoff from *B* than from *A*, the announcement is credible by conditions C0 – C3, and the equilibrium is not announcement proof.

Suppose now that $k = \infty$, and consider some equilibrium $e \in E^{\infty}$ which has outcome (*i*). Since players 1 and 2 play nonrevealing strategies, the candidate announcements that might upset this equilibrium entail one of them announcing his type in some round. But now the round of the announcement cannot be the final round, and the announcer will have to consider whether his announcement will lead to an unappealing equilibrium in which the other player reveals his type in a future round. Indeed, as we now argue, it does.

Because of the symmetry of the game, we may consider w.l.o.g. an announcement by player 1 of his type in some round *k*. Suppose he is type 1*a* and so announces. If the announcement is believed, player 3's beliefs are updated to some $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)$, where $\hat{\pi}_1(1a) = 1$ and $0 < \hat{\pi}_2(2a) < 1$.[25] The continuation game that follows, if the

---

[25] Player 3's beliefs about the type of player 2 are his equilibrium beliefs, since 2 is playing his equilibrium talking strategy in this round. Since the equilibrium action is *A*, player 2 is not revealing his type. Thus, $\hat{\pi}_2$, the updated beliefs at the end of round *k*, cannot be concentrated on $\{2a\}$ or $\{2b\}$.

announcement is believed, is $G^{\infty}(\hat{\pi})$.

Axiom 1 applies to $G^{\infty}(\hat{\pi})$. Given that player 3 believes that 1 is type $1a$, his best response actions are limited to $\{B, F, G\}$, depending on his beliefs about the type of player 2. One equilibrium is for player 2 to reveal his type, which has the outcome in which $2a$ induces F and $2b$ induces G. This equilibrium is in pure strategies, and it gives both types of player 2 their maximal payoffs in the set of equilibrium payoffs of $G^{\infty}(\hat{\pi})$ (since $G^{\infty}(\hat{\pi})$ has no equilibrium in which D or E are taken). Thus, Axiom 1 applies, and the unique announcement-proof outcome of $G^{\infty}(\hat{\pi})$ is that in which player 2 reveals his type, with $2a$ inducing F and $2b$ inducing G.[26]

By Axiom 2, this unique announcement-proof outcome of $G^{\infty}(\hat{\pi})$ provides the reference utility for determining the credibility of player 1's announcement. It give type $1a$ a payoff of 1, which is less than his equilibrium payoff (from A) of 2. Consequently, his announcement is not credible.

By symmetry, similar arguments show that no type of either player 1 or 2 can make a credible announcement in any round during the play of any equilibrium of G^{∞} that has the no-communication outcome (*i*). We conclude that the no-communication outcome is announcement proof if and only if the number of communication rounds is infinite.

This example bears some similarity to the repeated prisoners' dilemma. In a sense, revealing one's type is a "dominant strategy" for the players 2 and 3. For example, if 2 is not revealing, then by revealing 1 will shift the outcome from (*i*) to the more preferred (*ii*); if 2 is revealing, then by revealing 1 will shift the outcome from (*iii*) to the more preferred (*iv*). Of course, since a player is unable to *prove* to another that he is a certain type, he cannot really "choose" to reveal his type — his messages can always be ignored.

[26] Similar arguments show that the outcomes in which only one player reveals his type, (*ii*) and (*iii*), are not announcement proof for any $1 \leq k \leq \infty$. Given that one of players 1 and 2 has revealed his type in the past, or is currently revealing his type in the current round, a credible announcement for the other is to reveal his type. For example, in any equilibrium that supports (*ii*), it is credible for player 2 to announce his type in some round no earlier than the round in which player 1 reveals (as part of the equilibrium) his type, as this results in a payoff for player 2 of 1 (from F, G, H, or I) instead of his equilibrium payoff of 0 (from B or C).

But this is where the refinement criterion plays a role: announcements that are deemed credible must be believed, so that a player can use a credible announcement to prove, in essence, that he is a certain type. In the finite-round games, the dominant revealing strategy will be played in the last round (if not before). But in the infinite-round game, a player can be deterred from revealing by the threat of a retaliatory revelation by the other informed player in the future.

References

Austen-Smith, D. (1990), "Information transmission in debate," *American Journal of Political Science*, **34**, 124-152.

Blume, A. (1992a), "Equilibrium refinements in sender-receiver games," Iowa, mimeo.

Blume, A. (1992b), "Equilibrium refinement in perturbed games and in sender-receiver games," Iowa, mimeo.

Blume, A. and J. Sobel (1991), "Communication-proof equilibria in cheap-talk games," University of California at San Diego, mimeo.

Crawford, V. and J. Sobel (1982), "Strategic information transmission," *Econometrica* **50**, 1431-1452.

Farrell, J. (1985), "Credible neologisms in games of communication," MIT, mimeo.

Farrell, J. (forthcoming), "Meaning and credibility in cheap-talk games," in *Mathematical Models in Economics*, ed. M. Demster, Oxford Univ. Press.

Farrell, J. and R. Gibbons (1989), "Cheap talk can matter in bargaining," *Journal of Economic Theory*, **48**, 221-237.

Forges, F. (1990), "Equilibria with communication in a job market example," *Quarterly Journal of Economics* **105**, 375-398.

Fudenberg, D., and J. Tirole (1991), *Game Theory,* MIT Press, Cambridge, Massachusetts.

Kreps, D. and R. Wilson (1982), "Sequential equilibria," *Econometrica*, **50**, 863-894.

Green, J. and N. Stokey (1980), "Two-person games of information transmission," Harvard and Northwestern, mimeo.

Matthews, S. (1989), "Veto threats: Rhetoric in a bargaining game," *Quarterly Journal of Economics*, **104**, 347-369.

Matthews, S., M. Okuno-Fujiwara, and A. Postlewaite (1991),"Refining cheap-talk equilibria," *Journal of Economic Theory*, **55**, 247-273.

Matthews, S. and A. Postlewaite (1989), "Pre-play communication in a two-person sealed-bid double auction," *Journal of Economic Theory*, **48**, 238-263.

Myerson, R. (1982), "Optimal coordination mechanisms in principal-agent problems," *Journal of Mathematical Economics,* **10**, 67-81.

Myerson, R. (1989), "Credible negotiation statements and coherent plans," *Journal of Economic Theory*, **48**, 264-303.

Palfrey, T. R. and S. Srivastava (1991), "Efficient trading mechanisms with pre-play communication," *Journal of Economic Theory*, **55**, 17-40.

Rabin, M. (1990), "Communication between rational agents," *Journal of Economic Theory*, **51**, 144-170.

Seidmann, D. (1990), "Effective cheap talk with conflicting interests," *Journal of Economic Theory*, **50**, 445-458.

Stein, Jeremy C. (1990), "Cheap talk and the fed: a theory of imprecise policy announcements," *American Economic Review,* **79**, 32-42.