COMMUNICATION, COMPUTABILITY AND COMMON INTEREST GAMES*

LUCA ANDERLINI
(St. John's College, Cambridge)

October, 1995

ABSTRACT

This paper provides a theory of equilibrium selection for one-shot two-player finite-action strategic-form Common Interest games. A single round of costless unlimited pre-play communication is allowed. Players are restricted to use strategies which are *computable* in the sense of *Church's Thesis.* The equilibrium notion used involves perturbations which are themselves computable. The only equilibrium payoff vector which survives these strategic restrictions and the computable perturbations is the unique Pareto-efficient one.

JEL CLASSIFICATION: C72,C79.

KEYWORDS: Common Interest Games, Computability, Pre-play Communication.

PROPOSED RUNNING HEAD: COMMUNICATION IN COMMON INTEREST GAMES.

ADDRESS FOR CORRESPONDENCE: Luca Anderlini, University of Cambridge, Faculty of Economics & Politics, Sidgwick Avenue, Cambridge CB3 9DD, U.K. (e-mail la13@econ.cam.ac.uk).

---

* The initial research work for this paper was carried out while the author was visiting the Santa Fe Institute for Studies in the Sciences of Complexity. Earlier drafts were written during a one-year visit to the Department of Economics of Harvard University. The generous hospitality of both these institutions is gratefully acknowledged. I would also like to thank Robert Evans for pointing out a mistake in one of the previous formulations of these results, and Hamid Sabourian for his cooperation in a joint research effort closely related to this paper. I am grateful to two anonymous referees whose comments prompted this extensive revision of the paper which was prepared during a one-year visit to the Cowles Foundation for Research in Economics at Yale. Their generous support is also gratefully acknowledged. All remaining errors are my own responsibility.

1. Introduction

1.1. Motivation

Consider a pure coordination game like the following

	α_2	β_2
α_1	$3, 3$	$0, 0$
β_1	$0, 0$	$2, 2$

Figure 1

where player 1 chooses rows and player 2 columns. This game has two pure-strategy Nash equilibria: (α_1, α_2) and (β_1, β_2). Moreover, both equilibria survive just about any refinement of Nash equilibrium which has been put forward in the literature.[1] There is, however, a wide consensus that the Pareto-inferior equilibrium (β_1, β_2) is in some sense 'unlikely' to prevail. An informal argument often put forward in order to rule out the (β_1, β_2) equilibrium runs as follows. The pair (α_1, α_2) makes both players better off relative to (β_1, β_2), and it is self-enforcing (it is a Nash equilibrium). Hence 'rational' players will find a way to coordinate on (α_1, α_2) rather than (β_1, β_2). Very often, implicit in lines of reasoning like the above is that the natural 'way to coordinate' is pre-play communication. The trouble with modelling pre-play communication is an obvious one. Imagine a two-stage set-up in which players are first allowed to 'communicate' and then are asked to play the coordination game above. If it is the case that pre-play communication does not affect players' payoffs, then there is always a subgame-perfect equilibrium of the two-stage game (pre-play communication and play) in which players simply ignore whatever happened at the communication stage and play the 'wrong' equilibrium (β_1, β_2). The main aim of this paper is to propose a framework in which one round of unlimited, costless pre-play communication is effective in selecting the unique Pareto-efficient payoffs in the class of one-shot, two-player, finite-action, strategic (normal)-form games known as Common Interest games.

---

[1] But see our discussion of 'Cheap Talk' in Section 7.4 below.

1.2. Intuition

It is difficult to see how pre-play communication can matter in a game like the one in Figure 1 unless some 'noise' is introduced in the model. Without noise (or perturbations), any pure-strategy equilibrium will prescribe *only one* pair of messages to be sent by the players, and in the second stage an equilibrium of the underlying game to be played. If a message different from the one prescribed by the equilibrium strategies is observed by one of the players, the posterior probability of possible actions of the opponent is not defined by Bayes' rule. In other words, the interpretation of messages 'off the equilibrium path' is in the realm of the theory of refinements of the Nash equilibrium concept. A model of pre-play communication grounded in standard probability theory must allow for a non-trivial set of messages to be sent in equilibrium.

Introducing perturbations in the choice of strategies in a game as in Figure 1 with pre-play communication is *not* by itself sufficient to select the Pareto-efficient outcome, however. In a sense this is quite obvious, but dwelling on the point will help to see the intuition behind our main result.

It is useful to set up some notation which will be used for the rest of the paper. Let $\mathcal{M}$ be the message space of both players.[2] Player i's action set (which in the case of Figure 1 simply consists of $\{\alpha_i, \beta_i\}$) is $\mathcal{A}_i$, and $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2$. A strategy s_i for player i in the game with pre-play communication therefore consists of a pair $\{m_i \in \mathcal{M}\ ,\ \mu_i : \mathcal{M} \to \mathcal{A}_i\}$ — a message m_i and a map μ_i yielding an action for any possible message of the opposing player.[3] Let $\mathcal{S}_i$ be the set of strategies available to player i and $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$. The payoff to player i when the pair $s \in \mathcal{S}$ is played is denoted by $\Pi_i(s)$.

'Perturbations' are simply probability distributions P_i over $\mathcal{S}_i$. Consider now the set of strategies for player i which are a best response to player j playing strategy s_j

---

[2] What precisely the message space consists of is left deliberately vague. Of course, when the model is described in full in Section 3 this will be made precise. For the time being, we can think of $\mathcal{M}$ as any discrete set consisting of more that one element.

[3] We take the action of i in the second stage to be dependent only on j's message and not (as would be standard) on i's own message as well purely for notational simplicity. In Anderlini (1990) we allow the dependence of i's action on both messages and derive the same results as in this paper.

with probability $1-\varepsilon$ and playing according to the distribution P_j with probability ε. Formally

$$\mathcal{B}_i(s_j,\varepsilon,P_j) \equiv \arg\max_{s_i \in S_i} (1-\varepsilon)\Pi_i(s_i,s_j) + \varepsilon E_{P_j}\left[\Pi_i(s_i,s'_j)\right] \tag{1}$$

where $E_{P_j}[\cdot]$ denotes the expectation operator with respect to the distribution P_j over s'_j. Using Bayes' rule, we can also define the set of best response actions for player i given s_j, ε and P_j, conditional on the fact that a particular message m_j has been observed. Notice that since the action in G which each strategy for j yields in general depends on i's message m_i, the best response set depends on m_i as well as on m_j. Let this set be denoted by[4]

$$\mathcal{B}_i(s_j,\varepsilon,P_j|m_j,m_i) \tag{2}$$

A simple way to see why perturbations are not sufficient to rule out the inefficient equilibrium (β_1,β_2) in the game with pre-play communication is the following. It is clear that in general there exist values of (s_j,P_j) which induce the best response correspondence in (2) to be β_i *whatever* the message $m_j \in \mathcal{M}$ and for any value of ε. A formal way to put this which will become useful in a moment is to say that $\forall\ i,j=1,2$ and $j \neq i$

$$\exists(s_j,P_j) \text{ such that } \mathcal{B}_i(s_j,\varepsilon,P_j|m_j,m_i) \neq \alpha_i \quad \forall m_j \in \mathcal{M},\ \forall m_i \in \mathcal{M},\ \forall \varepsilon > 0 \tag{3}$$

From (3) it easily follows that there are Trembling-Hand Perfect (Selten 1975, Myerson 1978) equilibria of the game with pre-play communication in which the players play (β_1,β_2). Intuitively, pre-play communication is worthless in selecting the efficient equilibrium. This is because s_j and P_j may be such that after having observed any message of the opposing player it is still the case that action β_j has sufficiently higher probability than action α_j.

Suppose now that by careful choice of *message* and *strategy* spaces as well as of *allowed perturbations* we found that (3) did not hold. In other words suppose that in

---

[4] Of course message m_j must be one which is sent with positive probability for Bayes' rule to apply.

an appropriate model one could show that $\forall\ i, j = 1, 2$ and $j \neq i$, $\forall P_j$, and $\forall s_j$

$$\exists\ m_j^* \in \mathcal{M} \quad \text{such that} \quad \mathcal{B}_i(s_j, \varepsilon, P_j | m_j^*, m_i) = \alpha_i \quad \forall m_i \in \mathcal{M},\ \ \forall \varepsilon > 0 \tag{4}$$

then it would follow that pre-play communication is effective in selecting the good equilibrium of the game above in the sense that all Trembling-Hand Perfect[5] equilibria of the game with pre-play communication must involve the (α_1, α_2) equilibrium being played. The reason is easy to explain informally.

One way to read (4) is that for *any* pair (P_j, s_j) there exists a message m_j^* which *reveals* player j's intention to play cooperatively (to play α_j) to a sufficiently high degree so as to ensure that the optimal response for i once m_j^* has been observed is to play α_i. Let $s_j^* \in \mathcal{S}_j$ be a strategy for j which gives message m_j^* and subsequently plays α_j *whatever* the message m_i of player i. Consider now a perturbed equilibrium where the noise has almost vanished as we approach the actual Trembling-Hand Perfect equilibrium. The expected payoff to s_j^* in a game like the one in Figure 1 must be arbitrarily close to 3. The reason is that with very high probability s_j^* plays against the equilibrium strategy for player i which, by definition, must respond optimally after any message of player j. Since the equilibrium strategy for player j must do at least as well as any $s_j \in S_j$ in expected terms, it now follows that the expected payoff for the *equilibrium* strategy of player j must also be arbitrarily close to 3. Repeating the argument for the other player shows that the equilibrium payoffs to both players must approach 3 in the limit as the noise vanishes.

The purpose of this paper is to show that the intuition we have developed above holds when players' strategies and perturbations are both restricted to be *computable* in the sense of *Church's Thesis*, and the perturbations have sufficiently large support.[6] The players' ability to reveal their cooperative intentions through the pre-play communication stage implies cooperative play in the second stage whenever the underlying

---

[5] Note that, here and throughout the paper, since we are restricting the type of perturbations allowed and their support, we are in a way abusing the original meaning of the term Trembling-Hand Perfect equilibrium (Selten 1975, Myerson 1978).

[6] By 'sufficiently large support' we mean that we can identify a minimum support set which represents a 'lower bound' on the support of the perturbations. It follows that our main result holds for any perturbations with support *larger* than the minimum support set. See Theorem 1 and Remark 2 below.

game is a Common Interest game.

1.3. Overview

The plan of the paper is as follows. The next Section of the paper establishes basic notation and terminology concerning computable functions and computing devices (Turing machines). Section 3 describes the model in full detail. The main result of the paper is stated formally in Section 4. In Section 5 we first prove a lemma on communication on which the proof of the main result revolves and then present a formal proof of the main result stated in Section 4. In section 6, we present some further results which we believe are useful to evaluate correctly the force and interest of the main result of the paper. Section 7 concludes the paper with a discussion of our model and of some related literature. Finally, the Appendix contains some ancillary material which is removed from the main body of the paper for ease of exposition. In the numbering of equations, definitions and so on, a prefix 'A' indicates that the relevant item is to be found in the Appendix.

