Correlated Equilibrium Under Costly Disobedience

Ozdogan, Ayca and Saglam, Ismail

TOBB University of Economics and Technology

30 March 2020

Correlated Equilibrium Under Costly Disobedience*

Ayça Ozdogan[a] and Ismail Saglam[a,†]

[a] *Department of Economics, TOBB University of Economics and Technology, Sogutozu Cad. No:43, Sogutozu, 06560, Ankara, Turkey*

In this paper, we extend Aumann's (1974) well-known solution of correlated equilibrium to allow for a cost of disobedience for each player. Calling the new solution *costly correlated equilibrium* (CCE), we derive the necessary and sufficient conditions under which the set of CCE strictly expands when the players' cost of disobedience is increased by the mediator in any finite normal-form game. These conditions imply that for any game that has a Nash equilibrium (NE) that is unpure, the set of CCE strictly expands with the addition of even arbitrarily small cost of disobedience, whereas for games that have a unique NE in pure strategies, the set of CCE stays the same unless the cost gets sufficiently high. We also study the welfare implications and changes in the value of mediation with exogenous cost changes. We find that strictly better social outcomes can be attained and the value of mediation cannot decrease with an increase in the cost level. We also illustrate how our model can be integrated with a cost-selection game where players noncooperatively choose their costs of disobedience before mediation occurs. We show that there exist cost-selection games in which setting the cost of disobedience at zero is a strictly dominated strategy for each player as well as games this strategy becomes weakly dominant for everyone.

Keywords: Correlated equilibrium; cost of disobedience.

JEL Classification Numbers: C72.

*The authors have no conflicts of interests to declare. The usual disclaimer applies.

†Corresponding author. E-mail: isaglam@etu.edu.tr.

1 Introduction

One of the goals of non-cooperative game theory is to explore the means by which one can attain efficient and cooperative outcomes in a self-enforcing way in strategic interactions. A distinctly remarkable tool to achieve this goal in static environments where the Nash equilibrium (NE) concept is inadequate is the idea of mediation through self-enforcing correlation devices introduced by Aumann (1974). According to this idea, in strategic situations some outcomes that cannot arise in any NE can be implemented by appropriately chosen correlated recommendations of a credible mediator, forming a correlated equilibrium (CE).[1] The CE concept has been appealing as it proposes a correlated randomization over the set of strategy profiles that weakly expands the set of NE and NE payoffs.[2] However, while in some games (e.g. the Chicken, Stag-Hunt or Battle of the Sexes) the CE outcomes strictly improve upon the NE outcomes; in others (e.g. the Matching Pennies and the Prisoners' Dilemma) the set of CE is equal to the set of NE.[3] Moreover, in games where the set of CE strictly expands the set of NE, there may be a limit to achieve the total welfare maximizing efficient outcome.[4]

[1] As a matter of fact, the presence of a mediator is not always needed for correlation. Vanderschraaf (1995) shows that in games with at least three players correlations between the players' subjective probability distributions over their opponents' actions is possible without a mediator or an external event space.

[2] It is also appealing because of its behavioral justifications. For instance, Aumann (1987) shows that CE as an expression of Bayesian rationality. And, Hart and Mas-Colell (2000) and Hart (2005) prove the connection between the set of CE and the limit behavior of regret-based heuristics.

[3] Rosenthal (1974) calls that a CE is good if there is a player who prefers CE to NE for every NE of a two-person game. He shows that a game has no good CE if it is best-response equivalent to a two-person zero-sum game. Moulin and Vial (1978), on the other hand, propose a class of games called "strategically zero-sum games" for which no completely mixed NE can be improved upon. Moreover, in the infinite game setting, Liu (1996) and Yi (1997) show that the only CE in a large class of oligopoly games are mixtures of pure Nash equilibria. This result is extended to potential games with smooth and concave potential functions by Neyman (1997) and Ui (2008).

