GOVERNANCE AND THE EFFICIENCY
OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 438

Commitments, Intentions, Truth and Nash Equilibria

Karl H. Schlag*
Péter Vida**

* University of Vienna
** University of Mannheim

November 11, 2013

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Sonderforschungsbereich/Transregio 15 · www.sfbtr15.de
Universität Mannheim · Freie Universität Berlin · Humboldt-Universität zu Berlin · Ludwig-Maximilians-Universität München
Rheinische Friedrich-Wilhelms-Universität Bonn · Zentrum für Europäische Wirtschaftsforschung Mannheim

Speaker: Prof. Dr. Klaus M. Schmidt · Department of Economics · University of Munich · D-80539 Munich,
Phone: +49(89)2180 2250 · Fax: +49(89)2180 3510

Commitments, Intentions, Truth and Nash Equilibria*

Karl H. Schlag†
University of Vienna

Péter Vida,‡
University of Mannheim

November 11, 2013

Abstract

Games with multiple Nash equilibria are believed to be easier to play if players can communicate. We present a simple model of communication in games and investigate the importance of when communication takes place. Sending a message before play captures talk about intentions, after play captures talk about past commitments. We focus on equilibria where messages are believed whenever possible. Applying our results to Aumann's Stag Hunt game we find that communication is useless if talk is about commitments, while the efficient outcome is selected if talk is about intentions. This confirms intuition and empirical findings in the literature.

We develop a theory of credible communication under complete information and connect it to the notion of credibility in standard sender-receiver games.

Keywords: Pre-play communication, cheap talk, credibility, coordination, sender-receiver games.

JEL Classification Numbers: C72, D83.

*Thanks... Péter Vida received financial support from SFB/TR 15 which is gratefully acknowledged.

†University of Vienna, Department of Economics. E-mail: karl.schlag@univie.ac.at

‡Corresponding author. Tel.:+49(0)621 181 3059; University of Mannheim, Department of Economics, L7,3-5, D-68131 Mannheim, Germany. E-mail: vidapet@gmail.com

1 Introduction

Game theory is agnostic about how to play in games that have multiple Nash equilibria. Beliefs can be mutually self-confirming when all believe that others focus on an inefficient equilibrium even if there are alternative Nash equilibria where all are strictly better off. Yet, it is commonly believed that inefficient equilibria will not be played when players are allowed to communicate before they play the game. The reasoning is that it suffices that one player proposes an equilibrium outcome in which all players are better off to upset beliefs associated to inefficient play (see (Rabin,1994)).

At the same time Aumann (1990) claims that communication can be useless even in the simplest games, and illustrates this informally in a version of the Stag Hunt game. Farrell (1988)[1] objects and argues for this game that it depends on when communication takes place. If communication occurs after the person communicating has made a choice then he agrees. However, if communication occurs before making a choice then he argues that communication will lead all players to hunt the stag. Charness (2000) runs experiments for this game that reinforce the intuition of Farrell.

We present a simple formal framework to examine credible communication, where players are believed whenever possible, in two person normal form games. Simply adding cheap talk will not reduce the set of equilibrium outcomes. A necessary condition for upsetting beliefs supporting an inefficient equilibrium is that alternative proposals can be made. These would be initiated by sending unanticipated messages, naturally accompanied by an explanation of the circumstances surrounding the new proposal. One would also explain which messages one would have sent if one had other intentions or the circumstances would be different. For communication to then be successful the parties involved, both those that talk and those that listen, must be able and willing to rethink their intentions.

We embed these ingredients into a standard game theory analysis, by setting up the rules of communication and adding features to the strategic interaction that capture what happens when one explains behavior under alternative circumstances. In our analysis we then only consider those equilibria where messages are believed whenever possible.

We consider two different ways of adding communication to a two-player bimatrix game. In both cases the communication protocol is chosen to be as simple as possible, player one sends a single message and player two just listens. In the first variation, player one sends the message before either has chosen an action (referred to as Talk and Play, abbreviated by TP). In the second variation player one sends the message after he has chosen his own action, which is not observable by player two, and before player two has made a choice (Play then Talk, PT).

[1] This is based on earlier personal communication on this matter, see Farell (1988).

In both variations the message is sent from a given language, which is chosen by one of the players, who we call the interpreter. A language is defined as *a partition*[2] *of player one's action set.* The elements of the partitions are the messages among which player one can choose one and send it to player two. There are two languages of particular interest. We call the language *complete communication* if each action of player one is associated to a message, which means that the language is given by the finest partition of the set of actions. We refer to *no communication* if the language only has a single message, this message is then equal to the entire set of actions.

Given a fixed language, our solution concepts for the enlarged games stipulates that *all* (even out of equilibrium) messages of the language are truthful and believed by player two if it is *possible*, otherwise all messages are ignored. More precisely, if, given that player two believes that player one tells the truth (in PT about which action he has chosen, in TP which action he will choose) and player one has no strict incentives to lie, then we require that all messages from the given language are believed by player two and that player one tells the truth. However, as messages can be vague about which action will be (TP) or has been (PT) chosen, in games in which player one has more than two actions, credibility will depend on the beliefs of player two about the action chosen by player one that is consistent with truth-telling for the given message.

Truth-telling in TP means that player one chooses an action within the message he has sent. Truth-telling in PT means that player one sends the message which contains the action he has chosen. Player two believes a message if her belief is supported within that message. We say that a language is *credible* if there is a weak-perfect Bayesian equilibrium[3] of the enlarged game (in which that language is fixed) such that player one *always* tells the truth and player two believes *each* message of player one.