2. Computability

The formal notion of *algorithmic* which we adopt is that of *general recursive* functions or *effectively computable* (or simply *computable*) functions. Intuitively, a function is effectively computable if there exists a *finite* computing device which is capable of computing each of its values in a *finite* number of steps. There is a general consensus in the mathematical literature that the class of general recursive functions captures the widest possible intuitive notion of effective computability.[7] The class of general recursive functions coincides with the class of functions which can be computed by a class of abstract computing devices known as *Turing machines.*

A Turing machine is identified by its *program.* A program is a finite string of symbols obeying some syntactical rules which we do not specify here. It follows that Turing machines can be put in a one-to-one (computable) correspondence with the natural numbers. This is a standard technique known as *Gödel numbering.*[8] In the

---

[7] See for instance Davis (1958), Rogers (1967) or Cutland (1980).

[8] See for instance Cutland (1980), or for a brief exposition Anderlini (1989).

case of a finite set of symbols only being available the numbering procedure can be thought of in the following way. In the first place, assign an order to the set of symbols so that there is a first symbol, a second one and so on. Once this is done, it is clear that all finite strings of symbols can be ordered first by length and then 'alphabetically' in an obvious way. To each string then there corresponds its place in such a 'dictionary' of all finite strings. Given that Turing machines can be numbered in the way we have described, we can simply identify each machine with its Gödel number. The assignment of Gödel numbers to Turing machines *will remain fixed throughout the paper*. The possible inputs and outputs of Turing machines can also all be coded and decoded into the naturals and vice versa. In other words, this framework implies that there is never any loss of generality in considering only functions from the natural numbers into the natural numbers.

The set of natural numbers will be denoted by $\mathbb{N}$ throughout the paper. Given a pair $n \in \mathbb{N}$ and $e \in \mathbb{N}$, the notation $\varphi_n(e)$ will indicate the result of the computation of Turing Machine n given the input e. Not all computations are defined, since a Turing machine may not *halt* on some (or all) inputs. By $\varphi_n(e) \downarrow$ and $\varphi_n(e) \uparrow$ we will indicate that the computation $\varphi_n(e)$ does and does not halt respectively. Throughout most of the paper we will be concerned with Turing machines and computable functions of one variable. In some of the proofs we will need to consider the more general case of $m \geq 2$ variables. The result of the computation of Turing machine n on the m inputs $(e_1, \cdots, e_m)$ will be denoted by $\varphi_n^{(m)}(e_1, \cdots, e_m)$. When the superscript is omitted it will be understood that we are referring to Turing machines with one variable as input.

A function from $\mathbb{N}^m$ to $\mathbb{N}$ is general recursive or computable if and only if it can be computed by a Turing machine.[9] We let $\mathcal{F}^{(m)}$ be the set of all computable functions from $\mathbb{N}^m$ to $\mathbb{N}$ with typical element $f^{(m)}$. When the superscript is omitted, it will be understood that we are referring to computable functions of one variable.

---

[9] Notice that from this definition it follows immediately that some functions $\mathbb{N} \rightarrow \mathbb{N}$ are *not* computable. There are $2^{\aleph_o}$ functions $\mathbb{N} \rightarrow \mathbb{N}$, but only countably many are in fact computable. Notice also that we do not restrict attention to Turing machines which have any special properties as far as their computations halting or not halting. It follows, for instance, that the function 'nowhere defined' is a computable function. In general the set of inputs on which a computable function is defined — the set of inputs on which the corresponding program halts — only possesses the regularity property of being 'recursively enumerable' (see Definition 4 and Theorem A.4 below).

A computable function (of one or many variables) which is defined for all possible inputs is called a *total* computable function.

The symbol '$\simeq$' used between two Turing machines, two computable functions or any combination of these will mean 'defined on the same set of inputs and equal whenever defined'. This piece of notation is necessary purely because the output of a Turing machine need not be defined for all possible inputs.

3. THE MODEL

3.1. The Underlying Game

The underlying game is denoted by $G = (\mathcal{A}, \pi)$, where $\mathcal{A} = \mathcal{A}_1 \times A_2$ is the finite joint action space, and $\pi = (\pi_1, \pi_2)$ represents the players' payoff functions with $\pi_i : \mathcal{A} \to \mathrm{I\!R}$ $(i = 1, 2)$. The 'labels' of player i's actions are, for convenience, taken to be the first $||\mathcal{A}_i||$ natural[10] numbers so that effectively we have $\mathcal{A}_i = \{0, 1, \cdots, ||\mathcal{A}_i|| - 1\}$. The set of payoffs available in G is denoted by[11] $V \subset \mathrm{I\!R}^2$. We are now ready to define the class of games to which our main result applies.

DEFINITION 1: *A two-person strategic-form game is said to be a Common Interest game if and only if there exists one pair of payoffs — denoted π^e — (which may be associated with more than one pair of strategies) which strictly Pareto-dominates all other payoffs in the game.*

Throughout the rest of the paper, $a^e = (a_1^e, a_2^e)$ will denote one particular (arbitrarily fixed) pair of actions such that $\pi(a^e) = \pi^e$. We will refer to a_i^e as the cooperative action for player i. Moreover, purely for notational simplicity (see Definition 2 below), we assume that actions in $\mathcal{A}_i$ are labelled in such a way as to ensure that a_1^e and a_2^e have the same label. In other words, $a_1^e = a_2^e$. Lastly, throughout the paper, the set of best responses for player i to player j's action a_j^e will be denoted by $\mathcal{A}_i^e$. Formally $\forall i, j = 1, 2 \;\; j \neq i$

$$\mathcal{A}_i^e = \arg\max_{a_i \in \mathcal{A}_i} \pi_i(a_i, a_j^e) \tag{5}$$

---

[10] Throughout the paper, the notation $||\cdot||$ denotes the cardinality of a set.

[11] Throughout the paper, $A \subset B$ means $A \subseteq B$ and $A \neq B$.

3.2. Computable Strategies

The computability framework described in Section 2 makes $\mathbb{N}$ the natural choice for the message space of both players. Players' strategies in the game with pre-play communication are as in Section 1.2 except that they will now be required to be *algorithmic* or computable. A Turing machine n_i which computes a total computable function can be interpreted as a strategy for player i in the game with pre-play communication in the following way. For simplicity, we take the player's message to be the output of machine n_i on a special symbol — $\bullet$ — which will not be used for any other purpose. The function computed by machine n then defines player i's action in response to any possible message m_j of player j. To avoid the possibility that the action of player i may be outside $\mathcal{A}_i$ (if the output of $\varphi_{n_i}(m_j)$ exceeds $||\mathcal{A}_i|| - 1$) we take the action of i in G to be the output of $\varphi_{n_i}(m_j)$ modulus $||\mathcal{A}_i||$. In other words, given a pair of Turing machines (n_i, n_j), both computing a total computable function we take the payoff to player i to be equal to

$$\pi_i \left[\varphi_{n_i}(\varphi_{n_j}(\bullet)) \bmod ||\mathcal{A}_i||, \varphi_{n_j}(\varphi_{n_i}(\bullet)) \bmod ||\mathcal{A}_j|| \right] \tag{6}$$

For a variety of technical reasons, all somehow related to the so-called *halting problem* for Turing machines, we cannot exclude from the analysis all Turing machines which do not halt either at the communication stage or in the second stage of the game. We are, however able to restrict attention to Turing machines which either always halt or which never do so.[12]

To neutralize the role of non-halting machines we assume that using a non-halting machine defines a strictly dominated strategy in the game with pre-play communication.[13] Moreover we assume that playing any cooperative strategy is a best response to a non-halting Turing machine in the game with pre-play communication. We start by defining formally the sets of halting and non-halting strategies.

---

[12] This is purely for analytical convenience. In an earlier version of this paper we allowed for Turing machines which, for instance, halt at the communication stage but not in the second stage of the game. Given that we have to 'extend' the payoff functions to all types of Turing machines allowed in the support (cf. Assumption 1 below), restricting attention to two types of programs — those which always halt and those which never do — simplifies matters considerably.

[13] Note that we are therefore saying that a non-halting strategy performs 'badly' even against another non-halting strategy.

These are respectively

$$\mathcal{S}^{H} = \{n \in \mathbb{N} | \varphi_n(m) \downarrow \ \ \forall\, m \in \mathbb{N}\}$$

and

$$\mathcal{S}^{\overline{H}} = \{n \in \mathbb{N} | \varphi_n(m) \uparrow \ \ \forall\, m \in \mathbb{N}\}$$

The union of $\mathcal{S}^{H}$ and $\mathcal{S}^{\overline{H}}$ is the set of computable strategies which we will consider for either player in the game with pre-play communication. In other words, in the notation of Section 1.2, we set $\mathcal{M} = \mathbb{N}$ and $\mathcal{S}_i = \mathcal{S}^{H} \cup \mathcal{S}^{\overline{H}}$ for $i = 1, 2$. From now on, we let $\mathcal{S} = \mathcal{S}^{H} \cup \mathcal{S}^{\overline{H}}$, and $\Sigma = \mathcal{S} \times \mathcal{S}$.

We are now ready to define formally the three sets of computable strategies which will be used in the analysis that follows.

Definition 2: *The halting strategies which play action a_i^e irrespective of the message of the opposing player are called 'cooperative strategies'. Formally, let $\mathcal{C}$ be the set of cooperative strategies, then*

$$\mathcal{C} = \left\{n_i \in \mathcal{S}^{H} \mid \varphi_{n_i}(m_j) = a_i^e \quad \forall m_j\right\} \tag{7}$$

The set of halting strategies which fail to output the cooperative action for some (or all) message(s) of the opposing player is called the set of 'non-cooperative strategies'. Formally, let $\mathcal{N}$ be the set of non-cooperative strategies, then

$$\mathcal{N} = \{n \in \mathcal{S}^{H} \mid \exists m_j \in \mathbb{N} \ \text{ such that } \ \varphi_n(m_j) \neq a_i^e\} \tag{8}$$

the complement of $\mathcal{N}$ in $\mathcal{S}$ is called the set of 'quasi-cooperative strategies' and is denoted by $\mathcal{Q}$. Note that in fact $\mathcal{Q} = \mathcal{C} \cup \mathcal{S}^{\overline{H}}$.[14]

We can now be precise about the payoffs involving non-halting machines in the

---

[14] Notice that the sets $\mathcal{C}$, $\mathcal{N}$ and $\mathcal{Q}$ do not depend on the identity of the player, as in general they would, because we have assumed that the labelling of actions in G is such that $a_1^e = a_2^e$, *and* because we have omitted the 'mod' term from both (7) and (8). This simplifies the notation substantially in what follows, but is completely inessential to the substance of the arguments. All our results hold if the sets $\mathcal{C}$, $\mathcal{N}$ and $\mathcal{Q}$ are redefined in the obvious way to make them dependent on the players' identities.

game with pre-play communication.

Assumption 1: *The payoff to player i in the game with pre-play communication and computable strategies is a function* $\Pi_i : \Sigma \to \mathbb{R}$. *If* $n_i \in \mathcal{S}^H$ *for* $i = 1, 2$ *the value of* $\Pi_i(n_i, n_j)$ *is given by (6). In addition,* Π_i *satisfies the following two conditions:*

$$\forall n_i \in \mathcal{S}^{\overline{H}}\, \exists n_i' \in \mathcal{S}^H \;\; \textit{such that} \;\; \Pi_i(n_i', n_j) > \Pi_i(n_i, n_j) \quad \forall n_j \in \mathcal{S} \quad \text{(dominance)}$$

and

$$\Pi_i(n_i, n_j) \geq \Pi_i(n_i', n_j) \quad \forall n_i \in \mathcal{C}_i,\, n_i' \in \mathcal{S},\, n_j \in \mathcal{S}^{\overline{H}} \quad \text{(best response)}$$

For the rest of the paper $\Pi : \Sigma \to \mathbb{R}^2$ *denotes the pair* (Π_1, Π_2).