[4] For instance, in Aumann's example (see Example 2 in Section 3.1), the total payoff of the players cannot exceed 20/3 in a correlated equilibrium. Even though it is more than 6

In this paper, we examine if the set of CE can be strictly expanded through incorporating (even arbitrarily) small costs to disobeying the recommendations of the mediator in finite normal-form games. To this aim, we extend Aumann's (1974) well-known solution of CE to allow for a non-negative cost of disobedience for each player, calling the new solution *costly correlated equilibrium* (CCE). Our main finding indicates that for games that have a NE that is not pure, the set of CCE strictly expands even with an arbitrarily small increase in the cost of disobedience (if there is room for expansion). We also study the welfare implications of exogenous cost changes on the value of mediation. We find that socially more efficient outcomes can be attained and the value of mediation cannot decrease with an increase in the cost level. As these findings imply that players may find it to their interests to non-cooperatively commit to non-zero cost levels subsequently, we also introduce and briefly study a *cost-selection game under mediation*.

The CE is a randomization over the strategy profiles that is commonly known by all players and implemented by the recommendations of a reliable mediator who informs each player privately of her recommended action based on the realization of the lottery. It is ex ante optimal for each player to follow this recommendation if each player believes that the others are doing so. However, the findings of Cason and Sharma (2007) in laboratory experiments indicate that players do not always obey recommendations that implement CE outcomes because of the lack of mutual knowledge of beliefs. Incorporating some small costs to disobedience to the recommendations could be perceived as one way to induce players to sustain mutual trust for following the recommendations of the mediator (even though the costs are not realized in equilibrium).

It is well-known that traffic lights may be viewed as a mediator who sends private but correlated recommendations (based on the outcome of a commonly known lottery) to the drivers at an intersection. And, everyone follows the recommendations believing that every other is going to do so.

(the maximal total payoff that can be obtained by a Nash equilibrium), it is short of the total payoff of 8 that is provided by the symmetric efficient outcome.

However, in practice, there is often a cost of not obeying to the recommendations of the mediator (such as a fine of passing at a red light even though it is self-enforcing not to do so when everyone believes that everyone follows the recommendations). Hence, we believe that incorporating some small cost for disobeying the recommendations of the mediator could alleviate the issues related with players' trust to each other in following the recommendations and potentially expand the set of CE (and thus the NE) in the direction to attain Pareto improving outcomes. To illustrate a second benefit of disobedience costs in mediation, we may consider mediatory institutions like government agencies or independent international bodies (such as European Convention and Court of Human Rights) giving recommendations to all relevant parties that participate in issues such as environmental agreements, legal negotiations etc.[5] The CE notion puts no sanctions or punishments if a player chooses not to follow the recommendations and s/he would get information about the recommendations and thus what others may do even though s/he chooses not to follow. However, in many such contexts, there many be tangible or intangible costs of not following the recommendations of these agencies, and these costs seem to have been ignored by the economic theory so far to the best of our knowledge. In this study, we would like to capture the implications of incorporating these costs on the set of equilibrium outcomes.

There is a growing body of literature on how to expand the set of CE, which is essentially concerned with strengthening of the commitment of the players to follow the recommendations of the mediator. In particular, a "simple extension" of CE was introduced by Moulin and Vial (1978), which was later termed as "coarse correlated equilibrium" by Young (2004) and "weak correlated equilibrium" (WCE) by Forgó (2010). Like CE, the solution of WCE also picks the outcome of the game according to a commonly known probability distribution. The difference in WCE is that each player must first decide to commit or not to follow the strategy recommended by the mediator before the mediator implements the randomization and they are required to

[5]See Moulin, Ray and Sen Gupta (2014) for a further discussion.

