Using Contracts to Influence the Outcome of a Game[*]

Robert McGrew and **Yoav Shoham**
Computer Science Department
Stanford University
Stanford, CA 94305
{bmcgrew, shoham}@cs.stanford.edu

Abstract

We consider how much influence a center can exert on a game if its only power is to propose contracts to the agents before the original game, and enforce the contracts after the game if all agents sign it. Modelling the situation as an extensive-form game, we note that the outcomes that are enforceable are precisely those in which the payoff to each agent is higher than its payoff in at least one of the Nash equilibria of the original game. We then show that these outcomes can still be achieved without any effort actually expended by the center: We propose a mechanism in which the center does not monitor the game, and the contracts are written so that in equilibrium all agents sign and obey the contract, with no need for center intervention.

Introduction

There has been much interest in AI in mechanism design, the area of game theory devoted to designing protocols for self-interested agents. In the literature (Mas-Colell, Whinston, & Green 1995) it is generally assumed that the mechanism designer has complete freedom in designing the rules of the game. Yet the world is full of strategic situations with rules that already exist and cannot be changed arbitrarily. Recent work on *k-implementation* (Monderer & Tennenholtz 2003) restricts the capabilities of the mechanism designer in a particular way – it can add to any given cell in the payoff matrix, but it cannot subtract. (The interesting results in that line of work concern cases in which, despite that addition, the cost to the center in equilibrium is zero.) The opposite of this setting would be one in which the center can impose fines, rather than bonuses. This in and of itself is not interesting, because with sufficiently large fines any outcome can be enforced. However, suppose the mechanism cannot unilaterally impose fines, but it can do so in the context of a signed contract. Specifically, we consider the following class of mechanisms. Given a game G, the center can:

1. Propose a contract before G is played. This contract specifies a particular outcome, that is, a unique action for each agent, and a penalty for deviating from it.

2. Collect signatures on the contract and make it common knowledge who signed.

3. Monitor the players' actions during the execution of G.

4. If the contract was signed by all agents, fine anyone who deviated from it as specified by the contract.

Our setting is reminiscent of the work on social laws and conventions (Shoham & Tennenholtz 1997). There too the center can offer a *social convention* to the players, where each player agrees to a particular outcome so long as the other players play their part. The difference is that in that work it is assumed that, once all agents agree, the center has the power to enforce that outcome. Here we assume that players still have the freedom to choose whether or not to honor the agreement; the challenge is to design a mechanism such that, in equilibrium, they will.

The technical results of this paper will refer to games of complete information, but for intuition consider the example of online auctions, such as those conducted by eBay. Consider the complete game being played, including the decision after the close of the auction by the seller of whether to deliver the good and by the buyer of whether to send payment. Straightforward analysis shows that that the equilibrium is for neither to keep his promise, and the experience with fraud on eBay (Hafner 2004) demonstrates that the problem is not merely theoretical. It would be in eBay's interest to find a way to enable its customers to bind themselves to their promises.

Our first interest will be to characterize the achievable outcomes: What outcomes may the center suggest, with associated penalties, that the agents will accept? Our first result will be an observation that the center's power is quite broad: Any outcome will be accepted when accompanied by appropriate fines, so long as the payoffs of each agent in that outcome are better than that player's payoffs in *some* equilibrium of the original game.

Although the center can achieve almost any outcome, we note that the helpful center expends a large amount of effort to do so: suggesting an outcome, collecting signatures, observing the game, and enforcing the contracts. If this procedure happens not just for one game, but for hundreds or thousands per day, the center may wish to find a way to avoid this burden while still achieving the same effect.

The bulk of this paper concerns ways in which this reduction in effort can be achieved. We continue to assume that the center still needs to propose a contract. We also simply assume that it does not monitor the game. Nor does it participate in the signing phase; the agents do that among themselves using a broadcast channel. While we might imagine that the players could simply broadcast their signatures, this protocol allows a single player to learn the others' signatures and threaten them with fines. Nonetheless, we can construct a more complicated protocol - using a second stage of contracts - which does not require the center's participation. The only phase in which the center's protocol requires it to get involved under some conditions is the enforcement stage. However, our goal will be to devise contracts so that, *in equilibrium*, at this stage too the center sits idle. Our results here will be as follows. If the game play is *verifiable* (if the center can discover after the fact how the game played out), we can achieve all of the outcomes achievable by a fully engaged center. If the game is not verifiable then we can still achieve all previously achievable outcomes with some contract, but that contract might allow equilibria with additional outcomes.