The equilibrium concepts when the language is not fixed but is chosen by the interpreter are called TPE and PTE respectively. Both TPE and PTE are defined just simply as weak-perfect Bayesian Nash equilibria of the enlarged games such that if a credible language is chosen then player one must tell the truth and player two must (correctly) believe it. In case of non-credible languages player two plays as if the language no communication was chosen, that is, she ignores all messages coming from non-credible languages.

We now return to our motivating question, whether communication leads to

[2]See the discussion where we justify this assumption using Rabin's (1990) Message Profile Theory.

[3]In fact, under TP we require correct beliefs out of equilibrium, hence we require a subgame perfect equilibrium. See the discussion on this issue.

efficiency. Consider the Aumann's Stag Hunt game:

		Player two	
		S	R
Player one	S	$9,9$	$0,8$
	R	$8,0$	$7,7$

The analysis of this game reveals that communication helps players to coordinate on hunting the stag under TP but that it is useless, and hence unable to refine the set of Nash equilibrium outcomes, under PT. Consider TP. If player one says Stag and player two believes it then player two plays Stag and hence player one plays Stag. If player one says Rabbit and player two believes it then player two plays Rabbit and hence player one plays Rabbit. Hence complete communication is credible. No communication is always credible. Now no matter how players play when the language no communication was chosen, player one as the interpreter can choose the language complete communication send the message Stag and receives the payoff 9. In any TP equilibrium players go for the stag. On the contrary, complete communication is not credible under PT. If player one has chosen Rabbit, and player two believes him, then it is optimal for him to lie and to send the message Stag. Only no communication is credible and PT does not refine away any of the Nash equilibria.

This result confirms the intuition of Farrell (1988) and the findings of Charness (2000) and does not depend on which player is assigned as the interpreter (which is not true in general). In particular, communication does not necessarily lead to efficient outcomes under PT but it does in this game under TP.

Interestingly we also find that efficiency need not result under TP. We present a simple 3 by 3 game of common interest in which the action associated to the unique efficient outcome is contained in the support of any Nash equilibrium. The message that implies that player one will not choose this action is not believable. Consequently only no communication is credible and inefficient play can be supported. On the other hand, if one relaxes the definition of credibility and allows (see discussion in Section 7.1) that out of equilibrium beliefs of player two are different from what player one chooses then efficiency obtains. Yet neither game form TP or PT is superior for inducing efficient outcomes. For instance, efficiency emerges in Aumann's Stag Hunt game with talk about intentions (TP) but not about commitments (PT). On the other hand, efficiency emerges in this 3 by 3 game of common interest under talk about commitments but not under talk about intentions.

The structure of the paper is as follows. Section 2 introduces some basic notations, the notion of languages and messages. In section 3 we describe the TP game. In section 4 we describe the PT game. In section 5 we define credibility of a language under TP and PT and define our solution concepts TPE and PTE and prove their existence. Section 6 contains propositions about selecting Nash equilibria, sufficient condition for efficiency, the power of the sender and an example demonstrating the power of the interpreter. We also present a simple game in which TPE does not yield to efficient outcome but PTE does so. Section 7 contains the discussion. We give a weaker version of credibility under TP and show that it yield to efficient outcome in the previous example, but generally does not do so. We connect our notion of credibility under PT to the credibility notion of Rabin (1990) by weakening credibility under PT. We show examples in which Rabin's (1990) notion is too weak and in which it is too strong compared to our definition and suggest a stronger version of credibility under PT. We also discuss the related literature. Section 8 concludes.

2 Preliminaries

2.1 The Underlying Game

Let Γ be a two player (player one (he), player two (she)) simultaneous move game with finite action sets S_j and von Neumann-Morgenstern utility functions defined by the Bernoulli utilities $u_j : S_1 \times S_2 \to \mathbb{R}$ for player $j = 1, 2$. For a finite set X let ΔX be the set of probability distributions over X and let $C(\xi) = \{x \in X : \xi(x) > 0\}$ be the support of $\xi \in \Delta X$. $z \in \mathbb{R}^2$ is a Nash equilibrium outcome if there is a Nash equilibrium $\sigma \in \Delta S_1 \times \Delta S_2$ of Γ such that $u_j(\sigma) = z_j$ for $j = 1, 2$. z^* is the favorite (pure) Nash equilibrium outcome for player j if there is no (pure) Nash equilibrium outcome z such that $z_j > z_j^*$.

Formally, messages are elements of a partition L of S_1. This partition is called a language. Formally, $L = \{m | m \subseteq S_1\}$ is a *language* if $\forall s_1 \in S_1, \exists! m \in L$ such that $s_1 \in m$. The set of all languages is denoted by $\mathcal{L}$. Languages will be chosen by the interpreter who is one of the two players. We allow for randomizing over languages, hence choices in $\Delta\mathcal{L}$. A message from L is $m \in L$ and $L(s_1) \in L$ denotes the message which contains the action s_1. The degenerate language $\{S_1\}$ that contains a single element can be interpreted as there being no communication. At the opposite extreme, the language that contains only singletons, so $L(s_1) = \{s_1\}$ for all $s_1 \in S_1$, may be interpreted as complete communication. These two languages will thus be referred to as "no communication" and "complete communication".

We consider two scenarios for when communication takes place. In "first talk then play" player one first sends a message to player two and then both simultaneously play Γ. In "first play then talk" player one first privately chooses an action in Γ and then sends a message to player two after which player two chooses an action in Γ.

3 First Talk Then Play

We first model communication that occurs before either player chooses an action. First the interpreter chooses the language L. Then player one sends a message m from this language L. Conditional on the chosen language and the sent message player one chooses an action which is not observed by player two. Finally player two chooses an action.