do so if they choose to commit. A player who does not commit could choose any strategy of her own but s/he cannot receive any information about the outcome of the lottery. Again as in CE, it is ex-ante optimal to commit to the expected outcome of the lottery if a player believes that every other player is doing the same. Moulin and Vial (1978) show that it is possible to improve upon a completely mixed NE by WCE in strategically zero-sum games where CE cannot improve upon NE.[6] In a more recent study, Forgó (2010) proposes for finite games another generalization of CE, called soft correlated equilibrium (SCE). He shows that neither SCE nor WCE is a special case of the other, and in some normal-form games SCE can induce Pareto-superior outcomes than does WCE. The only difference of the two solutions is that in SCE players should either commit to the recommendations of the mediator or choose some action other than the one suggested by the mediator. Once again, it is ex-ante optimal for a player to commit to the recommendation if everyone beliefs that every other does so. Forgó (2010) shows that while WCE and CE cannot improve upon the unique NE in the Prisoners' Dilemma game, SCE could do so.

The experimental literature on third-party recommendations in overcoming coordination problems and studying the empirical validity of CE concept with a mediator is also blooming.[7] A common feature of these studies is that they aim to identify whether (experimental) subjects follow the recommendations and the factors which make them more or less likely to do so. They all find that subjects tend to follow recommendations but this tendency varies significantly with variations in the games and treatments. For instance, Ca-

[6]WCE is also studied in infinite strategic games e.g. Gerard-Varet and Moulin (1978), Ray and Sen Gupta (2013) and Moulin, Ray and Sen Gupta (2014). For instance, Moulin, Ray and Sen Gupta (2014) analyze the concept of WCE in a class of symmetric two-person games quadratic games (e.g. Cournot duopoly and the public good provision games) where WCE can strictly improve upon NE payoffs while CE cannot.

[7]To the best of our knowledge, Moreno and Woorders (1998) is the first experimental study which shows that subjects' behavior can be explained with the coalition-proof CE (incorporating the possibility that players could do small mistakes) when preplay communication is allowed (rather than incorporating a commonly known randomization device whose realization is privately recommended to each subject by a mediator).

son and Sharma (2007) find that recommendations were effective when the subjects played against robots that always followed the recommendations, rather than against other human subjects. They claim that the lack of mutual knowledge of conjectures is why subjects fail to play the CE when facing other human players, i.e. subjects do not want to choose the recommended action as they believe that their opponent will not do so. Duffy and Feltovich (2010) show that recommendations are more likely to be followed when they induce a CE that payoff-dominates the available (mixed-strategy) NE. Bone, Drouvelis and Ray (2012) similarly find that recommendations are typically followed when the CE is not payoff dominated by some other outcome. However, Anbarci, Feltovich and Gurdal (2018) show that it is also necessary for the CE either to be sufficiently payoff-equitable ex-post or for the cost of unilaterally disobeying recommendations to be low for the recommendations to be more effective. They examine different treatments where the equilibrium induced by the recommendations imply payoffs that are equal ex ante, but unequal ex post. They find that as either payoff asymmetry increases or the cost of disobeying an unfavorable recommendation decreases (meaning that the loss in the payoff s/he would receive by disobeying), subjects (who are sufficiently inequity-averse) are more likely to disobey recommendations after the ones that ex-post unfavor them. Georgalos, Ray and Sen Gupta (2019), on the other hand, investigates whether the subjects follow the WCE by asking subjects to commit to a device that randomizes between three symmetric outcomes (including the pure NE) with higher ex-ante expected payoff than the pure NE payoff. They find that players tend to avoid committing to the device and choose to play the game by coordinating on the pure NE. Their results also imply that the players do not like to commit to follow the recommendations that lead to ex-post unequal payoffs.

The findings in the experimental literature points out to the need for incentivizing players to follow the recommendations to coordinate on better outcomes. In our paper we provide the missing incentives in the CE model by introducing a non-negative cost of disobedience for each player, and thus generalize the notion of correlated equilibrium in normal-form games as "costly

correlated equilibrium" (CCE). In this setup, we study how the set of CCE is affected by increases in the cost of disobedience. We show that in case the cost of disobedience is uniform for all players and exogenously set by the mediator, an increase, however small, in its level expands the set of CCE if and only if the boundary of this set contains an unpure equilibrium.