The rest of the paper is organized as follows: we first formally define our setting. Then we characterize the set of outcomes which are achievable with a busy center using contracts in this game. Finally, we lighten the load on the center first in the enforcement stage, then in the signature exchange stage.

Formal Setting

The strategic situation the center wishes to influence can be characterized as a *strategic-form game* with consequences in O: $G = \langle N, A, O, g, V \rangle$. (We roughly follow the notation of (Osborne & Rubinstein 1994).) Here N is the set of players $\{1...n\}$. $A = A_1 \times A_2 \times ... \times A_n$, where A_i is the set of actions which can be taken by an individual agent. We will use a_i to refer to an action of i in G and a_{-i} to refer to the vector of actions of all other players. O is the space of outcomes; $g : A \to O$ determines the outcome after an action profile. We identify each outcome $o_{(a_i, a_{-i})}$ with a distinct action profile (a_i, a_{-i}), and assume $g(a_i, a_{-i}) = o_{(a_i, a_{-i})}$. $V = V_1 \times V_2 \times ... \times V_n$, where $V_i : A \to \mathbb{R}$ is the pay-off function for player i.

Before this strategic situation G occurs, the helpful center suggests a contract to the players. This contract specifies the outcome o suggested by the center and what actions h the center will take in response to different action profiles of the players. The center will not enforce this contract unless it is signed by all players. This contract defines the center's protocol in the enforcement stage H, as described below. We will denote a contract that describes a particular center protocol h as c_h.

Now we will describe the stages of the game, as initially formulated. We will adjust this formulation in later sections so that the center does no work in equilibrium.

Signature Exchange Stage F Each player who assents sends his signature on the contract to the center, who collects them. The center notifies all players of the identi-

ties of the signers. At the end of this stage, it is common knowledge whether or not the contract will be enforced.

Execution Stage G The players play the game G. Each player may take his action a_i to achieve o or he may not. The center observes the actions taken by the players.

Enforcement Stage H The center takes the actions specified in the contract in response to the actions he observed.

The outcomes are a consequence of the execution stage, but the only way the center can affect the players' actions in the is by fining them in the enforcement stage.

The extended game which arises from playing the stage games one after another we denote by $X = F \cdot G \cdot H$. Together, these define an *extensive-form game with simultaneous moves*. In general, an extensive-form game X can be defined as $X = \langle N, \Omega, A_\omega, P, U \rangle$, where N is again the players, Ω is the set of histories of actions taken, A_ω is the set of actions for all players that can be taken after history ω, $P : \Omega \to 2^N$ is the player function that defines which players get to move after a given history, and U is the utility function of players in the entire game.

In our particular setting, the history ω is just the set of actions taken in each stage game played so far, A_ω is the set of actions possible in each stage game following history ω, P is the set of all players (all players move simultaneously in each stage), and U is the (undiscounted) sum of the utilities of each stage game. We denote the subgame of $X = \langle N, \Omega_{|\omega}, A_{|\omega}, P_{|\omega}, U_{|\omega} \rangle$ that arises after history ω by $\Gamma(\omega)$, which simply refers to the play of the remaining stage games following the actions taken in ω. In later sections, we will refer to the strategy space of stage F as Γ^F, of stage G as Γ^G, and of stage H as Γ^H.

A *pure strategy* σ_i for player i in a strategic-form game corresponds to the choice of a single action $\sigma_i \in A_i$. A *mixed strategy* corresponds to the choice of a distribution over actions: $\sigma_i \in \Delta A_i$. A pure strategy in an extensive form game is defined as $\sigma_i : \Omega \to A_\omega$; a mixed strategy is defined accordingly. If σ_i is a strategy in X, then the strategy $\sigma_{i|\omega} : \Omega_{|\omega} \to A_{|\omega}$ induced by σ_i in the subgame $\Gamma(\omega)$ is $\sigma_{i|\omega}(\omega') = \sigma_i(\omega, \omega')$. Since it is unobservable whether a player has played a particular mixed strategy (only the realization of that strategy is observed), we will henceforth concentrate on enforcing outcomes that are the consequence of pure strategy profiles.

