VOLUNTARY COMMITMENTS LEAD TO EFFICIENCY

ADAM TAUMAN KALAI, EHUD KALAI, EHUD LEHRER, AND DOV SAMET

Abstract. Real world players often increase their payoffs by voluntarily committing to play a fixed strategy, prior to the start of a strategic game. In fact, the players may further benefit from commitments that are conditional on the commitments of others.

This paper proposes a model of conditional commitments that unifies earlier models while avoiding circularities that often arise in such models.

A commitment folk theorem shows that the potential of voluntary conditional commitments is essentially unlimited. All feasible and individually-rational payoffs of a two-person strategic game can be attained at the equilibria of one (universal) commitment game that uses simple commitment devices. The commitments are voluntary in the sense that each player maintains the option of playing the game without commitment, as originally defined.

1. Introduction

1.1. **The non-cooperative approach to cooperation.** In their monumental book, von Neumann and Morgenstern (1944) introduced both cooperative and non-cooperative approaches to games. In two-person zero-sum games, there is no conflict between the approaches, since the non-cooperative solutions of these are efficient, i.e., they cannot be mutually improved upon even if the two players cooperate.

However, for general games von Neumann and Morgenstern chose a cooperative approach, assuming that the players will find a way to cooperate and reach efficient outcomes. But later developments in game theory showed that this assumption needs justification. In particular, Nash's non-cooperative solutions to a game, the Nash equilibria, may be quite inefficient and cooperation may prove beneficial to all players.

The inefficiency of Nash equilibria resulted in several directions of research that address the discrepancy between the two approaches. The most effective, perhaps, is to consider games that are played repeatedly, instead of once. Indeed, the *Folk Theorem* of repeated games (see Aumann and Shapley (1976) and Rubinstein (1979) for early versions and Fudenberg and Maskin (1986) for later versions) states that when patient players play a one shot game repeatedly, the equilibrium payoffs of the repeated game are precisely all the individually-rational payoffs that are in the convex hull of all the feasible payoffs of the one-shot game. Thus all efficient outcomes are possible in non-cooperative play.

But what about games that are played only once? Several approaches in game theory address the issue of achieving cooperation in one-shot games.

Date: April 11, 2007.

This paper replaces "Meta-Games and Program Equilibrium," an earier version presented at the Second World Congress of the Game Theory Society in Marseilles (2004) and at the 15th Annual International Conference on Game Theory, Stony Brook (2004). The research of the first two authors is partially supported by the National Science Foundation Grant No. SES-0527656.

A certain degree of cooperation is obtained at correlated equilibria, where the players of the one-shot game are able to use exogenous correlation devices, see Aumann (1974, 1987). But while correlated equilibria are more efficient than Nash equilibria, they still fail to be fully efficient.

Going as far back as the Nash (1953) program, the implementation literature (see for example Jackson (2001)) has taken a brute force approach to the problem.[1] For a given game with inefficient equilibria, an implementor is granted the authority to design a replacement (implementation) game that the players are required to play. Such carefully-designed implementation games have new equilibria which are efficient in the original game.

This paper studies cooperation obtained through voluntary conditional commitments that may be made prior to the play of the game. We first formalize conditional commitments in a manner that is non-circular and well-defined. We then show that in a game that allows voluntary commitments full efficiency is possible under non cooperative Nash equilibria. In parallel to the folk theorem of repeated games, all feasible individually rational payoffs of the one-shot game are Nash equilibria of the one-shot game that allows voluntary commitments.

1.2. **Commitments and conditional commitments.** The idea that a player can improve his outcome in a game through the use of a commitment device goes back to Schelling (1956 and 1960). For example, when a player in a game delegates the play to an agent, with irreversible instruction to play strategy X, the agent may be viewed as a device that commits the player to the strategy X. The strategic delegation literature, see for example Katz and Shapiro (1985) and Fershtman and Judd (1987), studies multi-player delegation and shows that one may obtain partial folk theorems in one-shot games that are played through delegates, see Fershtman, Judd and Kalai (1991).

Indeed, real players often use agents and other commitment devices strategically. Sales people representing sellers, lawyers representing buyers, and sports agents representing athletes are only a few examples. Early price announcements, in newspapers, on the web and in store windows, are commitments to terms of sale by retailers. A limited menu of options posted on the web by an airline is a device that commits the airlines to not discuss certain options that customers may wish to raise.

But real life examples display the use of more sophisticated, conditional, commitment devices. For example, when placing an ad that states "we will sell TVs of brand X at a price of $500, but will match any competitor's price,"a retailer commits itself to a conditional pricing strategy. Such conditional commitment can be more efficient. For example, in oligopoly pricing games match-the-competitors clauses make the monopolist price be a dominant strategy for all sellers, see Kalai and Satterthwaite (1986) and Salop (1986).

Legal contracts are another example of effective conditional commitment devices. Each player's commitment to honor the contract is conditioned on his opponent's commitment to honor the contract. As Kalai (1981) Kalai and Samet (1985) show, under dynamic use of contracts, refined Nash equilibria must converge to partially efficient outcomes.

[1]This literature addresses a more substantial problem, since it also deals with inefficiencies that result from private information.