Moreover, we show that in situations where the mediator recommends a socially optimal CCE, a sufficiently large increase in the cost of disobedience may raise the social welfare unless the society of players is already enjoying an outcome with the highest attainable welfare. We quantify the performance of mediation under costly disobedience by extending the measures in Ashlagi et al (2008). We say that *the value of mediation* at any cost level is equal to the ratio between the total payoff obtained in any optimal CCE at that cost level and the maximal total NE payoff obtained in the absence of any mediation. Similarly, we say that *the value of enforcement* at any cost level is the ratio between the total payoff obtained in any optimal CCE at that cost level and the maximal total payoff in the game. We find that when the cost changes the value of mediation and the value of enforcement move in the same direction and they are always non-decreasing.

Lastly, we extend our model to a setup allowing for each player to choose his/her cost of disobedience prior to mediation. In this setup, the mediator first announces an optimal CCE rule that specifies a socially optimal CCE at each possible cost profile, before the players choose their costs strategically and non-cooperatively. This rule along with the game structure of the unmediated game induces for each player an expected utility function over the set of possible cost profiles, hence a strategic-form game that we call "the cost-selection game". We show that there exist cost-selection games where committing to zero (or some low levels of) cost is a strictly dominated strategy for each player as well as games where this extreme strategy becomes weakly dominant for each player.

The rest of the paper is organized as follows: Section 2 introduces the model. Section 3 gives our results and Section 4 concludes.

2 Model

We consider a normal-form (strategic-form) game $G = [N, \{S_i\}, \{u_i(.)\}]$ that specifies the set of players $N = \{1, \ldots, n\}$ where $n \geq 2$ and for each player i a set of pure strategies S_i and a payoff function $u_i : \times_{i=1}^{n} S_i \to \mathbb{R}$ giving the von Neumann-Morgenstern utilities. Let $S = \times_{i=1}^{n} S_i$. For each $s \in S$ and $i \in N$, we define s_i such that $s = (s_i, s_{-i})$. Similarly, we define for each $i \in N$, the set S_{-i} such that $S = S_i \times S_{-i}$.

For any integer $k \geq 1$ and any set $X \subseteq \mathbb{R}^k$, we denote by $\Delta(X)$ the probability distributions over X. A correlated equilibrium (Aumann 1974) is a probability distribution p over $\Delta(S)$ such that for all $i \in N$ and for all $r_i, t_i \in S_i$ the following is satisfied:

$$\sum_{s_{-i} \in S_{-i}} p(s_{-i}, r_i) u_i(r_i, s_{-i}) \geq \sum_{s_{-i} \in S_{-i}} p(s_{-i}, r_i) u_i(t_i, s_{-i}). \tag{1}$$

We use joint probabilities $p(s_{-i}, r_i)$, instead of conditional probabilities $p(s_{-i}|r_i)$, therefore (1) is valid even when $p(s_{-i}|r_i)$ is not defined for some r_i and s_i. Condition (1) requires that when a strategy profile $r \in S$ is randomly chosen by a mediator according to the probability distribution p and each player i is only informed about r_i and asked to play it, then no player can obtain higher payoffs if s/he disobeys the recommendation and plays another strategy t_i instead. As shown by Aumann (1974), in finite games each Nash equilibrium is a correlated equilibrium (with independent signals), and the existence result for Nash equilibrium ensures the existence of a correlated equilibrium.[8]

Now, suppose that disobedience to the recommendation of the mediator can be costly. Let $c \geq 0$ denote the common cost of disobedience for players.[9] For any $p \in \Delta(S)$, $i \in N$, $r_i, t_i \in S_i$, and $c \geq 0$, we define the difference between the expected payoff of a player from obeying to the rec-

[8] Hart and Schmeidler (1989) has a direct proof of existence based on linear duality.