Our chosen solution concept will be subgame perfect equilibrium. To define this, we must first define a *Nash equilibrium*: a profile of strategies σ is a Nash equilibrium in a game if $\forall i \in N, \sigma_i' \in \Delta A_i : U_i(\sigma_i, \sigma_{-i}) \geq U_i(\sigma_i', \sigma_{-i})$. A profile of strategies σ in an extensive form game is a *subgame perfect equilibrium* if for every $\omega \in \Omega$, $\sigma_{|\omega}$ is a Nash equilibrium of the subgame $\Gamma(\omega)$. A subgame perfect equilibrium is resistant to deviations by players even in subgames off the equilibrium path.

The Power of a Helpful Center

We wish to characterize the power of a helpful center without any resource limitations. In this section, the center is limited only by the voluntary consent required from all

agents and by its lack of desire to spend its own money. Specifically, we assume that it collects the signatures in F itself and that it monitors the players' actions in G. In later sections we will relax each of these two assumptions.

First, we must precisely define the game which is being played. We model the signature exchange stage as a game form F with players N and action space $\Gamma^F = \{0,1\}^n$. The player i assents to the contract if $\gamma_i^F \in \Gamma_i^F = 1$. Since the center broadcasts the identities of the signers, each player's action is common knowledge. The execution game G thus has an extended action space $\Gamma^G : \{0,1\}^n \to A$ in which players decide to take action based on the consequences of the signature exchange stage. In the initial formulation, the enforcement stage H requires no action on the part of the players, but only of the center. The center's protocol h sets the payoff function of the enforcement stage. The center observes the signatures it receives and the actions chosen by the players and chooses to fine or reward players. Formally, $h = h_1 \times h_2 \times ... \times h_n$ and $h_i : \{0,1\}^n \times O \to \mathbb{R}$.

We define the payoff function $U_i : \Gamma \to \mathbb{R}$ for each player in the extended game X given actions $v \in \{0,1\}^n$ and $(a_i, a_{-i}) \in A$ as $U_i(v, a_i, a_{-i}) = V(g(a_i, a_{-i})) + h_i(v, g(a_i, a_{-i}))$. Thus each player has a quasi-linear utility function over the outcome determined in G and the money taken or given by the center according to h.

We say that the center's protocol h is *voluntary* if the center neither fines nor rewards players if the contract is not signed by every player: for all $v \neq 1^n \in \{0,1\}^n$ and for all $o \in O$, it is the case that $h_i(v, o) = 0$. We say that h is *frugal* if the center never spends its own money: for all $v \in \{0,1\}^n$ and all $o \in O$, it is the case that $\sum_{i \in N} h_i(v, o) \leq 0$. As these capture the limitations on the helpful center in our setting, we will henceforth limit h to be frugal and voluntary.

We first wish to characterize what outcomes can occur in a subgame-perfect equilibrium of the extended game X. The outcome depends on two things: the contract c_h suggested by the center and the strategies of the players in X. We wish to find contracts to which the players will agree that ensure that our chosen outcome is played.

In order to characterize the space of possible outcomes which can be enforced, we must define the notion of a *punishment equilibrium*. ρ^i is a punishment equilibrium for i if ρ^i is the Nash equilibrium of G with minimal payoffs for i among all (mixed) Nash equilibria of G.