The above defines the following game, denoted by Γ_i^{TP} for $i = 1, 2$:

1. Player i (the interpreter) chooses a language $L \in \mathcal{L}$ and communicates it to the other player.

2. Player one sends a message $m \in L$ to player two.

3. Player one chooses an action s_1 (non-observable for player two)

4. Player two chooses an action s_2.

5. Payoffs are realized, where player j receives payoff $u_j(s_1, s_2)$, $j = 1, 2$.

Let us denote by $\Gamma^{TP}(L)$ the game in which L is given and starts with stage 2.

3.1 The Strategies in Γ_i^{TP}

We now introduce the notation for the possibly mixed strategies used in Γ_i^{TP}. Let L_i be the mixed language choice of the interpreter in stage 1, so $L_i \in \Delta\mathcal{L}$. We call L_i *degenerate* if L_i puts all weight on a single language. Given language $L \in \mathcal{L}$ chosen by the interpreter in stage 1 let $m_1^L \in \Delta L$ be the mixed message sent by player one in stage 2 and let $m_1 = (m_1^L)_{L \in \mathcal{L}}$. Let $\sigma_1^L(m)$ be the mixed action of player one in stage 3 after message $m \in L$ has been sent in stage 2, so $\sigma_1^L : L \to \Delta S_1$. Concerning player two, let $\sigma_2^L(m)$ be the mixed action of player two in stage 3 given the language L chosen by the interpreter in stage 1 and the message m received in stage 2, so $\sigma_2^L : L \to \Delta S_2$. We write $\sigma_j = (\sigma_j^L)_{L \in \mathcal{L}}$ for $j = 1, 2$. Hence, a strategy profile in the game Γ_i^{TP} is a tuple $(L_i, m_1, \sigma_1, \sigma_2)$.

4 First Play then Talk

In this scenario we model communication that takes place after player one has chosen an action. It is analogous to Γ_i^{TP} except the choice of player one is moved from stage 3 to stage 1. Consider the following game, denoted by Γ_i^{PT} for $i = 1, 2$:

1. Player one chooses a mixed action $\sigma_1 \in \Delta S_1$ and privately observes its realization, an action $s_1 \in C(\sigma_1)$.

2. Player i (the interpreter) publicly chooses a language $L \in \mathcal{L}$.

3. Player one sends a message $m \in L$ to player two.

4. Player two chooses an action $s_2 \in S_2$.

5. Payoffs are realized, where player j receives payoff $u_j(s_1, s_2)$, $j = 1, 2$.

Let us denote by $\Gamma^{PT}(L)$ the game above in which the interpreter *has to* choose L in stage 2, that is L is fixed.

4.1 The Strategies in Γ_i^{PT}

Let $\sigma_1 \in \Delta S_1$ be the mixed action of player one in stage 1. For $i = 1$ let $L_1(s_1)$ be the mixed language chosen in stage 2 after action s_1 has been realized in stage 1, $L_1 : S_1 \to \Delta\mathcal{L}$. If player two is the interpreter then $L_2 \in \Delta\mathcal{L}$ is independent from σ_1. In equilibrium we will concentrate on language choices which are independent of the realization of the equilibrium action and always put all weight on a single language. We say that L_1 is *degenerate and independent* of σ_1 if $L_1(s_1)$ is deterministic for all $s_1 \in S_1$ and for all $s_1', s_1'' \in C(\sigma_1)$ we have that $L_1(s_1') = L_1(s_1'')$. Notice that we allow that $L_1(s_1') \neq L_1(s_1'')$ for $(s_1', s_1'') \notin C(\sigma_1) \times C(\sigma_1)$. We say that L_2 is degenerate if L_2 is deterministic.

In stage 3, player one chooses a mixed message m_1^L belonging to the language L chosen in stage 2 given that action s_1 is the realization of σ_1 in stage 1, so $m_1^L : S_1 \to \Delta L$ and $m_1 = (m_1^L)_{L\in\mathcal{L}}$.

In stage 4, player two chooses a mixed action $\sigma_2^L(m)$ that depends on the language L chosen in stage 2 and on the message m received in stage 3, so $\sigma_2^L : L \to \Delta S_2$ and $\sigma_2 = (\sigma_2^L)_{L\in\mathcal{L}}$.

Hence a strategy profile in the game Γ_i^{PT} is described by $(\sigma_1, L_i, m_1, \sigma_2)$.

5 Solution Concepts

In this section we frequently refer to the notion of weak-Perfect Bayesian equilibrium (Mas Colell et al. (1995)). To fix notation let $\mu_2^L(m) \in \Delta S_1$ indicate player two's belief about player one's action after message $m \in L$. Let $\mu_2^L = (\mu_2^L(m))_{m\in L}$ and $\mu_2 = (\mu_2^L)_{L\in\mathcal{L}}$.

5.1 Credibility

Before defining equilibria in Γ_i^{TP} and Γ_i^{PT} we define the notion of credible languages under TP and PT.

Definition 1 *We say that a language L is* ***credible under TP*** *if there is a weak-Perfect Bayesian equilibrium $(m_1^L, \sigma_1^L, \sigma_2^L, \mu_2^L)$ of $\Gamma^{TP}(L)$ in which player one always tells the truth, and player two always correctly anticipates player one's action. That is:*

1. *for all $m \in L$, $C(\sigma_1^L(m)) \subseteq m$ and*
2. *for all $m \in L$, $\mu_2^L(m) \in \Delta m$,*
3. *for all $m \in L$, $\mu_2^L(m) = \sigma_1^L(m)$.*

Remark 1 *L is credible under TP if and only if there is a subgame perfect equilibrium $(m_1^L, \sigma_1^L, \sigma_2^L)$ of $\Gamma^{TP}(L)$ in which player one always tells the truth. Note that condition 2 is superfluous, however we keep it to clarify the role of condition 3, namely that we require in addition to telling the truth and believing that player two* always, *and not just on the equilibrium path, correctly anticipates player one's action (point 3). See the weaker definition without point 3 in the discussion in section 7.1.*

Definition 2 *We say that a language L is* ***credible under PT*** *if there is a weak-Perfect Bayesian equilibrium $(\sigma_1^L, m_1^L, \sigma_2^L, \mu_2^L)$ of $\Gamma^{PT}(L)$ in which player one tells the truth, and player two believes it. That is:*

1. *for all $s_1 \in S_1$, $L(s_1) \in argmax_{m \in L} u_1(s_1, \sigma_2^L(m))$ and*
2. *for all $m \in L$, $\mu_2^L(m) \in \Delta m$.*

The set of credible languages is denoted by $\mathcal{C}$.

