Centre for Economic Policy Research

STUDY CENTER GERZENSEE

EUROPEAN SUMMER SYMPOSIUM IN ECONOMIC THEORY

Generously hosted by
Study Center Gerzensee

Monday 1 – Friday 12 July 2013

Betrayal of Intentions in 2-Player Games

Beniamin Bachi (Tel Aviv University)
Sambuddha Ghosh (Boston University)
*Zvika Neeman (Tel Aviv University)

Betrayal of Intentions in 2-Player Games

Benjamin Bachi[†] Sambuddha Ghosh[‡] Zvika Neeman[§]

June, 2013

– Preliminary and Incomplete –

Abstract

We introduce communication and with it the ability to deceive other players about a player's own intentions into the standard model of 2-player strategic form games. Communication facilitates "commitment" that expands the set of equilibrium outcomes of the game. We show that successful costly deception may arise in equilibrium despite the inherent tension between successful costly deception and equilibrium play, which implies playing best response to the other player's *true* strategy. Our model accounts for the significant levels of cooperation and correlation that are observed in experimental Prisoner's Dilemma games with pre-play communication.

KEYWORDS: deception, communication, cheap talk, Prisoners' Dilemma.
JEL CLASSIFICATION CODES: C72, D83.

1 Introduction

A few salient facts emerge from a large body of experimental evidence on the Prisoner's Dilemma (henceforth PD). First, communication enlarges the range of possible payoffs, although theory predicts cheap talk should make no difference. Sally (1995) did a meta-analysis of experiments from 1958 to 1992. Combining data from 37 different experiments, he showed that communication increases the rate of cooperation by roughly 40%. Second, perhaps more interestingly, Frank (1998) reports experiments showing that when subjects are allowed to interact for 30 minutes before playing the PD, they are able to predict quite accurately their opponent's action. Moreover, roughly 84% of the subjects who predict that their opponent will cooperate (defect) respond with the same. A longer period of communication leads to a higher probability of cooperation; in the experiment,

Acknowledgements to be added.
[†]Tel Aviv University; Email bbbenj@gmail.com.
[‡]Boston University; Email sghosh@bu.edu; URL http://people.bu.edu/sghosh/Welcome.html
[§]Tel Aviv University; Email zvika@post.tau.ac.il; URL http://tau.ac.il/~zvika/.

both the level of cooperation and the accuracy of the predictions drop when players are allowed to interact for only 10 minutes.[1]

Standard game theory tends to ignore the fact that in strategic situations players often have the opportunity to communicate before choosing their actions. When communication is modeled it is usually taken to be either "cheap talk" that does not oblige the players in any way, or (costly) "signaling" about the player's private information (type) over which the player has no control. However, in many situations communication may lead players to either betray the manner in which they intend to play the game or learn the intentions of the other players.[2] Indeed, we believe this to be the main difference between 'cheap' and 'real' talk. When this is the case, a player may obviously want to condition his action or strategy on the information he learns from the other players during the interaction.

The presence of such learning can have a potentially big effect on the way games are played. To see this, consider the PD, where each player's action set is $\{C, D\}$ and payoffs are as follows.

	C	D
C	$3,3$	$0,4$
D	$4,0$	$1,1$

In game theoretic models pure strategies are the same as actions in a one-shot game. In order to incorporate the learning that is afforded by the type of communication mentioned above, we enrich this notion of a pure strategy to include not just actions but also richer notions about the other player's *intention* as follows.

Suppose each player has 3 strategies — C, D, and *nice*, to be defined shortly. Players simultaneously pick a strategy from $S_1 = S_2 = \{C, D, nice\}$ and then engage in talk. As explained above,

[1]Similar results were obtained by Kalay et al. (2003) who consider data obtained from a TV game similar to the PD, in which two players accumulate a substantial amount of money, and then divide it as follows: Players communicate for several minutes, and each player then chooses one of the two actions cooperate or defect. If they both cooperate, each obtains half of the money they accumulated. If one cooperates and the other defects, the one who defected receives everything and the other nothing. In case both defect, both receive nothing. The (weakly) dominant strategy is to defect. However, players cooperated 42% of the time. Moreover, the data reveals a correlation between the actions chosen by the two players (21% of the time both players cooperated, compared to 17.64% if there had been no correlation); this implies a correlation coefficient of 0.14.

More recently, Belot et al. (2010) and den Assem et al. (2010) have studied similar game-shows. Belot et al. find that making a promise to cooperate prior to the decision is positively related to the actual cooperation rate if the promise was voluntary, but not if it was elicited by the host. Using data from the TV game show 'Golden Balls' den Assem et al. find that while co-operation decreases with the stakes, players still cooperate about 50 percent of the time even at the higher end of the stakes. The authors also find evidence that people have reciprocal preferences. That is they prefer to cooperate with players who cooperate with them and defect against players who defect against them. Similar observations were made by Yaari (2011).

An interested reader should search the keywords 'split or steal' or 'share or shaft kilroy' in youtube. The results are both entertaining and instructive.

[2]For example, a recent paper by DeSteno et al. (2012) reports that subjects in an experiment are able to predict the untrustworthiness of other players with whom they are matched (as measured by their tendency to play defect in a PD-like game) by four visual cues that include leaning away, crossing arms in a "blocking fashion," touching, rubbing or grasping hands together, and touching oneself on the face, abdomen or elsewhere. Interestingly, the cues were predictive only in combination. Another interesting observation is that identification of someone as less trustworthy may be unconscious in the sense that subjects may not be aware that it is due to the observed visual cues.

when players talk to each other they may learn the other player's strategy with some probability. For now we make the unrealistic assumption that this probability is equal to one. This assumption is relaxed below. The strategies C and D imply play cooperate or defect regardless of what you learn about the other player's strategy. The *nice* strategy is a mapping from what a player has learned into an action as follows; it amounts to saying "If I learn that the other player picked *nice* then I will pick C, and pick D if he has picked the strategies C or D."

Each pair of *strategies* translates into a pair of *actions*. The interesting entries are contained in the third row and third column, corresponding to the choice of *nice* by at least one player.