A recent notion of sophisticated conditional delegation is the program equilibrium of Tennenholtz (2004). In his model, every player in a game delegates the choice of his strategy to a computer program. The program he selects reads the programs selected by his opponents and then outputs a (mixed) strategy that plays the game on his behalf. The equilibria in the game of choosing programs, called program equilibria, are more efficient than the unmodified Nash equilibria of the game. But they fail to reach full efficiency.

In general, however,conditioning requires caution, as conditional commitments may fail to uniquely determine the outcome, lead to circular reasoning, or generate programs that fail to terminate. For example, imagine that each of two retailers places the following ad in the paper: "we sell TVs at a price of $500, but will undercut any competitor's price by $50." Obviously, no pair of prices charged by the two competitors is consistent with their ads, because each of the prices should be $50 lower than the other price, and any dynamic process of changing the prices in an attempt to abide with the ads will never terminate.

Another example is the prisoners' dilemma game. If both players commit to matching the strategy of the opponent then there are two possible outcomes: both cooperate and both defect. But if one player commits to *match* and the other commits to *mismatch* then there are no possible outcomes consistent with such commitments.

Indeed, earlier models in game theory and economics, including metagames, see Howard (1971), and the literature on common agency, see for example Epstein and Peters (1999)[2], encounter such definitional difficulties. In order to deal with the difficulties above, they construct infinite hierarchical spaces in which higher levels of commitments are defined inductively over lower ones.

1.3. **Our approach and main finding.** The current paper offers two main contributions. First, it proposes a general model that encompasses the various approaches and questions above without getting trapped in the definitional difficulties of conditional commitments. Second, using this general model it shows that the potential of conditional commitment devices is essentially unlimited. Since this is a first attempt at such a model, we restrict the analysis to two-person games with complete information, avoiding the controversial modelling choices needed in more advanced settings (see discussion in the concluding section).

In our model a player may delegate her play to a conditioning device that selects her strategy in the game. To avoid circularities and timing issues, we require that the device conditions only on the *conditioning devices* chosen by opponents and not on the *strategies* that are finally used in the game. In the TV ad example, we require every ad to uniquely determine a selling price for every possible *ad* of the opponent. In particular, the introduction of a well-defined device space done in the model below bypasses the need to construct infinite hierarchies of commitments. (This is similar to Harsanyi's (1967) construction of type space that bypassed the need for infinite hierarchies of knowledge).

For a given (arbitrary) two-person strategic game we construct a simple voluntary universal device space. Being voluntary means, informally, that each individual player may choose to commit to a device ahead of time, but may also choose not to commit and just play the game as originally defined. Simplicity assures us that

[2]We thank Sandeep Baliga for pointing this connection.

the play of the game with commitments is possible, even if the number of possible commitment devices is large.

The universality of the space means that a full folk theorem is obtained through play of its (one-shot) induced commitment game. The equilibria of the one-shot commitment game span the convex hull of *all* the individually-rational feasible payoffs of the game without commitments. To generate distributions over payoff profiles obtained from mixed (as opposed to pure) correlated strategies, the universal device space uses jointly controlled lotteries, see Blum (1983) and Aumann and Maschler (1995).

2. A model of commitment devices

In what follows we restrict ourselves to a fixed 2-person game, defined by a triple $G \equiv (N = \{1,2\}, S = S_1 \times S_2, u = (u_1, u_2) : S \to \mathbb{R}^2)$. For simplicity, we assume that the game is finite.

$N = \{1,2\}$ is the set of players, each S_i is a non-empty set describing the feasible *strategies* of player i, and each u_i is the *payoff function* of player i. We use the standard convention where for every player i, player $-i$ denotes the other player.

A *mixed strategy* of player i is a probability distribution σ_i over S_i, with $\sigma_i(s_i)$ describing the probability that player i chooses the strategy s_i. A pair of independent mixed strategies $\sigma = (\sigma_1, \sigma_2)$ induces a probability distribution on S with $\sigma(s_1, s_2) = \sigma_1(s_1)\sigma_2(s_2)$. A *correlated strategy* is a probability distribution γ over S. Clearly, every pair of independent mixed strategies induces the product distribution described above, which is in particular a correlated strategy, but there are correlated strategies that cannot be obtained this way.

For a correlated strategy γ we define the (expected) payoffs in the natural way, $u(\gamma) = E_\gamma(u)$.

A *pure strategy Nash equilibrium* is a pair of strategies s, such that for every player i, $u_i(s)(= u_i(s_i, s_{-i})) \geq u_i(\overline{s}_i, s_{-i})$, for any alternative strategy $\overline{s}_i$ of player i. A *mixed strategy Nash equilibrium* is a vector of mixed strategies $\sigma = (\sigma_1, \sigma_2)$, with the same property, i.e., no player can increase his expected payoff by unilaterally switching to a different mixed strategy.

We say that a correlated strategy γ is *individually rational* if for all $i \in N$, $u_i(\gamma) \geq \min_{\sigma_{-i}} \max_{\sigma_i} u_i(\sigma_1, \sigma_2)$. For each player i let ψ_i be some fixed member of $\operatorname{argmin}_{\sigma_i}(\max_{\sigma_{-i}} u_{-i}(\sigma_1, \sigma_2))$, to be referred to as his *minmax strategy*. So when player i's strategy is ψ_i, then player $-i$'s payoff is at most her *individual rational payoff*.