Assumption 1 essentially defines a new game Γ obtained from the original G by adding a pre-play communication stage and constraining the players to a choice of computable strategies. The main result of the paper states that the all the equilibria of Γ which survive appropriately computable perturbations with sufficiently large support are cooperative. Before proceeding any further we therefore need to make precise the notion of computable perturbations.

3.3. Computable Perturbations

The perturbations we consider are probability distributions over $\mathbb{N}$ with support *at most* $\mathcal{S}$. The set of such probability distributions is denoted by Δ^∞, with typical element $P = \{P(0), P(1), \cdots, P(n), \cdots\}$. The support of a distribution $P \in \Delta^\infty$ is denoted by $\text{supp}(P) \subseteq \mathcal{S}$.

We impose two types of computability requirements on the perturbations. The first is to require the probabilities to be computable in the sense that some Turing machine must be able to compute at least all the positive values of $P(n)$ as a function of n. [15]

---

[15] In Anderlini (1990) the analysis is carried out assuming that all the probabilities (whether positive or zero) can be computed *approximately* by a Turing machine. The formulation of Definition 3 which we have chosen here simplifies considerably the details of some of the formal arguments below.

Definition 3: *A Probability distribution $P \in \Delta^\infty$ is said to be 'computable' if and only if there exists a Turing machine which computes (at least) all non-zero values of $P(n)$ as a function of n. Formally, $P \in \Delta^\infty$ is said to be computable if and only if $\exists\, p \in \mathbb{N}$ such that $n \in \operatorname{supp}(P)$ implies*

$$\varphi_p(n) = P(n) \tag{9}$$

and $\varphi_p(n) \downarrow \Rightarrow \varphi_p(n) = P(n)$.

The second computability requirement we impose concerns the feasibility of computing the cumulative probability of the set of strategies which fail to cooperate for some (or all) of the opponent's messages (the set $\mathcal{N}$ of Definition 1 above) and the probability of the complement of this set of strategies (the set $\mathcal{Q}$ of Definition 1 above). The role which this assumption plays in making pre-play communication effective in the selection of the efficient payoff pair will become apparent in Section 5.1. Here, we just present the assumption and discuss why it takes the form below.

Suppose that a computable probability distribution $P \in \Delta^\infty$ is given, and suppose that we are now interested in computing the probability (up to an arbitrary degree of precision) that it assigns to a certain set of strategies $\mathcal{Z} \subset \mathcal{S}$. Then our ability to compute the probability assigned to the set $\mathcal{Z}$ is clearly dependent on whether the set $\mathcal{Z}$ is well behaved in the sense that we can computably obtain a 'list' of the elements of $\mathcal{Z}$ and of the elements of its complement in $\mathcal{S}$ which are included in the support of the probability distribution P. Suppose that such a computable list cannot be obtained. Then trying to 'add up' the probabilities of the elements of $\mathcal{Z}$ so that we can approach its cumulative probability from below may simply not be possible, since a list of the elements of $\operatorname{supp}(P) \cap \mathcal{Z}$ cannot be obtained in a computable way. Indeed, if we desired to approach the probability assigned to $\mathcal{Z}$ both from below and from above in a computable way we need to be able to generate computable lists of both $\operatorname{supp}(P) \cap \mathcal{Z}$ and $\operatorname{supp}(P) \cap \overline{\mathcal{Z}}$. Sets of natural numbers which can be exhaustively listed in a computable way are known as *recursively enumerable* sets. Recursive enumerability is a standard 'regularity' property extensively studied in recursive function theory.[16]

---

[16] See, for instance, Cutland (1980) or for a fuller treatment Rogers (1967).

Definition 4: *An infinite set* $\mathcal{Z} \subseteq \mathbb{N}$ *is said to be 'recursively enumerable' (abbreviated r.e.) if and only if it is the range of a total one-to-one computable function. Formally, an infinite* $\mathcal{Z} \subseteq \mathbb{N}$ *is r.e. if and only if* $\exists s \in \mathbb{N}$ *such that* $\varphi_s(e) \downarrow \forall e$,

$$n \in \mathcal{Z} \quad \Leftrightarrow \quad \exists e \text{ such that } \varphi_s(e) = n$$

and $e \neq e' \Rightarrow \varphi_s(e) \neq \varphi_s(e')$. *Turing machine* s *is said to enumerate the set* $\mathcal{Z}$.[17] *Any finite set is recursively enumerable. The empty set is recursively enumerable.*

We are interested in probability distributions such that the cumulative probabilities of the two sets $\mathcal{N}$ and $\mathcal{Q}$ can be approximated in a computable way. For this reason we focus on probability distributions which give positive probability only to r.e. subsets of $\mathcal{N}$ and $\mathcal{Q}$. We call these probability distributions $\mathcal{N}$-regular distributions.

Definition 5: *A probability distribution* $P \in \Delta^\infty$ *is called* $\mathcal{N}$*-regular if and only if both* $\operatorname{supp}(P) \cap \mathcal{N}$ *and* $\operatorname{supp}(P) \cap \mathcal{Q}$ *are r.e. sets. Any set* $\mathcal{Z} \subset \mathcal{S}$ *such that both* $\mathcal{Z} \cap \mathcal{N}$ *and* $\mathcal{Z} \cap \mathcal{Q}$ *are r.e. sets is called an* $\mathcal{N}$*-regular set.*

Besides the fact that only machines in $\mathcal{S}$ should be given positive probability, computability and $\mathcal{N}$-regularity are the only two assumptions we will make about perturbations. It is convenient to establish a name for distributions which satisfy all of these properties.

Definition 6: *A probability distribution* $P_i \in \Delta^\infty$ *is 'admissible' if and only if it is both computable according to Definition 3 and* $\mathcal{N}$*-regular according to Definition 5. The set of admissible probability distributions will be denoted by* $\mathcal{P}$ *for the rest of the paper.*

As we anticipated above, a straightforward argument establishes formally that if a distribution is admissible then it is possible to compute up to any arbitrary degree of precision the probability of both $\mathcal{N}$ and $\mathcal{Q}$.

---

[17] The requirement that $\varphi_s(e) \neq \varphi_s(e')$ can be interpreted as saying that s enumerates $\mathcal{Z}$ *without repetitions.* Recursive enumerability is often defined without specifying the no repetitions clause. It is a standard observation that the two definitions are equivalent, however. See for instance Cutland (1980)

Remark 1: *Let P be an admissible probability distribution. Let also $P(\mathcal{N})$ denote $\sum_{n\in\mathcal{N}} P(n)$, and $P(\mathcal{Q}) = \sum_{n\in\mathcal{Q}} P(n)$. Then*[18] *there exist two Turing machines q and m such that $\forall x \in \mathbb{N}$*

$$|P(\mathcal{N}) - \varphi_q(x)| < \frac{1}{x} \qquad \text{and} \qquad |P(\mathcal{Q}) - \varphi_m(x)| < \frac{1}{x} \tag{10}$$

Proof: See Appendix.

Therefore, associated with each admissible probability distribution there are three Turing machines (among others). One which computes probabilities, another which enumerates $\mathrm{supp}(P) \cap \mathcal{N}$, and a third one which computes (approximately) $P(\mathcal{N})$. Since these Turing machines will be used extensively in the analysis that follows it is convenient to establish a name for the triple.

Definition 7: *A triple $(p, c, q) \in \mathbb{N}^3$ is called a 'basis' for an admissible distribution $P \in \mathcal{P}$ if and only if (a) p computes the values of $P(n)$ as in (9), (b) c enumerates $\mathrm{supp}(P) \cap \mathcal{N}$ as in*[19] *Definition 4, and (c) q computes any approximation to $P(\mathcal{N})$ as in (10).*

3.4. The Equilibrium Concept

The main result of this paper is to characterize as cooperative all the Trembling-Hand Perfect equilibria of the game with pre-play communication Γ, when the perturbations have sufficiently large support and are restricted to be admissible as in Definition 6.

The two differences between the standard definition of Trembling-Hand Perfect (Selten 1975, Myerson 1978) equilibrium and what follows are admissibility of the perturbations and the fact that we parameterize equilibria by the 'minimum size' of the support of the perturbations, instead of considering perturbations with full support. This is discussed at length in Section 6 below.

---

[18] In Anderlini (1990) we show the that the reverse implication also holds. In other words, a computable distribution which satisfies (10) must be admissible in the sense of Definition 6.

[19] If $\mathrm{supp}(P) \cap \mathcal{N}$ is empty we require c to compute the function which is 'nowhere defined'. If $\mathrm{supp}(P) \cap \mathcal{N}$ is finite we require c to enumerate it as in Theorem A.5.

Given any $\mathcal{R} \subseteq \mathcal{S}$ we denote by $\mathcal{P}(\mathcal{R})$ the set of admissible probability distributions with support equal to $\mathcal{R}$ or larger. Therefore,

$$\mathcal{P}(\mathcal{R}) \equiv \{P \in \mathcal{P} \quad | \quad \mathcal{R} \subseteq \operatorname{supp}(P)\}$$

We are now ready to define formally the equilibrium concept to be used in Section 4.

DEFINITION 8: *An $\varepsilon\mathcal{R}$ Computable Trembling-Hand Equilibrium (henceforth abbreviated ($\varepsilon\mathcal{R}$)-CTHE) of Γ is a pair of computable strategies, and a pair of admissible perturbations with support at least $\mathcal{R}$ such that each player's strategy is optimal against the other player's strategy — with probability $1-\varepsilon$ — and the opposing player's perturbation with probability ε. Formally, (n_1, n_2, P_1, P_2) with $n_i \in \mathcal{S}$ and $P_i \in \mathcal{P}(\mathcal{R})$ ($i = 1, 2$) is an $(\varepsilon, \mathcal{R})$-CTHE of Γ if and only if*

$$n_i \in \mathcal{B}_i(n_j, \varepsilon, P_j) \quad \forall\ i = 1, 2\ j \neq i \tag{11}$$

where $\mathcal{B}_i(\cdot)$ is defined as in (1) above. The set of equilibrium quadruples satisfying (11) is denoted by $E(\varepsilon, \mathcal{R})$ with typical element $(n_1^E, n_2^E, P_1^E, P_2^E)$.

The sets of equilibrium actions and equilibrium payoffs associated with equilibrium quadruples will be denoted by $\mathcal{A}^E(\varepsilon, \mathcal{R})$ and $\Pi^E(\varepsilon, \mathcal{R})$ respectively.

We are interested in the 'limit equilibrium' as the perturbations of a $\varepsilon\mathcal{R}$-CTHE vanish.

