

Fixed-Equilibrium Rationalizability in Signaling Games*

JOEL SOBEL

University of California at San Diego, La Jolla, California 92093

LARS STOLE

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

AND

IÑIGO ZAPATER

Brown University, Providence, Rhode Island 02912

Received May 27, 1989; revised April 13, 1990

This paper studies equilibrium refinements in signaling games through an examination of rationalizability in derived games obtained by replacing the equilibrium path with a sure outcome that yields the equilibrium payoff to all players. The informed player chooses between the sure payoff and sending an out-of-equilibrium signal from the original game. Whether or not the strategy of choosing the sure payoff is rationalizable is related to the iterated intuitive condition (divinity) when the original game is viewed as having imperfect (incomplete) information. Our results also demonstrate the significance of testing out-of-equilibrium signals as a set rather than individually. *Journal of Economic Literature* Classification Numbers: 021, 022, 026. © 1990 Academic Press, Inc.

1. INTRODUCTION

In this paper we attempt to unify some recent work on equilibrium refinements in signaling games by examining procedures which delete strategies that are dominated relative to some reference payoffs. We present three techniques, all of which are variations of rationalizability (Bernheim

* We thank Valentina Corradi, Eric van Damme, Eddie Dekel, Drew Fudenberg, Robert Gibbons, Matthew Rabin, and a referee for helpful comments. Sobel is grateful to the National Science Foundation and the Sloan Foundation for financial support. A National Science Foundation Graduate Fellowship supported Stole's research.

[3] and Pearce [15]), and relate them to the intuitive criterion (Cho and Kreps [5]) and divinity (Banks and Sobel [1]). These techniques take the original game and an equilibrium for that game, and derive a new signaling game. In the new game we replace the equilibrium path with a sure outcome that yields the equilibrium payoff of the original game to all players. The informed player may choose the sure payoff or may send a signal that was not used in the equilibrium of the original game. We then ask whether the strategy of choosing the sure payoff survives iterative deletion of dominated strategies.

Our different notions of rationalizability correspond to different ways of looking at signaling games. When we require that equilibrium outcomes survive iterative deletion of dominated strategies, it matters whether we treat the signaling game as a two-player game of incomplete information or an imperfect-information game in which there is a player for every type of informed player.¹ In the first case types of the informed player have common conjectures over strategy choices of the uninformed player. This requirement deletes more strategies than the imperfect-information treatment. It also matters whether we treat unreached information sets one at a time or all at once. The refinement ideas of Banks and Sobel [1], Cho [4], Cho and Kreps [5], Farrell [9], and Grossman and Perry [11] analyze behavior at unreached information sets one at a time. However, the existence of several possible unused signals may alter the way one of these signals can be interpreted.

The next section describes signaling games and defines fixed-equilibrium rationalizability. In Section 3, we show that fixed-equilibrium rationalizability for the imperfect-information game is equivalent to the iterated version of the Cho–Kreps [5] intuitive criterion. Section 4 demonstrates that fixed-equilibrium rationalizability for the incomplete-information game (when unreached information sets are treated one by one) is equivalent to co-divinity, which coarsens Banks and Sobel's [1] concept of divinity. If we apply rationalizability with respect to a fixed equilibrium, then we obtain a set of outcomes that are generally larger than if we require the rationalizability requirement to hold for each signal separately.

In Section 5 we look at what happens when we require outcomes of the derived game to satisfy a more stringent requirement than rationalizability. We relate the set of outcomes that survive a variation of Grossman and Perry's [11] test of Perfect Sequential Equilibrium due to van Damme [7] to the Nash Equilibria of the derived game in which the Sender refuses the sure outcome with positive probability.

Section 6 compares our use of auxiliary games to related work of Ben-Porath and Dekel [2] and van Damme [8].

¹ Fudenberg and Kreps [10] make a similar observation.

2. THE MODEL AND FIXED-EQUILIBRIUM RATIONALIZABILITY

Throughout the paper we limit attention to simple signaling games. In these games one player, the Sender, receives private information. We refer to this information as the Sender's type; we denote the type of the Sender by t ; t is drawn from a finite set T (we also use T to refer to the cardinality of the set of types). The Sender's type is drawn according to a probability distribution π over T . We assume that π is common knowledge and that $\pi(t) > 0$ for all $t \in T$. After the Sender learns his type, he sends a signal m to the other player, the Receiver. We denote the set of signals available to a Sender of type t by $M(t)$; $T(m)$ denotes the set of types that are able to send the signal m . The Receiver responds to the Sender's signal m by choosing an action, a , from a finite set of responses that we call $A(m)$. The players have von Neumann–Morgenstern utility functions defined over type, signal, and action. The Sender's payoff function is denoted $u(t, m, a)$ and the Receiver's payoff function is denoted $v(t, m, a)$; we extend these functions to the set of all mixed strategies by linearity and use $u(\cdot)$ and $v(\cdot)$ to refer to these extensions. It is convenient to introduce notation for the set of best responses of the Receiver. Let μ be a probability distribution over $T(m)$. Let

$$\text{BR}(\mu, m) = \arg \max_{a \in A(m)} \sum_{t \in T(m)} v(t, m, a) \mu(t).$$

If the Receiver thinks that $\mu(t)$ is the probability that the Sender is type t given the signal m , then $\text{BR}(\mu, m)$ is the set of best responses to m . Let $\text{BR}(S, m)$ denote the set of the Receiver's best responses to probabilities concentrated on a subset S of the set of all probability distributions on $T(m)$, $\text{BR}(S, m) = \bigcup_{\{\mu: \mu \in S\}} \text{BR}(\mu, m)$; we write $\text{MBR}(\mu, m)$ and $\text{MBR}(S, m)$ for the sets of mixed best responses corresponding to $\text{BR}(\mu, m)$ and $\text{BR}(S, m)$, respectively. On occasion, we abuse notation and write $\text{BR}(S, m)$ when S is a subset of types. At these times we identify S with the set of probability distributions on S ; hence, when $S \subset T(m)$, $\text{BR}(S, m) = \bigcup_{\{\mu: \mu(S)=1\}} \text{BR}(\mu, m)$.

We will investigate the effect of imposing rationalizability requirements on a fixed equilibrium outcome. We begin with a particular sequential equilibrium to a signaling game. The equilibrium consists of a behavior strategy for the Sender, denoted by $\sigma(m|t)$, which specifies the probability that the Sender of type t sends the signal $m \in M(t)$; a behavior strategy for the Receiver, denoted by $\rho(a|m)$, which specifies the probability that the Receiver takes the action $a \in A(m)$ in response to the signal m ; and assessments, denoted by $\mu(t|m)$, such that $\mu(\cdot|m)$ is a probability distribution over $T(m)$ for each $m \in M$. A triple (σ, ρ, μ) is a sequential equilibrium (Kreps and Wilson [14]) to a signaling game if and only if σ is a

best response to ρ ($\sigma(m'|t) > 0$ only if m' maximizes $u(t, m, \rho(\cdot|m))$ over all $m \in M(t)$); the Receiver responds optimally to his assessment ($\rho(\cdot|m) \in \text{MBR}(\mu(\cdot|m), m)$ for all $m \in M$); and the assessments are consistent with the equilibrium strategy of the Sender and the prior whenever possible (if $\sum_{t' \in T(m)} \pi(t') \sigma(m|t') > 0$, then $\mu(t|m) = [\pi(t) \sigma(m|t)] / [\sum_{t' \in T(m)} \pi(t') \sigma(m|t')]$). Given the strategies (σ, ρ) , we can identify the equilibrium payoffs of the players,

$$u^*(t) = \sum_{m \in M(t)} \sum_{a \in A(m)} u(t, m, a) \sigma(m|t) \rho(a|m)$$

for the Sender of type t , and

$$v^* = \sum_{t \in T} \sum_{m \in M(t)} \sum_{a \in A(m)} v(t, m, a) \sigma(m|t) \rho(a|m) \pi(t)$$

for the Receiver. In addition, we can define the set of unsent signals,

$$M^* = \left\{ m \in M : \sum_{t \in T(m)} \pi(t) \sigma(m|t) = 0 \right\},$$

and the equilibrium outcome, which is the probability distribution $\sigma(m|t) \rho(a|m) \pi(t)$ on the terminal nodes (t, m, a) of the game induced by the equilibrium strategies.

Given a sequential equilibrium and a subset M_0 of the set of unsent signals M^* , we define a new signaling game. The set of possible types of the Sender is T , as in the original game. The set of pure strategies available to the type t Sender is $\{m^*\} \cup [M_0 \cap M(t)]$. If the Sender's signal is m^* , then the Receiver's action set $A(m^*)$ is a single point; call it a^* . Otherwise, the Receiver has precisely the same actions available as he had in the original game. The preferences for the new game, denoted by $\tilde{u}(\cdot)$ for the Sender and $\tilde{v}(\cdot)$ for the Receiver, satisfy

$$\begin{aligned} \tilde{u}(t, m, a) &= u(t, m, a) && \text{for } m \in M_0 \text{ and } a \in A(m), \\ \tilde{v}(t, m, a) &= v(t, m, a) && \text{for } m \in M_0 \text{ and } a \in A(m), \\ \tilde{u}(t, m^*, a^*) &= u^*(t), && \text{and } \tilde{v}(t, m^*, a^*) = v^*. \end{aligned}$$

We denote the new game by $G(\sigma, \rho, M_0)$. We refer to this game (often without specifying the choices of σ, ρ , and M_0) as the derived game. Grossman and Perry [11, p. 111] and van Damme [7, p. 287] use derived games of the form $G(\sigma, \rho, \{m\})$ to study properties of Perfect Sequential Equilibria. Grossman and Perry attribute the idea to David Kreps.

The derived game replaces the equilibrium of the original game with a signal m^* that gives the players the payoff they would have received in the

original equilibrium. In the derived game, the Sender has the option of selecting the payoff to the original game or selecting a signal from a subset of the original signals. If the Sender decides not to send the signal m^* , then the game continues as the original.