2.1. Commitment devices and commitment games.

In the model below, sophisticated players choose their conditioning devices optimally against each other. For example, for a pair of devices (d_1^*, d_2^*) to be an equilibrium, d_1^* must be the best device that player 1 can select against the device d_2^* of player 2, taking into account the known responses of d_2^* to hypothetical alternatives to d_1^*.

Non-empty set D_i describes the *conditional commitment devices* (or just devices) available to player i. With every device $d_i \in D_i$ there is an associated *device response function*: $r_{d_i} : D_{-i} \to S_i$ where $r_{d_i}(d_{-i})$ denotes the strategy that d_i selects for player i, if it plays against the device d_{-i} of the opponent.

However, to ease the discussion we use a more compact way of representing the response functions. The responses of the various devices of player i are aggregated into one *(grand) response function* $R_i : D_1 \times D_2 \to S_i$, where $R_i(d_i, d_{-i}) = r_{d_i}(d_{-i})$

describes the strategy chosen by the device d_i of player i when matched against the device d_{-i} of the opponent. The two response functions together describe a *joint response function* $R(d_1, d_2) = (r_{d_1}(d_2), r_{d_2}(d_1))$ where $R(d_1, d_2)$ describe the pair of strategies selected by the devices when they respond to each other.

Note, however, that any function $R\colon D_1 \times D_2 \to S$ is a possible joint response function. This lead to a simple definition of a commitment space.

Definition 1 (Device Space)**.** *A* space of commitment devices *(also a device space) of G is a pair $\mathfrak{D} \equiv (D = D_1 \times D_2, R\colon D \to S)$.*

Each D_i is a none empty set describing the possible devices of player i*, and R is the* joint response function*. The associated* device response functions *are defined (as above) by $r_{d_i}(d_{-i}) = R_i(d_i, d_{-i})$.*

A device space $\mathfrak{D}$ induces a two person *commitment game* $G^{\mathfrak{D}}$ (or device game) in the following natural way. The feasible pure strategies of player i are the devices in the set D_i and the payoff functions are defined by $u(d) = (u_1(R(d)), u_2(R(d)))$ (we abuse notation by using the letter u to denote both, the payoffs in G and the payoffs in $G^{\mathfrak{D}}$).

Definition 2 (Device Equilibrium)**.** *A* commitment equilibrium *(or device equilibrium) of the game G is a pair $(\mathfrak{D}, \sigma)$, consisting of a device space $\mathfrak{D}$ and an equilibrium σ of the device game $G^{\mathfrak{D}}$.*

Clearly, the pair of payoffs of any pair of mixed strategies in the device game, including any device equilibrium, are the payoffs of some correlated strategy in G.

Of special interest to us are the equilibrium payoffs in *voluntary* commitment spaces. These allow each player i to play the game G as scheduled, without making any advanced commitment. In other words, he can choose any G strategy $s_i \in S_i$ without conditioning on the opponent's choices and with the opponent not being able to condition on s_i. Formally, we incorporate this into a device space by adding to it *neutral* (non committal) devices.

Definition 3 (Voluntary)**.** *The device space $\mathfrak{D}$ is* voluntary for player i *if for every strategy $s_i \in S_i$, his set of devices, D_i, contains one designated* neutral device $s_i^{\mathfrak{D}}$ *with the following properties:*

1. *For every $d_{-i} \in D_{-i}$, $r_{s_i^{\mathfrak{D}}}(d_{-i}) = s_i$, and*
2. *For every $d_{-i} \in D_{-i}$, and $s_i, \overline{s}_i \in S_i$, $r_{d_{-i}}(s_i^{\mathfrak{D}}) = r_{d_{-i}}(\overline{s}_i^{\mathfrak{D}})$.*

A voluntary device space *is one that is voluntary for both players.*

3. Elaboration on the model.

A trivial example of voluntary commitment space is with each $D_i = S_i$, where $G^{\mathfrak{D}} = G$. But all the examples discussed in the introduction, delegation to agents, newspaper ads, contracts, program equilibrium, etc., and many more can be effectively described by the model above. The next example illustrates this point.

Example 1 (Price competition)**.** *Consider two retailers,* 1 *and* 2*, preparing to compete in the sales of TVs of brand X in the upcoming weekend. The game G is described by the (per-unit) prices that each retailer may charge, and the payoff of each retailer is the profit realized after informed buyers choose who to buy from. Assume, for simplicity, that there is a known demand curve, that buyers buy from the less expensive retailer, and that if their prices are the same, the demand is equally split.*

As discussed in the introduction, this game lends itself to the use of commitment devices in the form of newspaper ads posted in Friday's newspaper. To fit into the formal model above, we may let D_1 and D_2 describe (respectively) all the ads that the two retailers are allowed to post. With the D_i's specified, it is straightforward to verify that ads lead to well-defined prices: one must check that for every ad of player i, d_i, there is a well define price of retailer i, $r_{d_i}(d_{-i})$, resulting from every competitor's ad, $d_{-i} \in D_{-i}$. This formulation disallows vague ads, like "I will undercut opponents' prices by \$50," which fail to specify a response price to an identical competitor's ad.