[9] We assume a common cost of disobedience for the sake of simplicity. In general, each player $i \in N$ may bear a (possibly) distinct cost $c_i(r_i, t_i)$ when it deviates from a recommended strategy $r_i \in S_i$ to another strategy $t_i \in S_i$.

ommended strategy r_i and from deviating to the strategy t_i, subject to the cost of disobedience c, as follows

$$D_i(p, c, r_i, t_i) = \begin{cases} \sum_{s_{-i} \in S_{-i}} p(s_{-i}|r_i) \left[u_i(r_i, s_{-i}) - u_i(t_i, s_{-i}) + c\right] \\ \qquad\qquad \text{if } \sum_{s_{-i} \in S_{-i}} p(s_{-i}|r_i) > 0, \\ 0 \qquad\quad \text{if } \sum_{s_{-i} \in S_{-i}} p(s_{-i}|r_i) = 0. \end{cases} \tag{2}$$

We say that the probability distribution $p \in \Delta(S)$ is a *costly correlated equilibrium* (CCE) under the cost profile c if $D_i(p, c, r_i, t_i) \geq 0$ for all $i \in N$ and $r_i, t_i \in S_i$. Let $\mathcal{P}(c)$ denote the set of all CCE under c. Clearly $\mathcal{P}(c) \subseteq \Delta(S)$ for any $c \geq 0$. We use the signs $\subset$ and $\subseteq$ for strict and weak inclusion, respectively. Analogously, for $\supset$ and $\supseteq$.

For some of our results and discussions, we will refer to the following definitions. Let $s_i \in S_i$ be a possible strategy of player $i \in N$. A strategy s_i is strictly dominated if there exists a mixed strategy $\sigma_i \in \Delta(S_i)$ such that for any possible combination of the other players' strategies, $s_{-i} \in S_i$, player i obtains strictly lower payoff from s_i than from σ_i, i.e., $u_i(s_i, s_{-i}) < u_i(\sigma_i, s_{-i})$ for all $s_{-i} \in S_{-i}$. Similarly, a strategy s_i is weakly dominated if there exists a mixed strategy $\sigma_i \in \Delta(S_i)$ such that for any possible combination of the other players' strategies, $s_{-i} \in S_i$, player i obtains weakly lower payoff from s_i than from σ_i, i.e., $u_i(s_i, s_{-i}) \leq u_i(\sigma_i, s_{-i})$ for all $s_{-i} \in S_{-i}$, and s/he obtains strictly lower payoff for some $s_{-i} \in S_{-i}$. On the other hand, a strategy s_i is strictly (weakly) dominant if any other strategy in $\Delta(S_i)$ is strictly (weakly) dominated by s_i.

Finally, the following definition will be helpful. A probability distribution p is called a vertex of $\Delta(S)$ if there exists $i \in \{1, 2, \ldots, |S|\}$ such that $p_i = 1$ and $p_j = 0$ for all $j \in \{1, 2, \ldots, |S|\} \setminus \{i\}$. We let $\mathcal{V}(S)$ denote the set of vertices of S.

3 Results

We will study in Section 3.1 how the set of CCE of a normal-form game may change when the cost of disobedience increases for all players. We will consider welfare effects in Section 3.2 and a strategic game of cost selection (as an extension for future research) in Section 3.3.

3.1 The Effect of Cost of Disobedience on the Set of CCE

We will first present several examples to gain some insight about when and how the set of CCE is affected by a change in the players' cost of disobedience. We will use these examples also to discuss our theoretical results.

Example 1. Consider the following normal-form game, known as the Matching Pennies game.