Theorem 1 *Let ρ^i be the punishment equilibrium for i. For all $o_{(a_i, a_{-i})}$, if $V_i(a_i, a_{-i}) \geq V_i(\rho^i)$, then there exists a voluntary and frugal center protocol h and a subgame perfect equilibrium π^* in which all players agree to c_h and play (a_i, a_{-i}), and in no subgame perfect equilibrium do players agree to c_h and then fail to play (a_i, a_{-i}). Furthermore, for all i, $U_i(\pi^*) = V_i(a_i, a_{-i})$. If $V_i(a_i, a_{-i}) < V_i(\rho^i)$, then there is no subgame perfect equilibrium in which (a_i, a_{-i}) is played.*

Proof: First, suppose $V_i(a_i, a_{-i}) < V_i(\rho^i)$. Since player i will get at least $V_i(\rho^i)$ in any subgame perfect equilibrium without fines, i can profit by withholding his assent. As (a_i, a_{-i}) cannot be a Nash equilibrium by assumption and no fines are assessed in H, there can be no subgame

perfect equilibrium in which (a_i, a_{-i}) is played.

Second, suppose $V_i(a_i, a_{-i}) \geq V_i(\rho^i)$. We choose $h_i(a_i) = 0$ and $h_i(a_i' \neq a_i) = -M$. If we choose M so that for all i, a_i', and a_{-i}', it is the case that $M > V_i(a_i', a_{-i}') - V_i(a_i, a_{-i})$, then (a_i, a_{-i}) will be the only subgame perfect equilibrium of the subgame $G \cdot H$, supposing all players agree to c_h. We also require that all players assent in F. If any player does not assent, all players coordinate on his punishment equilibrium ρ^i in G. If more than one player fails to assent, we break ties arbitrarily to see which ρ^i is played. No matter which player fails to assent, ρ^i will be a subgame perfect equilibrium of $G \cdot H$, since the center will not assess fines. Thus i will not profit by withholding his assent. $\square$

Thus we show that, with a fully engaged center that takes part in the protocol and monitors the players' actions, we can achieve any payoffs for the players which are at least as good for every player as some Nash equilibrium of G. Furthermore, once a contract for o is mutually signed, the unique subgame perfect equilibrium achieves o.

We notice that, already, the center takes no action in H in equilibrium. Yet as the center takes action in every other stage, we shall consider how to lighten the load on the center.

Removing the Center From the Enforcement Stage

In this section, we will drop the assumption that the center does not monitor the players' actions in the execution stage G. Instead, we assume that actions and outcomes are common knowledge among the players but are not observed by the center. The center must therefore encourage the players to tell him if there has been a deviation. We will distinguish two cases. In the *verifiable* case, the center can verify that a particular player played a given action if he chooses to do so once the game G has been played. Specifically, we require that the center be able to verify, for each player i, whether i played the correct action a_i or some other action $a_i' \neq a_i$. The center saves effort by not paying attention to G; we merely require that he can determine the truth after the fact, if necessary. In the *unverifiable* case, the center has no information about players' actions whatsoever.

Because we now require the center to be notified by the players of deviations, the enforcement games we now consider will be of the following form: first, the players observe the outcome and send messages to the center. The center publishes any messages he receives to all players. The players then have the chance to respond to the center's messages. This repeats for some number of rounds. Finally, the center makes monetary transfers between the players based on the messages sent.

For our purposes, this full generality is not needed. Our enforcement stage H is a single-round stage game where each player chooses whether or not to *complain* about other players by sending their names to the center, and the center chooses a fine to impose on each player: $H = \langle N, \Gamma^H, \mathbb{R}^n, h \rangle$. $\gamma_i^H \in \Gamma_n^H : O \to 2^N$ specifies which complaints player i will send to the center after each outcome. As before, h is the center's protocol which maps out-

comes and complaints received to monetary consequences in $\mathbb{R}^n$. The center may make payments based on the outcome (if he can verify it), the identities of the complainers, and the target of their complaints. In the verifiable case, $h : O \times (2^N)^n \to \mathbb{R}^n$, while in the unverifiable case, $h : (2^N)^n \to \mathbb{R}^n$.

Now that we have specified an enforcement game, we wish to characterize the set of outcomes obtainable thereby in the extended game corresponding to this enforcement game.

We define a protocol h_o for the center, which will induce an equilibrium under which the center takes no action in the enforcement stage. Let M and m be a large and small amount of money, respectively. In h_o, the center punishes each player who deviated by a large-enough amount M, but also rewards each player who sent in a correct complaint by m for each correct complaint. The center also punishes any player who sent in an incorrect complaint by m. The contract that specifies center protocol h_o we call c_{h_o}.