	C	D	*nice*
C	C,C	C,D	C,D
D	D,C	D,D	D,D
nice	D,C	D,D	C,C

This leads to the following payoff matrix:

	C	D	*nice*
C	3,3	0,4	0,4
D	4,0	1,1	1,1
nice	4,0	1,1	3,3

There are two desirable strategy profiles that give the best payoff – (C, C) and $(nice, nice)$. The first is not an equilibrium but the second is. In fact the conclusion is stronger: *nice* weakly dominates both C and D. In particular, D is not a weakly dominant strategy in the augmented game. This paper shows that the observation that communication expands the range of equilibrium outcomes applies also to other games, even when players may actively attempt to deceive the other players.

Indeed one of the key elements of communication is deception. Deception is a tricky idea to model in a standard game theoretic framework because the fact that in equilibrium players best-respond to each other's (true) strategy implies that successful deception is essentially precluded by the notion of Nash equilibrium.[3]

We propose a model that expands the scope of the simple model above to include the possibility of costly deception. Instead of choosing a single strategy, each player chooses a probability of deception p_i at a cost $c_i p_i$ for some $c_i \geq 0$ and two strategies, an 'actual' strategy to be played and a 'deception' strategy. The other player observes the player's actual and deception strategies with probabilities $1 - p_i$ and p_i, respectively. The other player cannot tell just by looking at a strategy if it is the actual strategy or a deception strategy. Hence, player i can completely deceive player j and make player j believe whatever player i wants j to believe about i's strategy or intention at a cost c_i. The cost of deception may be due to guilt or to the mental exertion that is required to fabricate a lie.

[3]The only attempts we are aware of to model deception in a game theoretic framework are Crawford (2003) and Ettinger and Jehiel (2010). Both are discussed in the next section.

As we show below, if the costs of deception c_1 and c_2 are commonly known among the players, then there can be no deception that benefits a player at the expense of the other player in equilibrium, in line with the general intuition about deception in equilibrium described above. We provide an example that demonstrates that deception that benefits *both* players may be sustained in equilibrium. However, deception does not expand the set of equilibrium payoffs.

We prove a "folk theorem" that characterizes the set of equilibrium payoffs. We show that if deception is costly, then the set of equilibrium payoffs encompasses all feasible and individually rational payoffs. The set of equilibrium payoffs shrinks as the costs of deception decrease. If the costs of deception are sufficiently small, then the set coincides with the set of equilibrium payoffs of the game without communication.

If however the costs of deception are privately known by the players, then there may exist equilibria with costly deception on the equilibrium path that benefits certain players' types at the expense of some of the other players' types in a way that is consistent with the observed data described above.

The paper proceeds as follows. Section 2 consists of a review of related literature. In Section 3 we present the model and basic definitions. In section 4 we find what payoffs may be sustained in equilibrium of general two player games with commonly known costs of deception. In Section 5 we present a Folk Theorem and in Section 6 we consider the case of privately known costs of deception. Brief concluding remarks are offered in Section 7.

2 Related Literature

This paper concerns true information that is transferred between two people during communication. The fact that this information can be fully controlled only at a cost makes it very different from "cheap talk," which Farrell and Rabin (1996) describe as "costless, non-binding, non-verifiable messages that may affect the listener's beliefs." The information that is transferred need not be truthful; in our model the player can convey a different false impression, but at a certain cost. As we show below, the possibility of communication and the difficulty to fully control it may expand the set of equilibria in games when cheap talk fails to. The best example is the Prisoners' Dilemma game: cheap talk does not add any equilibrium to the game, whereas with the type of communication that is considered here players can achieve full cooperation. The notion of communication that is developed in this paper is also different from Aumann's (1974) notion of a correlated equilibrium that employs the communication of information among the players as a correlating device, because cooperation cannot emerge even in a correlated equilibrium of the Prisoners' Dilemma.[4]

Frank (1998) describes an informal model of commitment without any external mechanisms where the players' emotions serve as commitment devices. Since psychological research shows that emotions are both observable and hard to fake (see Frank (1988) and references within), an agent can use them as signals in a game.

[4]But see Forgó (2010) for a generalization of correlated equilibrium that admits some cooperation in PD-like games.

Gauthier (1986) describes an environment in which there are two types of agents: straightforward maximizers and constrained maximizers. Straightforward maximizers simply maximize their utility; constrained maximizers are more sophisticated. They take into account the utilities of the other players and base their actions on a joint strategy: "A constrained maximizers is conditionally disposed to cooperate in ways that, followed by all, would yield nearly optimal and fair outcomes, and does cooperate in such ways when she may actually expect to benefit." Gauthier assumes that an agent's type is known to everybody else (or at least with some positive probability). Thus, in the Prisoners' Dilemma, when a constrained maximizer meets another constrained maximizer, they will both cooperate. In any other interaction between two players, they will both defect.

These last two works resemble ours but are not posed in a formal game theoretic framework.[5] Neither considers the possibility of deception. In contrast, this paper provides a formal game-theoretic model that captures the intuition above. Moreover it does not rely on perfect commitment; indeed it allows for active deception.

A small game theoretic literature establishes a folk theorem when players' strategies may depend on the other players' strategies. We discuss the connection between the results of this literature and those obtained here after we present our own folk theorem (Proposition 4) in Section 5 below .

Finally, there is also a small literature that attempts to model deception as an equilibrium phenomenon. As in this paper, the focus of Crawford (2003) is on active misrepresentation rather than less-than-full disclosure, and on signaling intentions rather than private information. Crawford uses the case of the allied invasion to Normandy in World War Two as his motivating example. He considers a 2×2 sender-receiver game in which the sender has several different types: a truthful type whose action is identical to its signal, several "wily" types whose actions and signals may differ, and a "sophisticated" type that plays optimally given its beliefs. Importantly, deception cannot be sustained in Crawford's model without the truthful and wily types, and so hinges on bounded rationality.

Ettinger and Jehiel (2010) also rely on bounded rationality to model deception. Their motivating example is that of a seller of a house who reveals a damning fact about his property to induce the prospective buyer to believe that the house does not suffer from a much more serious defect. Agents in their model only have coarse knowledge of their opponent's strategy. Equilibrium requires the coarse knowledge available to agents to be correct, and the inferences and optimizations to be made on the basis of the simplest theories compatible with the available knowledge. This predisposes their agents to the so called Fundamental Attribution Error and allows them to be deceived.

Notably, in contrast to these two papers, our model provides a fully rational model of equilibrium deception.[6]

[5] Binmore (1994), for example, criticized Gauthier for logical inconsistency.

[6] Another related paper is Kartik (2009) who considers the possibility of costly lying in a sender-receiver game. He shows that all the sender's types inflate their reports in equilibrium and that some types pool on the highest possible message, regardless of the intensity of lying costs.