5.2 TPE

We now present our equilibrium concept for TP. We search for a weak-PBE of Γ_i^{TP} in which communication is truthful and believed when the language is credible, and where messages are ignored otherwise. We denote by $\mu_2 = (\mu_2^L)_{L \in \mathcal{C}}$, where $\mu_2^L : L \to \Delta S_1$ and $\mu_2^L(m)$ indicates player two's belief about player one's action after language L and message $m \in L$.

Definition 3 (TPE) *$(L_i, m_1, \sigma_1, \sigma_2, \mu_2)$ is called a* ***talk then play equilibrium*** (TPE) *of Γ if it is a weak-Perfect Bayesian equilibrium of Γ_i^{TP} and:*

1. *L_i is degenerate and credible,*
2. *if L is credible then for all $m \in L$: $C(\sigma_1^L(m)) \subseteq m$ and $\mu_2^L(m) = \sigma_1^L(m)$ (truth-telling and correctly believing),*
3. *if L is not credible then: $\sigma_2^L(m) = \sigma_2^{\{S_1\}}$ for all $m \in L$ (ignorance).*

Remark 2 *A TPE is a subgame perfect equilibrium of of* Γ_i^{TP} *with truth-telling for credible languages and ignorance for non-credible languages.*

It is straightforward to show:

Proposition 1 (Existence of TPE) *For any* Γ *for* $i = 1, 2$ *there exists a TPE of* Γ*.*

5.3 PTE

Our equilibrium concept for PT is analogous to the one for TP. Communication is truthful and believed for credible languages, otherwise all messages are ignored.

Definition 4 (PTE) $(\sigma_1, L_i, m_1, \sigma_2, \mu_2)$ *is called a* ***play then talk equilibrium*** (PTE) *of* Γ *if it is a weak-Perfect Bayesian equilibrium of* Γ_i^{PT} *and:*

1. L_i *is degenerate, independent of* σ_1 *and credible,*
2. *if* L *is credible then: for all* $s_1 \in S_1$*,* $m_1^L(s_1) = L(s_1)$ *and for all* $m \in L$*,* $\mu_2^L(m) \in \Delta m$ *(truth-telling and believing),*
3. *if* L *is not credible then:* $\sigma_2^L(m) = \sigma_2^{\{S_1\}}$ *for all* $m \in L$ *(ignorance).*

Proposition 2 (Existence of PTE) *For any* Γ *for* $i = 1, 2$ *there exists a PTE of* Γ*.*

Proof: Let $i = 1$ and consider the favorite Nash equilibrium $(\sigma_1, \zeta_2) \in \Delta S_1 \times \Delta S_2$ of player one in Γ. Let no communication be the candidate for the equilibrium language. Consider the set $\mathcal{C}$ of credible languages. For any credible language $L \in \mathcal{C}$ which is different from no communication and for any $m \in L$ we define $\mu_2^L(m)$ in the following way. Consider one of the μ_2^L, σ_2^L associated to the weak-Perfect Bayesian equilibrium of $\Gamma^{PT}(L)$ under which L is credible. Consider the payoff $max_{s_1 \in S_1, m \in L} u_1(s_1, \sigma_2^L(m))$. This is a Nash equilibrium payoff for player one given credibility. In fact, this is the weak-Perfect Bayesian outcome of $\Gamma^{PT}(L)$

Since (σ_1, ζ_2) is the favorite Nash equilibria of player one, we can conclude that player one cannot benefit from deviating to another credible language than no communication, while choosing a different action than σ_1 and sending some message from that language.

Finally, we can set $\sigma_2^L(m) = \zeta_2$ for all $L \notin \mathcal{C}$ and for all $m \in L$ and $\sigma_2^{\{S_1\}} = \zeta_2$. Hence deviations to non credible languages is not profitable. For any $s_1 \notin C(\sigma_1)$ we can set $L_1(s_1) \in argmax_{L \in \mathcal{C}} u_1(s_1, \sigma_2^L(L(s_1)))$.

If $i = 2$ we can simply choose no communication as the equilibrium language and (σ_1, ζ_2) to be the favorite equilibrium of player two in Γ. ■

6 Further Propositions

Proposition 3 (Nash equilibrium) *For any Γ the PTE (TPE) outcomes of Γ, in which σ_1 (m_1) is pure, are Nash equilibria of Γ.*

Proof: Straightforward. ■

Remark 3 *If σ_1 (m_1) can be mixed then any PTE (TPE) outcome is in the convex hull of Nash equilibria. The proof is straightforward. Player one may choose a mixed action in PT and depending on the outcome of his randomization may send different messages. Player two on the equilibrium path correctly believes player one's action, hence, it must be that Nash equilibria are played after the different messages. This can be the case only if player one is indifferent between the two (or more) Nash equilibria. Similar argument shows that under what circumstances would player one choose random messages on the equilibrium path in TP.*

Some qualifications about the games Γ for which we state our propositions are needed. Given Γ let $NE(\Gamma)$ be the set of Nash equilibrium payoffs of Γ. When we say that a payoff profile or equilibrium is efficient we mean that there are no payoff profile in $NE(\Gamma)$ which (weakly) Pareto dominates it. Let us call $max_{s_1 \in S_1} u_1(s_1, b_2(s_1))$ the Stackelberg payoff of player one, where $b_2 : S_1 \to S_2$ is player two's best response function which is assumed to be unique. Assume that there is a unique favorite Nash equilibrium of player one.