Theorem 2 (Contracts for Verifiable Games) *Let G be a game with verifiable consequences in O and let $o_{(a_i, a_{-i})} \in O$ be the desired outcome. Assume that the center has suggested contract c_{h_o} defined above and consider the subgame $G \cdot H$ that follows unanimous agreement to this contract. Then there is a strategy profile π^* such that π^* is the unique subgame perfect equilibrium of $G \cdot H$, $o_{(a_i, a_{-i})}$ is the equilibrium outcome of π^*, and π^* has payoffs $V(a_i, a_{-i})$. The center takes no action if π^* is played.*

[Proof omitted.]

We now consider the unverifiable case. As before, we first define a particular center protocol h'_o. In h'_o, the center punishes the target of each complaint by a large-enough amount M, but does not reward or punish players for complaints. After all, the center cannot distinguish valid complaints from invalid ones. The contract that specifies center protocol h'_o we call $c_{h'_o}$.

Theorem 3 (Contracts for Unverifiable Games) *Let G be a game with unverifiable consequences in O, and let $o_{(a_i, a_{-i})} \in O$ be the desired outcome. Assume that the center has suggested contract $c_{h'_o}$ and consider the subgame $G \cdot H$ that follows unanimous agreement to this contract. Then there is a strategy profile $\pi^{*\prime}$ such that $\pi^{*\prime}$ is a subgame perfect equilibrium of $G \cdot H$, $o_{(a_i, a_{-i})}$ is the equilibrium outcome of $\pi^{*\prime}$, and $\pi^{*\prime}$ has payoffs $V(a_i, a_{-i})$. The center takes no action if $\pi^{*\prime}$ is played.*

[Proof omitted.]

We have seen that even without verifiability, it is possible to achieve almost any outcome in equilibrium. Unfortunately, these equilibria are no longer unique. As we shall see, in the unverifiable case, a given signed contract may have many possible equilibrium outcomes rather than just the intended one.

Given a game G, define the *shortfall* s_i^σ of pure-strategy profile $\sigma = (a_i, a_{-i})$ for i as $s_i^\sigma = \max_{a_i'} V_i(a_i', a_{-i}) - V_i(a_i, a_{-i})$. The shortfall of i in σ is the amount i's payoffs would need to rise so that i would have no incentive to deviate from σ, all else held constant. We can see that

there must be some equilibrium of the enforcement game in which an agent i is punished by at least $s_i^{(a_i, a_{-i})}$ whenever he deviates from his action a_i. Yet, in an unverifiable game, there is nothing in the center's protocol which makes (a_i, a_{-i}) special. The players could just as well coordinate on this equilibrium in the enforcement game when the actions are not some other action (a_i', a_{-i}). This implies that any enforcement scheme for the unverifiable case will not in general have a unique outcome. Here we consider not only our chosen center protocol h'_o, but in fact any center protocol h in any form of enforcement game H.

Theorem 4 (Spurious Equilibria) *Consider an unverifiable enforcement game with a frugal and voluntary h under which $G \cdot H$ has a subgame-perfect equilibrium π in which the center does no work, where (a_i, a_{-i}) is the strategy profile that π plays in G. Then if σ' is a pure strategy profile of G and $\forall i : s_i^{\sigma'} \leq s_i^{(a_i, a_{-i})}$, then there exists a subgame perfect equilibrium π' of $G \cdot H$ such that σ' is the strategy profile that π' plays in G.*

[Proof omitted.]

A consequence of this theorem is that any Nash equilibrium of G can be played in $F \cdot G \cdot H$ regardless of the contract signed.

Corollary 5 (No Deletion) *If σ is a pure or mixed Nash equilibrium in the unverifiable game G, then, for any frugal and voluntary center protocol h that has a subgame perfect equilibrium π where the center does no work, there is a subgame perfect equilibrium π' of $G \cdot H$ such that σ is the strategy profile that π' plays in G.*

[Proof omitted.]

Thus, if the center cannot verify the players' actions, he cannot in general enforce any outcomes uniquely. After signing the contracts, the players might arrive at an outcome different from the one the center suggested. In a real-world setting, this would substantially weaken the case that the players should sign the contract.