3 The Model

Let $G = \langle A_1, A_2, \Pi_1, \Pi_2 \rangle$ be a two-person game in normal form, where A_i is a finite set of actions for player i ($i = 1, 2$), and $\Pi_i : A_1 \times A_2 \to \mathbb{R}$ is the payoff function for player i. The set of mixed strategies of G is denoted $S = S_1 \times S_2$.

Definition. *A Nash equilibrium of G is a pair of strategies (s_1, s_2) such that for $i, j \in \{1, 2\}, i = j$: $\Pi_i\left(s_i, s_j\right) \geq \Pi_i\left(s_i, s_j\right)$ for any $s_i \in S_i$.*

We define an augmented game $\hat{G}$ that is induced by G. The augmented game $\hat{G}$ can be thought of as the game G with pre-play communication, or the "game with real talk" that is induced by G. We be gin with the definition of a strategy in the augmented game $\hat{G}$.

Definition. *A strategy $(p_i, \widehat{s}_i, \widehat{s}_i) \in [0, 1] \times \widehat{S}_i \times \widehat{S}_i$ for player i in the augmented game $\widehat{G}$ that is induced by G consists of:*

1. *a probability $p_i \in [0, 1]$; and*
2. *a pair of mappings from $\widehat{S}_j$ to $\Delta(A_i)$, denoted $\widehat{s}_i$ and $\widehat{s}_i$, respectively, where $\widehat{S}_j$ is the other player's strategy set, and*
$$\widehat{S}_i \in \left\{f : \widehat{S}_j \to \Delta\left(A_i\right)\right\}.$$

The mapping $\widehat{s}_i \in \widehat{S}_i$ describes player i's "actual strategy," or how i plays the augmented game $\widehat{G}$; the mapping $\widehat{s}_i \in \widehat{S}_i$ describes player i's "deception strategy," or what player i wants player j to think about player i's actual strategy in the augmented game. Player j learns player i's actual and deception strategies with the probabilities $1 - p_j$ and p_j, respectively.

The augmented game $\widehat{G} \equiv \widehat{G}\left(G, \left(\widehat{S}_1, \widehat{S}_2\right), (c_1, c_2)\right)$ that is induced by G consists of four stages and is played as follows:

1. The two players each choose their strategies $(p_1, \widehat{s}_1, \widehat{s}_1) \in [0, 1] \times \widehat{S}_1 \times \widehat{S}_1$ and $(p_2, \widehat{s}_2, \widehat{s}_2) \in [0, 1] \times \widehat{S}_2 \times \widehat{S}_2$ simultaneously.
2. Each player i observes a signal σ_j that is equal to $\widehat{s}_j$ with probability $1 - p_j$ and $\widehat{s}_j$ with probability p_j.
3. The two players each play their actual strategies $\widehat{s}_i$ and $\widehat{s}_j$ depending on the signals they each observed.
4. Payoffs are realized. The payoff to player i is given by:
$$\Pi_i\left(\widehat{s}_i\left(\sigma_j\right), \widehat{s}_j\left(\sigma_i\right)\right) - c_i p_i$$
where $c_i \geq 0$ denotes player i's cost of deception.

Remark. The strategy sets $\widehat{S}_1$ and $\widehat{S}_2$ are given exogenously as part of the description of the augmented game $\widehat{G}$. $\widehat{S}_1$ and $\widehat{S}_2$ can include all the pure strategies in the underlying game G as constant mappings in the augmented game $\widehat{G}$. In this case, if we abstract from the possibility of deception, then the augmented game coincides with the underlying game G. The sets $\widehat{S}_1$ and $\widehat{S}_2$ can also be either smaller or larger. For example, $\widehat{S}_1$ and $\widehat{S}_2$ can be singleton sets that contain just one mapping for each player. For example, $\widehat{S}_i = \{s_i\}$ where for each player, s_i is a mapping that always plays some pure action $a_i \quad A_i$. At the other extreme, and perhaps also more naturally, $\widehat{S}_1$ and $\widehat{S}_2$ can be taken to include all the mappings from $\widehat{S}_j$ to $\Delta(A_i)$ that can be described in finite sentences (using Gödel encoding, see Peters and Szentes (2012) for details).[7] We henceforth assume that $\widehat{S}_1$ and $\widehat{S}_2$ indeed include all these mappings. This assumption is more than what we need to obtain our results, but it saves us the rather tedious need to clarify exactly what mappings should be contained in $\widehat{S}_1$ and $\widehat{S}_2$.

4 Equilibrium

Once the sets of mappings $\widehat{S}_1$ and $\widehat{S}_2$ are specified, the augmented game $\widehat{G}$ can be analyzed in the same way as any other normal form game with strategy sets that are given by $S = S_1 \times S_2$ where $S_i = [0,1] \times \widehat{S}_i \times \widehat{S}_i$ and payoffs that are given by (with a slight abuse of notation):

$$\begin{aligned}\widehat{\Pi}_i(s_i, s_j) &= \Pi_i\left[\widehat{s}_i(\widehat{s}_j), \widehat{s}_j(\widehat{s}_i)\right](1-p_i)(1-p_j) + \Pi_i\left[\widehat{s}_i(\widehat{s}_j), \widehat{s}_j(\widehat{s}_i)\right] p_i(1-p_j) \\ &\quad + \Pi_i\left[\widehat{s}_i(\widehat{s}_j), \widehat{s}_j(\widehat{s}_i)\right](1-p_i)\, p_j + \Pi_i\left[\widehat{s}_i(\widehat{s}_j), \widehat{s}_j(\widehat{s}_i)\right] p_j p_i.\end{aligned}$$

where $s_i = (p_i, \widehat{s}_i, \widehat{s}_i) \quad S_i$ for $i \quad \{1,2\}$.

Definition. *A Nash equilibrium of the augmented game $\widehat{G}$ is a pair of strategies $(s_1, s_2) \quad S$ such that for $i, j \quad \{1,2\}, i = j$: $\widehat{\Pi}_i(s^*{}_i, s_{-i}) \geq \widehat{\Pi}_i(s_i, s_{-i})$ for any $s_i \quad S_i$.*

Thus, the augmented game $\widehat{G}$ admits the existence of a Nash equilibrium under the usual conditions on strategies and payoffs under which strategic form games admit the existence of a Nash equilibrium. In particular, the augmented game $\widehat{G}$ "inherits" all the Nash equilibria of the game G.