We apply the concept of rationalizability in the extensive form to the derived game. In the context of simple signaling games, the definition below is equivalent to the definition introduced by Pearce [15]. Let $\text{co}(X)$ denote the convex hull of the set X . Let $R\mathcal{S}^0$ be the pure strategy set of the Sender and $R\mathcal{R}^0$ be the pure strategy set of the Receiver. Consider the iterative procedure that determines $R\mathcal{S}^{k+1}$ and $R\mathcal{R}^{k+1}$ given $R\mathcal{S}^k$ and $R\mathcal{R}^k$ using the following steps:

$$\text{R0. } \text{RJ}^k = \{m \in M : \exists s \in R\mathcal{S}^k \text{ such that } s(t) = m \text{ for some } t \in T(m)\}.$$

$\text{R1. } R\mathcal{S}^{k+1} = \{s \in R\mathcal{S}^k : \exists \rho \in \text{co}(R\mathcal{R}^k) \text{ such that } s \text{ is a best response to } \rho\}.$

$\text{R2. } R\mathcal{R}^{k+1} = \{r \in R\mathcal{R}^k : \forall m \in \text{RJ}^{k+1} \exists \sigma \in \text{co}(R\mathcal{S}^{k+1}) \text{ such that } \sigma(m|t) > 0 \text{ for some } t, \text{ and } r(m) \text{ is a best response to } \sigma\}.$

The sets $R\mathcal{S}^k$ and $R\mathcal{R}^k$ represent admissible strategies at the k th step of the process. A strategy s of the Sender is included in $R\mathcal{S}^{k+1}$ if and only if it is a best response to an element of the convex hull of the pure strategies of the Receiver. If it is common knowledge that the Receiver uses only the strategies in $R\mathcal{R}^k$, then elements of $\text{co}(R\mathcal{R}^k)$ represent possible conjectures that the Sender could have. R1 restricts attention to strategies that respond optimally to some conjecture over the Receiver's admissible set of pure strategies.² If it is common knowledge that the Receiver uses only the strategies in $R\mathcal{R}^k$, then it is sensible to restrict the Sender to strategies in $R\mathcal{S}^{k+1}$.

The condition that defines $R\mathcal{R}^{k+1}$ is a bit more complicated. RJ^{k+1} consists of the set of signals that can be sent using strategies in $R\mathcal{S}^{k+1}$. R2 states that strategies in $R\mathcal{R}^{k+1}$ are optimal responses to a conjecture that the Receiver can hold over the possible strategies of the Sender. R2 allows the Receiver to have a different conjecture depending on what signal is sent: The conjecture σ could depend upon m . If there exists a strategy in $R\mathcal{S}^{k+1}$ in which at least one type of Sender uses the signal m , then R2 requires that the Receiver's reply to m must be an optimal response to a conjecture that the Sender uses m with positive probability. That is, the conjecture should explain why the Receiver hears the signal m . Without this restriction, R2 would never delete any strategies since any strategy of

² Pearce [15] shows that any conjecture over a subset S of a player's mixed strategy set can be represented as a point in $\text{co}(S)$. Since we can identify the set of mixtures of pure strategies in a set X with $\text{co}(X)$, allowing conjectures over sets of mixed strategies does not change the analysis.

the Receiver is an optimal response to the conjecture that the Sender uses m^* with probability one. If $m \notin \text{RJ}^{k+1}$, then there is no strategy in RS^{k+1} that uses the signal with positive probability. R2 deletes no further responses to m .

Form the sets $\text{RS}^* = \bigcap_{k \geq 0} \text{RS}^k$ and $\text{RR}^* = \bigcap_{k \geq 0} \text{RR}^k$ of rationalizable strategies for the Sender and Receiver respectively.³ Because RS^0 and RR^0 are nonempty and finite, RS^k and RR^k are nonempty for all k . Let s^* denote the strategy in which $s^*(t) = m^*$ for all t .

DEFINITION. The equilibrium (σ, ρ) determines a *fixed-equilibrium rationalizable outcome* (FERO) of the original game if $s^* \in \text{RS}^*$.

That is, (σ, ρ) determines a FERO if it is a rationalizable strategy in $G(\sigma, \rho, M^*)$ for every Sender type to use m^* .

R0, R1, and R2 describe a particular way in which to delete a subset of the set of weakly dominated strategies in the derived game. Deleting strictly dominated strategies has no cutting power in the derived game because all of the Receiver's strategies are best responses to s^* ; consequently none are strictly dominated. We require, if the Receiver hears the signal m , that he respond optimally to some conjecture over the Sender's strategies. This requirement prevents the Receiver from using a behavior strategy that is strictly dominated given m . In fact we shall see that it is even more restrictive. Nevertheless, our procedure does not rule out strategies which specify that the Receiver take an action that is weakly dominated given a signal. We present a variety of iterative procedures in the paper. The procedures differ in how they describe the strategies of the Sender and what criterion is used to delete these strategies. The intuitive criterion restricts the set of Sender types that use a particular message, which corresponds

³ Our definition differs from Pearce's in two ways. First, an iteration of Pearce's procedure simultaneously deletes strategies of all players. Our procedure alternates between deleting the Sender's strategies and the Receiver's strategies. When weakly dominated strategies are deleted, this difference could lead to a different set of rationalizable strategies. However, it does not affect whether $s^* \in \text{RS}^*$. The second difference in the definitions is that Pearce requires that in order to be an element in, say, RS^{k+1} , a strategy \bar{s} need only satisfy

$$\begin{aligned} & \sum_{t \in T} \sum_{a \in A(m)} \tilde{u}(t, \bar{s}(t), a) \rho(a | \bar{s}(t)) \pi(t) \\ & = \max \left\{ \sum_{t \in T} \sum_{a \in A(m)} \tilde{u}(t, s(t), a) \rho(a | s(t)) \pi(t) : s \in \text{RS}^k \right\}. \end{aligned}$$

That is, \bar{s} need only be a best response in RS^k . We require that strategies in RS^{k+1} be best responses in the larger set RS^0 (that is, the maximum above is taken over $s \in \text{RS}^0$). A simple induction argument shows that the two approaches are equivalent in our context. However, the approaches lead to different answers if one uses closely related equilibrium concepts. See the discussion of the "Never a Weak Best Response" Criterion in Cho and Kreps [5, p. 207].

directly to deleting a set of the Sender's strategies. Co-divinity and divinity restrict the set of beliefs that the Receiver may hold. Since each set of admissible strategies for the Sender gives rise to a set of beliefs for the Receiver, fixed-equilibrium rationalizability and the intuitive criterion can be viewed as techniques that restrict the Receiver's set of admissible beliefs.

We construct $G(\sigma, \rho, M^*)$ using only the equilibrium path and payoffs induced by (σ, ρ) . The equilibrium (σ, ρ) determines the payoff for choosing m^* and a set of unexpected signals (for which $\sum_{t \in T(m)} \pi(t) \sigma(m|t) = 0$) that are not available in the derived game. Consequently $G(\sigma, \rho, M^*)$ does not depend upon the responses of the Receiver to unexpected signals. In this way, the equilibrium path plays a different role in our construction than the specification of off-the-equilibrium-path behavior. We do not have formal justification for this asymmetric treatment of reached and unreached information sets. However, the approach has been useful in providing an intuitive framework for the Kohlberg–Mertens [13] notion of forward induction. It has been used in a number of places to define equilibrium refinements or restrict outcomes in extensive games (see, for example, Banks and Sobel [1], Cho [4], Cho and Kreps [5], Cho and Sobel [6], Farrell [9], and Grossman and Pery [11, 12]).

Analyzing a derived game assumes a particular view of off-the-equilibrium-path behavior. According to this approach, when the Receiver hears an unexpected signal, he acts as if he is playing a new game. The new game contains only a subset of the strategies of the original game, and our analysis hinges on the idea that the Sender only “chooses” to play this game when he does not expect to lose (relative to reference payoffs determined by an equilibrium) by doing so. The work of Fudenberg and Kreps [10] provides a dynamic motivation for studying fixed-equilibrium rationalizability. If players arrive at equilibrium behavior following a period of learning and experimentation, it may be that a player learns about the equilibrium path earlier (or in more detail) than about his opponents' out-of-equilibrium behavior. In this case, a player may make comparisons between a “sure” equilibrium payoff and a conjecture about what would happen if he strayed from the equilibrium path. The analysis of Fudenberg and Kreps [10] is quite different from ours. They present a complete theory of play in the sense that all actions are taken with positive probability and therefore beliefs are not specified arbitrarily. The experimentation that takes place in the Fudenberg and Kreps framework contrasts with the behavior in our model, which we justify using rationalizability. Nevertheless, our approaches complement one another because the limits of the Fudenberg–Kreps dynamics satisfy fixed-equilibrium rationalizability.

We need to introduce two variations of our definition. One variation involves testing unsent signals one at time.

DEFINITION. The equilibrium strategies (σ, ρ) determine a *fixed-equilibrium signal-by-signal rationalizable outcome* (FESSO⁴) of the original game if s^* is a rationalizable strategy in $G(\sigma, \rho, \{m\})$ for all $m \in M^*$.

Section 4 contains an example that demonstrates that a FERO need not be a FESSO.

Our definition of fixed-equilibrium rationalizability treats signaling games as two-player games of incomplete information. One can also think of signaling games as $(T+1)$ -player games in which there is a player for each type of Sender. This distinction modifies the definition of rationalizability. If Sender types are treated as separate players, then rationalizability does not require different types to have the same conjecture about the strategy choice of the Receiver. The difference makes it more difficult to rule out strategies and provides another way to test outcomes in signaling games.

DEFINITION. The equilibrium (σ, ρ) determines a *fixed-equilibrium rationalizable outcome for the imperfect information game* (FERIMO) if s^* is a rationalizable strategy in $G(\sigma, \rho, M^*)$ viewed as a $(T+1)$ -player game.