Let us define $\bar{u}_1^{TP}(L), \bar{u}_1^{PT}(L)$ player one's worst subgame perfect, weak Perfect Bayesian equilibrium payoff in $\Gamma^{TP}(L), \Gamma^{PT}(L)$ (respectively) for some credible L in which player one tells the truth and player two believes him. Let $u_1^{TP} = max_{L \in \mathcal{C}} \bar{u}_1^{TP}(L)$ and $u_1^{PT} = max_{L \in \mathcal{C}} \bar{u}_1^{PT}(L)$.

Remark 4 *For example u_1^{TP} (u_1^{PT}) equals player one's favorite Nash equilibrium payoff if there is a credible language under TP (PT) and a message such that the unique equilibrium supported within that messages is player one's favorite Nash equilibrium.*

Proposition 4 (Sender's Power) *For any Γ if $i = 1$, in any TPE (PTE) player one must get at least $u_1^{TP}, (u_1^{PT})$ and there are TPE (PTE) in which player one's payoff is equal to his Nash equilibrium payoff if this payoff is larger or equal to u_1^{TP} (u_1^{PT}).*

Proof: For PT it follows from the equilibrium constructed in the proof of existence. For TP it is straightforward. ■

The power of the interpreter $i = 2$: To demonstrate the power of the interpreter in TP (PT) we exhibit an example where $i = 2$ and player two, by choosing complete communication forces player one to communicate all the details of his choice, splits the support of the favorite (mixed) equilibrium of player one and player two gets her best payoff. Here the Stackelberg payoff is not the favorite equilibrium of player one.

		Player two			
		L	M	R	RR
Player one	U	$1, -1$	$-1, 1$	$-2, -3$	$-1, 2$
	M	$-1, 1$	$1, -1$	$-1, 2$	$-2, -3$

Both languages are credible in TP (PT), player one's favorite equilibrium payoff is 0, obtained by mixing equally likely between U and M. But player two will choose complete communication and get a payoff of 2 in all TPE (PTE).

Proposition 5 (Class of games where all TPE are efficient) *If Γ is supermodular and player one's favorite equilibrium is in pure strategies then all TPE are efficient if $i = 1$. If the game is supermodular*[4] *and exhibits diminishing return and non-degenerate (see Berger (2008)) then all TPE are efficient if $i = 1$. (For example in games with positive spill-over, or Cournot with linear demand. This is not necessarily true for PTE.)*

Proof: The first part is straightforward along the lines of Milgrom and Roberts (1990), Shannon (1990). All one has to show is that there is a credible language under TP and a message such that the unique equilibrium supported within that message is player one's favorite Nash equilibrium. This is the case if there is another equilibrium of the game such that its support does not contain player one's favorite Nash equilibrium action, or the game has a unique pure equilibrium. For the second part, Berger (2008) and Krishna (1992) shows that any mixed strategy equilibrium can have at most two actions in its support given diminishing returns. It follows, that player one's favorite equilibrium cannot have both extreme pure Nash equilibrium in its support hence there is a credible language with a message containing only the favorite Nash equilibrium of player one. ■

We say that a game is *self-choosing* if for all s_1, s_1' it is true that $u_1(s_1, b_2(s_1)) \geq u_1(s_1, b_2(s_1'))$, where $b_2 : S_1 \to S_2$ is player two's best response function. This is weaker than Baliga and Morris's (2002) notion of *self-signalling*: for all $(s_1, s_2) \in S$ it is true that $u_1(s_1, b_2(s_1)) \geq u_1(s_1, s_2)$. We say that a game is of *common interest* if for all $(s_1, s_2), (s_1', s_2') \in S$ it is true that $u_1(s_1, s_2)) \geq u_1(s_1', s_2')$ if and only if $u_2(s_1, s_2)) \geq u_2(s_1', s_2')$. Common interest games are self-signalling and self-choosing.

Proposition 6 (Class of games where all PTE are efficient) *If player one's favorite equilibrium is in pure strategies and the game is self-choosing or the game is self-signalling then complete communication is credible and all PTE are efficient and player one receives his favorite Nash equilibrium payoff.*

[4] Weaker condition might suffice.

Proof: In self-choosing games complete communication is credible. In self-signalling games the favorite equilibrium of player one is in pure strategies. ■

6.1 A 3 by 3 Common Interest Game (TP)

The game shown below demonstrates how communication can be useless in TP even if the game has common interests because only no-communication is credible under TP; but it yields to efficiency in PT as complete communication is credible under PT.

		Player two		
		L	N	R
	T	5,5	0,0	-3,-3
Player one	M	-1,-1	1,1	2,2
	B	4,4	-2,-2	3,3

(T, L) is a pure strategy Nash equilibrium that leads to the unique efficient outcome.[5] It is natural that player one wants to say "I will play T". However, each of the other two Nash equilibria[6] of this game have T in the support of the corresponding equilibrium strategy of player one.[7] This means that player one cannot truthfully (in TP) communicate that she will not be playing T. Consequently, only $\{\{T, M, B\}\}$ is a credible language. Regardless of who is the interpreter, nontrivial information about intentions cannot be transmitted under credible communication in this game.

7 Discussion

7.1 Efficiency in Common Interest Games with Weak Credibility under TP

Now we give a weaker version of credibility under TP (definition 1) which does not require that player two guesses correctly player one's action after out of equilibrium messages.