Proposition 1. *If (s_1, s_2) is a Nash equilibrium of the game G, then the strategies $(0, \widehat{s}_1, \widehat{s}_1)$ and $(0, \widehat{s}_2, \widehat{s}_2)$ where there is no deception and $\widehat{s}_i$ is the constant mapping that assigns s_i to each strategy of the other player is a Nash equilibrium of the augmented game $\widehat{G}$.*

Proof. Follows immediately from the fact that (s_1, s_2) is a Nash equilibrium of the game G. □

The next four examples illustrate the ways in which players can utilize communication and deception to increase their payoffs in the underlying game. The first example shows that commu-

[7] Notice that for at least one player i it has to be that $\widehat{S}_i$ is a strict subset of the set of mappings $\{f : \widehat{S}_j \to \Delta(A_i)\}$. Otherwise $\widehat{S}_1 = |\Delta(A_1)|^{|S_2|}$ and $\widehat{S}_2 = |\Delta(A_2)|^{|S_1|}$, which is impossible by Cantor's Theorem according to which no set has the same cardinality as its power set.

nication plus the fact that deception is costly allows a player to commit to an induced strategy that is not a best response in the underlying game so as to increase its payoff.

Example 1. Consider the following strategic form game.

	L	R
U	$1,3$	$1,4$
D	$0,1$	$2,2$

Strategy L is dominated by strategy R for Player 2 and so the only Nash equilibrium of the game is (D,R) with payoffs $(2,2)$.

But, if the cost of deception for Player 2 is high enough, then the strategies $(0,\widehat{s}_1,\widehat{s}_1)$ and $\left(0,\widehat{L},\widehat{L}\right)$ where

$$\widehat{L}(\sigma_1) = \left\{ \begin{array}{ll} L & \text{for any } \sigma_1 \end{array} \right.$$

and

$$\widehat{s}_1(\sigma_2) = \left\{ \begin{array}{ll} U & \text{if } \sigma_2 = \widehat{L} \\ D & \text{if } \sigma_2 = \widehat{L} \end{array} \right.$$

constitute an equilibrium of the augmented game. In this equilibrium player 2 "commits" to playing L and player 1 best responds to player 2's strategy. The equilibrium outcome is (U,L) with payoffs $(1,3)$.

The next example shows that if deception is costly, then communication involuntarily reveals a player's strategy. This implies that a player may be penalized for playing certain induced strategies and induced to play a strategy that is favorable to the other player, or alternatively, that a player can commit to penalize the other player if he doesn't behave in a certain way.

Example 2. Consider the following strategic form game.

	L	R
U	$3,3$	$0,0$
D	$2,4$	$1,0$

Strategy R is dominated by strategy L for Player 2 and so the only Nash equilibrium of the game is (L,U) with payoffs $(3,3)$.

But, if player 1's cost of deception is sufficiently high, then the strategies $\left(0,\widehat{D},\widehat{D}\right)$ and $(0,\widehat{s}_2,\widehat{s}_2)$ where

$$\widehat{D}(\sigma_2) = \left\{ \begin{array}{ll} D & \text{for any } \sigma_2 \end{array} \right. ;$$

and

$$\widehat{s}_2(\sigma_1) = \left\{ \begin{array}{ll} L & \text{if } \sigma_1 = \widehat{D} \\ R & \text{if } \sigma_1 = \widehat{D} \end{array} \right.$$

constitute an equilibrium of the augmented game. This equilibrium is sustained by player 2 "commitment" to penalize player 1 by playing R if player 1 deviates from playing D. If the cost of deception is sufficiently high, then player 1 cannot fool player 2 into believing that player 1 is playing D when he plays the induced strategy U, which is better for him. Thus, it is in the best interest of player 1 to surrender to player 2's threat and to play D.

The next example shows that contrary to the intuition that, since in equilibrium players anyway best respond to each other's true strategy deception cannot arise in equilibrium, equilibrium deception is possible.

Example 3. Consider the following strategic form game.

	C	N
C	$1,1$	$0,0$
N	$0,0$	$0,0$

If the players' deception costs are not too high, then the pair of strategies $\left(1, \widehat{s}_C, \widehat{N}\right)$ and $\left(1, \widehat{s}_C, \widehat{N}\right)$ where $\widehat{N}$ is the constant mapping that plays the action N regardless of what is learned about the other player, and the mapping $\widehat{s}_C$ is defined as:

$$\widehat{s}_C\left(\sigma_{-i}\right) = \begin{cases} N & \text{if } \sigma_{-i} = \widehat{N} \\ C & \text{if } \sigma_{-i} = \widehat{N} \end{cases}$$

constitutes a symmetric equilibrium of the augmented game. In this equilibrium both players deceive the other player to believe they play the strategy $\widehat{N}$ only to play the actual strategy $\widehat{s}_C$. Observe that in this equilibrium, both players benefit from deception because it allows them to play the cooperative outcome (C, C). However, this benefit does not come at the expense of the other player.

The next proposition shows that Example 3 illustrates a general result, namely,

Proposition 2. *An equilibrium of the augmented game $\widehat{G}$ in which both players engage in equilibrium deception induces a Nash equilibrium of the underlying game G in the sense that the players' induced strategies in G are a Nash equilibrium of G for generic deception costs.*[8]

Thus, the players' payoffs in an equilibrium of the augmented game in which both players deceive each other are identical to their payoffs in some Nash equilibrium of the underlying game G, net of the deception costs.

Proposition 2 is an immediate corollary of the next lemma that shows that a player who engages in equilibrium deception, necessarily plays a strategy in the induced game G that is a best response to the other player's induced strategy.

[8]The term "generic" refers to all but one specific value of the deception cost for each player.