	H	T
H	$1,-1$	$-1,1$
T	$-1,1$	$1,-1$

Note that $S_1 = S_2 = \{H, T\}$, and $S = \{(H,H),(H,T),(T,H),(T,T)\}$. For any $p \in \Delta(S)$, let $p = (p_{11}, p_{12}, p_{21}, p_{22})$ where $p_{11} = p((H,H))$, $p_{12} = p((H,T))$, $p_{21} = p((T,H))$, and $p_{22} = p((T,T))$. For any $c \geq 0$, we can calculate

$$\mathcal{P}(c) = \left\{ \begin{array}{rl} p \in \Delta(S): & (2+c)p_{11} \geq (2-c)p_{12}, \quad (2+c)p_{12} \geq (2-c)p_{22}, \\ & (2+c)p_{22} \geq (2-c)p_{21}, \quad (2+c)p_{21} \geq (2-c)p_{11}. \end{array} \right\}$$

Clearly, $\mathcal{P}(0) = \{(0.25, 0.25, 0.25, 0.25)\} = \partial\mathcal{P}(0)$. For any $c > 0$, we have $(0.25, 0.25, 0.25, 0.25) \in \mathcal{P}(c)$, implying $\mathcal{P}(c) \supseteq \mathcal{P}(0)$. On the other hand, for any $c > 0$, $\mathcal{P}(c)$ also contains the distribution $\hat{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ such that $p_{11} = ap_{21}$, $p_{12} = a^2 p_{21}$, $p_{22} = a^3 p_{21}$ along with $p_{21} = 1/(1 + a + a^2 + a^3)$ and $a = (2+c)/(2-c)$. Apparently, $\mathcal{P}(0)$ does not contain $\hat{p}$. So, $\mathcal{P}(c) \supset \mathcal{P}(0)$.

More generally, for any $c', c'' \geq 0$ such that $c'' > c'$, one can easily check that $\mathcal{P}(c'') \supset \mathcal{P}(c')$. Moreover, for any $c \geq 2$, we have $\mathcal{P}(c) = \Delta(S)$. ■

Example 1 suggests that there exist normal-form games in which any increase in the costs of disobedience, however small, always expands the set of CCE, unless this set is already as wide as $\Delta(S)$. What is most peculiar about Example 1 is that the set of CCE, which consists of $\{(0.25, 0.25, 0.25, 0.25)\}$ at the cost level $c = 0$, starts to contain infinitely many probability distributions once the cost of disobedience is increased even infinitesimally. $\mathcal{P}(0)$ does not need to be a singleton set to observe these results, which we illustrate in the next example.

Example 2. Consider the following normal-form game, borrowed from Aumann (1974).

	L	R
U	$5, 1$	$0, 0$
D	$4, 4$	$1, 5$

Note that $S_1 = \{U, D\}$, $S_2 = \{L, R\}$, and $S = \{(U, L), (U, R), (D, L), (D, R)\}$. For any $p \in \Delta(S)$, let $p = (p_{11}, p_{12}, p_{21}, p_{22})$ where $p_{11} = p((U, L))$, $p_{12} = p((U, R))$, $p_{21} = p((D, L))$, and $p_{22} = p((D, R))$. For any $c \geq 0$, we can calculate

$$\mathcal{P}(c) = \left\{ p \in \Delta(S) : \begin{array}{ll} (1+c)p_{11} \geq (1-c)p_{12}, & (1+c)p_{22} \geq (1-c)p_{21}, \\ (1+c)p_{11} \geq (1-c)p_{21}, & (1+c)p_{22} \geq (1-c)p_{12}. \end{array} \right\}$$

Clearly, for any $c > 0$ the set $\mathcal{P}(c)$ contains $\mathcal{P}(0)$ as well as the distribution $\hat{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ such that $p_{11} = p_{22} = (1-c)/(3-c)$, $p_{12} = 0$, and $p_{21} = (1+c)/(3-c)$, whereas $\hat{p} \notin \mathcal{P}(0)$. So, $\mathcal{P}(c) \supset \mathcal{P}(0)$. More generally, for any $c', c'' \geq 0$ such that $c'' > c'$, one can check that $\mathcal{P}(c'') \supset \mathcal{P}(c')$. Moreover, for any $c \geq 1$, we have $\mathcal{P}(c) = \Delta(S)$. ■