Formally, denote the set of admissible strategies for the Sender and Receiver at the k th stage of the process by \mathcal{I}^k and \mathcal{R}^k , respectively. If we replace step R1 in the definition of rationalizability for (incomplete information) signaling games with

R1'. $\mathcal{I}^k = \{s \in \mathcal{I}^k : \forall t \exists \rho_t \in \text{co}(\mathcal{R}^k) \text{ such that } s(t) \text{ is a best response to } \rho_t\}$,

then we obtain the definition of rationalizability for signaling games treated as $(T+1)$ -player games. (We construct the set of admissible strategies for the Receiver at stage $k+1$, \mathcal{R}^{k+1} , from \mathcal{I}^{k+1} using R2.) If we write $\mathcal{I}^* = \bigcap_{k \geq 0} \mathcal{I}^k$ and $\mathcal{R}^* = \bigcap_{k \geq 0} \mathcal{R}^k$, then an outcome is a FERIMO if and only if $s^* \in \mathcal{I}^*$.

The FERIMO procedure places fewer restrictions on admissible beliefs at each stage of the deletion process than the FERO procedure because FERIMOs do not require that each type of Sender have the same conjecture over the Receiver's responses. Our first result follows for this reason.⁵

PROPOSITION 1. *Any FERO is a FERIMO.*

⁴ Soltano le persone che conoscono l'italiano e la teoria dei giochi possono valutare se questa terminologia é appropriata.

⁵ A separate argument, similar to the one we use to prove Proposition 5, is needed to prove Proposition 1 when the FERO procedure reaches a step in which all strategies that send $m \in M^*$ are deleted.

While Pearce shows that the set of rationalizable strategies is nonempty in every finite game, this result does not guarantee that fixed-equilibrium rationalizable outcomes exist in signaling games. However, existence of FEROs follows from the relationship between fixed-equilibrium rationalizability and divinity that we establish in Section 4. In Section 4 we show that any FESSO must be a FERo and that FESSO is equivalent to an equilibrium concept called co-divinity. The set of divine outcomes is a subset of the set of co-divine outcomes. Since Banks and Sobel [1] show that every signaling game has a divine outcome, every signaling game has a FESSO and a FERo. Section 3 demonstrates that the set of FERIMOs coincides with the set of outcomes that survive the iterated intuitive criterion. Therefore, results of Cho and Kreps [5] show that any signaling game has a FERIMO.

3. THE ITERATED INTUITIVE CRITERION

In this section we treat the signaling game as a $(T + 1)$ -player game of imperfect information. We show that a sequential equilibrium outcome passes the iterative intuitive criterion of Cho and Kreps [5, p. 202] if and only if it is a fixed equilibrium rationalizable outcome in the imperfect-information game.

The intuitive criterion of Cho and Kreps is a procedure by which signal-action pairs (t, m) are deleted from the game if the Sender of type t prefers his equilibrium utility to sending m no matter what admissible response the Receiver takes when he hears m . The iterative version of the text strikes signal-action pairs as above, restricts the set of admissible responses of the Receiver to optimal responses to beliefs concentrated on the remaining Sender types, and then repeats the process. Formally, fix a sequential equilibrium that yields expected utility $u^*(t)$ to the type t Sender, and let m be a signal that is sent with probability zero in the equilibrium. The iterated intuitive criterion takes $IC\mathcal{T}^0(m) = T(m)$ and $IC\mathcal{R}^0(m) = A(m)$. IC1 and IC2 use $IC\mathcal{T}^k(m)$ and $IC\mathcal{R}^k(m)$ to determine $IC\mathcal{T}^{k+1}(m)$ and $IC\mathcal{R}^{k+1}(m)$.

$$IC1. \quad IC\mathcal{T}^{k+1}(m) = \{t \in IC\mathcal{T}^k(m) : u^*(t) \leq \max_{a \in IC\mathcal{R}^k(m)} u(t, a)\}$$

$$IC2. \quad IC\mathcal{R}^{k+1}(m) = \begin{cases} BR(IC\mathcal{T}^{k+1}(m), m) & \text{if } IC\mathcal{T}^{k+1}(m) \neq \emptyset \\ IC\mathcal{R}^k(m) & \text{if } IC\mathcal{T}^{k+1}(m) = \emptyset. \end{cases}$$

IC1 deletes those types that do worse than their equilibrium payoff if the Receiver responds to m by taking actions in $IC\mathcal{R}^k(m)$. IC2 limits attention

to actions in $A(m)$ that respond optimally to some conjecture placing probability one on types in $IC\mathcal{T}^k(m)$.

Let $IC\mathcal{T}^*(m) = \bigcap_{k \geq 0} IC\mathcal{T}^k(m)$ and $IC\mathcal{R}^*(m) = \bigcap_{k \geq 0} IC\mathcal{R}^k(m)$. IC2 guarantees that $IC\mathcal{R}^*(m)$ is not empty.

DEFINITION. An equilibrium outcome satisfies the *iterated intuitive criterion with respect to the signal* $m \in M^*$, if and only if for all $t \in T(m)$ there exists $a \in IC\mathcal{R}^*(m)$ such that $u(t, m, a) \leq u^*(t)$. An equilibrium satisfies the *iterated intuitive criterion* if and only if it satisfies the criterion for all $m \in M^*$.⁶

Fixed-equilibrium rationalizability in the imperfect-information game is little more than a restatement of the iterated intuitive criterion. Condition IC1 in the definition of the intuitive criterion serves to delete Sender types who prefer their equilibrium payoff to sending another signal. This restriction is then used to limit Receiver responses in IC2. The same process of deletion is carried out by first deleting Sender strategies in condition R1' of the definition of FERIMO.

Before we prove that outcomes that pass the iterated intuitive criterion coincide with the set of FERIMOs, we introduce some notation: Let $I\mathcal{R}^k(m) = \{a \in A(m) : a = r(m) \text{ for some } r \in I\mathcal{R}^k\}$; $I\mathcal{R}^k(m)$ is the set of actions that are admissible for the Receiver at the k th stage of the FERIMO procedure. Similarly, we use $I\mathcal{R}^*$ to define $I\mathcal{R}^*(m)$. It is also useful to observe that if $A \subset A(m)$, then

$$u^*(t) > u(t, m, r) \quad \text{for all } r \in \text{co}(A) \tag{1}$$

is equivalent to

$$u^*(t) > u(t, m, r) \quad \text{for all } r \in W \tag{2}$$

provided that $A \subset W \subset \text{co}(A)$.

PROPOSITION 2. *A sequential-equilibrium outcome satisfies the iterated intuitive criterion if and only if it is a fixed-equilibrium rationalizable outcome in the imperfect information game.*

Proof. We show by induction that if the outcome either passes the iterated intuitive criterion or is a FERIMO, then

$$I\mathcal{R}^k(m) = IC\mathcal{R}^k(m) \quad \text{for all } k \text{ and all } m \in M^*. \tag{3}$$

(3) holds for $k = 0$ from the definitions of $I\mathcal{R}^0(m)$ and $IC\mathcal{R}^0(m)$.

⁶ We have modified the definition of Cho and Kreps in order to guarantee that $IC\mathcal{R}^*(m)$ is nonempty. This change is not significant. If there exists a stage k at which $IC\mathcal{R}^k(m) = \emptyset$, then the outcome will always satisfy the iterated intuitive criterion with respect to m .

Assume that (3) holds for all $k = 0, 1, \dots, n - 1$. We claim that

$$\{t : s(t) = m \text{ for some } s \in \mathcal{I}^n\} = \text{IC}\mathcal{T}^n(m). \tag{4}$$

To establish (4) note that if $s(t) = m$ for some $s \in \mathcal{I}^n$, then R1' implies that

$$u^*(t) \leq u(t, m, \alpha) \quad \text{for some } \alpha \in \text{co}(\mathcal{I}^{n-1}(m)). \tag{5}$$

By (1) and (2), (5) is equivalent to

$$u^*(t) \leq u(t, m, a) \quad \text{for some } a \in \mathcal{I}^{n-1}(m) = \text{IC}\mathcal{R}^{n-1}(m). \tag{6}$$

(6) implies that $t \in \text{IC}\mathcal{T}^n(m)$. It follows that the left-hand side of (4) is contained in the right-hand-side of (4). Moreover, it follows from the induction hypothesis and the assumption that the outcome either is a FERIMO or satisfies the iterated intuitive criterion that for all $m' \in M(t) \cap M^*$,

$$u^*(t) \geq u(t, m', a'(m')) \quad \text{for some } a'(m') \in \mathcal{I}^{n-1}(m') = \text{IC}\mathcal{R}^{n-1}(m'). \tag{7}$$

Hence, if (7) holds, then $s(t) = m$ for some $s \in \mathcal{I}^n$ because the Sender can conjecture that the response to m' will be $a'(m')$ for all $m' \neq m$. Consequently, claim (4) follows from IC1.

If $\text{IC}\mathcal{T}^n(m) = \emptyset$, then, by (4), $m \notin \text{IJ}^n$ and so R2 implies that $\mathcal{I}^{n-1}(m) = \mathcal{I}^n(m)$. Also, it follows from IC2 that $\text{IC}\mathcal{R}^{n-1}(m) = \text{IC}\mathcal{R}^n(m)$. So in this case (3) holds for $k = n$. To show that (3) holds for $k = n$ when $\text{IC}\mathcal{T}^n(m) \neq \emptyset$, it suffices to show that

$$\mathcal{I}^n(m) = \text{BR}(\{t : s(t) = m \text{ for some } s \in \mathcal{I}^n\}, m). \tag{8}$$

When $\text{IC}\mathcal{T}^n(m) \neq \emptyset$, it follows from the definition of $\mathcal{I}^n(m)$ that $\mathcal{I}^n(m) = \text{BR}(S, m)$, where S is the set of probability distributions over $T(m)$ induced by strategies in \mathcal{I}^n . In symbols $S = \{\mu : \mu(t) = [\pi(t)\sigma(m|t)] / [\sum_{t' \in T(m)} \pi(t')\sigma(m|t')]\}$ for some $\sigma \in \text{co}(\mathcal{I}^n)$ such that $\sigma(m|\cdot) \neq 0$. S is contained in the set of probabilities on $\{t : s(t) = m \text{ for some } s \in \mathcal{I}^n\}$. Therefore, the left-hand side of (8) is contained in the right-hand side. (7) implies that for each t there exists a $\rho_t \in \text{co}(\mathcal{I}^{n-1}(m))$ such that $s(t) = m^*$ is a best response to ρ_t in the derived game. Consequently, \mathcal{I}^n contains all strategies of the form s_L , where

$$s_L(t) = \begin{cases} m^* & \text{if } t \notin L \\ m & \text{if } t \in L \end{cases} \quad \text{and } L \text{ is a subset of } \{t : s(t) = m \text{ for } s \in \mathcal{I}^n\}.$$

It follows that R2 implies that the left-hand side of (8) contains the right-hand side of (8).