Notice that in parallel to the above mathematical need for coherent responses, there is a legal real-life need for coherence. This could be accomplished through a variety of restrictions. For example, the paper may insist that an ad consist of two items, a posted price, p, and a rule, h, that responds to posted (not computed) prices of the opponent. In this case, the device set of player i consists of all such pairs (p_i, h_i), and if retailers 1 and 2 place the ads $d_1 = (p_1, h_1)$ and $d_2 = (p_2, h_2)$ then the selling prices are $R(d_1, d_2) = (h_1(p_2), h_2(p_1))$.

3.1. **More effective model.** Earlier attempts to deal with sophisticated conditional commitments (without the use of well defined commitment device spaces) lead to difficult models. Howard (1971) wanted to describe a notion of a meta strategy, one that conditions its choice of an action based on the action chosen by the opponent. For example, a player in a one shot prisoners' dilemma game should be able to match-the-opponent, and in effect induce a tit-for-tat strategy in the one shot game.

But this plan proved to be difficult due to the issue of timing. How can a player react to his opponent's choice, if they play simultaneously?

Howard's solution was to construct an infinite hierarchical structure of reaction rules: At the lowest level each player chooses a strategy in the underlying game, and at level $t+1$ he specifies response rules to his opponent level t rules.[3]

But hierarchical structures have not proven useful in dealing with applications, and the space of commitment devices used in this paper bypasses the need for such complex structures. This is similar to Harsanyi's (1967) use of a space of types that bypasses the need for a complete hierarchical structure of knowledge about knowledge a la Mertens and Zamir (1985).[4]

The following example may help in illustrating this point.

Example 2 (Divorce-settlement)**.** *This game is a simple model of divorce between two players, he and she. The underlying game is exactly like the standard Prisoners' Dilemma game with cooperative (c) and aggressive (a) strategies.*

But assume now that each player has the option of choosing a lawyer to represent him in the game and that lawyers are of two possible types: flexible (fl) and tough (tl) (and lawyers know the types of other lawyers).

No matter who they face, tl's choose the strategy a. But fl's choose the strategy c when they face an opponent of type fl, and choose the strategy a against all others.

[3]Similar issues are addressed in the economic literature on common agencies , see for example Epstein and Peters (1999), where each firm is allowed to condition its strategies on the strategies chosen simultaneously by its rival firms.

[4]Unlike Harsanyi's model that superseded the hierarchical structure of Mertens and Zamir, here the simple device model comes after the existence of the hierarchical models.

The lawyers fit the description of commitment devices in the model above, and the table below describes the game in which players have the choice of delegating the play to a lawyer or playing on their own using the (non committal) neutral devices, cd, or ad.

Notice that the restrictions made on neutral devices are satisfied. For example, when Player 1 "delegates" to cd, regardless of what device is used by the opponent, Player 1 ends up with the action c. Moreover, the devices of Player 2, in choosing an action for Player 2, never differentiate between the devices cd and ad of Player 1 (the second entries in the two bottom cells of every column are identical).

If one substitutes the prisoners' dilemma payoffs in the sixteen cells in the table (assuming that the lawyers fees are negligible :), it is easy to see that fl,fl is a dominant strategy equilibrium. In effect, this equilibrium employs a tit-for-tat type of strategy to get cooperation in this one shot prisoners' dilemma game: a player deviating from fl causes the opponent's device to switch from c to a.

		Pl 2			
		fl	tl	cd	ad
	fl	c,c	a,a	a,c	a,a
Pl 1	tl	a,a	a,a	a,c	a,a
	cd	c,a	c,a	c,c	c,a
	ad	a,a	a,a	a,c	a,a

4. A folk theorem in a universal device space

In the Divorce Settlement Example above, it is easy to generate cooperation through the use of commitment devices. But this task is more difficult in examples of the following type.

Example 3. *(2-person Cournot entry game)*

		Pl 2	
		in	out
Pl 1	in	1, 1	10,0
	out	0,10	0 ,0

Here, the minmax strategy for both players guarantees each player a payoff of at least 1. But unlike in the prisoners' dilemma game, there is no pair of pure strategies that simultaneously yield each player a payoffs greater than 1. Yet a full folk theorem should attain every payoff in the convex hull of $\{(1,1),(1,9),(9,1)\}$, for example $(5,5)$, as an equilibrium payoff.

In the repeated-game folk theorem, this is not a problem since the players can alternate in playing the cells (in, out) and (out, in), and a trigger strategy will induce the correct incentives to do so. But such alteration is impossible if the game is played only once.

As it turns out, however, such alterations may be replaced by jointly controlled lotteries, a la Blum (1983) and Aumann and Maschler (1995). The folk theorem below illustrates how this can be done with the appropriate incentives.

Theorem 1 (Commitment-device folk-theorem)**.** *For the two player game G, there exists a voluntary commitment device space $\mathfrak{U} = (U, L)$ with a commitment game $G^{\mathfrak{U}}$ that has the following property. Every individually rational correlated strategy in the*

game G can be obtained as a (mixed strategy) Nash equilibrium of the commitment game $G^{\mathfrak{U}}$.