DEFINITION 9: *An $\mathcal{R}$ Computable Trembling-Hand Equilibrium (henceforth abbreviated $\mathcal{R}$-CTHE) of Γ is the set of 'limit points' of any sequence of $\varepsilon\mathcal{R}$-CTHE actions as the noise vanishes. Formally, an action pair a^E is an $\mathcal{R}$-CTHE of Γ if and only if there exists two sequences, $\varepsilon_n \to 0$ and a_n^E such that $a_n^E \in \mathcal{A}^E(\varepsilon_n, \mathcal{R}) \, \forall\, n \in \mathbb{N}$ and $\lim_{n\to\infty} a_n^E = a^E$. The set of $\mathcal{R}$-CTHE of Γ is denoted by $\mathcal{A}^E(\mathcal{R})$. The set of payoff pairs associated with each element of $\mathcal{A}^E(\mathcal{R})$ is denoted by $\Pi^E(\mathcal{R})$.*

4. RESULTS

This section reports the main findings of the paper. The proof of Theorem 1 is quite lengthy. Therefore it is deferred until Section 5, which is entirely devoted to it.

Our model selects the cooperative payoff vector as the unique equilibrium outcome in a Common Interest game with pre-play communication, computable strategies and computable perturbations with sufficiently large support.

THEOREM 1: *There exists* $\mathcal{R}^* \subset \mathcal{S}$, *such*[20] *that the set of* $(\varepsilon, \mathcal{R}^*)$*-CTHE is not empty and all equilibria are cooperative. More precisely,* $\exists \mathcal{R}^* \subset \mathcal{S}$ *and* $\overline{\varepsilon} > 0$ *such that*

$$\mathcal{A}^E(\varepsilon, \mathcal{R}^*) \neq \emptyset \quad \forall \varepsilon \in [0, 1]$$

and $\varepsilon < \overline{\varepsilon}$ *implies*

$$\pi(a) = \pi^e \quad \forall\, a \in \mathcal{A}^E(\varepsilon, \mathcal{R}^*)$$

A trivial consequence of Theorem 1 which we state without proof is the following.

COROLLARY 1: *There exists* $\mathcal{R}^* \subset \mathcal{S}$ *such that* $\Pi^E(\mathcal{R}^*) = \pi^e$

We conclude this Section with an observation.

REMARK 2: *Recall that in the definition of a* $(\varepsilon, \mathcal{R})$*-CTHE the set* $\mathcal{R}$ *is a 'lower bound' on the support of the perturbation. It follows trivially that* $\mathcal{R}^* \subseteq \mathcal{R}'$ *implies* $\mathcal{A}^E(\mathcal{R}') \subseteq \mathcal{A}^E(\mathcal{R}^*)$*. Therefore Theorem 1 implies that all* $\mathcal{R}$*-CTHE of* Γ *with* $\mathcal{R}^* \subseteq \mathcal{R}$ *are cooperative.*

5. PROOFS

The proof of Theorem 1 can be divided into three separate arguments. The first shows that in our model each player can use the pre-play communication stage to convincingly reveal his intention to cooperate in the second stage of the game. We call this the Communication Lemma (Lemma 3 below). The second part of the argument shows that, assuming the equilibrium set is not empty, since the stage game is a Common Interest game, the Communication Lemma implies that all equilibrium payoffs are in fact cooperative. We present this argument as Lemma 4 below. The

---

[20] As is clear from Remark 4 below, Theorem 1 can only hold with $\mathcal{R}^*$ a strict subset of $\mathcal{S}$ since there are no admissible distributions with support precisely equal to $\mathcal{S}$.

last part of the argument is devoted to showing that the equilibrium set is not empty. This is the content of Lemma 5 below. Theorem 1 is then an immediate consequence of Lemma 4 and Lemma 5.

In the arguments which follow and in the Appendix, we make use of a technique accepted as standard in this area of mathematics known as *proof by Church's Thesis*. The claim which underlies this methodology is that whenever a 'clear procedure' exists for computing a function in a finite number of 'steps' then it follows that such function is computable by a Turing machine. Extreme care must be taken of course in the definition of 'step' and of a 'clear procedure'. Thorough discussions of this way of proceeding are in, for instance, Cutland (1980) or Rogers (1967).

5.1. Effective Communication: Intuition

It is useful to start with an intuitive outline of why our assumptions of computable strategies and computable trembles are sufficient for pre-play communication to be effective in conveying a player's cooperative intentions up to any arbitrary degree of precision; in other words, an intuitive account of why (4) holds in our model.

Our computability assumptions ensure that the players' strategy sets are countable in the game with pre-play communication (even though countably many messages are possible in principle). It follows that the set of non-cooperative strategies (the set $\mathcal{N}$ of Definition 2 above) is also countable in our model.

In Figure 2 below, we represent $\mathcal{N}$ on the horizontal axis, with the probabilities assigned to its elements by the perturbation on the vertical axis. Let us now construct a computable strategy n^* which performs the following computations. First n^* enumerates sufficiently many non-cooperative strategies so that the probability of the 'tail' of non-cooperative strategies which have not been enumerated is 'small' relative to the probability assigned to n^* itself. Since the set $\mathcal{N}$ is countable, this can be done enumerating *finitely many* non-cooperative strategies. Let $\overline{n}$ be the number of strategies in $\mathcal{N}$ which need to be enumerated in this way — in Figure 2 we have set $\overline{n} = 10$.

Strategy $\overline{n}$ now goes on to 'simulate', one by one, all the non-cooperative strategies which have been enumerated, in order to compute the message that each one of them outputs at the communication stage. Once all these messages have been computed, strategy n^* can set its own message to be *different* from all of them. For simplicity, assume that this message also turns out to be different from the message produced by the equilibrium strategy.

Once arrived at the second stage of the game, strategy n^* plays the cooperative action, regardless of the message of the opponent. Thus n^* is a cooperative strategy.

It is now apparent how playing strategy n^* enables a player to reveal convincingly his cooperative intentions. Observing the message which n^* produces at the communication stage tells the opposing player that, if he is facing a non-cooperative strategy, then it can only be a strategy from the 'tail' of Figure 2. Since the probability of n^* is 'large' relative to this tail, Bayes' rule now tells the opposing player that he is facing a quasi-cooperative strategy (a strategy which is not in $\mathcal{N}$ — see Definition 2 above) with probability arbitrarily close to one.

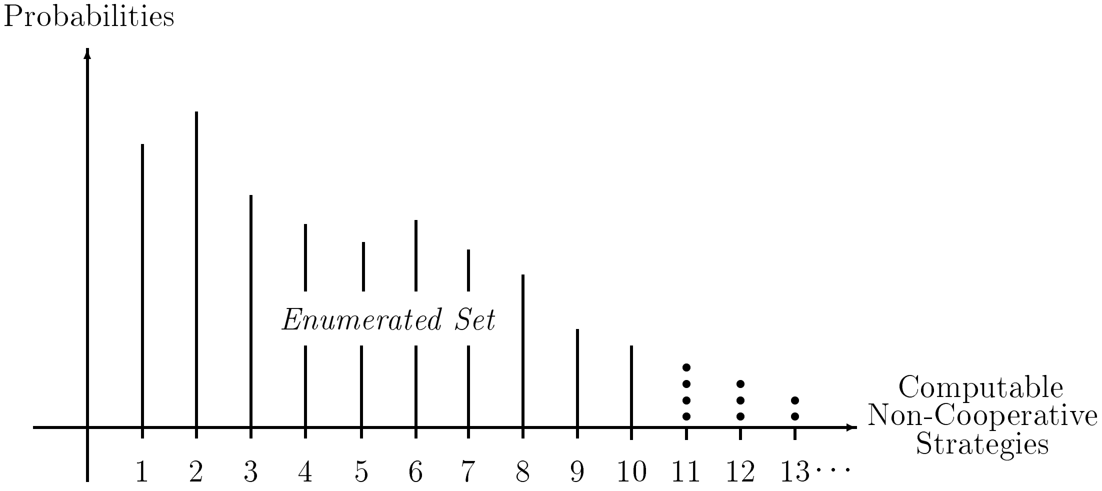

Figure 2

The simple intuition we have outlined is what drives the main result of this paper. The formal analysis encounters one further difficulty, however. In the construction above we are 'keeping fixed' the probability which the perturbation assigns to n^* and at the same time constructing the strategy n^*. Clearly as we construct n^*, its 'number' and hence its probability may change. Below, we are able to rely on a pseudo-fixed point result (Theorem A.3) which ensures that this potential circularity is avoided, thus ensuring that our strategy n^* is well defined.

5.2. Effective Communication: A Lemma

It is convenient to divide the formalization of the intuitive argument outlined above into three distinct Lemmas. We start by showing that, given a basis for an admissible distribution, an arbitrary n such that $P(n) > 0$ and a 'level of precision' parameter x, it is possible to compute a cut-off value like $\overline{n}$ in Figure 2, such that the 'tail' of P after $\overline{n}$ has probability smaller than $(1/x)P(n)$.

Lemma 1: *There exists a computable function $t^{(5)} \in \mathcal{F}^{(5)}$ such that $\forall (n, p, c, q, x) \in \mathbb{N}^5$, whenever (p, c, q) form a basis for some admissible probability distribution $P \in \mathcal{P}$ and $P(n) > 0$, we have $t^{(5)}(n, p, c, q, x) = \overline{n}$, where $\overline{n}$ is such that*

$$\frac{1}{x}\varphi_p(n) > P(\mathcal{Q}) - \sum_{t=0}^{\overline{n}} \varphi_p(\varphi_c(t))$$

Proof: See Appendix.

The second step is to show that given that a cut-off value like $\overline{n}$ of Lemma 1 can be computed, it is possible to construct a computable function which given a basis for $P \in \mathcal{P}$, an arbitrary n, a degree of precision parameter x, and arbitrary $\hat{m}$ and m, outputs the cooperative action a_i^e if $m \neq \bullet$, and a message m^*, which is *different* from both the messages of *all* the machines in $\mathcal{N}$ which are enumerated before $\overline{n}$, and from the arbitrary given $\hat{m}$.

Lemma 2: *There exists a computable function $d^{(7)} \in \mathcal{F}^{(7)}$ such that: (a) $\forall$ $(n,\ p,\ c,\ q,\ \hat{m},\ x) \in \mathbb{N}^6$ either $d^{(7)}(n, p, c, q, \hat{m}, x, m) \downarrow \forall m \in \mathbb{N}$ or $d^{(7)}(n, p, c, q, \hat{m}, x, m) \uparrow$ $\forall m \in \mathbb{N}$, and moreover (b) $\forall (n, p, c, q, \hat{m}, x, m) \in \mathbb{N}^7$, whenever (p, c, q) form a basis for some admissible probability distribution $P \in \mathcal{P}$ and $P(n) > 0$ we have*

$$d^{(7)}(n, p, c, q, \hat{m}, x, m) = \begin{cases} a_i^e & \textit{if } m \neq \bullet \\ \\ m^* = 1 + \hat{m} + \sum_{t=0}^{\overline{n}} \varphi_{\varphi_c(t)}(\bullet) & \textit{if } m = \bullet \end{cases} \tag{12}$$

Where $\overline{n}$ is as in Lemma 1.

PROOF: See Appendix.

Before we formalize the next step, notice that the second line of (12) implies that both $m^* \neq \varphi_{\varphi_c(t)}(\bullet)$ $\forall 0 \leq t \leq \overline{n}$ and $m^* \neq \hat{m}$ hold. In other words, m^* is different from each of the messages given by the first $\overline{n}$ strategies in the enumeration of $\mathcal{N}$, and is also different from $\hat{m}$.