From (4) and (8) we conclude that if the outcome satisfies the iterated intuitive criterion or is a FERIMO and $IC\mathcal{T}^n(m) \neq \emptyset$, then

$$I\mathcal{R}^n(m) = \text{BR}(IC\mathcal{T}^n(m), m) \quad \text{for all } m \in M^*. \quad (9)$$

By IC2, $IC\mathcal{R}^n(m) = \text{BR}(IC\mathcal{T}^n(m), m)$. Consequently, (9) implies that (3) is satisfied for $k = n$. It follows by induction that $IC\mathcal{R}^*(m) = I\mathcal{R}^*(m)$ for all $m \in M^*$, which suffices to complete the proof.

Cho and Kreps [5, p. 204] propose the equilibrium dominance test, a slight variation on the intuitive criterion. To pass the test of equilibrium dominance, an outcome must be a sequential equilibrium outcome with beliefs concentrated on $IC(\mathcal{T}^*(m))$ for all out-of-equilibrium signals m . Both the test of equilibrium dominance and the iterated intuitive criterion require that there be some justification for the Sender to use m^* when the Receiver believes that only types in $IC\mathcal{T}^*(m)$ would use the signal m . It is harder for an equilibrium to pass the equilibrium dominance test than the iterated intuitive criterion because the intuitive criterion allows the Sender to consider more possible responses to m (and therefore a greater possibility of finding some justification for avoiding the signal). An important difference is that equilibrium dominance requires all types of the Sender to hold a common conjecture about the response of the Receiver (see Cho and Kreps [5, p. 197]). In the next section, we view signaling games as two-player games of incomplete information. As a result, all types of Sender will hold common beliefs about the Receiver's behavior.

4. CO-DIVINITY

This section relates FEROs to divine and co-divine outcomes. First, we define divinity. Fix a sequential equilibrium to the signaling game. As usual, let $u^*(t)$ be the equilibrium expected utility of the Sender of type t and let M^* be the set of unsent signals in the equilibrium. The motivation for the concept is the following. In equilibrium, the Receiver forms a probability assessment about which types of Sender use a particular signal. For the signals that are sent with positive probability in equilibrium, Bayes' Rule determines this assessment. Divinity attempts to describe what admissible assessments are when the Receiver hears an unexpected signal. We compute these beliefs, which we denote below by $D\mathcal{B}^*(m)$, assuming that all types of the Sender have a common conjecture about how the Receiver will respond to a signal, and that the Receiver thinks that the Sender would not use an unanticipated signal unless the Sender expects to (weakly) gain utility (relative to the equilibrium payoff) by doing so.

For each behavior strategy α that the Receiver can take in response to m , let

$$P(\alpha, m) = \{t \in T(m) : u^*(t) < u(t, m, \alpha)\}$$

be the set of types that strictly prefer the response α to their equilibrium payoff,

$$I(\alpha, m) = \{t \in T(m) : u^*(t) = u(t, m, \alpha)\}$$

be the set of types for which sending m and inducing the response α yields the same payoff as the equilibrium, and, for nonempty subsets K of T , denote by π^K the conditional distribution of π given that $t \in K$. That is,

$$\pi^K(t) = \begin{cases} \pi(t) / \sum_{s \in K} \pi(s) & \text{if } t \in K \\ 0 & \text{if } t \notin K. \end{cases}$$

If $P(\alpha, m) \cup I(\alpha, m) \neq \emptyset$, then let $\bar{I}(\alpha, m) = \text{co}\{\pi^K : P(\alpha, m) \subset K \subset P(\alpha, m) \cup I(\alpha, m)\}$; otherwise, set $\bar{I}(\alpha, m) = \emptyset$.⁷ Finally, if A is a subset of $\text{MBR}(T(m), m)$, then let $\bar{I}(A, m) = \text{co}[\bigcup_{\alpha \in A} \bar{I}(\alpha, m)]$. If it were common knowledge that the Receiver would respond to the signal m with the behavior strategy α , then the set K of Sender types who would prefer to send m rather than to follow the equilibrium path would necessarily satisfy

$$P(\alpha, m) \subset K \subset P(\alpha, m) \cup I(\alpha, m). \tag{10}$$

Hence, the set $\bar{I}(\alpha, m)$ describes the beliefs that the Receiver might hold after hearing the signal m . $\bar{I}(A, m)$ describes those beliefs when the set of actions that the Receiver may take in response to m is the set A .

Now consider the iterative process that begins with the set $D\mathcal{B}^0(m)$ of all probability distributions on $T(m)$ and $D\mathcal{R}^0(m) = \text{co}(A(m))$, the set of responses the Receiver can take given m . Given $D\mathcal{B}^k(m)$ and $D\mathcal{R}^k(m)$, we define $D\mathcal{B}^{k+1}(m)$ and $D\mathcal{R}^{k+1}(m)$ by

- D1. $D\mathcal{B}^{k+1}(m) = \bar{I}(D\mathcal{R}^k(m), m)$ and
- D2. $D\mathcal{R}^{k+1}(m) = \begin{cases} \text{MBR}(D\mathcal{B}^{k+1}(m), m) & \text{if } D\mathcal{B}^{k+1}(m) \neq \emptyset \\ D\mathcal{B}^k(m) & \text{if } D\mathcal{B}^{k+1}(m) = \emptyset. \end{cases}$

We set $D\mathcal{B}^*(m) = \bigcap_{k \geq 0} D\mathcal{B}^k(m)$ and $D\mathcal{R}^*(m) = \bigcap_{k \geq 0} D\mathcal{R}^k(m)$.

DEFINITION. An equilibrium outcome is *divine* if for each $m \in M^*$ there exists $r \in D\mathcal{R}^*(m)$ such that $u^*(t) \geq u(t, m, r)$ for all $t \in T(m)$.

⁷ Banks and Sobel [1] write the definition of $\bar{I}(\alpha, m)$ in a different, but equivalent, form.

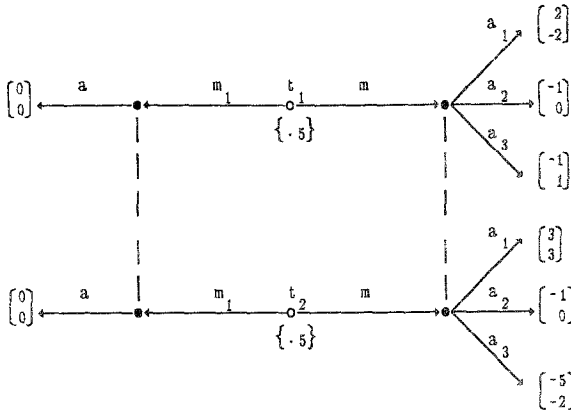


FIGURE 1

Banks and Sobel [1] show that there exists a divine equilibrium outcome in all finite signaling games.

In rationalizability the Sender forms conjectures over the set of possible pure-strategy best responses of the Receiver. Consequently the convex hull of the set of pure-strategy responses possible for the Receiver determines the strategies that the Sender may use. In contrast, divinity computes the set of allowable beliefs for the Receiver using a set of possible mixed-strategy best responses of the Receiver. Since the set of mixed strategy best responses may be strictly smaller than the convex hull of pure-strategy best responses, divinity may be a harder test to pass than fixed-equilibrium rationalizability.

Consider the example in Fig. 1,⁸ which is based on a similar example in Fudenberg and Kreps [10]. Let us test the equilibrium in which the Sender plays strategy m_1 with probability one, and the Receiver responds to m with a_2 . It is straightforward to check that $D\mathcal{B}^1(m) = D\mathcal{B}^0(m)$ is equal to the set of probability distributions on $\{t_1, t_2\}$. The mixed-strategy best-response set of the Receiver to $D\mathcal{B}^1(m)$ contains all mixtures of a_1 and a_2 , and all mixtures of a_2 and a_3 . As a result, we have $D\mathcal{B}^2(m) = \text{co}\{(\frac{1}{2}, \frac{1}{2}), (0, 1)\}$, since any action in $D\mathcal{B}^1(m)$ that yields a nonnegative utility to t_1 yields a positive utility to t_2 . Therefore, $D\mathcal{B}^2(m)$ is equal to $\{a_1\}$. Direct applications of D1 and D2 show that $D\mathcal{B}^*(m) = \{(\frac{1}{2}, \frac{1}{2})\}$ and $D\mathcal{R}^*(m) = \{a_1\}$, and therefore the outcome is not divine. However, fixed-equilibrium rationalizability deletes no strategies, and therefore the outcome is a FERRO. To see that fixed-equilibrium rationalizability deletes no strategies, argue as follows. The strategy s^* is a best response for the

⁸ In all of the examples the Sender's payoffs are written above the Receiver's.

Sender provided he conjectures that the Receiver responds to m with a_2 . The strategy $(s(t_1), s(t_2)) = (m, m)$ is an optimal response to the conjecture that the Receiver responds to m with a_1 . If the Sender conjectures that the Receiver will respond to m with equal mixtures of the pure strategies a_1 and a_3 , then he responds optimally by using m if t_1 and m^* if t_2 . Alternatively, if the Sender conjectures that the Receiver will respond to m with a $(.3, .7)$ mixture of the pure strategies a_1 and a_2 , then he responds optimally by using m^* if t_1 and m if t_2 . Consequently, $R\mathcal{S}^1 = R\mathcal{S}^0$. It is straightforward to verify that R2 deletes none of the Receiver's strategies. In particular, a_2 is a best response to the conjecture that the Sender is equally likely to play (m^*, m) and (m, m^*) .