[5] In fact, T is self-committing and the game satisfies self-signalling (Farrell, 1986, 1993).

[6] The Nash equilibria of the examples are computed using a program written by Rahul Savani. The program is based on the algorithm described in Avis, Rosenberg, Savani, and von Stengel (2009), and can be found at http://banach.lse.ac.uk/form.html.

[7] The other two mixed Nash equilibria τ and ρ are given by

$$\tau_1(T) = 2/7, \tau_1(M) = 5/7, \tau_1(B) = 0, \tau_2(L) = 1/7, \tau_2(N) = 6/7, \tau_2(R) = 0$$

$$\rho_1(T) = 4/15, \rho_1(M) = 43/60, \rho_1(B) = 1/60, \rho_2(L) = 4/15, \rho_2(N) = 31/60, \rho_2(R) = 13/60$$

with corresponding outcomes 5/7 and 41/60.

Definition 5 *We say that a language L is weakly-credible under TP if there is a $(m_1^L, \sigma_1^L, \sigma_2^L, \mu_2^L)$ weak-Perfect Bayesian equilibrium of $\Gamma^{TP}(L)$ in which player one always tells the truth, and player two always believes it. That is:*

1. *for all $m \in L$, $C(\sigma_1^L(m)) \subseteq m$ and*
2. *for all $m \in L$, $\mu_2^L(m) \in \Delta m$.*

Requiring only weak perfect Bayesian is weaker than subgame perfection and hence allows for more credible languages. In the common interest game in section 6.1 $\{\{T\}, \{M, B\}\}$ is weakly-credible. If player one says $\{M, B\}$ player two can believe it by putting not too much weight on B and play R. Player one then is telling the truth because he plays B. Hence weak-TPE yields efficiency if we allow incorrect out of equilibrium beliefs when defining credibility under TP and in point 2 of definition of TPE. In fact, this is true in general:

Remark 5 *In common interest games weak-TPE are efficient. It can namely be shown that the language which contains two messages $\{T\}$ and $S_1 \setminus \{T\}$ (where T is the action of player one yielding the best outcome) is weakly credible or the game has a single pure strategy equilibrium.*

7.2 Inefficiency with Weak Credibility under TP

If we change the payoff (-3,-3) to (4,-3) after (T, R) in the common interest game above we still have multiple Nash equilibria each containing T in its support. The game is still of self-choosing hence PT yields to efficient outcome. However, $\{\{T\}, \{M, B\}\}$ is not weakly-credible under TP anymore. It is natural that player one wants to say "I will play T". But he cannot do so in equilibrium, because after the message $\{M, B\}$ player two either plays L or R no matter what he believes in $\{M, B\}$. But then in both cases player one plays T which is out of $\{M, B\}$. This means that player one cannot truthfully (in TP) communicate that she will not be playing T. Similarly after message M player two must believe M and play R but then player one plays T. After message B player two must play L and then player one plays T. Consequently, only $\{\{T, M, B\}\}$ is a weakly-credible language. Regardless of who is the interpreter, nontrivial information about intentions cannot be transmitted under weakly-credible communication in this game.

7.3 Rabin's Credibility and Credibility under PT

We compare our notion of credibilty in PT to that of Rabin (1990). Rabin (1990) defines the notion of a Credible Message Profile (CMP) for simple communication games (sender-receiver games) with prior p over the types T of the sender. He does so by starting from a large enough message set M such that for each $X \subseteq T$

there is an exclusive set of messages $M(X) \subseteq M$ such that $M(X_i) \cap M(X_j) = \emptyset$ holds for all $X_i \neq X_j$. To simplify exposition identify subsets of T with messages, thus sending X has the meaning that "my type is in X". So the language is described by the power set of T as opposed to a partition as in our case. Focus is on a subset of messages called a message profile $\mathcal{X} = \{X_1, \ldots, X_D\}$ where $X_i \cap X_j = \emptyset$. Through definitions 1 till 6 Rabin (1990) defines when $\mathcal{X}$ is a CMP. Broadly speaking, a message profile is a CMP if for each message belonging to $\mathcal{X}$ received by the receiver, given that the receiver believes that she faces the types in the message, each type in the message gets his best payoff. In particular, messages within $\mathcal{X}$ are believed even if they are sent by types outside $\cup_{i=1}^{D} X_i$.

Now consider PT as a sender receiver game in which the sender, player one can choose his own type. We identify the sender with player one, T with S_1 and the receiver with player two. Rabin (1990) investigates which types can tell the truth, allowing others to lie. Our approach however builds on an understanding of communication in which all types can be believed. One reason is that types are endogenous in this paper. We consider credibility of a single sender while Rabin has many different senders, identified by their types. Hence, we only concentrate on CMP-s which are partitions (languages in our sense) of T.

Our definition of credible languages uses *equilibria* of the enlarged games $\Gamma^{PT}(L)$ which relies on disciplined beliefs on the equilibrium path. Consider an alternative definition that does not refer to an equilibrium in which a language is called *credible** if player two can form beliefs within the messages such that no matter which action was chosen by player one, it is optimal for player one to tell the truth, given that player two plays optimally given her beliefs. Clearly, if a language is credible under PT then it is *credible** under PT. Moreover, complete communication is credible if and only if it is *credible**.

Beliefs after messages containing a single action are fixed. Hence, one could hope that communication leads Nash equilibrium play when this involves player one choosing a single action. But player one may want to choose a different action and still tell the truth (by sending a vague message) and deviate from the candidate equilibrium. It is easy to construct examples which show that *crediblility** is too weak in the sense that player one can manipulate player two and achieve his best (non equilibrium) payoff in the game.