Lemma. *For generic deception costs,*[9] *if player i engages in equilibrium deception in the augmented game* $\widehat{G}$*, then i's induced strategy in the underlying game G*

$$(1 - p_j)\,\widehat{s}_i\left(\sigma_j = \widehat{s}_j\right) + p_j\widehat{s}_i\left(\sigma_j = \widehat{s}_j\right)$$

is a best response to player j's induced strategy in G

$$(1 - p_i)\,\widehat{s}_j\left(\sigma_i = \widehat{s}_i\right) + p_i\widehat{s}_j\left(\sigma_i = \widehat{s}_i\right).$$

Proof. Suppose that $(p_i, \widehat{s}_i, \widehat{s}_i)$ and $\left(p_j, \widehat{s}_j, \widehat{s}_j\right)$ is a Nash equilibrium of the augmented game $\widehat{G}$ in which player i deceives player j, or such that $p_i > 0$ and $\widehat{s}_i = \widehat{s}_i$. The fact that the cost of deception is linear in the probability of deception implies that either $p_i = 1$ or player i is indifferent between the strategies $(p_i, \widehat{s}_i, \widehat{s}_i)$ and $(1, \widehat{s}_i, \widehat{s}_i)$ in $\widehat{G}$. If $p_i = 1$ then a deviation by i to a different actual strategy $\widehat{s}_i$ cannot be detected by player j. It therefore follows that player i's induced strategy in the underlying game G must be a best response to player j's induced strategy in the underlying game G. Indifference between the strategies $(p_i, \widehat{s}_i, \widehat{s}_i)$ and $(1, \widehat{s}_i, \widehat{s}_i)$ occurs for only one specific value of player i's cost of deception c_i and is thus nongeneric. □

Example 4 below shows that deception may also hurt a player in the sense that the player's payoff in an equilibrium in which he deceives the other player may be lower than in any equilibrium in the game with no communication (in the next section, we show that the set of equilibria of the underlying game (with no communication) coincides with the set of equilibria of the augmented game with costless communication).

Example 4. Consider the following strategic form game.

[9] That is, for all but one specific value of the deception cost for each player. The following counter-example demonstrates the necessity of this qualification. Let the underlying game G be given by

	L	R
T	$0,1$	$0,2$
M	$0,2$	$0,0$
B	$0,0$	$0,0$

Suppose that the costs of deception are $c_1 = c_2 = 1$. Consider the strategies $(0, \widehat{s}_1, \widehat{s}_1)$ and $\left(p_2, \widehat{L}, \widehat{R}\right)$ where

$$\widehat{s}_1(\sigma_1) = \begin{cases} T & \text{if } \sigma_2 = \widehat{L} \\ M & \text{if } \sigma_2 = \widehat{R} \\ B & \text{otherwise} \end{cases}$$

and the mappings $\widehat{L}$ and $\widehat{R}$ denote the constant mappings that play L and R, respectively, regardless of what is learned about the other player.

The strategies $(0, \widehat{s}_1, \widehat{s}_1)$ and $\left(p_2, \widehat{L}, \widehat{R}\right)$ are a Nash equilibrium of the augmented game for for every $p_2 \quad [0,1]$. We need not worry about player 1. As for player 2, his induced payoffs is 1, which is maximal given $\widehat{s}_1$ because player 2 can get a payoff of 2 only when he deceives player 1 but this costs him 1 for any "unit" of deception.

	L	M	R
U	$1,3$	$0,3$	$1,4$
D	$0,1$	$-1,1$	$2,2$

Strategies L and M are dominated by strategy R for Player 2 and so the only Nash equilibrium of the game is (D, R) with payoffs $(2,2)$.

But, if the cost of deception for Player 2 is high enough, then the strategies $\left(1, \widehat{U}, \widehat{D}\right)$ and $\left(0, \widehat{LM}, \widehat{LM}\right)$ where

$$\widehat{LM}\left(\sigma_1\right) = \begin{cases} L & \text{if } \sigma_1 = \widehat{D} \\ M & \text{if } \sigma_1 = D \end{cases};$$

$$\widehat{U}\left(\sigma_2\right) = \begin{cases} U & \text{if } \sigma_2 = \widehat{LM} \\ D & \text{if } \sigma_2 = \widehat{LM} \end{cases};$$

and

$$\widehat{D}\left(\widehat{s}_2\right) = \begin{cases} D & \text{for any } \sigma_2 \end{cases};$$

constitute an equilibrium of the augmented game. In this equilibrium player 2 "commits" to playing the action L but only if player 1 engages in deception, and "threatens" to play the action M, which player 1 dislikes, otherwise. Player 1 deceives player 2 into believing he plays the constant strategy $\widehat{D}$ while playing the actual strategy $\widehat{U}$, which is a best response to player 2's induced strategy in the underlying game above. The equilibrium outcome is (U, L) with payoffs $(1,3)$, respectively.

Finally, the next proposition shows that deception is "redundant" in that it does not expand the range of equilibrium outcomes.

Proposition 3. *For any equilibrium with deception in of the augmented game $\widehat{G}$ there exists another equilibrium with no deception in $\widehat{G}$ that induces the same strategies in the underlying game G and so generates the same or better payoffs to the players.*

The Proof of Proposition 3 is constructive. The constructed equilibrium employs the players' minmax strategies as disciplinary devices.

Let w_i denote the minmax value for player i in the strategic form game G, i.e.

$$w_i = \min_{s_j \quad S_j} \max_{a_i \quad A_i} \Pi_i(a_i, s_j).$$

A payoff π_i for player i is said to be individually rational if $\pi_i \geq w_i$. Note that the players' equilibrium payoffs are necessarily individually rational, because otherwise they have a profitable deviation.

Let ψ_i denote the strategy of player of i that minmaxes the other player; i.e. when ψ_i is played, player j can achieve a payoff of at most w_j. Formally, $\psi_i \quad \arg\min_{s_i \quad S_i} \max_{a_j \quad A_j} \Pi_j(s_i, a_j)$. For any $s_i \quad S_i$, let $s_i(a_i)$ be the probability of the action a_i.

Proof. Let $(p_1, \widehat{s}_1, \widehat{s}_1)$ and $(p_2, \widehat{s}_2, \widehat{s}_2)$ be a Nash equilibrium of the augmented game $\widehat{G}$ that induces play of the strategies $s_i \equiv \left(1 - p_j\right)\widehat{s}_i\left(\sigma_j = \widehat{s}_j\right) + p_j\widehat{s}_i\left(\sigma_j = \widehat{s}_j\right)$, $i = 1, 2$, $j = i$, in the underlying game G. Recall that ψ_i denotes a strategy of player i that minmaxes the other player, and let $\bar{s}_i \quad \widehat{S}$ be a constant mapping that always plays s_i. Define

$$\widehat{s}_1(\sigma_2) = \begin{cases} \bar{s}_1 & \text{if } \sigma_2 = \widehat{s}_2 \\ \psi_1 & \text{otherwise} \end{cases};$$

$$\widehat{s}_2(\sigma_1) = \begin{cases} \bar{s}_2 & \text{if } \sigma_1 = \widehat{s}_1 \\ \psi_2 & \text{otherwise} \end{cases}.$$

Clearly, if $(0, \widehat{s}_1, \widehat{s}_1)$ and $(0, \widehat{s}_2, \widehat{s}_2)$ are played, then the induced actions and payoffs are the same as in the original equilibrium. We show that $(0, \widehat{s}_1, \widehat{s}_1)$ and $(0, \widehat{s}_2, \widehat{s}_2)$ is also a Nash equilibrium in $\widehat{G}$.