The difference between divinity and rationalizability in the example arises because the convex hull of the set of pure-strategy best responses and the set of mixed-strategy best responses are not the same. Divinity is a refinement of sequential equilibrium obtained by imposing additional conditions on beliefs held off the equilibrium path. Hence, in a divine outcome it must be common knowledge that the Receiver responds to each out-of-equilibrium signal with a best response to some belief. When the convex hull of the set of pure-strategy best responses is strictly larger than the set of mixed-strategy best responses, fixed-equilibrium rationalizable outcomes can be supported by conjectures that do not correspond to an optimal response to beliefs. The same distinction can be made between the test of equilibrium dominance and the intuitive criterion: The equilibrium dominance test requires that the response to an unexpected signal is known to be a best response to beliefs; the intuitive criterion does not. The distinction between the sets of mixed-strategy best responses and the convex hull of pure-strategy best responses did not play a role in Section 3 because an outcome fails the iterated intuitive criterion if and only if there is $t \in T(m)$ such that $u^*(t) < u(t, m, a)$ for all pure strategy responses $a \in \text{BR}(\text{IC}\mathcal{T}^*(m), m)$. As we have noted ((1) and (2)), this condition is equivalent to $u^*(t) < u(t, m, r)$ for all $r \in \text{MBR}(\text{IC}\mathcal{T}^*(m), m)$ or for all $r \in \text{co}(\text{BR}(\text{IC}\mathcal{T}^*(m), m))$.

A minor variation of divinity does correspond to (a form of) fixed-equilibrium rationalizability.

Begin with the set $C\mathcal{B}^0(m)$ of all probability distributions on $T(m)$ and $C\mathcal{R}^0(m) = A(m)$, the set of responses the Receiver can take given m . Given $C\mathcal{B}^k(m)$ and $C\mathcal{R}^k(m)$, we define $C\mathcal{B}^{k+1}(m)$ and $C\mathcal{R}^{k+1}(m)$ by

$$\begin{aligned} \text{C1. } C\mathcal{B}^{k+1}(m) &= \bar{F}(\text{co}(C\mathcal{R}^k(m)), m) \quad \text{and} \\ \text{C2. } C\mathcal{R}^{k+1}(m) &= \begin{cases} \text{BR}(C\mathcal{B}^{k+1}(m), m) & \text{if } C\mathcal{B}^{k+1}(m) \neq \emptyset \\ C\mathcal{R}^k(m) & \text{if } C\mathcal{B}^{k+1}(m) = \emptyset. \end{cases} \end{aligned}$$

We set $C\mathcal{B}^*(m) = \bigcap_{k \geq 0} C\mathcal{B}^k(m)$ and $C\mathcal{R}^*(m) = \bigcap_{k \geq 0} C\mathcal{R}^k(m)$.

DEFINITION. An equilibrium outcome is *co-divine* if for each $m \in M^*$ there exists $\rho(m) \in \text{co}(C\mathcal{R}^*(m))$ such that $u^*(t) \geq u(t, m, \rho(m))$ for all $t \in T(m)$.⁹

Co-divinity differs from divinity in two respects. Co-divinity treats the set of possible actions of the Receiver, $C\mathcal{R}^k(m)$, as a set of pure strategies. More importantly, the set of allowable beliefs, $C\mathcal{B}^k(m)$, is derived as the set of probability assessments available to the Receiver if the Sender responds optimally to arbitrary conjectures on the (pure) strategies in $C\mathcal{B}^k(m)$, instead of the smaller set of mixed-strategy best responses.

It is a simple matter to check that $D\mathcal{R}^k(m) \subset C\mathcal{R}^k(m)$ for all $k \geq 0$. Consequently, we have the next result.

PROPOSITION 3. *Any divine outcome is co-divine.*

Next we compare FESSO to co-divinity. We denote by $S\mathcal{R}^k(m)$ and $S\mathcal{R}^*(m)$ ($S\mathcal{S}^k(m)$ and $S\mathcal{S}^*(m)$) the sets of admissible strategies for the Receiver (Sender) obtained by applying R0, R1, and R2 to $G(\{m\})$. Divinity and co-divinity differ from rationalizability because the first concepts compute the Receiver's possible responses as a set of best responses to beliefs, while rationalizability constructs $S\mathcal{R}^k(m)$ by allowing the Receiver to best respond to arbitrary conjectures over the strategies in $S\mathcal{S}^k(m)$. To facilitate a comparison between co-divinity and rationalizability, we will represent $S\mathcal{R}^k(m)$ as the set of best responses to a collection of probability distributions over $T(m)$. If $s \in S\mathcal{S}^k(m)$ and $K(s, m) = \{t : s(t) = m\}$ is the subset of types that send m when the Sender uses s , then the Receiver's set of optimal responses to s is equal to $\text{BR}(\pi^{K(s, m)}, m)$. More generally, responding optimally to a conjecture over strategies in a set $S\mathcal{S}^k(m)$ is equivalent to making an optimal response to a probability distribution in

$$\text{co}\{\pi^K : \exists s \in S\mathcal{S}^k(m) \text{ such that } K = K(s, m)\}. \tag{11}$$

PROPOSITION 4. *An equilibrium outcome is co-divine if and only if it is a fixed-equilibrium signal-by-signal rationalizable outcome.*

Proof. Both co-divinity and fixed-equilibrium signal-by-signal rationalizability treat unselected signals one at a time. Therefore, we can fix a signal $m \in M^*$ and show that the Receiver's set of admissible actions at each step of the iterative process for co-divinity agrees with the set of admissible actions for the Receiver defined for FESSO. That is, we will show that

$$C\mathcal{R}^k(m) = S\mathcal{R}^k(m) \quad \text{for all } k \geq 0. \tag{12}$$

⁹ Fudenberg and Kreps [10] independently develop and analyze a related concept.

Condition (12) holds for $k=0$ by the definition of the two sets. Assume that (12) holds for $k=0, 1, \dots, n-1$. We prove that it holds for $k=n$. Notice that it is sufficient to show that

$$\begin{aligned} & \{K \subset T(m) : K \neq \emptyset \text{ and } \exists s \in S\mathcal{S}^n(m) \text{ such that } K = K(s, m)\} \\ &= \{K \subset T(m) : K \neq \emptyset \text{ and} \\ & \quad \exists \alpha \in \text{co}(\mathcal{C}\mathcal{R}^{n-1}(m)) \text{ such that } P(\alpha, m) \subset K \subset P(\alpha, m) \cup I(\alpha, m)\}. \end{aligned} \tag{13}$$

By (11), $S\mathcal{R}^n(m)$ is the set of best responses to probability measures determined by the first set in (13). The second set in (13) determines $\mathcal{C}\mathcal{R}^n(m)$ in the same way. However, $s \in S\mathcal{S}^n(m)$ if and only if it is an optimal response to a conjecture over strategies in $S\mathcal{R}^{n-1}(m)$. Since $\mathcal{C}\mathcal{R}^{n-1}(m) = S\mathcal{R}^{n-1}(m)$ by the induction hypothesis, R1 implies that $s \in S\mathcal{S}^n(m)$ if and only if $\exists \alpha \in \text{co}(\mathcal{C}\mathcal{R}^{n-1}(m))$ such that s is a best response to α . However, s is a best response to α if and only if

$$P(\alpha, m) \subset K(s, m) \subset P(\alpha, m) \cup I(\alpha, m). \tag{14}$$

Equation (13) follows from (14). Therefore, we have established that (12) holds for $k=n$. By induction, we can conclude that

$$\mathcal{C}\mathcal{R}^*(m) = S\mathcal{R}^*(m). \tag{15}$$

In order to have a FESSO, we must have that $m^* \in S\mathcal{S}^*(m)$, or, equivalently, that for all $t \in T(m)$,

$$\exists \alpha \in \text{co}(S\mathcal{R}^*(m)) \quad \text{such that} \quad u^*(t) \geq u(t, m, \alpha). \tag{16}$$

In view of (15), (16) is equivalent to the outcome being co-divine (relative to the unsent signal m).

Co-divinity, divinity, and fixed-equilibrium, signal-by-signal rationalizability treat unreached information sets one at a time. Implicit in this approach is an assumption that whatever causes a particular “unexpected” signal to be sent is independent of the other available unsent signals. The assumption that deviations are independent across signals is restrictive, as the next example shows.

Figure 2 describes a signaling game.¹⁰ The outcome in which both types of Sender send m_3 with probability one is a sequential equilibrium outcome since the Receiver could respond to m_1 with a_3 (a_3 is a best response to $\mu(t_1 | m_1) \in [\frac{1}{3}, \frac{2}{3}]$). Similarly, the Receiver could have $\mu(t_1 | m_2) \in [\frac{1}{2}, \frac{2}{3}]$ and

¹⁰ Eric van Damme has told us of a similar example constructed by Georg Noldeke.

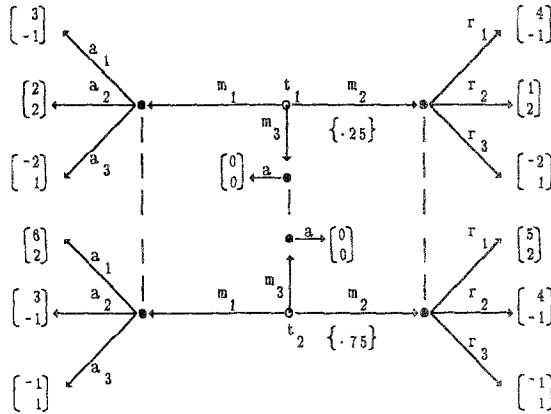


FIGURE 2

respond to m_2 with r_3 . However, the outcome fails to be co-divine. To see this, note that $u(t_2, m_2, r_j) > u(t_1, m_2, r_j)$ for $j = 1, 2$, and 3 . Therefore, t_2 prefers m_2 to m_3 whenever t_1 prefers m_2 to m_3 . Since m_2 is not dominated for either player in the first round of deletion, co-divinity requires that $\mu(t_2 | m_2) \geq \pi(t_2)$. Consequently, the Receiver must respond to m_2 with r_1 , and both types of Sender would deviate from their equilibrium strategy. Moreover, a similar analysis applies to the out-of-equilibrium play of m_1 , giving us an alternative basis for co-divinity to reject the outcome.