Suppose now that we were lucky enough to find that the *actual* perturbation of a player's equilibrium strategy gives probability $P(n)$ (or larger) to a strategy which is 'like' $d^{(7)}$ as a function of m for 'large' x, (p, c, q) a basis of the *actual* perturbation, and $\hat{m}$ equal precisely to the message produced by the equilibrium strategy. This is a cooperative strategy by construction. By the construction in Lemma 2 it also reveals itself to be cooperative up to the arbitrary degree of precision implied by x. Therefore it would fulfil the revelation requirements which we have outlined above.

The proof of the Communication Lemma 3 below exploits the construction of $d^{(7)}$ to yield a set of revealing strategies. Firstly, we use the parameterization Theorem A.1 known as the *s-m-n* Theorem to show that it is feasible to treat all inputs of $d^{(7)}$ but m as 'parameters' and hence define a function of m only, which behaves in the same way as $d^{(7)}$. Secondly, we use a pseudo-fixed point result (Theorem A.3) to ensure that the probability used to compute the cut-off value $\overline{n}$ of Lemma 1 is precisely the probability of the parameterized version of $d^{(7)}$ we have obtained in the first step. In this way we obtain one potential revealing strategy for each possible configuration of parameters $(p, c, q, \hat{m}, x)$, and hence for any possible $P \in \mathcal{P}$, $\hat{m} \in \mathbb{N}$ and $x \in \mathbb{N}$. The proof is then concluded by stipulating that the 'minimum support set' $\mathcal{R}^*$ must be the set of all such potential revealing strategies. Hence for any $P \in \mathcal{P}(\mathcal{R}^*)$ we can be sure that the relevant potential revealing strategy has in fact positive probability. Therefore for any such $P \in \mathcal{P}(\mathcal{R}^*)$ and message of the equilibrium strategy, the player's cooperative intentions can be revealed up to any arbitrary degree of precision.[21]

---

[21] It is worth pointing out one intuitively appealing feature of the procedure yielding the revealing messages used in the proof of Lemma 3 below. The length of the computation yielding such 'smart' messages, it is easily verified, is increasing in the precision parameter x of the statement of the Lemma. In other words the 'complexity' of such revealing messages is increasing in the level of precision with which we require the machine to reveal itself. In order to select the Pareto-efficient equilibrium, such level of precision needs to be higher as the Pareto-efficient equilibrium becomes 'riskier' in the sense of Harsanyi and Selten (1988). Hence more 'complex' messages will be required

As we anticipated, the Communication Lemma, says that the analogue of (4) holds in the model we have developed.

Lemma 3 *(Communication Lemma): There exists* $\mathcal{R}^* \subset \mathcal{S}$ *such that,* $\forall\ i, j = 1, 2$ *and* $j \neq i$, $\forall P_j \in \mathcal{P}(\mathcal{R}^*)$, *and* $\forall n_j \in \mathcal{S}$

$$\exists\ n_j^* \in \mathcal{R}^* \text{ such that } \mathcal{B}_i \left[n_j, \varepsilon, P_j | \varphi_{n_j^*}(\bullet), m_i \right] = \mathcal{A}_i^e \qquad \forall m_i,\ \ \forall \varepsilon > 0$$

where $\mathcal{A}_i^e$ *is defined as in (5). Notice that* n_j^* *and therefore* $m_j^* = \varphi_{n_j^*}(\bullet)$ *depend on both* n_j *and* P_j *itself. We suppress this from the notation since it does not cause any ambiguity.*

Proof: The *s-m-n* Theorem A.1 guarantees that there exists a total computable function $s^{(6)} \in \mathcal{F}^{(6)}$ such that

$$\varphi_{s^{(6)}(n,p,c,q,\hat{m},x)}(m) \simeq \varphi_n^{(6)}(p, c, q, \hat{m}, x, m) \qquad \forall\ (n, p, c, q, \hat{m}, x, m) \in \mathbb{N}^7 \tag{13}$$

By Theorem A.2 and by Church's thesis, $f^{(7)} \in \mathcal{F}^{(7)}$ defined by

$$f^{(7)}(n, p, c, q, \hat{m}, x, m) \equiv d^{(7)}(s^{(6)}(n, p, c, q, \hat{m}, x), p, c, q, \hat{m}, x, m) \tag{14}$$

where $d^{(7)}$ is as in Lemma 2, is a computable function. By the pseudo-fixed point Theorem A.3 we then have that $\exists\ \hat{n} \in \mathbb{N}$ such that

$$\varphi_{\hat{n}}^{(6)}(p, c, q, \hat{m}, x, m) \simeq f^{(7)}(\hat{n}, p, c, q, \hat{m}, x, m) \quad \forall\ (p, c, q, \hat{m}, x, m) \in \mathbb{N}^6 \tag{15}$$

Substituting (13) and (14) in (15) we finally obtain that $\forall\ (p, c, q, \hat{m}, x, m) \in \mathbb{N}^6$

$$\varphi_{s^{(6)}(\hat{n},p,c,q,\hat{m},x)}(m) \simeq d^{(7)}(s^{(6)}(\hat{n}, p, c, q, \hat{m}, x), p, c, q, \hat{m}, x, m) \tag{16}$$

Consider now a *fixed* $P_j \in \mathcal{P}$ and its 'parameterization' (its basis of Definition 7)

---

in order to destroy an equilibrium which does not yield Pareto-efficient payoffs but which risk-dominates the Pareto-efficient equilibrium. I owe this observation to a conversation with Andreu Mas-Colell.

$(p, c, q) \in \mathbb{N}^3$. Suppose that for such given basis, and given $\hat{m}$ and x we have

$$\varphi_p(s^{(6)}(\hat{n}, p, c, q, \hat{m}, x)) > 0 \tag{17}$$

where $\hat{n}$ is the pseudo-fixed point of equation (15). Now set

$$n^*_{\hat{m}x} \equiv s^{(6)}(\hat{n}, p, c, q, \hat{m}, x) \quad \text{and} \quad m^*_{\hat{m}x} \equiv \varphi_{n^*_{\hat{m}x}}(\bullet)$$

notice that from (16) and by construction of $d^{(7)}$ in Lemma 2, we have that $n^*_{\hat{m}x} \in \mathcal{C}$. Moreover,

$$m^*_{\hat{m}x} \neq \varphi_{\varphi_c(t)}(\bullet) \quad \forall 0 \leq t \leq \overline{n}$$

where $\overline{n}$ is as in Lemma 1. Therefore,

$$\frac{1}{x}\varphi_p(n^*_{\hat{m}x}) > P_j(\mathcal{N}|m^*_{\hat{m}x}) \tag{18}$$

where the right hand side of (18) is the probability which $P_j \in \mathcal{P}$ assigns to $\mathcal{N}$, *conditional* on message $m^*_{\hat{m}x}$ having been observed.

Since $n^*_{\hat{m}x} \in \mathcal{C}$, (18) implies that

$$P_j(\mathcal{C}|m^*_{\hat{m}x}) \geq \varphi_p(n^*_{\hat{m}x}) > xP_j(\mathcal{N}|m^*_{\hat{m}x})$$

Since $P_j(\mathcal{C}|m^*_{\hat{m}x}) + P_j(\mathcal{N}|m^*_{\hat{m}x}) = 1$,[22] (18) implies that for all x and $\hat{m}$

$$P_j(\mathcal{C}|m^*_{\hat{m}x}) > \frac{x}{x+1} \tag{19}$$

To close the first part of the argument we must now find $\mathcal{R}^* \subseteq \mathcal{S}$ such that $\forall P_j \in \mathcal{P}(\mathcal{R}^*)$, if (p, c, q) is a basis for P_j, then (17) is satisfied for all $\hat{m}$ and x. It is sufficient

---

[22] Note that in general we should write $P_j(\mathcal{C}|m) + P_j(\mathcal{N}|m) + P_j(\mathcal{S}^{\overline{H}}|m) = 1$. However, since $\mathcal{S}^{\overline{H}}$ is the set of non-halting strategies which do not output a message at all, if m is any message which is observed with positive probability, we have that $P_j(\mathcal{S}^{\overline{H}}|m) = 0$. Therefore $P_j(\mathcal{C}|m) + P_j(\mathcal{N}|m) = 1$.

to set

$$\mathcal{R}^* \equiv \operatorname*{Range}_{(p,c,q,\hat{m},x)\in \mathbb{N}^6} s^{(6)}(\hat{n},p,c,q,\hat{m},x) \tag{20}$$

Clearly, (20) together with (19) gives that $P_j \in \mathcal{P}(\mathcal{R}^*)$ implies that for all $\hat{m}$ and x

$$P_j(\mathcal{C}|m^*_{\hat{m}x}) > \frac{x}{x+1} \tag{21}$$

Consider now a pair (n_j, P_j) with $n_j \in \mathcal{S}$ and $P_j \in \mathcal{P}(\mathcal{R}^*)$. Let $P_j^{\varepsilon n_j}$ be the probability distribution obtained by combining the degenerate distribution which puts probability 1 on n_j with weight $1-\varepsilon$ with the distribution P_j with weight $\varepsilon > 0$. This is the effective distribution facing player i which defines implicitly the best response correspondence $\mathcal{B}_i(\cdot)$ of the statement of the Lemma. Suppose now that $\varphi_{n_j}(\bullet) \neq m$ where m is any message which has positive probability given P. Then, by Bayes' rule

$$P_j^{\varepsilon n_j}(\mathcal{C}|m) = P_j(\mathcal{C}|m) \quad \forall \varepsilon > 0 \tag{22}$$

Recall now that by construction of $d^{(7)}$ in Lemma 2, we have that $\varphi_{n_{\hat{m}x}}(\bullet) \neq \hat{m}$. Therefore, combining (21) and (22) we get that $P_j \in \mathcal{P}(\mathcal{R}^*)$ implies that for all $x \in \mathbb{N}$, setting $\hat{m} = \varphi_{n_j}(\bullet)$ ensures

$$\forall n_j \in \mathcal{S} \quad P_j^{\varepsilon n_j}(\mathcal{C}|m^*_{\hat{m}x}) > \frac{x}{x+1} \qquad \forall \varepsilon > 0 \tag{23}$$

Since all strategies in $\mathcal{C}$ are certain to play a_j^e, regardless of the message of player i, and since the underlying game is a Common Interest game, the statement of the Lemma follows immediately from (23) as we let x become arbitrarily large. Finally, it remains to check that $\mathcal{R}^*$ is indeed a subset of $\mathcal{S}$. From the construction in (a) of Lemma 2 it is clear that in fact $\mathcal{R}^* \subseteq \mathcal{Q}$, which is enough to prove the claim.

Q.E.D.

5.3. Optimality Proof

We are now ready to prove the first part of Theorem 1: provided that the set of

$(\varepsilon, \mathcal{R}^*)$-CTHE is not empty, then all equilibria are cooperative. The intuition behind the formalities which follow is exactly the one described at length in Section 1.2 above; we do not repeat it here.