Non-existence of equilibrium with *credible** languages:

		Player two			
		L	M	R	RR
	U	$1,1$	$0,0$	$0,0$	$0,0$
Player one	M	$0,-1$	$-3,1$	$5,0$	$0,-2$
	D	$-1,-1$	$-2,-2$	$0,0$	$-3,1$

(U,L) is the unique Nash equilibrium of the game. But (U,L) cannot be the outcome of any PTE*. The language $L = \{\{U\},\{M,D\}\}$ is *credible** under PT if we choose $\mu_2^L(\{M,D\}) = (\alpha, 1-\alpha) \in \Delta\{M,D\}$ so that $\sigma_2^L(\{M,D\})(R) = 1$ is a best response to this belief, that is player two plays R optimally after message $\{M, D\}$. But then player one will choose M instead of U and receive a payoff of 5. No other languages, but no communication is *credible**. However, $\{\{U\}, \{M, D\}\}$ is not credible under PT. It follows that one must further restrict the set of *credible** languages.

Rabin (1990) offers a stronger[8] notion of credibility for sender receiver games in terms of credible message profiles, described above. His definition is clearly not applicable directly to our setting because player one (the sender) can choose his type. However, we immediately have the following observation. *If L is a CMP then it is credible**. Notice also, that $\{\{U\}, \{M, D\}\}$ is a CMP for an open set of priors in the example above and so CMP appears weaker than credibility under PT.

Further interesting comparisons can be made when considering complete communication. In particular, if complete communication is a CMP then it is also credible under PT. For self-signalling games complete communication is a CMP. There are games where complete communication is credible under PT but it is not a CMP (see Example 2 in Rabin (1990)). This and the observation above suggests that if player one can choose his type optimally before communication takes place then it allows for more precise communication compared to standard sender receiver setups. However, CMP is neither weaker nor stronger than credibility under PT (see the example above). This is because, the choice of action in PT gives more possibility to communicate but the requirement that beliefs must be correct on the equilibrium path (which is not an issue in CMP) restricts the possibilities of credible communication. It is easy to find a condition which guarantees that whenever a language is a CMP then the language is credible under PT, though we have found it too restrictive.

Our framework gives interesting possibility to analyze situations in which player one wants to "pool" some of his actions when playing a mixed equilibrium. In particular, in Γ_1^{PT} we require that the language is chosen optimally after each action of player one. Mixing can be interesting out of equilibrium as well, once we further restrict player two's out of equilibrium beliefs for credible languages, for example by requiring the existence of a proper[9] equilibrium of $\Gamma^{PT}(L)$ in which player one tells the truth.

7.4 Related Literature

Farrell (1986, 1993) pioneered the communication literature in which messages have an intrinsic meaning. Typically communication is about private information, the stereotypical model is a sender-receiver game introduced by Crawford and

[8]Rabin (1990) argues that in some situation it is rather weak. Indeed, $\{\{U\}, \{M, D\}\}$ is a CMP for an open set of priors.

[9]No other equilibrium concept has bite on out of equilibrium beliefs in $\Gamma^{PT}(L)$. Properness in $\Gamma^{PT}(L)$ is very similar to subgame perfection in $\Gamma^{TP}(L)$.

Sobel (1992). In the literature on neologisms, unexpected messages are checked in terms of their credibility (self-signalling), with reasoning becoming more involved when more than one message passes this test (e.g. see Matthews et al., 1991). Baliga and Morris (2002) conduct a formal game theoretic analysis, thus avoiding plausibility checks. In contrast to Baliga and Morris (2002), we incorporate choice of language and allow for partial information revelation. Moreover, under "first play then talk", private information is endogenous.

There are only few papers where communication is about intentions and messages have meaning, as we model in "first talk then play". Farrell (1988) investigates communication about intentions in the light of rationalizability, albeit adding additional plausibility requirements and not formally defining beliefs. Lo (2007) formally analyzes elimination of weakly dominated strategies for a rich class of messages, providing intricate conditions for ruling out messages that are "opposite" to each other. She finds that a unique outcome is selected in Battle of Sexes but not in Aumann's Stag Hunt game, the latter result being difficult to interpret. Farrell and Rabin (1996) first treat intentions as if they are private information, requiring self-signalling, and then add a condition (self-committing) that ensures that players behave according to their intentions. According to our formalization, self-signalling is not relevant for communication about intentions. Ellingsen and Ostling (2010) show for the level k model that there is always more coordination on pure Nash equilibria when there is one way communication. Demichelis and Weibull (2008) consider evolution in symmetric games under two-sided communication.

Truth can be incorporated in different ways, as seen in the papers highlighted above. Neologisms build on informal plausibility arguments. Baliga and Morris (2002) restrict attention to equilibria in which all information is transmitted. Other approaches include Chen (2004) who assumes that senders tell the truth with positive probability and Kartik et al. (2007) where there is a cost of telling a lie. In our paper we assume that the receiver believes that the sender tells truth, provided this is possible under the given language. Otherwise both behave as if there is a single message when truth-telling trivially holds. In contrast to Baliga and Morris (2002) this also puts discipline on out of equilibrium behavior.[10]

There is a closely related paper by Zultan (2012), albeit where messages have no meaning, in which a game with multiple selves is proposed to account for the findings of Charness (2000). Informally it is claimed that a standard game-theoretic model will not suffice. The focus is on sequential equilibria in which information is transmitted. These do not exist if the action is chosen before the message is sent, but exist if the message is sent first. Note that this does not mirror the findings of Charness (2000), even if one assumes that players select

[10]Note that Baliga and Morris (2002) do not to consider the complete information setting (talk about intentions) as they find it difficult to formalize their intutions in that context (see page 467 in their paper).

among those equilibria in which information is transmitted. This is because inefficient equilibria exist in which information is transmitted when the message is sent first.[11]

There is also experimental evidence that adding one-sided pre-play communication increases efficiency (see Cooper et al. (1989, 1992), Blume and Ortmann (2007)).