We have shown that a helpful center who neither monitors the player's actions nor fines any player in equilibrium can enforce every outcome that a fully engaged center can enforce with more burdensome contracts. In an unverifiable game, however, the center must generally accept spurious equilibria. Our next task is to remove the center from the signature exchange stage.

Exchanging Signatures Without The Center

Under the original contract, the center collected signatures on the contract c_h and enforced the contract if every player signed. We now show how the players can exchange signatures on the contract by use of a broadcast channel without requiring any action from the center in equilibrium. In this, our goal is similar to the goal of *optimistic signature exchange* (Garay & MacKenzie 1999), but with rational actors instead of computationally-bounded ones.

If players may communicate without being observed by others, F would be a game of imperfect information. As these games are difficult to analyze and generally admit of

many solutions, we require the players to use a broadcast channel, on which all messages sent are common knowledge.

When the center no longer monitors the signature exchange stage, he no longer knows in the enforcement stage H whether the contracts have been signed or not. Therefore, we now require that each complaint sent to the center in the enforcement stage H include a fully signed copy of the contract.

The Naive Broadcast Protocol

We might hope that the signature collection service performed by the center was superfluous: that we will achieve the same results if we simply require players to broadcast their agreement or disagreement. Unfortunately, this will not be so. Consider the naive broadcast protocol where all players simultaneously broadcast their signatures. Let us formally define F to be the one-round stage game $F = \langle N, \Gamma^F, S, f \rangle$, where N is the set of players and $\Gamma^F = \{0, 1\}^n$, where 0 represents the decision not to broadcast one's signature, while 1 represents the decision to do so. $S = (\{0, 1\}^n)^n$ is the set of outcomes of the game. Each outcome specifies the set of signatures (represented by $\{0, 1\}^n$) possessed by each player in N. $f : \Gamma^{\hat{F}} \to S$ is the outcome function of F: each player knows his own signature and every signature which is broadcast.

The complete set of signatures is thus common knowledge if and only if every player chooses to broadcast his signature. Consider what occurs if exactly one player i fails to reveal his signature: i has received all the signatures of the other players, and he can produce his own. Thus i is the only player to possess all signatures on the contract, and this fact is common knowledge among the players. The center, on the other hand, cannot distinguish this case from the case where all players know all signatures, but only i chooses to complain. Therefore i is able to unilaterally enforce the contract, unlike in the original formulation.

Recall that, in H, every message sent by one player to the center is broadcast to all other players. Thus, once one player has sent a complaint about another (which includes a fully signed contract), every player will know all signatures on c_h and be able to complain. A player who deviates in F cannot choose to punish other players while remaining unscathed himself, but he can choose unilaterally whether or not to enforce the contract. Unfortunately, this power implies that our previously specified equilibrium for the extended game $F \cdot G \cdot H$ is no longer an equilibrium.

The equilibrium for $F \cdot G \cdot H$ discussed above requires that, if i fails to reveal his signature on the contract, all players coordinate on i's punishment equilibrium. Consider the case, for instance, where the punishment equilibrium for some i is (a_i, a'_{-i}), where a_i is the action i is contractually obliged to play. Suppose i alone fails to reveal his signature and all players play i's punishment equilibrium (a_i, a'_{-i}). In stage H, then, i will profit by choosing to enforce the contract: the center will punish the other players and reward i. Knowing this, the other players will not in general wish to play their part of the punishment equilibrium, so our previous strategy fails.

The Pre-Contract Protocol

Although the naive broadcast protocol did not allow us to guarantee all the payoffs we wanted, we shall see that we can use a more complicated signature exchange stage F to ensure that either each player receives all signatures on c_h, or no players receive all signatures on c_h. Our exchange scheme is modelled on the contracts mechanism of the rest of the paper: we will add a *pre-contract* $\hat{c}_h$ that the players will sign before signing c_h. This contract authorizes the center to fine players who do not reveal their signature on c_h. Surprisingly, this does not lead to infinite regress: this one pre-contract is sufficient to allow for signature exchange.