Proof. We first construct $\mathfrak{U}$. The game will have infinite sets of strategies. The strategies of a player are a triple, where the first part is an encoding of a correlated strategy, the second part is a number in the interval $[0,1]$, and the third part is a fall-back strategy in S_i. Let $M = |S|$ and let $[M]$ denote $\{1, 2, \ldots, M\}$.

We now describe a method for encoding any correlated strategy γ over S by a unique $x \in \Delta_M = \{x \in [0,1]^M | \sum_i x_i = 1\}$, the simplex of dimension $M-1$. The important property is that there is a function $f : \Delta_M \times [0,1] \to S$ such that the probability that $f(x,r) = s$ for a uniformly random $r \in [0,1]$ is the same as the probability assigned to s by γ. (There are several ways to achieve this, and any other method of achieving it would be satisfactory.) For completeness, we give one such encoding now. Any $x \in \Delta_M$ corresponds to a probability distribution over $[M]$ by choosing r uniformly from $[0,1]$ and the following map $g : \Delta_M \times [0,1] \to [M]$,

$$g(x,r) = \min\big\{i \in [M] \;\big|\; x_1 + x_2 + \ldots + x_j \geq r\big\}.$$

Finally, let $\pi : [M] \to S$ denote an arbitrary bijection from $[M]$ to S. The map π should be fixed and known in advance to all players. Hence, Δ_M gives a unique encoding of correlated strategies over S, where the correlated strategy corresponding to $x \in \Delta_M$ is chosen by picking r uniformly at random from $[0,1]$ and taking $f(x,r) = \pi(g(x,r))$.

We can now specify $\mathfrak{U} = (U, L)$. $U = (\Delta_M \cup \{\perp\}) \times [0,1] \times S_i$. The special symbol $\perp$ is necessary to make the game voluntary, and indicates that the player wants to play the fall-back strategy, and L is defined by,

$$L\big((x_1, r_1, s_1), (x_2, r_2, s_2)\big) = \begin{cases} f(x_1, r_1 + r_2 - \lfloor r_1 + r_2 \rfloor) & \text{if } x_1 = x_2 \text{ and } x_1 \neq \perp \\ (s_1, s_2) & \text{otherwise} \end{cases}$$

The expression $r_1 + r_2 - \lfloor r_1 + r_2 \rfloor$ above computes the fractional part of $r_1 + r_2$.

Now let γ be an individually rational correlated strategy of G. We will see that there is a mixed device equilibrium of $\mathfrak{U}$ with an outcome distribution that coincides with the correlated strategy γ. Let x be the unique encoding of γ so that, for any $s \in S$, the probability that $f(x,r) = s$ is equal to the probability that γ assigns to s. Take the mixed device for each player μ_i that chooses (x_i, r_i, s_i) by taking $x_i = x$ (with probability 1), $r_i \in [0,1]$ uniformly at random and, independently, s_i according to the mixed minmax strategy of player i.

To see that $\mu = (\mu_1, \mu_2)$ has the desired properties, notice first that for any r_i chosen by player i, the equilibrium strategy of the opponent induces the distribution γ on S. In other words, player i cannot gain by deviating from the uniform distribution on his r_i's. Moreover, deviating by submitting a vector $x_i' \neq x$, makes him face the minmax distribution of his opponent, which can only decrease his payoff.

The game is voluntary because player i has neutral strategy $(\perp, 0, s_i)$ for any strategy $s_i \in S_i$. □

4.1. **Finite number of devices.** The above commitment space is infinite. It is important to note that a finite version approximation of the above folk theorem can be made where correlated strategies have coefficients that are integer multiples of $1/n$, meaning that the probabilities assigned to the different strategies $s \in S$ are in

the set $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$. While this does not give a full-folk theorem, it is sufficient for many practical purposes and has the advantages of being finite.

Theorem 2 (Finite commitment-device folk-theorem)**.** *For the two player game G and any $n \geq 1$, there exists a* finite *voluntary commitment device space $\mathfrak{U}_n = (U_n, L_n)$ with a commitment game $G_n^{\mathfrak{U}}$ that has the following property. Every individually rational correlated strategy in the game G whose coefficients are in $\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ can be obtained as a (mixed strategy) Nash equilibrium of the commitment game $G_n^{\mathfrak{U}}$. Moreover, the function L_n can be computed in time polynomial in $\log(n)$.*

The proof of the above theorem is nearly the same as that of Theorem 1. The only difference is that the correlated strategies (and simplex) are discretized to an accuracy of $1/n$ and the players choose $r_i \in \left\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\right\}$ uniformly at random. Such numbers are represented using $O(\log n)$ bits. The function L_n is straightforward to efficiently compute, i.e., compute in time polynomial in the input length.

In some applications, a finite number of commitment devices may be sufficient to achieve a full folk theorem. It may be useful to know, however, that for the folk theorem with the generality above (one universal commitment space that achieves all equilibria of the game G) one needs infinitely many devices, unless the game is of a very narrow form. The following is a sketch of such a theorem and its proof.