8 Conclusion

Interestingly, despite the large literature on communication in games, we seem to be the first to use an equilibrium analysis to investigate the impact of truthful communication under pre-play communication (as modelled in our "first talk then play" scenario). Truthful does not mean that players are forced to tell the truth. It means that the sender is able to convince the receiver whenever he can be believed. We call this *credible communication*. Our findings show that efficiency is not guaranteed in common interest games that have more than two strategies per player. The debate raised by Aumann also necessitates that we present a model in which communication occurs during play, called "first play then talk". This model has its own value as it is the first step to understanding communication while playing extensive form games of imperfect information. Results in the two models are very different and are useful to highlight how communication influences outcomes. They are both very tractable when analyzing specific games and can help understand in applications which equilibria have good properties. After all, parties will typically communicate and this should be considered formally when making predictions, instead of using it only as a motivation like in the literature on renegotiation.

Clearly communication as modelled in this paper is very specific. Once our modelling approach is well received we believe it to be important to tackle various extensions. We find it valuable, thereby contrasting the modelling of Baliga and Morris (2002), to allow for general messages and to identify all equilibria with truth-telling, and not just those where all information is transmitted. In other words, we wish to predict outcomes in games, not to understand when all information can be transmitted. Other extensions that are easy to implement include considering the case where player two is uncertain about whether or not player one has already committed to an action and considering an n player game where only player one communicates to the others. Extensions that require more thought in terms of making the right modelling choice include two-sided communication.

[11] Let players coordinate on the mixed Nash equilibrium when message m is sent. If any other message is sent assume that they coordinate on the inefficient pure strategy Nash equilbrium.

References

[1] R.J. Aumann, (1990), "Nash-Equilibria are not Self-Enforcing", *in Economic Decision Making: Games, Econometrics and Optimisation* (J. Gabszewicz, J.-F. Richard, and L. Wolsey, Eds.), Amsterdam, Elsevier 201-206.

[2] D. Avis, G. Rosenberg, R. Savani, and B. von Stengel (2009), "Enumeration of Nash equilibria for two-player games", *Economic Theory* **42,1**, 9-37.

[3] S. Baliga and S. Morris (2002), "Co-ordination, Spillovers, and Cheap Talk", *Journal of Economic Theory* **105**, 450-468.

[4] U. Berger (2008), "Learning in games with strategic complementarities revisited ", *Journal of Economic Theory* **143**, 292-301.

[5] A. Blume and A. Ortmann (2007), "The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria", Journal of Economic Theory **132**, 274-290.

[6] G. Charness, (2000), "Self-Serving Cheap Talk: A Test of Aumann's Conjecture", *Economic Theory* **33**, 177-194.

[7] Y. Chen (2004), "Perturbed Communication Games with Honest Senders and Naive Receivers", Journal of Economic Theory **146**, 401-424.

[8] R. Cooper, D.V. DeJong, R. Forsythe and T.W. Ross (1989), "Communication in the Battle of the Sexes Game: Some Experimental Results", *The RAND Journal of Economics* **20**, 568-587.

[9] R. Cooper, D.V. DeJong, R. Forsythe and T.W. Ross (1992), "Communication in Coordination Games", *The Quarterly Journal of Economics* **107**. 739-771.

[10] V.P. Crawford and J. Sobel (1982), "Strategic Information Transmission", *Econometrica* **50,6**. 1431-1451.

[11] S. Demichelis and J.W. Weibull (2008), "Language, Meaning, and Games: A Model of Communication, Coordination, and Evolution", *American Economic Review* **98**, 1292-1311.

[12] T. Ellingsen and R. Ostling (2010), "When Does Communication Improve Coordination? " *American Economic Review* **100**, 1695-1724.

[13] J. Farrell (1986), "Meaning and Credibility in Cheap Talk Games," University of California, Berkeley, Department of Economics working paper 8609.

[14] J. Farrell (1988), "Communication, Coordination, and Nash Equilibrium", *Economic Letters* **27**, 209-214.

[15] J. Farrell and M. Rabin (1996), “Cheap Talk”, *The Journal of Economic Perspectives* **10**, 103-118.

[16] N. Kartik, M. Ottaviani, and F. Squintani (2007), “Credulity, Lies, and Costly Talk”, *Journal of Economic Theory* **134**, 93–11 6.

[17] V. Krishna (1992), “Learning in games with strategic complementarities ”, *HBS Working Paper* 92-073, Harvard University.

[18] S.A. Matthews, M. Okuno-Fujiwara, and A. Postlewaite (1991), “Refining Cheap-Talk Equilibria”, *Journal of Economic Theory* **55**, 247-273.

[19] A. Mas-Colell, M.D. Whinston, and J.R. Green (1995), Microeconomic Theory, *Oxford University Press.*

[20] P. Milgrom and J. Roberts (1990), “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities ”, *Econometrica* **58,6**, 1255-1277.

[21] Pei-yu Lo (2007), “Language and Coordination Games”, unpublished manuscript.

[22] M. Rabin (1990), “Communication between Rational Agents”, *Journal of Economic Theory* **51**, 144-170.

[23] M. Rabin (1994), “A Model of Pre-game Communication”, *Journal of Economic Theory* **63**, 370-391.

[24] C. Shannon (1990), “An Ordinal Theory of Games with Strategic Complementarities ”, *Working Paper* Department of Economics, Stanford University.

[25] R. Zultan (2012), “Timing of messages and the Aumann conjecture: a multiple-selves approach”, *International Journal of Game Theory.*