The argument for rejecting the equilibrium depends upon testing the play of one out-of-equilibrium signal at a time while ignoring the existence of other unsent signals. In particular, the test we described for co-divinity (and FESSO) looked at the play of m_2 by (at least) one type of Sender while ignoring the possibility that another type of Sender may be playing m_1 with positive probability. We can use the example to demonstrate that requiring the players to treat deviations one signal at a time may restrict equilibria in an unreasonable way.

Consider the game derived from the equilibrium in which the Sender always uses m_3 . The strategies $(s(t_1), s(t_2)) = (m_1, m^*)$ are strictly dominated for $i = 1$ and 2 (for example by a $(\frac{2}{3}, \frac{1}{3})$ mixture of (m^*, m^*) and (m^*, m_i)) because any conjecture that makes m^* optimal for t_2 must make m^* uniquely optimal for t_1 , but no other strategies for the Sender are dominated in the derived game. Furthermore, when only (m_1, m^*) and (m_2, m^*) are deleted from the Sender's strategy set, all of the Receiver's strategies in the derived game remain undominated. Consequently, the outcome is a FERRO even though it fails the test of fixed-equilibrium rationalizability when either message is taken alone.

Co-divinity does not allow the Receiver to believe that t_1 is likely to send

m_2 . However, the discussion above provides a context in which there is some justification for this belief. If the Sender conjectures that a_1 is the response to m_1 and r_1 is the response to m_2 , then t_1 would send m_2 and t_2 would send m_1 . Hence, t_2 can use the signal m_1 ; t_2 's best response to a conjecture about the Receiver's strategy need not be m_2 whenever t_1 's best response to the same conjecture is m_2 . When all unsent signals are taken into account, it may be inappropriate to conclude that " t_1 is less likely than t_2 to defect to m_2 ." The same type of argument demonstrates that we cannot conclude that " t_1 is less likely than t_2 to defect to m_2 " even though co-divinity requires that $\mu(t_2|m_1) \geq .75$.

The example demonstrates that a FERRO need not be a FESSO. The next proposition shows that the set of FERROs is always as large as the set of FESSOs.

PROPOSITION 5. *Any FESSO is a FERRO.*

A proof of Proposition 5 is in the appendix. Here is an informal argument. The strategies of a signal-by-signal (FESSO) game are a subset of the strategies for a multiple-signal (FERRO) derived game: Any strategy for the multiple-signal game in which the Sender uses only one signal other than m^* with positive probability is also a strategy for a signal-by-signal game. We can show that for any FESSO, the set of strategies for the Sender that remain after k iterations of R0, R1, and R2 for the signal-by-signal game is always contained in the set of strategies that remain for the multiple-signal game. Therefore, if the strategy of sending only m^* is rationalizable in all of the signal-by-signal games, then it is rationalizable in the multiple-signal game. To see this, suppose that the strategy s remains after k iterations in the signal-by-signal game. If $s(t) \in \{m^*, m'\}$ for all t , and $s(t)$ is also an optimal response to a conjecture $\alpha(m')$, then let the Sender conjecture that the Receiver will make a response $\alpha(m)$ that satisfies $u^*(t) \geq u(t, m, \alpha(m))$ for all t and for all $m \in M^* \setminus \{m'\}$. If the Sender's conjectures over all $m \in M^*$ are given by $\alpha(\cdot)$, then s is an optimal response in the multiple-signal game. The only detail missing in this argument is a verification that the Sender's conjectures are made over strategies that remain in the Receiver's possible response set at the k th step of the iterative process. The verification is straightforward provided that for each $m \in M^*$ there remains a strategy $s \in S^k(m)$ such that $s(t) = m$ for some $t \in T(m)$. In this case one can show that any behavior strategy of the Receiver that the multiple-signal procedure deletes must also be deleted by the signal-by-signal procedure. Therefore, any conjecture feasible for the signal-by-signal procedure is feasible for the multiple-signal procedure also. The argument is more complicated if there is a step in the signal-by-signal procedure at which all strategies that send $m \in M^*$ are deleted. At that point the signal-

by-signal procedure does not delete any more of the Receiver's responses to m . The multiple-signal procedure may continue to delete responses to m . To show that a FESSO is a FERRO we must show that the multiple-signal procedure does not delete all actions that are worse than m^* for some Sender type. The appendix contains the details of an argument that shows there always remains a strategy for the Receiver that yields a payoff that no Sender prefers to the fixed-equilibrium payoff.

Banks and Sobel [1] demonstrate that all finite signaling games have at least one divine outcome and that divine outcomes survive the equilibrium dominance and iterated intuitive criteria. Therefore, the first five propositions imply the following existence result.

COROLLARY. Every finite signaling game has a FESSO, a FERRO, and a FERIMO.

We can now position our concepts within the current refinement hierarchy. Propositions 1 and 2 demonstrate that the set of FERIMOs is precisely the outcomes that satisfy the iterated intuitive criterion, and contains the set of FERROs. In turn, the set of FERROs contains the FESSOs by Proposition 5. FESSOs coincide with the set of co-divine outcomes (Proposition 4). We cannot place outcomes that survive the equilibrium dominance test inside this hierarchy. Although any divine outcome passes the equilibrium dominance test and anything that passes equilibrium dominance is a FERIMO, there is no general relationship between FERROs or FESSOs and the outcomes that pass equilibrium dominance. The test of equilibrium dominance can be harder to pass than FERRO or FESSO because it requires all types of the Sender to conjecture that the Receiver choose an element in the set of mixed strategy best responses rather than in the possibly larger convex hull of pure strategy best responses. The equilibrium dominance test may be easier to pass because it does not require all Sender types to hold the same beliefs over the Receiver's response. It is not difficult to construct examples in which an outcome that passes the test of equilibrium dominance is not a FERRO (and hence not a FESSO), and examples in which an outcome that is a FESSO (and hence a FERRO) fails to pass the equilibrium dominance test.

The choice of which fixed-equilibrium refinement to select (FERIMO, FERRO, or FESSO) is determined by an evaluation of whether the economic environment is best modeled as a $(T+1)$ -player game of imperfect information or a two-player game of incomplete information, and ranges from the iterated intuitive criterion to co-divinity. If one is troubled by tests that treat unreached information sets independently, then co-divinity may be too strong.

5. PERFECT SEQUENTIAL EQUILIBRIA

Thus far we have focused on the set of outcomes to signaling games that are rationalizable in a derived game. One could also ask what would happen if we required our outcomes to satisfy another game-theoretical restriction in the derived game. A natural approach is to replace rationalizability with Nash Equilibrium. In this section, we discuss one way to do this.

One possible question to ask is: Under what conditions will s^* be a Nash Equilibrium strategy for the Sender in the derived game? This is not an interesting question because (s^*, ρ) is always a Nash Equilibrium strategy profile for the derived game $G(\sigma, \rho, M_0)$ for $M_0 \subset M^*$. Next, we could ask whether s^* is still a Nash Equilibrium strategy after we have deleted strategies from the derived game using one of the procedures introduced earlier. For example, we could find conditions under which s^* is a Nash Equilibrium strategy for the Sender when strategies outside of $R\mathcal{S}^*$ and $R\mathcal{R}^*$ are deleted. Here the requirement that s^* be a Nash Equilibrium strategy is equivalent to the outcome being fixed-equilibrium rationalizable. This follows because if $s^* \in R\mathcal{S}^*$, then there exists a conjecture ρ^* over $R\mathcal{R}^*$ such that s^* is an optimal response to ρ^* . Since all strategies in $R\mathcal{R}^*$ are optimal responses to s^* , (s^*, ρ^*) is a Nash Equilibrium.

One other interpretation of the derived game leads to a refinement of Grossman and Perry's [11] (see also Farrell [9] who introduced a related concept) idea of Perfect Sequential Equilibrium (PSE). In this section we introduce PSE and PSE*, a refinement of PSE due to van Damme [7]. We then provide an interpretation of the derived game that provides a characterization of PSE*. Finally, we use the derived game to interpret the solution concepts.

We now define PSE. Fix a sequential equilibrium (σ, ρ) to the original signaling game. Let $m \in M^*$ be a signal that is not used by the Sender in equilibrium. Recall from Section 4 that for any behavior strategy α that the Receiver may take in response to the signal m , $\bar{I}(\alpha, m)$ is the set of beliefs that the Receiver may hold if the Sender of type t chooses between $u^*(t)$ and $u(t, m, \alpha)$. Grossman and Perry call the belief μ consistent if there exists a behavior strategy $\alpha \in \text{MBR}(\mu, m)$ such that $\mu \in \bar{I}(\alpha, m)$. Let $\text{CB}(m)$ be the set of consistent beliefs given the signal m . A Perfect Sequential Equilibrium is a sequential equilibrium in which off-the-equilibrium path beliefs are consistent whenever possible. Hence, the Nash Equilibrium (σ, ρ) gives rise to a PSE outcome if and only if for all $m \in M^*$ such that $\text{CB}(m) \neq \emptyset$, there exists $\mu \in \text{CB}(m)$ and $\alpha \in \text{MBR}(\mu, m)$ such that $u^*(t) \geq u(t, m, \alpha)$ for all $t \in T(m)$. Van Damme [7] introduces a refinement of PSE, which he calls PSE*. Van Damme calls a behavior strategy α consistent with an equilibrium given the signal m if there exists $\mu \in \bar{I}(\alpha, m)$

such that $\alpha \in \text{MBR}(\mu, m)$. Denote by $\text{CBS}(m)$ the set of consistent behavior strategies given m . A sequential equilibrium is a PSE* if and only if the off-the-equilibrium actions are consistent whenever possible. Hence, the Nash Equilibrium (σ, ρ) gives rise to a PSE* outcome if and only if for all $m \in M^*$ such that $\text{CBS}(m) \neq \emptyset$, there exists $\alpha \in \text{CBS}(m)$ such that $u^*(t) \geq u(t, m, \alpha)$ for all $t \in T(m)$. Note that for any consistent behavior strategy of the Receiver there exists a consistent belief. Hence any PSE* outcome is a PSE outcome. However, there may be a behavior strategy α' that is an optimal response to a consistent belief, but fails to be a consistent behavior strategy. So the PSE* test is logically more difficult to pass than the PSE test. Indeed, van Damme shows by an example that there exist perfect sequential equilibria that fail to be PSE*.