Theorem 3. *For any two player game G the following two conditions are equivalent:*

1. There exists a finite device space $\mathfrak{U}$ in which every individually rational correlated strategy in G can be obtained as a Nash equilibrium of $G^{\mathfrak{U}}$,

2. The feasible payoffs set of G is a rectangle with facets parallel to the axes.

Proof. If (2) holds, then there are four payoffs in the game which are the extreme points of the feasible set. Thus, one can define a 2×2 device game in which each player controls the payoff of the other and has no say over her own payoff. The equilibrium payoffs in this game are the entire feasible set of G.

As for the converse, assume (1) and that (contrary to (2)) one the facets of G's payoffs, say F, is not parallel to one of the axes. Since F is a facet of the feasible set, in order to obtain a (correlated) payoff in F, all the payoffs involved should be also in F.

Let $\sigma = (\sigma_1, \sigma_2)$ be any equilibrium of $G^{\mathfrak{U}}$ whose payoff is in F and let $\mathfrak{U}^{\sigma} = (D_1^{\sigma} \times D_2^{\sigma}, T)$ where each D_i^{σ} denotes the supports of σ_i. The payoffs of $G^{\mathfrak{U}^{\sigma}}$ are all in F. Moreover, σ induces a full-support equilibrium of $G^{\mathfrak{U}^{\sigma}}$.

Consider any subspace $\mathfrak{U}' = (D_1' \times D_2', T)$ of $\mathfrak{U}$ where all payoffs of $G^{\mathfrak{U}'}$ are in F. By a linear transformation of the payoffs of player 1, $G^{\mathfrak{U}'}$ can be transformed to a zero-sum game, say $G_0^{\mathfrak{U}'}$. As a zero-sum game $G_0^{\mathfrak{U}'}$ has only one equilibrium payoff. In particular, all full-support equilibria of $G_0^{\mathfrak{U}'}$ induce the same payoff.

Since $G_0^{\mathfrak{U}'}$ is derived from $G^{\mathfrak{U}'}$ by a linear transformation (of the payoffs of one of the players), any full-support equilibrium of $G_0^{\mathfrak{U}'}$ is a full-support equilibrium of $G^{\mathfrak{U}'}$. Consequently, any $G^{\mathfrak{U}'}$ has only one full-support equilibrium payoff. Since there are finitely many subgames $G^{\mathfrak{U}'}$ in $G^{\mathfrak{U}}$ with payoffs in F, and each has at most one full-support equilibrium payoff, there are only finitely many equilibrium payoffs of $G^{\mathfrak{U}}$ in F. Thus, the equilibrium payoffs of $G^{\mathfrak{U}}$ cannot cover all the correlated equilibrium strategies payoffs in F. This contradiction leads to the conclusion that

if $G^{\mathfrak{U}}$ is finite, then all the facets of the feasible set of $\mathfrak{U}$ are parallel to the axes. □

4.2. **Contrasts between commitment and correlated equilibria.** The notion of a commitment device, introduced here, is different from Aumann's (1974, 1987) notion of correlation device in some important ways. Given a strategic game G, a correlation device outputs, prior to the start of the game, a vector of individual private messages generated by a commonly known probability distribution. The players proceed to play G after learning their private messages. Once a player received a signal he has no way to affect the distribution over other players' strategies.

To generate a correlated equilibrium one needs an external impartial mediator, or, alternatively, use a system of devices that produces signaling that induce the desired correlated distribution over the game outcome (see, Barany (1992), Lehrer (1996), Lehrer and Sorin (1997), Ben-Porath (1998), Gossner (1998), and Urbano and Vila (2002)).

In a commitment space, on the other hand, there is only common knowledge of the individual devices that the players may use. Players may delegate the play to their devices prior to the start of the game. In a commitment equilibrium there is a commonly known probability distribution over commitments, which in turn induces commonly known probability distribution over outcomes. However, by changing commitments a player may change the probability distribution over other players' strategies (in the original game).

As it turns out, the set of commitment equilibrium payoffs is significantly larger than that of correlated equilibrium. In the Cournot Entry game above the *only* correlated equilibrium is for both players to choose *in*, with the payoffs $(1, 1)$. But as the folk theorem above illustrates any pair of payoffs in the convex hull of $\{(1, 1), (1, 9), (9, 1)\}$, including $(5, 5)$, can be obtained at a commitment equilibrium.

4.3. **Contrasts between commitment and program equilibrium.** Tennenholtz (2004) presents a *partial* folk theorem using program equilibria: The program equilibrium payoffs of a game G consist of all the individually-rational payoff pairs that can be obtained through *independent* (not correlated) mixes strategies of G. Applying the result of Tennenholtz to the Cournot entry game above the largest symmetric program-equilibrium payoffs are $(2\frac{7}{9}, 2\frac{7}{9})$, short of the efficient payoffs $(5, 5)$ that can be obtained at a commitment equilibrium.

Tennenholtz's programs may be viewed as commitment devices, but there are important differences between the formal models. A commitment device, as defined in this paper, outputs a pure strategy for a player. A program, in Tennenholtz's model, outputs a mixed strategy for a player. Thus, Tennenholtz's programs offer more flexibility then our commitment devices.

Given this added flexibility, one would expect Tennenholtz to get a larger, rather than the obtained smaller, set of equilibrium payoffs. But there is another important difference. Tennenholtz's analysis is restricted to the payoffs obtained through the use of *pure*-strategy program equilibria, while our model allows for *mixed*-strategy commitment equilibria.