LEMMA 4: *Let* $\mathcal{R}^* \subset \mathcal{S}$ *be as in the statement of the Communication Lemma (Lemma 3). Assume that* $\mathcal{A}^E(\varepsilon, \mathcal{R}^*) \neq \emptyset \; \forall \varepsilon > 0$. *Then there exists an* $\bar{\varepsilon} > 0$ *such that* $\varepsilon < \bar{\varepsilon}$ *implies*

$$\pi(a) = \pi^e \qquad \forall \; a \in \mathcal{A}^E(\varepsilon, \mathcal{R}^*)$$

PROOF: Let π_i^L and π_i^H be respectively the *lowest* and *second highest* payoff which player i can achieve in *any outcome* of the extended game Γ. Notice that by Assumption 1 (dominance) $\pi_i^e > \pi_i^L$ for $i = 1, 2$, which of course implies that $\pi_i^e > \pi_i^H \geq \pi_i^L$ for any $i = 1, 2$. Consider an equilibrium quadruple $(n_1^E, n_2^E, P_1^E, P_2^E) \in E(\varepsilon, \mathcal{R}^*)$ with

$$\varepsilon < \bar{\varepsilon} = \min_{i \in \{1,2\}} \left\{ \frac{\pi_i^e - \pi_i^H}{2\pi_i^e - \pi_i^H - \pi_i^L} \right\} \tag{24}$$

Let n_i^* $(i = 1, 2)$ be the revealing strategy of the Communication Lemma, corresponding to the pair (n_i^E, P_i^E), and let $m_i^* = \varphi_{n_i^*}(\bullet)$. Since equilibrium strategies must be optimal in expected terms, they must play an action in the best response set conditional on any message which is given with positive probability. It then follows immediately from the Communication Lemma that

$$\varphi_{n_i^E}(m_j^*) \in \mathcal{A}_i^e$$

From the fact that n_j^* is a cooperative strategy it then follows that

$$\Pi_j(n_j^*, n_i^E) = \pi_j^e$$

Since n_j^* plays against the equilibrium strategy n_i^E with probability at least $1 - \varepsilon$, it follows that the *expected* payoff to strategy n_j^* in this equilibrium is at least

$$(1 - \varepsilon)\pi_j^e + \varepsilon \pi_j^L$$

Since equilibrium strategies must be optimal in expected terms, it must be that the expected payoff to strategy n_j^E is also at least as large as $(1-\varepsilon)\ \pi_j^e + \varepsilon\ \pi_j^L$. Suppose now that

$$\Pi(n_1^E, n_2^E) \neq \pi^e$$

then the expected payoff to n_j^E cannot be greater than

$$(1-\varepsilon)\pi_j^H + \varepsilon\pi_j^e$$

However, since ε is small enough to satisfy (24), we have that

$$(1-\varepsilon)\pi_j^e + \varepsilon\pi_j^L > (1-\varepsilon)\pi_j^H + \varepsilon\pi_j^e$$

From this contradiction it follows that (24) implies that

$$\Pi(n_1^E, n_2^E) = \pi^e$$

and this is enough to prove the claim.

Q.E.D.

5.4. Existence Proof

We are now in a position to complete the proof of Theorem 1.

LEMMA 5: *Let $\mathcal{R}^* \subset \mathcal{S}$ be as in Lemma 3. Then*

$$\mathcal{A}^E(\varepsilon, \mathcal{R}^*) \neq \emptyset \quad \forall \varepsilon \in [0,1]$$

PROOF: By (20) and by Theorem A.6, $\mathcal{R}^*$ is an r.e. set. It is also straightforward to see that $\mathcal{R}^*$ is infinite. We omit the proof of this claim. Therefore, $\mathcal{R}^*$ can be enumerated as in Definition 4 by a Turing machine r which computes a one-to-one total computable function. Consider now the following $P^r \in \Delta^\infty$

$$P^r(n) \equiv \begin{cases} 1/2^e & \text{if } n = \varphi_r(e) \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

The distribution P^r is computable according to Definition 3 since, given any n, successively larger values of e can be tried in the computation $\varphi_r(e)$, until one is found (which will be the case if $n \in \mathcal{R}^*$) such that $\varphi_r(e) = n$. Given such e it is feasible to compute the value of $1/2^e$. By (25), the support of P^r is exactly $\mathcal{R}^*$.

Recall now that $\mathcal{R}^* \subseteq \mathcal{Q}$. Since the support of P^r is exactly $\mathcal{R}^*$ it follows that $\text{supp}(P) \cap \mathcal{Q} = \mathcal{R}^*$ and $\text{supp}(P) \cap \mathcal{N} = \emptyset$. Therefore, since both $\mathcal{R}^*$ and the empty set are r.e., the distribution P^r is $\mathcal{N}$-regular and hence it is admissible.

Now let n^E be some arbitrarily chosen element of $\mathcal{C}$, and set $n_1^E = n_2^E = n^E$. We claim that the quadruple (n_1^E, n_2^E, P^r, P^r) is an element of $\mathcal{A}^E(\varepsilon, \mathcal{R}^*)$ for any $\varepsilon \in [0,1]$. Indeed, this is trivial from the fact that the underlying game G is a Common Interest game and Assumption 1 (best response) since all strategies in the support of P^r and strategy n^E are quasi-cooperative.

Q.E.D.

6. Admissibility and Large Support

While Lemma 5 is clearly enough to conclude the proof of Theorem 1, the argument is unsatisfactory in the sense that it relies on the special case of equilibrium strategies and perturbations including only cooperative strategies.

To evaluate the strength and interest of Theorem 1, it is natural to ask just how special is the construction in the proof of Lemma 5. In other words exactly what 'type' of perturbations turn out to be possible under the assumption of admissibility. This question can roughly be divided into two further questions. The first is what 'shape' of perturbations are possible, and the second is 'how large' can their support be? Answering these questions will also answer the questions of 'how many' non-cooperative Turing machines can be included in the support of an admissible perturbation and how much probability can be assigned to them.

Before attempting to analyze these questions formally, it is useful to reason as to what type of answers we may expect, given the model at hand. The first reason for caution, is that the Communication Lemma already partly characterizes the 'shape' of any admissible distribution with support at least $\mathcal{R}^*$. The answer to the first question is therefore already unlikely to be that any distribution is possible. The second reason for caution concerns the question of 'how many' non-cooperative strategies can we find in the support of an admissible perturbation. Since the answer will turn out to be neither 'none', nor 'finitely many', nor 'all but finitely many', nor 'all' (the same turns out to be true if we ask the same question about quasi-cooperative strategies), it will be difficult to go any further. Since our strategy spaces are countably infinite, any attempt to use some 'uniform measure' would give answers which we know to be meaningless.

In a standard 'topological' sense, admissibility is a weak restriction, even when coupled with our assumption of large support.

Remark 3: *Consider the 'sup' distance between probability distributions over* $\mathbb{N}$ *defined by*

$$\mathrm{d}(P', P'') \equiv \sup_{n \in \mathbb{N}} |P'(n) - P''(n)|$$

Given any probability distribution $P' \in \Delta^\infty$ *(whether admissible or not) and any* $\varepsilon > 0$*, there exists an admissible probability distribution* $P'' \in \mathcal{P}(\mathcal{R}^*)$ *such that* $\mathrm{d}(P', P'') < \varepsilon$*.*

The claim in Remark 3 follows easily from the fact (Theorem A.8) that any total function which is constant after a finite n is computable, and the fact that probability distributions over $\mathbb{N}$ must have a 'tail'. The details are omitted.

There is an important way in which the standard topological view of Remark 3 is misleading about the restrictiveness of the assumption of admissibility.

Remark 4: *Recall that the support* $\mathcal{Z} \subseteq \mathcal{S}$ *of any admissible distribution must be* $\mathcal{N}$*-regular in the sense that both* $\mathcal{Z} \cap \mathcal{N}$ *and* $\mathcal{S} \cap \mathcal{Q}$ *must be r.e. sets. Let* $\overline{\mathcal{Z}} = \mathcal{S}/\mathcal{Z}$ *be the complement of* $\mathcal{Z}$ *in* $\mathcal{S}$*. It is possible to show that if* $\mathcal{Z} \subseteq \mathcal{S}$ *is any* $\mathcal{N}$*-regular set, then both* $\overline{\mathcal{Z}} \cap \mathcal{N}$ *and* $\overline{\mathcal{Z}} \cap \mathcal{Q}$ *are infinite sets. In other words, the support of* *any admissible probability distribution must exclude infinitely many non-cooperative Turing machines and infinitely many quasi-cooperative Turing machines.*

PROOF: We have to show that both $\overline{\mathcal{Z}} \cap \mathcal{N}$ and $\overline{\mathcal{S}} \cap \mathcal{Q}$ are infinite whenever $\mathcal{Z}$ is an $\mathcal{N}$-regular set. The argument is identical for both cases, so we only show that $\overline{\mathcal{Z}} \cap \mathcal{N}$ must be infinite.

From Theorem A.9 we know that $\mathcal{N}$ is not r.e. Suppose now that $\overline{\mathcal{Z}} \cap \mathcal{N}$ is either empty or finite. Then by Definition 4, $\overline{\mathcal{Z}} \cap \mathcal{N}$ is r.e. Since $\mathcal{Z}$ is $\mathcal{N}$-regular, by definition $\mathcal{Z} \cap \mathcal{N}$ is r.e. Since by Theorem A.7 the union of two r.e. sets is r.e., it follows that $[\overline{\mathcal{Z}} \cap \mathcal{N}] \cup [\mathcal{Z} \cap \mathcal{N}]$ is r.e. But the latter implies that $\mathcal{N}$ is r.e. This contradiction establishes the claim.

Q.E.D.

The following shows that the limit to how 'large' the support of an admissible distribution can be is entirely determined by the requirement that its support be $\mathcal{N}$-regular, and hence by the characterization in Remark 4 above.

REMARK 5: *Given any $\mathcal{Z} \subset \mathcal{S}$ which is $\mathcal{N}$-regular, there exists an admissible probability distribution $P \in \mathcal{P}$ such that* $\text{supp}(P) = \mathcal{Z}$.

PROOF: By Lemma A.1, if $\mathcal{Z}$ is $\mathcal{N}$-regular, then it is r.e. Given a machine which enumerates it, a construction like the one in (25) in the proof of Lemma 5 is enough to show that a computable probability distribution with support precisely equal to $\mathcal{Z}$ exists. This is enough to prove the claim.

Q.E.D.

Putting together Remarks 3 and 5 immediately gives the following characterizations of the class of support sets to which our optimality results apply. We state the following two without proof since they are direct consequences of Remarks 3 and 5, Definition 4 and Theorem A.7.

REMARK 6: *Let $\mathcal{H}$ be any r.e. subset of either $\mathcal{N}$ or of $\mathcal{Q}$. Then there exists an admissible probability distribution $P \in \mathcal{P}(\mathcal{R}^*)$ such that $\mathcal{H} \subseteq$* $\text{supp}(P)$.

REMARK 7: *Let $\mathcal{H}$ be any subset of $\mathcal{S}$, and ε' and ε'' be two real numbers in* $[0,1]$ *with $\varepsilon' > \varepsilon''$. Then there exists an admissible probability distribution $P \in \mathcal{P}(\mathcal{R}^*)$ such that $\varepsilon'' < P(\mathcal{H}) < \varepsilon'$.*

7. RELATED LITERATURE AND DISCUSSION OF RESULTS

7.1. Computing Devices

The idea that players in a game can be thought of as computing devices is not new. Aumann (1981), Neyman (1985), Rubinstein (1986) and Abreu and Rubinstein (1988) (to name a few) are early contributions which, in the context of repeated games, consider strategies which can be implemented by a class of computing devices known as finite automata (Moore machines). This literature concentrates mostly on the Nash equilibrium of the machine game: a choice of machine for each player such that no player can improve his payoff by choosing a different machine, given the machines attributed to other players and a lexicographic cost of using larger automata.