PSE* outcomes can be described using derived games. Delete the pure strategy s^* in which all types of the informed player use m^* from the game $G(\sigma, \rho, \{m\})$. For notational convenience we do not refer to (σ, ρ) , and we call this game $G'(\{m\})$. In $G'(\{m\})$ at least one type of Sender uses m with positive probability. Consider the set $E(m)$ of Nash Equilibria (σ', α') to $G'(\{m\})$. Because we have deleted the pure strategy s^* , but allow the Sender to use s^* with positive probability, $E(m)$ may be empty. However, for all $(\sigma', \alpha') \in E(m)$,

$$u^*(t) \leq u(t, \sigma'(t), \alpha') \quad \text{for all } t. \tag{17}$$

If (17) failed to hold for some type t' , then t' could increase his payoff by increasing the probability he sends m^* (provided that he did not increase the probability to one).

PROPOSITION 6. *A sequential-equilibrium strategy profile (σ, ρ) determines a PSE* outcome if and only if for all $m \in M^*$, either $E(m) = \emptyset$ or there exists $(\sigma', \alpha') \in E(m)$ such that*

$$u^*(t) \geq u(t, m, \alpha') \quad \text{for all } t \in T(m). \tag{18}$$

In view of (17), (18) is equivalent to $u^*(t) = u(t, \sigma'(t), \alpha')$ for all t .

To prove the proposition, one need only check that α is a consistent behavior strategy if and only if there exists a σ' such that $(\sigma', \alpha) \in E(m)$. Consequently $E(m) = \emptyset$ if and only if $\text{CBS}(m) = \emptyset$. We omit the straightforward verification.

While Farrell, Grossman and Perry, and van Damme provide arguments to motivate PSE, Proposition 6 provides a different perspective. When the Receiver hears the unexpected signal m , he could act as if he is playing the game $G'(\{m\})$, since every type of Sender could not have used the strategy s^* with probability one in this case. If the Sender prefers to send the signal m rather than to play as specified by the equilibrium, then it is plausible

to assume that he expects to gain (weakly) relative to the equilibrium. According to Proposition 6, PSE* requires that the players coordinate on a Nash Equilibrium of $G'(\{m\})$ that is at least as good as the original equilibrium for each type. (If no such an equilibrium exists, then the outcome is a PSE*.) Consequently, an outcome fails to be a PSE* if every Nash Equilibrium of $G'(\{m\})$ is better for at least one Sender type than the original outcome.

Proposition 6 suggests that PSE* requires a higher level of coordination than does co-divinity. Both concepts depend on the assumption that unexpected signals are treated independently, and that the Sender only sends them if he does not expect to lose relative to his equilibrium expected utility. However, beyond these restrictions, co-divinity demands only that there be a rationalizable outcome of $G'(\{m\})$ that is no better than the equilibrium outcome for each type of Sender; PSE* goes further by asking that there be a Nash Equilibrium outcome of $G'(\{m\})$ that is no better than the equilibrium outcome for each type of Sender. The additional restriction helps to explain why there exist signaling games with no PSE* (Grossman and Perry [11, p. 112] give an example, due to Joseph Farrell and Eric Maskin, of a signaling game without a PSE).

6. AN ALTERNATE INTERPRETATION OF THE DERIVED GAME

This paper studies equilibrium refinements for signaling games by examining the outcomes that survive iterated deletion of weakly dominated strategies in an auxiliary game. Ben-Porath and Dekel [2] and van Damme [8] do a similar exercise. These two papers add to a given game a stage in which one of the players may publicly burn money before they play the original game. This strategy lowers the burner's payoffs uniformly. When weakly dominated strategies are iteratively deleted (or the more restrictive strategic stability of Kohlberg and Mertens [15] is applied), the new game may have a unique outcome, which is an equilibrium outcome of the original game, even if the original game has multiple equilibria.¹¹ The framework of these papers is quite different from ours. Not only do they treat a different class of games, but the derived games of Ben-Porath and Dekel and van Damme can be defined without reference to an equilibrium outcome in the original game.

¹¹ Ben-Porath and Dekel study finite games which contain an outcome that every player prefers to all other outcomes. For this class of games, they show that only the preferred outcome of the original game survives deletion of weakly dominated strategies in an auxiliary game in which one player has the option to burn money (in sufficiently small denominations). No money is burned in the undominated outcome of the auxiliary game.

A reinterpretation of our auxiliary game brings it closer to these papers. View a signaling game as a decision problem for the Sender. Let the subjective expected utility of the Sender of type t be $u^*(t)$. Now consider an auxiliary game, call it G'' , in which a Sender of type t may send the signal m^* in addition to the strategies in M . If he uses m^* , then he receives $u^*(t)$; otherwise, G'' is no different from the original game. Hence, in the auxiliary game the Sender has an opportunity to collect with certainty (what he claims is) the value of the original game. Under what conditions is the value of the auxiliary game equal to the value of the original game? If one requires solutions to survive the iterative deletion procedures that we have discussed, then our paper comes close to answering this question.

This game differs from the derived games considered in the paper because the signal m^* does not replace the equilibrium path. However, the concepts are related. Let $u^*(t)$, $t = 1, 2, \dots, T$, be expected utilities from a sequential equilibrium to the original game. It is easy to check that m^* is a signal-by-signal rationalizable strategy for the corresponding auxiliary game G'' if and only if the original equilibrium outcome is a FESSO. Also, if the original equilibrium outcome is a FERRO, then m^* is a rationalizable strategy in G'' . The converse is not true, as the next example demonstrates. (See Fig. 3).

Consider the sequential equilibrium outcome in which the Sender sends m_1 with probability one and the Receiver responds to m_1 by choosing a_2 with probability one. This outcome can be supported as a sequential equilibrium provided that $\mu(t_1 | m_2) \leq \frac{1}{3}$, so that the Receiver can respond to m_2 by playing r_2 with enough probability to discourage the Sender from using m_2 . However, since $u(t_1, m_2, r_1) > u(t_2, m_2, r_1)$, $u(t_1, m_2, r_2) = u(t_2, m_2, r_2)$, and $u^*(t_1) = u^*(t_2) = 0$, co-divinity requires that $\mu(t_1 | m_2) \geq \frac{1}{2}$. Therefore,

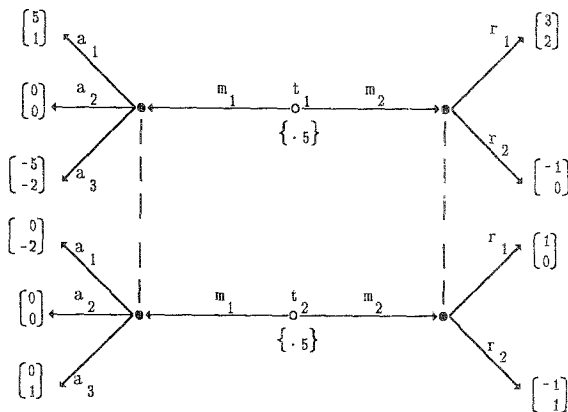


FIGURE 3

the outcome is not co-divine. (Equivalently, note that the outcome is not a FERO since the strategy $(s(t_1), s(t_2)) = (m^*, m_2)$ is deleted in the first iteration.) Nevertheless, if the strategy m^* were added to the game without deleting the signal m_1 , we could delete only the strategy (m^*, m_2) of the Sender. In particular, the strategy (m_1, m_2) would survive because the Sender could conjecture that the Receiver would respond to m_1 with a_1 and respond to m_2 with r_1 . Consequently, the set of values that are rationalizable in G^n is strictly larger than the set of values that can be obtained through FEROS.

APPENDIX

This appendix contains a proof of Proposition 5. We start with some definitions.

Define RJ^k and SJ^k by $RJ^k = \{m : \exists s \in R\mathcal{S}^k \text{ such that } s(t) = m \text{ for some } t \in T(m)\}$ and $SJ^k = \{m : \exists s \in S\mathcal{S}^k(m) \text{ such that } s(t) = m \text{ for some } t \in T(m)\}$. RJ^k and SJ^k are the sets of signals that are reached by strategies in $R\mathcal{S}^k$ and $S\mathcal{S}^k$; we say that a conjecture σ on the Sender's strategy set reaches m if there exists $t \in T(m)$ such that $\sigma(m|t) > 0$.

Next, we define a procedure that iteratively defines the artificial sets $A\mathcal{S}^k(m)$ and $A\mathcal{R}^k(m)$. Set $A\mathcal{S}^0(m) = S\mathcal{S}^0(m)$ and $A\mathcal{R}^0(m) = S\mathcal{R}^0(m)$ and let AJ^k , $A\mathcal{S}^{k+1}(m)$, and $A\mathcal{R}^{k+1}(m)$ be determined from $A\mathcal{S}^k(m)$ and $A\mathcal{R}^k(m)$ using the following steps.

A0. $AJ^k = \{m \in M^* : \exists s \in A\mathcal{S}^k(m) \text{ such that } s(t) = m \text{ for some } t \in T(m)\}$.

A1. $A\mathcal{S}^{k+1}(m) = \{s : \exists \alpha \in \text{co}(A\mathcal{R}^k(m)) \text{ such that } s \text{ is a best response to } \alpha\}$.

A2. $A\mathcal{R}^{k+1}(m) = \{a \in A\mathcal{R}^k(m) : \exists \sigma \in \text{co}(A\mathcal{S}^{k+1}(m)) \text{ such that } \sigma(m|t) > 0 \text{ for some } t, \text{ and } a \text{ is a best response to } \sigma\}$, if $m \in AJ^{k+1}$ and

$A\mathcal{R}^{k+1}(m) = \{a \in \Lambda(m) : \exists r \in R\mathcal{R}^{k+1} \text{ such that } r(m) = a\}$ if $m \notin AJ^{k+1}$.