It may be tempting to work within Tennenholtz's model, and to study the set of mixed-strategy program-equilibrium payoffs. But then, as our model shows, one can achieve a folk theorem with devices that output deterministic strategies.

5. Additional remarks

To avoid controversial modeling choices, the discussion above was restricted to a simple model rich enough for meaningful positive results. But there are several important directions to investigate.

5.1. **Extensions to n-players.** When dealing with more than two players, repeated-game folk theorems bring about some modeling choices. For example, if player i deviates from the equilibrium, can the remaining players secretly correlate their future strategies in order to achieve a more effective punishment against him? Different answers to the above question lead to different possible equilibrium sets.

Similar related choices must be faced when dealing with commitment devices of more than 2 players. For example, in the two player case above we assume that every player's device can condition on (e.g., see) the device used by his opponent. When we deal with more players, are all devices fully visible to all the players' devices, or should we allow each coalition to have devices that are only visible to the devices of its own members?

What equilibrium payoffs can be obtained under a various visibility assumptions? Can the results of Aumann (1961) on Alpha and Beta cores in repeated games be reproduced in one shot games with devices?

5.2. **Commitment in Bayesian games.** Restricting ourselves to complete information games, the folk theorem above shows that strategic inefficiencies may be removed by commitments. The following example shows that one may expect similar improvements with regards to informational inefficiencies.

Example 4 (Hunting a hidden stag)**.** *Consider two players,* 1 *and* 2*, and three locations,* H_1, H_2*, and* H_3*. A prize is located at random in one of the three locations (with probability* 1/3 *for each), and each player* i *, who is initially located at* H_i *, is told whether the prize is at* his location*, or not. Following this, in one simultaneous move, each player chooses one of the three locations. If both players choose the location with the prize they are paid \$1 each, otherwise zero. Assuming no communication, the highest achievable equilibrium payoff is* 2/3 *each.*

When dealing with commitments in Bayesian games, there are several modeling choices. For example, are the individual commitments done before or after the private information is revealed. Assuming the latter, the example above illustrates that commitments may be used as means for efficient communication.

Consider a commitment space in which each player i has two devices, s_i (for stubborn) and f_i (for flexible). The device s_i chooses the location H_i no matter what device is used by the opponent. The device f_i chooses the location H_{-i} against the device s_{-i} of the opponent, but chooses H_3 against the device f_{-i} of the opponent. Consider the strategy profile where each player i chooses s_i when the prize is at his location and f_i otherwise. It is easy to see that this is an equilibrium that guarantees that they both show up at the right place, whichever one it is.

5.3. **Uncertain, partial, and dynamic commitment.** What can be achieved by devices that are not fully observable? This issue was partially studied in the delegation literature. For example, Katz and Shapiro (1985) argued that unobserved delegation could not really change the equilibrium of a game. On the other hand Fershtman and Kalai (1997) have shown that under restriction to *perfect* Nash equilibrium, even unobserved delegation may drastically affect payoffs.

Another important direction is partial commitments. What if the commitment devices do not fully determine the strategies of their owners, but only restrict the play to subsets of strategies, to be completed in subsequent play by the real players?

It seems that a fully developed model of commitments should allow for the options above and more. It should be dynamic, with gradually increasing levels of commitments that are only partially observable.

5.4. **Commitment and Implementation.** There are some important differences between these two areas. In much of the implementation literature, an outsider is in charge of constructing the alternative game to be played. In this paper, we think of the players as volunteering to make their own individual commitments. But our model of commitments lends itself to other interpretations.

In this regard, it is useful to note that the device space constructed in the folk theorem above does not require any knowledge of the players' payoffs, but only knowledge of their feasible strategies. This means that the construction of the device space may be accomplished by an uninformed implementor, as is the case in many models of the implementation literature.

5.5. **Commitment devices for self control.** Commitments against one's future self were studied in Ferreira, Maschler and Gilboa (1995). What is the best thing to do now, if I know that my future actions will be dictated by new preferences that contradict my current ones (I don't want and don't plan to eat the peanuts in tonight's party, but I know that I will once I am there)? In the current paper, players conditioned their choices on the choices of opponents. But a special interesting case, in a dynamic commitment model, would allow conditioning on a player's own future selves.

5.6. **Coordination, contracts, and cryptography.** The commitments we have described are different than contracts in the sense that both sides make their own commitments. Thus less communication is required than in designing a contract, where both sides must agree on the same document. As we have shown, it is possible to achieve a folk theorem without agreeing upon a contract. However, our construction requires substantial coordination in order to achieve anything but minmax play.[5]

Finally, an alternative would be to achieve coordination using *cryptography* and contracts. Players could describe a single correlated strategy in the contract and agree to play it based upon a joint lottery (Blum, 1983; Aumann and Maschler, 1995). The random numbers required for the joint lottery could be securely encoded within the contract using a computational commitment device, see, e.g., Naor (1991). For example, a player can commit to a random bit $b \in \{0, 1\}$ by writing down a number n which is either the product of two ($b = 0$) or three ($b = 1$) large prime numbers. Since it is believed that counting the number of factors of a large number is computationally hard, their choices remain essentially secret until the factorization is revealed. (This can be repeated to commit to as many bits as

[5]Alternative schemes could allow plays to quickly commit to play a "default" correlated strategy, where the default would presumably be an individually rational correlated strategy that maximizes some quantity (e.g., sum of payoffs). In order to play this strategy, they would simply need to commit to it and submit a single random number, rather than explicitly stating the correlated strategy.

necessary.) Finally, after the contract is signed, both parties reveal the factorizations of their numbers and the play will be determined and legally enforceable.