Following some pioneering work of Binmore (1987), we adopt the stance that Turing machines are an important 'benchmark' class of computing devices to consider. This is defensible in fairly general terms, but we do not attempt this here. Binmore (1987) and Anderlini (1989) elaborate at length on the fact that the general computability framework is an interesting one in many game-theoretic situations since it embodies the widest possible intuitive notion of effective computability. McAfee (1984), Megiddo (1986), Megiddo (1989), Megiddo and Widgerson (1987), Howard (1988), Spear (1989), Canning (1992) and Anderlini and Sabourian (1995a) also model players as Turing Machines.

7.2. Is Computability Essential?

Whether the computability framework we have used in this paper is essential to the results is clearly a matter for conjectures only at this stage. It is interesting however to attempt to isolate the features of this framework which are essential to our proof of the main result of the paper. First of all, computability affords us countable strategy spaces and therefore trembles which have the all the strong properties of probability distributions over the natural numbers.

In the proof of the Communication Lemma we have leaned heavily on three properties of the computability framework. The simulation possibility (Theorem A.2), which makes it possible for our revealing strategies to construct a message which is different from the message of 'sufficiently many' non-cooperative strategies. The parameterization result (Theorem A.1), which allows us to consider meaningfully the entire set of potential revealing strategies, as we change the parameters of the trembles of the CTHE. And finally the pseudo-fixed point result (Theorem A.3) which we have used to close the proof of the Communication Lemma.[23]

It is possible that other restrictions on strategies and perturbations of a game with pre-play communication which ensure that these features hold may yield equilibrium selection results analogous to the ones we have presented here.

7.3. The Work of Howard and McAfee

McAfee (1984) and more extensively Howard (1988), analyze a framework in which Turing machines are required first to elicit their correct Gödel number to their opponent, and then to play a strategic form game — a version of the Prisoners' Dilemma. Their results indicate that cooperation will emerge as the only outcome of evolutionarily stable strategies (ESS) in such set up. Loosely speaking, the same pseudo-fixed point result we have used in this paper guarantees the existence of 'mutants' which will be able to recognize another identical mutant when they meet. By cooperating with identical mutants (which being identical obviously reciprocate) and defecting against any other machine, such mutants are able to destroy as ESS any non-cooperative equilibrium.

Their results are close but fundamentally different from the results of this paper. The key difference is that in their model machines are required to elicit their *correct* Gödel number, while here the messages which the machines exchange at least in principle may or may not reveal anything about the machines which have produced

---

[23] It is interesting to note that (cf. Rogers (1967)), the Recursion Theorem (of which Theorem A.3 is a corollary) is true in any system in which Theorem A.1 and Theorem A.2 hold.

them. The central result of this paper is that even when the messages are at least in principle *ambiguous*, if certain conditions are satisfied, they will be able to convey enough information to trigger coordination on the good equilibrium of a Common Interest game. Precisely because the messages in this paper are at least in principle ambiguous, our results do not apply to an underlying game like Prisoner's Dilemma, while the model used by McAfee (1984) and Howard (1988) does generate cooperation when the underlying game is a Common Interest game.

7.4. Cheap Talk

The Pareto-inefficient equilibrium of the strategic-form game discussed in the introduction (Figure 1) survives all the 'standard' refinements of Nash equilibrium put forward in the literature, with the important exception of the 'Cheap Talk' literature.

Farrel (1983) and Farrell (1988) and a number of subsequent contributions, propose a refinement of Nash equilibrium which can, broadly speaking, be justified in terms of the following intuitive story. Before the game is played, one player can make a suggestion to the other player on how to play the game. Only suggestions which are consistent (self-enforcing) can be expected to be followed. In addition one imagines that if a suggestion is both consistent and to the advantage of both players, then it will necessarily be followed. It is then relatively easy to conclude that in the game discussed in the introduction the only viable equilibrium is in fact (α_1, α_2).

One crucial point to note about the cheap-talk literature is that the appeal to a pre-play communication stage is only an intuitive and suggestive way to justify a particular refinement of the Nash equilibrium concept. Pre-play communication is not explicitly modeled, but rather its effects embodied in an appealing set of axioms which are then used to single out a subset of the Nash equilibria as the viable ones.

More recently (and mostly after the original version of our results) the analysis of solution concepts related to evolutionary stability has shown that cheap talk in general matters, and yields cooperation when the underlying game is a Common Interest game (or satisfies some stronger version of this property). We recall the contributions by Matsui (1991), Kim and Sobel (1995), Sobel (1993) and Wärneryd (1993). The cooperation results in these papers rely on set valued solution concepts which require the solution to be robust to an ‘invasion’ of ‘mutants’ which take up un-used messages to recognize each other and trigger cooperation. These models are fundamentally different from ours in the sense that cooperation is supported (as in the work of McAfee (1984) and Howard (1988)) by some type of ‘secret hand-shake’ argument, to which the set valued solution concepts give bite; by contrast, our results are driven by the possibility of genuine revelation of a player’s cooperative intentions.

7.5. Repeated Games

In an infinitely repeated game with little or no discounting any finite portion of the history of play is unimportant as far as the players’ long-run payoffs are concerned. One question that naturally arises from the analysis carried out in this paper is whether in the context of an infinitely repeated game with little or no discounting, is it possible that the players use the ‘early stages’ of the game as a means of communicating to each other information about their long-run strategy in the remainder of the game? Suppose moreover that the stage game of the infinitely repeated interaction is a Common Interest game. Can such use of the early stages of the game as a communication device yield the selection of the Pareto-efficient payoffs in the infinitely repeated game?

The results in Anderlini and Sabourian (1995a) provide an affirmative answer. Using a solution concept close to the one we used here, the only equilibrium long-run payoffs possible in an infinitely repeated Common Interest game with little or no time discounting are the Pareto-efficient payoffs of the stage game. The statement of this result is indeed very close to that of the main theorem of Aumann and Sorin (1989). It is interesting to note that the equilibrium selection result in Anderlini and Sabourian (1995a) holds for discounted infinitely repeated Common Interest games as the discount factor approaches one. This points to the fact that while the our results in this paper hinge on the extreme assumption of communication being costless, they are also valid in the limit as the cost of communication approaches zero.

The results of this paper and those in Anderlini and Sabourian (1995a) can be viewed as the application of the same techniques to two different, but related, problems. The message spaces in this paper are unrestricted, while in a repeated game the players are restricted to choose their signals from the finite action sets of the stage game. It is interesting to observe that a countable message space gives the players the same ‘signaling power’ as an arbitrary sequence of actions in the early stages of the repeated game. This is due to the fact that players are restricted to computable strategies in both models.

The two-stage structure (pre-play communication and play) of our model does not allow for any ambiguity as to when communication ends, and (cooperative) play-begins. In a repeated game, the signalling protocol has to be capable of revealing a player’s cooperative intention, *and* of effectively conveying information about when the signalling ends and the cooperative play starts.

The limit results as the noise vanishes are sharper in this paper than in the context of a repeated game. Because the underlying game is finite, we find that for small but still positive noise, the equilibrium strategies in this paper are cooperative. This is because any non-cooperative outcome entails a utility loss which is bounded away from zero. In a repeated game this is not the case since the cooperative long-run payoffs can be approximated arbitrarily closely by strategies which play cooperatively ‘most of the time’. As a result, in Anderlini and Sabourian (1995b) the long-run payoffs of the equilibrium strategies are only guaranteed to approach the cooperative level as the weight of the perturbations actually approaches zero.

APPENDIX

We start with some Theorems and Definitions which are standard in the computability literature. All the results which are stated without proof can be found, for instance, in Cutland (1980).

THEOREM A.1 *(s-m-n): For each $m \geq 0$ and $n \geq 1$ there exists a total computable function of $m+1$ variables $s^{(m+1)} \in \mathcal{F}^{(m+1)}$ such that $\forall\ e \in \mathbb{N}$ and $\forall\ (h_1, \cdots, h_m, \cdots, h_{m+n}) \in \mathbb{N}^{m+n}$ we have*

$$\varphi_e^{(m+n)}(h_1, \cdots, h_{m+n}) \simeq \varphi_{s^{(m+1)}(e,h_1,\cdots,h_m)}(h_{m+1}, \cdots, h_{m+n})$$

THEOREM A.2 *(Universal Turing Machine): Given any $m \geq 1$, there exists a number u_m, such that*

$$\varphi_{u_m}^{(m+1)}(n, e_1, \cdots, e_m) \simeq \varphi_n^{(m)}(e_1, \cdots, e_m)\ \ \forall\ (n; e_1, \cdots, e_m) \in \mathbb{N}^{m+1}$$

THEOREM A.3 *(Pseudo-Fixed Point): For any computable function $f^{(m+1)} \in \mathcal{F}^{(m+1)}$ of $m+1$ variables, there exists $\hat{x} \in \mathbb{N}$ such that*

$$\varphi_{\hat{x}}^{(m)}(e_1, \cdots, e_m) \simeq f^{(m+1)}(\hat{x}, e_1, \cdots, e_m)\ \ \forall\ (e_1, \cdots, e_m) \in \mathbb{N}^m$$

THEOREM A.4: *A set is r.e. if and only if it is equal to the domain of a computable function. Formally, $\mathcal{Z} \subseteq \mathbb{N}$ is r.e. if and only if for some $m \in \mathbb{N}$ we have $\varphi_m(n) \downarrow \Leftrightarrow\ n \in \mathcal{Z}$.*

THEOREM A.5: *Any finite set $\mathcal{Z} \subset \mathbb{N}$ can be enumerated by a Turing machine $\varphi_s(\cdot)$ in the following way. $\varphi_s(e) \downarrow \Leftrightarrow 0 \leq e \leq ||\mathcal{Z} - 1||$, $\forall n \in \mathcal{Z}\ \exists e \leq ||\mathcal{Z} - 1||$ such that $\varphi_s(e) = n$.*

THEOREM A.6: *A set $\mathcal{Z} \subseteq \mathbb{N}$ is r.e. if and only if it is the range of a computable function of $m \geq 1$ variables. Formally $\mathcal{Z} \subseteq \mathbb{N}$ is r.e. if and only if there exists a Turing machine $\varphi_s^{(m)}(\cdot)$ such that*

$$n \in \mathcal{Z}\ \ \Leftrightarrow\ \ \exists(e_1, \cdots, e_m) \text{ such that } \varphi_s^{(m)}(e_1, \cdots, e_m) = n$$

THEOREM A.7: *The intersection of two r.e. sets is r.e. The union of two r.e. sets is r.e.*

THEOREM A.8: *Suppose that* $f : \mathbb{N} \to \mathbb{N}$ *is such that* $\exists \overline{n} \in \mathbb{N}$ *and* $\exists c \in \mathbb{N}$ *such that* $f(n) = c \ \forall n > \overline{n}$. *Then* f *is computable.*

The three claims in the following Theorem are an immediate consequence of a standard result known as the Rice-Shapiro Theorem (Cutland 1980, Thm.7.2.16).