Any deviation from these strategies that does not involve deception leads to a payoff smaller or equal to the minmax and hence not profitable. Furthermore, the most profitable strategy that does involve deception has $\widehat{s}_i$ as the deception strategy.

Let $\overline{BR}_{i,s_j} \quad \widehat{S}_i$ denote the constant mapping that is the best response against the strategy s_j in the underlying game G. Thus, the most profitable deviation from $(0, \widehat{s}_i, \widehat{s}_i)$ is $(1, \overline{BR}_{i,s_j}, \widehat{s}_i)$, and player i's payoff when he plays this strategy against $(0, \widehat{s}_j, \widehat{s}_j)$ is $\widehat{\Pi}_i\left(\overline{BR}_{i,s_j}, \widehat{s}_j\right) = \Pi_i\left(BR_{i,s_j}, s_j\right) - c_i$ where $BR_{i,s_j} \quad S_i$ is the best response against the strategy s_j in the underlying game G.

Distinguish between the following two cases:

(1) Player i did not deceive in the original equilibrium ($p_i = 0$): player i could have played $(1, \overline{BR}_{i,s_j}, \widehat{s}_i)$ against $\left(p_j, \widehat{s}_j, \widehat{s}_j\right)$ in the first place and obtain $\Pi_i\left(BR_{i,s_j}, s_j\right) - c_i$ but preferred not to. This means that his original equilibrium payoff is larger than or equal to $\Pi_i\left(BR_{i,s_j}, s_j\right) - c_i$.

(2) Player i did deceive in the original equilibrium ($p_i > 0$): by Proposition 2, s_i is a best response against s_j, and therefore the deviation reduces player i's profits by c_i. □

Is it the case then that players cannot benefit from being able to easily deceive other players? As we show in Section 6 below, the answer to this question is negative, but it hinges on asymmetric information about the players' costs of deception.

5 A Folk Theorem

We characterize the set of equilibrium payoffs as a function of the costs of deception c_1 and c_2.

Definition. *A payoff profile (π_1, π_2) of a strategic form game G is feasible if there exist two strategies*

s_1 S_1 *and* s_2 S_2 *such that*[10]

$$(\pi_1, \pi_2) = \sum_{a_1 \ A_1} \sum_{a_2 \ A_2} s_1(a_1) s_2(a_2) \Pi(a_1, a_2).$$

Proposition 4. *For any strategic form game G there exists a cost of deception c such that any individually rational and feasible payoff profile (π_1, π_2) of G can be sustained as an equilibrium payoff of the same game $\widehat{G}$ provided that $c_1, c_2 > c$.*

Proof. Let $\pi \equiv (\pi_1, \pi_2)$ be an individually rational and feasible payoff profile. Let s_1 and s_2 be *any* pair of probability distributions over A_1 and A_2 respectively such that

$$\pi = \sum_{a_1 \ A_1} \sum_{a_2 \ A_2} s_1(a_1) s_2(a_2) \Pi(a_1, a_2)$$

Define $\widehat{s}^{\pi} := (\widehat{s}_1^{\pi}, \widehat{s}_2^{\pi})$ as follows:

$$\widehat{s}_i^{\pi}(\sigma_j) = \begin{cases} s_i \text{ if } \sigma_j = \widehat{s}_j^{\pi} \\ \psi_i \text{ otherwise} \end{cases}.$$

Let $\hat{S}_1$ and $\hat{S}_2$ be arbitrary mutually consistent strategy sets such that $\widehat{s}_1^{\pi}$ $\hat{S}_1$ and $\widehat{s}_2^{\pi}$ $\hat{S}_2$ for all individually rational and feasible payoffs π.

The strategy profile $(0, \widehat{s}_i^{\pi}, \widehat{s}_i^{\pi})_{i=1,2}$ is a Nash equilibrium of the augmented game $\widehat{G}$ for any π if

$$c_i \geq c \ := \max_{(a_1,a_2),(b_1,b_2) \ A_1 \times A_2} \{\Pi_i(a_1, a_2) - \Pi_i(b_1, b_2)\}.$$

To see this, suppose player 2 plays s_2^{π}. Further suppose that $p_1 = p_2 = 0$. If player 1 plays $\widehat{s}_1^{\pi}$ then the players will play (s_1, s_2), yielding player 1 a payoff of π_1. If player 1 deviates to any other strategy with $p_1 = 0$, player 2 will play ψ_2 against him, giving player 1 a payoff of no more than w_1. However, since π is individually rational, $\pi_1 \geq w_1$. The same argument holds for player 2. If player 1 chooses $p_1 = \epsilon > 0$ and possibly different actual and deception strategies, then his payoff is at most

$$(1 - \epsilon) w_1 + \epsilon \left(\max_{a_1,a_2 \ A_1 \times A_2} \Pi_1(a_1, a_2) - c_1 \right),$$

which falls short of the equilibrium payoff π_1 if

$$\epsilon \max_{a_1,a_2 \ A_1 \times A_2} \Pi_1(a_1, a_2) - \epsilon c_1 \leq \pi_1 \epsilon$$

[10]Note that our definition of feasibility requires independent mixing. This is different from the standard definition used in the theory of repeated games according to which a feasible payoff profile (π_1, π_2) is a convex combination of all outcomes in G, that is $\sum_{\alpha \ A} \alpha(a_1, a_2) \Pi(a_1, a_2)$ where α is a probability distribution over the joint action space A.

if

$$c_1 \geq \max_{a_1,a_2 \quad A_1 \times A_2} \Pi_1(a_1, a_2) - \pi_1.$$

Therefore, the strategy profile $(0, \widehat{s}_i^{\pi}, \widehat{s}_i^{\pi})_{i=1,2}$ is a Nash equilibrium of the augmented game $\widehat{G}$ if $c_i \geq c := \max_{i \ \{1,2\}} \max_{a_1,a_2 \quad A_1 \times A_2} \{\Pi_i(a_1, a_2) - \pi_i\}$. □

Tennenholtz (2004) proves a similar folk theorem for the case of a game that is played by computer programs that may condition their strategy on the other computer program with which they are matched to play the game. Kalai et al. (2010) offer a similar model and folk theorem for the case where each player chooses a "commitment device" that may depend on the commitment devices chosen by other players rather than a computer program. Peters and Szentes (2009) explore games where each player writes a contract that obliges a player to respond with a specified action depending on the opponent's contract. They prove a similar folk theorem to Kalai et al. and further show that this result does not hold in an environment with incomplete information.[11] As we show below, our model produces the same equilibrium payoffs as Tennenholtz (2004) and Kalai et al. (2010) if the costs of deception are high. But if the costs of deception are low, then our model produces the same equilibrium payoffs as standard Nash equilibrium analysis (intermediate deception costs produce intermediate equilibrium outcomes).