6. References

Aumann, R.J. (1961), "The core of a cooperative game without side payments," *Transactions of the American Mathematical Society,* 98, 539–552

Aumann, R.J. (1974), "Subjectivity and correlation in randomized strategies." *Journal of Mathematical Economics*, 1, 67-96.

Aumann, R.J. (1987) "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica*, 55(1), 1-18.

Aumann, R.J. and M. Maschler (1995), *Repeated Games with Incomplete Information*, MIT Press.

Barany, I. (1992), "Fair distribution protocols or how players replace fortune," *Mathematics of Operations Research*, 17, 327-340.

Ben-Porath, E. (1998), "Correlation without mediation: expanding the set of equilibrium outcomes by cheap pre-play procedures,' *Journal of Economic Theory*, 80, 108-122.

Blum, M. (1983), "Coin Flipping by Telephone: A Protocol for Solving Impossible Problems," SIGACT News, 15(1), 23-27.

Epstein, L., and M. Peters (1999), "A Revelation Principle for Competing Mechanisms,"*Journal of Economic Theory*, 88, 119-161.

Ferreira, J.L., I. Gilboa, and M. Maschler M. (1995), "Credible Equilibria in Games with Utilities Changing during the Play," *Games and Economic Behavior*, 10(2), 284-317.

Fershtman, C., and K. Judd (1987), "Equilibrium Incentives in Oligopoly," *American Economic Review*,77(5), 927-940.

Fershtman, C., K. Judd and E. Kalai (1991), "Observable Contracts: Strategic Delegation and Cooperation," *International Economic Review*, 32(3), 551-59.

Fershtman, C. and E. Kalai, "Unobserved Delegation," *International Economic Review*, 38(4), 763-74.

Fudenberg, D., and E. Maskin (1986), "Folk Theorem for Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54(3), 533-554.

Gossner, O (1998) "Secure Protocols or How Communication Generates Correlation," *Journal of Economic Theory*, 83(1), 69-89.

Harsanyi, J. (1967), "Games with incomplete information played by Bayesian players", *Management Science*, 14, 159-82.

Howard, N. (1971), Paradoxes of Rationality: Theory of Metagames and Political Behavior, The MIT Press, Cambridge.

Jackson, M.O. (2001), "A Crash Course in Implementation Theory," *Social Choice and Welfare*, 18(4), 655-708.

Lehrer, E. (1996), "Mediated talk," *International Journal of Game Theory*, 25, 177-188.

Lehrer, E. and S. Sorin (1997), "One-Shot Public Mediated Talk," *Games and Economic Behavior*, 20(2), 131-148.

Kalai, E. and M. Satterthwaite (1986) "The Kinked Demand Curve, Facilitating Practices and Oligopolistic Competition," DP 677, Center for Math Studies in Econ and Mgt Science, published also in Imperfection and Behavior in Economic

Organizations, R. P. Gilles and P. H. M. Ruys (eds.), Kluwer Academic Publishers, 1994, pp. 15-38

Kalai, E. (1981), "Preplay Negotiations and the Prisoner's Dilemma," Mathematical Social Sciences, 1(4), 375-379.

Kalai, E. and D. Samet (1985), "Unanimity Games and Pareto Optimality," International Journal of Game Theory, 14(1), 41-50.

Katz, M.L., and C. Shapiro (1985) "Network Externalities, Competition, and Compatibility," The American Economic Review, 75(3), 424-40.

Salop, S.C. (1986), "Practices that (Credibly) Facilitate Oligopoly Coordination," Analysis of Market Structure, Cambridge: MIT Press, 265-90

Mertens, J.F., and S. Zamir (1985), "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, 14(1), 1–29.

Naor, M. (1991), "Bit Commitment Using Pseudo-Randomness," *Journal of Cryptology*, 4(22), 151-158.

Schelling, T.C. (1956), "An Essay on Bargaining," *The American Economic Review*, 46(3), pp. 281-306

Schelling, T. C. (1960), The Strategy of Conflict. Cambridge, Mass.: Harvard University Press.

Tennenholtz, M., (2004), "Program Equilibrium", *Games and Economic Behavior*, 49, 363–373.

Urbano, A. and J. E. Vila (2002), "Computational complexity and communication: coordination in two-player games,' *Econometrica*, 70, 1893- 1927.

College of Computing, Georgia Institute of Technology
E-mail address: atk@cc.gatech.edu

Kellogg School of Management, Northwesetern University
E-mail address: kalai@kellogg.northwestern.edu

School of Mathematical Sciences, Tel Aviv University
E-mail address: lehrer@tau.ac.il

Faculty of Management, Tel Aviv University
E-mail address: samet@tau.ac.il