The artificial sets differ from the rationalizable signal-by-signal sets only in the way in which $A\mathcal{R}^k(m)$ is defined when no strategy in $A\mathcal{S}^k(m)$ reaches m . In this case, the set of admissible strategies for the Receiver is defined to be the projection of $R\mathcal{R}^k$ onto the signal m . So if there is ever a step when the signal-by-signal iterative process fails to reach m , then we artificially enlarge the set of allowable responses to m to include any action permitted by the fixed-equilibrium deletion procedure. We do not require that $A\mathcal{S}^{k+1}(m) \subset A\mathcal{S}^k(m)$. This containment need not hold if $A\mathcal{R}^k(m)$ is not equal to $S\mathcal{R}^k(m)$.

We show in the proof of Proposition 5 that if $m \in \text{SJ}^k$, then $\text{S}\mathcal{R}^k(m) \subset \text{R}\mathcal{R}^k(m)$. If $m \notin \text{SJ}^k$, then we use artificial sets to establish the result.

Proof of Proposition 5. Note that the strategies for the Sender in the signal-by-signal game are naturally included in the Sender's strategy set for $G(M^*)$. We will view $\text{A}\mathcal{S}^k(m)$ as a subset of the Sender's strategy space for $G(M^*)$.

In order to prove the proposition we will use induction to show that if we start with a FESSO, then for all $k \geq 0$

$$s^* \in \text{A}\mathcal{S}^k(m) \quad \text{for all } m \in M^*, \tag{19}$$

$$\text{A}\mathcal{S}^k(m) \subset \text{R}\mathcal{S}^k \quad \text{for all } m \in M^*, \quad \text{and} \tag{20}$$

$$\prod_{m \in M^*} \text{A}\mathcal{R}^k(m) \subset \text{R}\mathcal{R}^k. \tag{21}$$

Recall that s^* is the strategy of the Sender for which $s^*(t) = m^*$ for all t . (19), (20), and (21) hold for $k = 0$ from the definitions of the sets involved. Assume that (19), (20), and (21) hold for $k = 0, 1, \dots, n - 1$. We claim that s^* is an element of $\text{A}\mathcal{S}^n(m)$ for all $m \in M^*$. If $\text{A}\mathcal{S}^n(m) = \text{S}\mathcal{S}^n(m)$, then the claim follows from the definition of a FESSO. Otherwise, $m \notin \text{SJ}^k$ for some $k < n$ and therefore $\text{S}\mathcal{S}^*(m) = \{s^*\}$. Let h be the largest value of $k < n$ for which $m \notin \text{AJ}^h$. Such an h exists because $m \notin \text{SJ}^k$ for some $k < n$. By the induction hypothesis, s^* is an element of $\text{A}\mathcal{S}^h(m)$. Now construct the sets $\tilde{\text{A}}\mathcal{S}^k(m)$, $\tilde{\text{A}}\mathcal{R}^k(m)$, $\tilde{\text{A}}\mathcal{S}^*(m)$, and $\tilde{\text{A}}\mathcal{R}^*(m)$ for $k \geq h$ using R1 and R2 starting with $\tilde{\text{A}}\mathcal{S}^h(m) = \text{A}\mathcal{S}^h(m)$ and $\tilde{\text{A}}\mathcal{R}^h(m) = \text{A}\mathcal{R}^h(m)$. By the definition of h and A1 and A2, we have $\tilde{\text{A}}\mathcal{S}^k(m) = \text{A}\mathcal{S}^k(m)$ and $\tilde{\text{A}}\mathcal{R}^k(m) = \text{A}\mathcal{R}^k(m)$ for $h \leq k \leq n$. Therefore, $s^* \in \tilde{\text{A}}\mathcal{S}^n(m)$ implies that $s^* \in \text{A}\mathcal{S}^n(m)$. To prove (19) for $k = n$, it suffices to show that $s^* \in \tilde{\text{A}}\mathcal{S}^*(m)$. In order to prove this, we show that

$$\text{if } s^* \notin \tilde{\text{A}}\mathcal{S}^*(m), \quad \text{then } \tilde{\text{A}}\mathcal{S}^*(m) \subset \text{S}\mathcal{S}^*(m). \tag{22}$$

Since $\tilde{\text{A}}\mathcal{S}^*(m) \neq \emptyset$, (22) contradicts $\{s^*\} = \text{S}\mathcal{S}^*(m)$ and establishes (19). We prove (22) by induction. Plainly,

$$\tilde{\text{A}}\mathcal{S}^*(m) \subset \text{S}\mathcal{S}^k(m) \quad \text{and} \tag{23}$$

$$\tilde{\text{A}}\mathcal{R}^*(m) \subset \text{S}\mathcal{R}^k(m) \tag{24}$$

hold when $k = 0$. If (24) holds when $k = j - 1$, then (23) holds for $k = j$ by R1, since every element in $\tilde{\text{A}}\mathcal{S}^*(m)$ responds optimally to a conjecture over strategies in $\tilde{\text{A}}\mathcal{R}^*(m)$, and hence responds optimally to a conjecture over strategies in the larger set $\text{S}\mathcal{R}^{j-1}(m)$. Similarly, if (23) holds for $k = j$, then (24) holds for $k = j$. This claim follows from R2 because if

$s^* \notin \tilde{A}\mathcal{S}^*(m)$, then every strategy in $\tilde{A}\mathcal{S}^*(m)$ reaches m . Consequently every strategy in $\tilde{A}\mathcal{R}^*(m)$ responds optimally to a conjecture over strategies in $\tilde{A}\mathcal{S}^*(m)$ that reach m . If (23) holds for $j = k$, then this conjecture is also over strategies in $S\mathcal{S}^k(m)$. Hence (24) must hold for $j = k$. Therefore, (22) follows by induction, and by our remarks, (19) holds when $k = n$.

Next we show that (20) holds when $k = n$. If $s \in A\mathcal{S}^n(m')$, then

$$s \text{ is an optimal response to } \alpha \in \text{co}(A\mathcal{R}^{n-1}(m')). \tag{25}$$

It follows from (19) that for all $m \in M^*$, there exist conjectures $\rho^*(m) \in \text{co}(A\mathcal{R}^{n-1}(m))$ such that

$$u^*(t) \geq u(t, m, \rho^*(m)) \quad \text{for all } t \in T(m). \tag{26}$$

Therefore, (21) implies that for $k = n - 1$ the conjecture $\tilde{\rho}$ defined by

$$\tilde{\rho}(m) = \begin{cases} \rho^*(m) & \text{if } m \in M^* \setminus \{m'\} \\ \alpha & \text{if } m = m' \end{cases}$$

is an element of $\text{co}(R\mathcal{R}^{n-1})$. Together, (25) and (26) imply that s is an optimal response to the conjecture $\tilde{\rho}$ in the game $G(M^*)$. Hence (20) holds for $k = n$.

We must also show that (21) holds for $k = n$. It suffices to show, for all $m \in RJ^n$,

if $a(m) \in A\mathcal{R}^n(m)$, then $\exists \sigma_m \in \text{co}(R\mathcal{S}^n)$ such that

$$\sigma_m \text{ reaches } m \text{ and } a(m) \text{ is an optimal response to } \sigma_m. \tag{27}$$

If $m \in AJ^n$, then there exists $\sigma_m \in \text{co}(A\mathcal{S}^n(m))$ such that σ_m reaches m and $a(m)$ is an optimal response to σ_m . (27) now follows from (20). If $m \notin AJ^n$, then (27) is an immediate consequence of A2.

We have shown that (19), (20), and (21) hold for all $k \geq 0$. Therefore, $s^* \in R\mathcal{S}^*$ and the proof is complete.

Note added in proof. After completing this paper we learned that Peter DeMarzo has applied similar techniques to study the relationship between Perfect Sequential Equilibria and Divine Equilibria.

REFERENCES

1. J. S. BANKS AND J. SOBEL, Equilibrium selection in signaling games, *Econometrica* **55** (1987), 647–661.
2. E. BEN-PORATH AND E. DEKEL, “Coordination and the Potential for Self Sacrifice,” Stanford GSB Discussion Paper, 1988.

3. B. D. BERNHEIM, Rationalizable strategic behavior, *Econometrica* **52** (1984), 1007–1028.
4. I.-K. CHO, A refinement of sequential equilibrium, *Econometrica* **55** (1987), 1367–1389.
5. I.-K. CHO AND D. M. KREPS, Signaling games and stable equilibria, *Quart. J. Econ.* **102** (1987), 179–221.
6. I.-K. CHO AND J. SOBEL, Strategic stability and uniqueness in signaling games, *J. Econ. Theory* **50** (1990), 381–413.
7. E. v. DAMME, “Stability and Perfection of Nash Equilibria,” Springer-Verlag, Berlin/New York, 1987.
8. E. v. DAMME, Stable equilibria and forward induction, *J. Econ. Theory* **49** (1989), 476–496.
9. J. FARRELL, Meaning and credibility in cheap-talk games, in “Mathematical Models in Economics” (M. Demster, Ed.), Oxford Univ. Press, Oxford, in press.
10. D. FUDENBERG AND D. M. KREPS, A theory of learning, experimentation, and equilibrium in games, mimeo, Massachusetts Institute of Technology, 1988.
11. S. J. GROSSMAN AND M. PERRY, Perfect sequential equilibrium, *J. Econ. Theory* **39** (1986), 97–119.
12. S. J. GROSSMAN AND M. PERRY, Sequential bargaining under asymmetric information, *J. Econ. Theory* **39** (1986), 120–154.
13. E. KOHLBERG AND J.-F. MERTENS, On the strategic stability of equilibria, *Econometrica* **54** (1986), 1003–1037.
14. D. M. KREPS AND R. WILSON, Sequential equilibrium, *Econometrica* **50** (1982), 863–894.
15. D. PEARCE, Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52** (1984), 1029–1041.