We obtain $\mathcal{P}(c) \supset \mathcal{P}(0)$ for any $c > 0$, as in Example 1. On the other hand, in Example 2 neither $\mathcal{P}(0)$ nor $\partial\mathcal{P}(0)$ is a singleton set. For instance,

the distributions $(1,0,0,0)$, $(0,0,0,1)$, $(1/3,0,1/3,1/3)$ are all in $\mathcal{P}(0)$ and $\partial\mathcal{P}(0)$. What is common in both examples is that $\partial\mathcal{P}(0)$ contains a probability distribution that is not a vertex of $\Delta(S)$.

Example 3. We will consider a modified form of Matching Pennies game, in which player 2 (column player) has the additional strategy of not showing (N) its coin to the other player. If s/he chooses this new strategy, s/he pays player 1 a penalty fee of 2. The payoff matrix of this modified game is shown below.

	H	T	N
H	$1,-1$	$-1,1$	$2,-2$
T	$-1,1$	$1,-1$	$2,-2$

Note that $S_1 = \{H,T\}$, $S_2 = \{H,T,N\}$, and $S = \{(H,H),(H,T),(T,H),(T,T),(N,H),(N,T)\}$. For any $p \in \Delta(S)$, let $p = (p_{11},p_{12},p_{13},p_{21},p_{22},p_{23})$ where $p_{11} = p((H,H))$, $p_{12} = p((H,T))$, $p_{13} = p((H,N))$, $p_{21} = p((T,H))$, $p_{22} = p((T,T))$, and $p_{23} = p((T,N))$. For any cost $c \geq 0$, we can calculate

$$\mathcal{P}(c) = \left\{ \begin{array}{lll} p \in \Delta(S): & (2+c)p_{11} \geq (2-c)p_{12}, & (2+c)p_{12} \geq (2-c)p_{22}, \\ & (2+c)p_{22} \geq (2-c)p_{21}, & (2+c)p_{21} \geq (2-c)p_{11}, \\ & p_{13} = p_{23} = 0. & \end{array} \right\}$$

Clearly, $\mathcal{P}(0) = \{(0.25,0.25,0,0.25,0.25,0)\}$, and for any $c > 0$ we have $\mathcal{P}(c) \supseteq \mathcal{P}(0)$. On the other hand, for any $c > 0$, the set $\mathcal{P}(c)$ also contains the distribution $\hat{p} = (p_{11},p_{12},p_{13},p_{21},p_{22},p_{23})$ such that $p_{13} = p_{23} = 0$, $p_{11} = ap_{21}$, $p_{12} = a^2p_{21}$, $p_{22} = a^3p_{21}$ along with $p_{21} = 1/(1+a+a^2+a^3)$ and $a = (2+c)/(2-c)$. Apparently, $\mathcal{P}(0)$ does not contain $\hat{p}$. So, $\mathcal{P}(c) \supset \mathcal{P}(0)$. More generally, for any $c', c'' \geq 0$ such that $c'' > c'$, one can check that $\mathcal{P}(c'') \supset \mathcal{P}(c')$. Moreover, for any $c \geq 2$, we have $\mathcal{P}(c) = \Delta(S)$. ■

In Example 3, we should note that $\mathcal{P}(0)$ does not contain any probability distribution that is strictly positive. This is because of the fact that player 2 has a strictly dominated strategy (N) that is never recommended

by the mediator, hence the outcomes (N,H) and (N,T) are never realized, implying that p_{13} and p_{23} are always zero. The absence of a strictly positive equilibrium in $\mathcal{P}(0)$, or in $\mathcal{P}(c)$ for any $c \geq 0$, does not prevent, however, any change in the cost of disobedience to affect the set of CCE. We should also note that neither in Example 1 nor Example 3, any vertex of $\Delta(S)$ can become a CCE unless the cost of disobedience is sufficiently large, i.e., $c \geq 2$. On the other hand, in Example 2, two vertices of $\Delta(S)$ (corresponding to two pure Nash equilibria) are contained by $\mathcal{P}(0)$. It seems that the lack or the presence of a vertex element in $\mathcal{P}(0)$ is inconsequential. Our final example shows what happens when the unique element of $\mathcal{P}(0)$ (and also $\partial\mathcal{P}(0)$) is a vertex of $\Delta(S)$, which implies that the only NE and CE is a pure strategy.