Denote the set of Nash equilibrium payoffs of a strategic game G by N_G. Denote the maximal set of equilibrium payoffs of any augmented game $\widehat{G}$ that is induced by G with costs of deception c_1 and c_2 by $N_{\widehat{G}}(c_1, c_2)$. If $c_1, c_2 \geq c$, then the previous proposition establishes a type of a "folk theorem" result. The next two propositions describe how the set of equilibrium payoffs varies with the costs of deception c_1 and c_2. Proposition 5 shows that if deception is costless ($c_1 = c_2 = 0$), then the only equilibrium payoffs of the augmented game $\widehat{G}$ are those of the strategic game G, or $N_{\widehat{G}}(0, 0) = N_G$. Proposition 6 shows that the set of equilibrium payoffs increases monotonically with the costs of deception.

Proposition 5. *If deception is costless, then the set of equilibrium payoffs of any strategic form game G and coincides with the maximal set of equilibrium payoffs of the augmented game $\widehat{G}$ that is induced by G, or $N_{\widehat{G}}(0, 0) = N_G$.*

Proof. By Proposition 1, $N_G \quad N_{\widehat{G}}(0, 0)$. We show that $N_{\widehat{G}}(0, 0) \quad N_G$. Suppose that $(p_1, \widehat{s}_1, \widehat{s}_1)$ and $(p_2, \widehat{s}_2, \widehat{s}_2)$ is a Nash equilibrium of the augmented game $\widehat{G}$ that induces play of the strategies $(1 - p_2)\widehat{s}_1(\sigma_2 = \widehat{s}_2) + p_2\widehat{s}_1(\sigma_2 = \widehat{s}_2)$ and $(1 - p_1)\widehat{s}_2(\sigma_1 = \widehat{s}_1) + p_1\widehat{s}_2(\sigma_1 = \widehat{s}_1)$ in the game G.

By Proposition 2, if $p_i > 0$, then $(1 - p_j)\widehat{s}_i(\sigma_j = \widehat{s}_j) + p_j\widehat{s}_i\left(\sigma_j = \widehat{s}_j\right)$ is a best response against $(1 - p_i)\widehat{s}_j(\sigma_i = \widehat{s}_i) + p_i\widehat{s}_j(\sigma_i = \widehat{s}_i)$ in G.

Assume in contradiction that $p_i = 0$ and that $(1 - p_j)\widehat{s}_i(\sigma_j = \widehat{s}_j) + p_j\widehat{s}_i\left(\sigma_j = \widehat{s}_j\right)$ is not a best response against $\widehat{s}_j(\sigma_i)$ in G. Let a_i be a best response against $\widehat{s}_j(\sigma_i)$ (in G) and let $\widehat{a}_i \quad \widehat{S}_i$ be the constant mapping that always plays a_i. Player i can costlessly deviate to playing $(1, \widehat{a}_i, \widehat{s}_i)$, leaving player j's induced strategy as $\widehat{s}_j(\sigma_i)$, and increasing his own payoffs, a contradiction. □

[11] See Forges (2013) for a generalization of Kalai et al.'s model where it does.

Inspection of the proof of Proposition 5 reveals that it also holds for any costs of deception that are sufficiently small, and not just for the case where deception is costless.

Proposition 6. For any strategic form game G and for any augmented game $\widehat{G}$ that is induced by G, if $c_1 \leq c_1$ and $c_2 \leq c_2$ then $N_{\widehat{G}}(c_1, c_2) \quad N_{\widehat{G}}(c_1, c_2)$.

Proof. Let $\widehat{G}$ be an augmented game with costs (c_1, c_2). By Proposition 3 for any Nash equilibrium in $\widehat{G}$ there exists another equilibrium with no deception that yields the same payoffs. This last equilibrium still holds when costs are increased to (c_1, c_2): the same deviations are available in both cases, and the higher costs of deception make each deviation strictly less profitable with (c_1, c_2).

When there is no deception in equilibrium, the payoffs are independent of deception costs, and therefore any possible equilibrium payoff vector under (c_1, c_2) remains possible under (c_1, c_2). □

To summarize, Proposition 4 shows that the commitment that can be achieved through communication in which players involuntarily reveal their strategies expands the set of equilibria and generates a folk theorem. Propositions 5 and 6 show that deception has the power to undermine this commitment: the ability to commit is monotonically decreasing with the cost of deception; and easy deception completely undermines communication and commitment.

Credible Strategies. The equilibrium in Example 2 above is sustained by player 2's "threat" to play the dominated action R if player 1 does not play in the way that player 2 wants her to. This raises the question of what is the set of equilibria payoffs in the augmented game in which players are constrained to only use "credible" strategies. It turns out that this is a tricky question. It is natural to define a *credible* strategy to be such that it plays a best response to any strategy of the other player. Player 2's augmented equilibrium strategy in Example 2 is not credible in this sense. But the strategy *nice* in the Prisoners' Dilemma that was described in the Introduction is credible. We leave further study of this question of future work.[12]

6 Successful Equilibrium Deception

In this section we show by example that if the cost of deception is privately known by the players, or equivalently, if the cost of deception is heterogenous and unobservable and players are matched to play the game in pairs, then deception can be successfully practiced in equilibrium, and that those players who engage in deception benefit from it, even compared to other players.

Example 5. Consider again the example of the Prisoner's Dilemma with the following payoffs.

	C	D
C	$3,3$	$0,4$
D	$4,0$	$1,1$

[12] This question is similar in spirit to the question of what is the subgame perfect folk theorem for infinitely repeated games.