THEOREM A.9: *The sets* $\mathcal{S}$, $\mathcal{N}$, $\mathcal{Q}$ *and* $\mathcal{C}$ *of Definition 2 are not r.e.*

LEMMA A.1: *Let* $\mathcal{Z} \subseteq \mathcal{S}$ *be a* $\mathbb{N}$*-regular set in the sense of Definition 5. Then* $\mathcal{Z}$ *is r.e.*

PROOF: Notice that $\mathcal{Z} = [\mathcal{N} \cap \mathcal{Z}] \cup [\mathcal{Q} \cap \mathcal{Z}]$ and both $\mathcal{N} \cap \mathcal{Z}$ and $\mathcal{Q} \cap \mathcal{Z}$ are r.e. by the fact that $\mathcal{Z}$ is $\mathcal{N}$-regular. Therefore, by Theorem A.7, $\mathcal{Z}$ is r.e.

Q.E.D.

PROOF OF REMARK 1: If either $\text{supp}(P) \cap \mathcal{N}$ or $\text{supp}(P) \cap \mathcal{Q}$, or both, are finite sets, the claim is a trivial consequence of the fact that P is computable. We omit the details for this case and assume that both sets are infinite.

Let c and d be two Turing machines which respectively enumerate the infinite sets $\mathcal{N} \cap \text{supp}(P)$ and $\mathcal{Q} \cap \text{supp}(P)$ as in definition 4. Let p compute the values of $P(n)$ as in (9). By Church's thesis we can then construct two Turing machines a and b such that $\forall e \in \mathbb{N}$

$$\varphi_a(e) = \sum_{i=1}^{e} \varphi_p(\varphi_c(i)) \ \text{ and } \ \varphi_b(e) = \sum_{i=1}^{e} \varphi_p(\varphi_d(i))$$

Let $\varphi_s(e) = \varphi_a(e) + \varphi_b(e)$. Clearly $\varphi_s(e)$ is a monotonic function of e and $\lim_{e \to \infty} \varphi_s(e) = 1$. Therefore, by Church's thesis again, we can find $g \in \mathbb{N}$ such that, $\forall x \in \mathbb{N}$

$$\begin{aligned} \varphi_g(x) \quad &= \quad \min n \in \mathbb{N} \\ & \quad\quad \text{s.t. } \varphi_s(e) > 1 - \frac{1}{x} \end{aligned} \tag{A.1}$$

To complete the proof it is now enough to set $\varphi_q(x) = \varphi_a(\varphi_g(x))$ and $\varphi_m(x) = \varphi_b(\varphi_g(x))$, which by (A.1) clearly have the desired property.

Q.E.D.

Proof of Lemma 1: By Theorem A.2 and by Church's thesis we can find $a \in \mathbb{N}$ such that $\forall (n,p,c,q,x,t) \in \mathbb{N}^6$, with (p,c,q) forming a basis for some $P \in \mathcal{P}$, and such that $\varphi_p(n) \downarrow > 0$ we have

$$\varphi_a^{(6)}(n,p,c,q,x,t) = \frac{1}{x}\varphi_p(n)$$

Let $\hat{t} \in \mathbb{N}$ be a number such that $\hat{t} > 2x/\varphi_p(n)$, and note that such a number can easily be effectively computed from the inputs (n,p,c,q,x,t). By Church's thesis and applying Theorem A.2 twice, we can find $b \in \mathbb{N}$ such that (note that the computation need not be always defined)

$$\varphi_b^{(6)}(n,p,c,q,x,t) = \begin{cases} \varphi_q(\hat{t}) + \frac{1}{\hat{t}} & \text{if } t = 0 \\ \\ \varphi_q(\hat{t}+t) + \frac{1}{\hat{t}+t} - \sum_{\tau=0}^{t} \varphi_p(\varphi_c(\tau)) & \text{if } t \geq 1 \end{cases}$$

Notice that $\varphi_b^{(6)}(n,p,c,q,x,t) \downarrow$ provided that $\varphi_c(\tau) \downarrow \forall \tau \leq t$. Set now the function $t^{(5)} \in \mathcal{F}^{(5)}$ of the statement of the Lemma to be as follows

$$t^{(5)}(n,p,c,q,x) \equiv \overline{n} \equiv \min\Big\{ t \in \mathbb{N} \mid \varphi_a^{(6)}(n,p,c,q,x,t) > \varphi_b^{(6)}(n,p,c,q,x,t) \Big\} \qquad \text{(A.2)}$$

By Church's thesis $t^{(5)}$ is clearly a computable function since successively larger values of t can be tried in the computations $\varphi_a^{(6)}(\cdot)$ and $\varphi_b^{(6)}(\cdot)$ until $\overline{n}$ as in (A.2) is found (if one exists). It remains to show that $t^{(5)}$ is defined whenever (p,c,q) form the basis for some admissible probability distribution $P \in \mathcal{P}$ and $\varphi_p(n) > 0$, and that it has the desired property. That $\overline{n}$ has the desired property is obvious once we notice that $\varphi_q(1/(\hat{t}+t)) + 1/(\hat{t}+t) > P(\mathcal{N})$ $\forall t \in \mathbb{N}$. Notice further that because of the way $\hat{t}$ is chosen we also have that

$$\varphi_q(\hat{t}+t) + \frac{1}{\hat{t}+t} - P(\mathcal{N}) < \frac{1}{x}\varphi_p(n) \;\; \forall\, t \in \mathbb{N} \qquad \text{(A.3)}$$

To see that $t^{(5)}$ is defined for the appropriate input values, consider three distinct cases. Suppose first that $\mathcal{N} \cap \operatorname{supp}(P)$ is empty. Then $\varphi_c(\tau) \uparrow \forall \tau \in \mathbb{N}$. However, because of the way $\hat{t}$ is chosen (see (A.3)), it must be that $\varphi_a^{(6)}(n,p,c,q,x,0) > \varphi_6(b,n,p,c,q,x,0)$, so that $t = 0$ will satisfy (A.2). Suppose next that $\mathcal{N} \cap \operatorname{supp}(P)$ is a finite set. Then c enumerates it without repetitions as in Theorem A.5. Using (A.3) again, it is clear that the inequality in (A.2) must be satisfied for some $t \leq ||\mathcal{N} \cap \operatorname{supp}(P)|| - 1$, and therefore $t^{(5)}$ is defined. The third and last case is when $\mathcal{N} \cap \operatorname{supp}(P)$ is an infinite set. In this case, c computes a total function, and therefore $\varphi_b^{(6)}(n,p,c,q,x,t) \downarrow \forall t \in \mathbb{N}$. Since $\lim_{t\to\infty} \varphi_b^{(6)}(n,p,c,q,x,t) = 0$, it follows that $t^{(5)}$ as in (A.2) is defined in this case.

Q.E.D.

Proof of Lemma 2: A machine a which computes $d^{(7)}$ can be constructed as follows. Start by computing the value of $\overline{n}$ as in Lemma 1. If this computation does not halt, leave the output of a undefined. If this computation halts, proceed further as follows.

Applying Theorem A.2 twice, it is feasible to compute the result of $\varphi_{\varphi_c(t)}(\bullet)$, $\forall t \leq \overline{n}$. If any of these computations do not halt, leave the output of a undefined. If all the computations required to obtain $\varphi_{\varphi_c(t)}(\bullet)$ halt, proceed to check whether $m = \bullet$ or not. If $m \neq \bullet$, it is clearly feasible to simply output the cooperative action a_i^e, irrespective of the other inputs. Suppose now that $m = \bullet$. Then by Church's thesis it is clearly feasible to set the output of $\varphi_a^{(7)}(n,p,c,q,\hat{m},x,m)$ to be equal to $m^* = 1 + \hat{m} + \sum_{t=0}^{\overline{n}} \varphi_{\varphi_c(t)}(\bullet)$.

Q.E.D.

REFERENCES

Abreu, D., and A. Rubinstein (1988): "The Structure of Nash Equilibrium in Repeated Games with Finite Automata," *Econometrica*, 56, 1259–81.

Anderlini, L. (1989): "Some Notes on Church's Thesis and the Theory of Games," *Theory and Decision*, 29, 19–52.

——— (1990): "Communication, Computability and Common Interest Games," Economic Theory Discussion Paper 159, Department of Applied Economics, University of Cambridge.

Anderlini, L., and H. Sabourian (1995a): "Cooperation and Effective Computability," *Econometrica,* forthcoming.

——— (1995b): "The Evolution of Algorithmic Learning Rules: A Global Convergence Result," University of Cambridge, mimeo.

Aumann, R. J. (1981): "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern*, pp. 11–42. Mannheim: Bibliographisches Institüt.

Aumann, R. J., and S. Sorin (1989): "Cooperation and Bounded Recall," *Games and Economic Behavior*, 1, 5–39.

Binmore, K. (1987): "Modeling Rational Players: Part I," *Economics and Philosophy*, 3, 179–214.

Canning, D. (1992): "Average Behavior in Learning Models," *Journal of Economic Theory*, 57, 442–72.

Cutland, N. J. (1980): *Computability: An Introduction to Recursive Function Theory.* Cambridge: Cambridge University Press.

Davis, M. (1958): *Computability and Unsolvability.* New York: Dover Publications.

Farrel, J. (1983): "Meaning and Credibility in Cheap-Talk Games," University of California Berkeley, mimeo.

Farrell, J. (1988): "Communication, Coordination and Nash Equilibrium," *Economics Letters*, 27, 209–14.

Harsanyi, J. C., and R. Selten (1988): *A General Theory of Equilibrium Selection in Games.* Cambridge, Mass. and London: MIT Press.

Howard, J. V. (1988): "Cooperation in the Prisoners' Dilemma," *Theory and Decision*, 24, 203–213.

Kim, Y. G., and J. Sobel (1995): "An Evolutionary Approach to Pre-Play Communication," *Econometrica*, 63, 1181–1193.

Matsui, A. (1991): "Cheap Talk and Cooperation in a Society," *Journal of Economic Theory*, 54, 245–58.

McAfee, R. P. (1984): "Effective Computability in Economic Decisions," University of Western Ontario, mimeo.

Megiddo, N. (1986): "Remarks on Bounded Rationality," Research Report RJ5270, IBM Almaden Research Center, San Jose, California.

——— (1989): "On Computable Beliefs of Rational Machines," *Games and Economic Behavior*, 1, 144–69.

Megiddo, N., and R. Widgerson (1987): "Turing Machines Play the Finitely Repeated Prisoners' Dilemma," IBM Almaden Research Center, San Jose California, mimeo.

Myerson, R. (1978): "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7, 59–90.

Neyman, A. (1985): "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma," *Economics Letters*, 19, 227–229.

Rogers, H. (1967): *Theory of Recursive Functions and Effective Computability.* London: McGraw-Hill Book Company.

Rubinstein, A. (1986): "Finite Automata Play the Repeated Prisoner's Dilemma," *Journal of Economic Theory*, 39, 83–96.

Selten, R. (1975): "A Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

Sobel, J. (1993): "Evolutionary Stability and Efficiency," *Economics Letters*, 42, 301–12.

Spear, S. E. (1989): "Learning Rational Expectations under Computability Constraints," *Econometrica*, 57, 889–910.

Wärneryd, K. (1993): "Cheap Talk, Coordination, and Evolutionary Stability," *Games and Economic Behavior*, 5, 532–46.