We will divide F itself into stages: a miniature contract exchange stage $\hat{F}$, a miniature execution stage $\hat{G}$, and a miniature enforcement stage $\hat{H}$. The players will bind themselves in contract $\hat{c}_h$ to reveal their signatures on the contract c_h in such a way that, if they fail to reveal them, they can be fined by the center. We will allow them, however, to recoup that fine by revealing their signatures on c_h to the center and all players after the fact. The after-the-fact alteration of the outcome allows us to use the naive broadcast protocol for $\hat{F}$ where we could not use it for F.

Formally, let the signature stage $F = \hat{F} \cdot \hat{G} \cdot \hat{H}$. Let us call the contract signed in $\hat{F}$ the pre-contract $\hat{c}_h$, which binds players to release their signatures on the real contract c_h. $\hat{F}$ is the naive broadcast protocol defined above as the stage game $\hat{F} = \langle N, \Gamma^{\hat{F}}, S, \hat{f} \rangle$. S is the set of signatures on $\hat{c}_h$ that each player knows, and $\Gamma^{\hat{F}}$ is each player's choice to broadcast or not broadcast his signature on $\hat{c}_h$. $\hat{G}$ is also the naive broadcast protocol defined above for the signatures on c_h: $\hat{G} = \langle N, \Gamma^{\hat{G}}, \Phi, \hat{g} \rangle$, with Φ the set of known signatures on c_h, and $\Gamma^{\hat{G}} : S \to 0, 1^n$ the decision of the players to broadcast their signatures on c_h given what signatures each player knows on $\hat{c}_h$.

$\hat{H} = \langle N, \Gamma^{\hat{H}}, \mathbb{R}^n, \hat{h} \rangle$ is a miniature enforcement stage that is substantially different from the enforcement stage H. $\hat{H}$ has two rounds. In the first round, a player i is allowed to complain to the center that he has not received some signature on c_h. To do so, i must submit the contract $\hat{c}_h$, all signatures on $\hat{c}_h$, and i's signature on c_h. When the center rebroadcasts this message to all players, both the signatures on $\hat{c}_h$ and i's signature on c_h become known to all players. In the second round, each player who did not complain in the first round is given a chance to complain. $\Gamma_i^{\hat{H}}$ simply characterizes whether i will complain to the center after each history.

We now state our chosen center protocol $\hat{h}$ in $\hat{H}$. If the center received a complaint in the first round, then the center fines all players who did not complain in either the first or the second round by a large-enough amount M. If the center does not receive any complaints, he does not fine any players. Note that if all players complain, all signatures on c_h become common knowledge and no fines are assessed.

We now specify the strategy π^* we expect the agents to play in F. In $\hat{F}$, each player first reveals his signature on $\hat{c}_h$. In $\hat{G}$, he will reveal his signature on c_h if and only if

he has received the signatures of every other player on $\hat{c}_h$. In the enforcement stage $\hat{H}$, he complains to the center if and only if he has received all signatures on $\hat{c}_h$, but he has not received all signatures on c_h, or if some other player has complained.

The remainder of π^* for G and H is simple. If all signatures on c_h become known to all players, then the players play (a_i, a_{-i}) in G to achieve o, just as before, and then complain to the center in H if some player deviates. If, however, some player i deviates from the equilibrium in F (whether by choosing not to reveal in $\hat{F}$, choosing not to reveal in $\hat{G}$, or failing to complain in $\hat{H}$), in such a way that the signatures on c_h do not become commonly known, then the agents coordinate on that player's punishment equilibrium ρ^i. If several players deviate during F, the agents coordinate on the punishment equilibrium of the last player to deviate.

Our strategy π^* is now an equilibrium. Thus, we can achieve any outcome achievable with a busy center with a center that does no work in equilibrium.

Theorem 6 *Let X be the extended game $\hat{F} \cdot \hat{G} \cdot \hat{H} \cdot G \cdot H$, and let ρ^i be the punishment equilibrium for i in G. Then, for any $o_{(a_i, a_{-i})}$ such that for all i, $V_i(a_i, a_{-i}) \geq V_i(\rho^i)$, there exists a contract c_h for which there is an strategy profile π^* such that π^* is a subgame perfect equilibrium of the extended game and $o_{(a_i, a_{-i})}$ is the equilibrium outcome of π^*. Furthermore, in π^*, the center takes no action during any stage.*

Proof: Let us sketch why this will be a subgame perfect equilibrium. We will proceed by backwards induction.