Suppose that the cost of deception is independently drawn from the uniform distribution on the interval $[0,2]$. The cost of deception is the private information of the players so that each player knows its own cost of deception and believes the cost of the other player to be uniform on the interval $[0,2]$. The following strategies are an equilibrium of the augmented game that is based on the Prisoner's Dilemma

$$s_i(c_i) = \begin{cases} (1, D, Nice) & \text{if } c_i < 1 \\ (0, Nice, Nice) & \text{if } c_i \geq 1 \end{cases}$$

where *Nice* is defined as in the Introduction to be "if the other player plays *Nice* then *Cooperate* and if the other player plays anything else, play $Defect$."

In this equilibrium, a player with a high cost of deception ($c_i \geq 1$) commits to playing *Nice*, which since the other player plays *Nice* and $Defect$ with equal probabilities yields the payoff $\frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 0 = \frac{3}{2}$. This is higher than or equal to the payoff that the player can get from playing $Defect$, which is 1; or from the payoff that the player can get by deceiving the other player, which is no more than $\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 1 - 1 = \frac{3}{2}$ because with probability one half the other player would play *Nice*, be deceived and therefore cooperate, but with probability one half the other player would anyway play $Defect$, and the player needs to deduct the cost of deception, which is at least one.

A player with a low cost of deception ($c_i < 1$) successfully deceives players with a high cost of deception into playing cooperate, and therefore obtains an equilibrium payoff of $\frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 1 - c_i = \frac{5}{2} - c_i$, which is larger than the payoff that the player would obtain if it just played *Nice*, which is $\frac{3}{2}$ as shown above.[13]

This example shows how the model developed in this paper can account for the experimental evidence described in the introduction. Namely, the opportunity to communicate increases the extent of cooperation; but nevertheless, some players still succeed in inducing the other player to play cooperatively against their own defection.

The observation that is more difficult to account for is the players' correlated play, or the fact that players are more likely to both play cooperatively than can be accounted for by independent mixing. Indeed, the probabilities with which the players play each strategic combination in this example are:

	C	D
C	$\frac{1}{4}$	$\frac{1}{4}$
D	$\frac{1}{4}$	$\frac{1}{4}$

which indicates independent mixing.

However, correlated play can easily be accounted for by our model if we assume that the players' costs of deception are correlated, which seems sensible enough because all it requires is that players with a high cost of deception would be slightly more inclined to believe that the other player is also

[13] Such behavior is consistent with Gneezy's (2005) finding that unless they stand to gain a lot from it, people generally avoid deception.

likely to have a high cost of deception and players with a low cost of deception would be slightly more inclined to believe that the other player is also likely to have a low cost of deception.

The example shows that successful equilibrium deception depends on the number of cheaters, or the probability of deception, to not be too large. This has to hold more generally as well.

Proposition 7. *In a model with privately known costs of deception, if players are sufficiently likely to have a sufficiently low cost of deception, then the set of equilibrium payoffs of the augmented game coincides with that of the basic game (where no deception is possible).*

Proof. The proof follows from the same argument given in the proof of Proposition 5. By continuity, this argument also holds if the costs of deception are sufficiently low with a sufficiently high probability. □

7 Conclusion

This paper proposes a fully rational model of equilibrium deception. It may be argued that unlike in the standard model, in our augmented games, players do not "know" the equilibrium they are playing. But notice that "knowledge of the equilibrium" is an interpretation, not a formal property, of the standard model. And in our augmented game, players best respond to the information or signal they obtain about the other player. In this sense, players are as rational as they can possibly be.

References

[1] Belot, M., V. Bhaskar, J. Van de Ven, G. (2010). "Social Preferences in the Public Arena: Evidence from a Prisoner's Dilemma Game on a TV Show," *mimeo.*

[2] Aumann, R. (1974). "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics* 1, 67-96.

[3] Binmore, K. (1994). *Playing Fair: Game Theory and the Social Contract,* MIT Press, Cambridge, Massachusetts.

[4] Camerer, Colin F.; Spezio, Michael; Wang, Joseph Tao-yi (2010). "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games," *American Economic Review* 100, 984-1007.

[5] Crawford, Vincent P. (2003) "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions," *American Economic Review* 93, 133-149.

[6] den Assem, M. J. V., D. V. Dolder, R. H. Thaler (2010) "Split or Steal? Cooperative Behavior When the Stakes are Large," *mimeo.*

[7] DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L, & Lee, J. (2012). "Detecting the trustworthiness of novel partners in economic exchange," *Psychological Science,* to appear.

[8] Ettinger, D. and P. Jehiel (2010) "A Theory of Deception" *American Economic Journal: Microeconomics* 2, 1-20.

[9] Farrell, J. and M. Rabin (1996) "Cheap Talk," *Journal of Economic Perspectives* 10, 103-118.

[10] Forges, F. (2013) "A folk theorem for Bayesian games with commitment," *Games and Economic Behavior* 78, 64-71.

[11] Forgó, F. (2010) "A generalization of correlated equilibrium: A new protocol," *Mathematical Social Sciences* 60, 186-190.

[12] Frank, R. H. (1998). "Passions within Reason," New York: Norton.

[13] Frank, R. H. (1987). "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?," *American Economic Review* 77, 593-604.

[14] Gauthier, D. (1986). "Morals by Agreement," Oxford: Clarendon Press.

[15] Gneezy, U. (2005) "Deception: The Role of Consequences," *American Economic Review* 95, 384-394.

[16] Kalay A., A. Kalay, and A. Kalay (2003). "Friends or Foes? Empirical Test of a Simple One-Period Nash Equilibrium," *mimeo*.

[17] Kalai, A. T., E. Kalai, E. Lehrer, and D. Samet (2010). "A Commitment Folk Theorem," *Games and Economic Behavior* 69, 127-137.

[18] Kartik, N. (2009) "Strategic Communication with Lying Costs," *The Review of Economic Studies* 76, 1359-1395.

[19] Peters, M., and B. Szentes (2012). "Definable and Contractible Contracts," *Econometrica* 80, 363-411.

[20] Sally, D. (1995). "Conversation and Cooperation in Social Dilemmas," *Rationality and Society* 7, 58-92.

[21] Tennenholtz, M., (2004) "Program Equilibrium," *Games and Economic Behavior* 49, 363-373.

[22] Yaari, M., (2011). "Coping with Correlation," mimeo.