So long as no single player gains complete knowledge of the signatures on c_h, then π^* is a subgame perfect equilibrium in G and H. This does not occur if π^* is played in F, so it is sufficient to prove that π^* is a subgame perfect equilibrium in F. We will show that, even after one deviation, either all players know all the signatures on c_h or no player knows all signatures on c_h.

Consider the second round of $\hat{H}$. If no player complained in the first round, second-round complaints have no effect. If a player complained in the first round of $\hat{H}$, then it will be dominant for every other player to complain according to π^* to avoid the punishment of M from the center. Thus, all players will know all signatures on c_h.

Consider the first round of $\hat{H}$. There are three cases to distinguish. First, if every player knows all signatures on both $\hat{c}_h$ and c_h, then complaining will have no effect. Second, if every player knows all signatures on $\hat{c}_h$, but only one player knows all signatures on c_h. Every other player will complain in the first round and i must therefore complain in the first or second rounds to avoid losing M. Every player will learn all signatures. Third, if only one player i knows all the signatures on $\hat{c}_h$, then i will not know all the signatures on c_h. If i does not complain, no player will learn all signatures on c_h and players will coordinate on i's punishment equilibrium. If i does complain, all others will complain in the second round and all players will learn all signatures.

Consider the stage $\hat{G}$. There are now two cases to consider. First, suppose all players know all signatures on $\hat{c}_h$. Then no player i can benefit by failing to reveal his signature on c_h, since the other players will complain, i will complain to avoid punishment, and all players will end up learning all signatures on c_h. Second, suppose only one player i knows the signatures on $\hat{c}_h$ because he failed to reveal in $\hat{F}$. Then no other players will reveal their signatures, and, whether i reveals or not, all players will coordinate on i's punishment equilibrium.

Finally, consider the stage $\hat{F}$. Suppose one player i deviates in the stage $\hat{F}$ by failing to reveal his signature on the pre-contract $\hat{c}_h$. Then he alone will have all the signatures on $\hat{c}_h$, and no one else will reveal their signatures on c_h in $\hat{G}$. According to the equilibrium, i will complain to the center in stage $\hat{H}$, resulting in complete knowledge of c_h and the decision to play (a_i, a_{-i}). $\square$

Conclusion

We have discussed the power of a helpful center in enabling a group of players to make contracts which require them to play a certain strategy or face penalties. Even if the center brings no money to the system and transfers money from the players only after receiving permission, the center is able to help the players achieve nearly any outcome of the game. Moreover, we find that the center is still able to help the players achieve these outcomes in equilibrium, even if he does not monitor the game and does not participate on the equilibrium path - in other words, even when the center does no work in equilibrium beyond suggesting a contract.

In fact, if the contracts the center would suggest are common knowledge or determined by a negotiation stage between the agents, the center does no work whatsoever in equilibrium. Incidentally, we notice that the center makes a profit to cover his costs whenever his services are used. These two properties are very important for a third party who wishes to influence outcomes in strategic settings that occur frequently, such as in the online auction setting.

References

Garay, J. A., and MacKenzie, P. D. 1999. Abuse-free multi-party contract signing. In *International Symposium on Distributed Computing*, 151–165.

Hafner, K. 2004. Vigilantes attack eBay fraud. *The New York Times*. Available at http://news.com.com/2100-1038_3-5176525.html.

Mas-Colell, A.; Whinston, M.; and Green, J. 1995. *"Microeconomic Theory"*. Oxford University Press.

Monderer, D., and Tennenholtz, M. 2003. k-implementation. In *Proceedings of the 4th ACM Conference on Electronic Commerce*, 19–28.

Osborne, M., and Rubinstein, A. 1994. *A Course in Game Theory*. MIT Press.

Shoham, Y., and Tennenholtz, M. 1997. On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence* 94:139–166.