Example 4. Consider the following Prisoners' Dilemma game.

	C	D
C	$2, 2$	$0, 3$
D	$3, 0$	$1, 1$

Note that $S_1 = S_2 = \{C, D\}$ and $S = \{(C, C), (C, D), (D, C), (D, D)\}$. For any $p \in \Delta(S)$, let $p = (p_{11}, p_{12}, p_{21}, p_{22})$ where $p_{11} = p((C, C))$, $p_{12} = p((C, D))$, $p_{21} = p((D, C))$, and $p_{22} = p((D, D))$. Given any $c \geq 0$,

$$\mathcal{P}(c) \ = \{p \in \Delta(S) : \ (1 - c)(p_{11} + p_{12}) \leq 0 \text{ and } (1 - c)(p_{11} + p_{21}) \leq 0\}.$$

It is easy to check that

$$\mathcal{P}(c) = \begin{cases} \{(0, 0, 0, 1)\} & \text{if } c < 1, \\ \Delta(S) & \text{if } c \geq 1. \end{cases}$$

Clearly, $\mathcal{P}(c) \supset \mathcal{P}(0) = \{(0, 0, 0, 1)\}$ if and only if $c \geq 1$. Similarly, for any $c', c'' \geq 0$ such that $c'' > c'$, $\mathcal{P}(c'') \supset \mathcal{P}(c')$ if and only if $c'' \geq 1 > c'$. ■

Note in Example 4 that $\partial\mathcal{P}(0) = \{(0, 0, 0, 1)\}$ and $\partial\mathcal{P}(0) \setminus \mathcal{V}(\Delta(S)) = \emptyset$. It seems that the lack of a non-vertex element in $\partial\mathcal{P}(0)$ (and thus in $\mathcal{P}(0)$) prevents any cost change to have expansionary effects, unless it is sufficiently

large. Notice that in Example 1-3, $\partial\mathcal{P}(0) \setminus \mathcal{V}(\Delta(S)) \neq \emptyset$, which implies that the CCE with zero cost of disobedience has at least one element that is not pure. After these observations, we are ready to present our results.

Lemma 1. *Consider any normal-form game* $G = [N, \{S_i\}, \{u_i(.)\}]$. *For any* $c', c'' \geq 0$ *such that* $c'' \geq c$, $\mathcal{P}(c'') \supseteq \mathcal{P}(c')$.

Proof. Simply follows from (2).

Lemma 1 states that when the cost of disobedience increases, the set of CCE weakly expands in any normal-form game. We will characterize conditions under which this expansion is strict in Proposition 1 below. But, first we need to present the following lemma that is to be used in its proof.

Lemma 2. *Consider any normal-form game* $G = [N, \{S_i\}, \{u_i(.)\}]$. *Suppose there exists* $c' \geq 0$ *such that* $\mathcal{P}(c') \neq \Delta(S)$ *and there exist* $\hat{p} \in \partial\mathcal{P}(c')$ *and* $\hat{S} \subseteq S$ *such that* $|\hat{S}| > 1$ *and* $\sum_{s_{-i} \in \hat{S}_{-i}} \hat{p}(s_i', s_{-i}) > 0$ *if and only if* $s' \in \hat{S}$. *Then, for any* $c'' > c'$, $i \in N$, *and* $r_i, t_i \in S_i$ *it is true that* $D_i(\hat{p}, c'', r_i, t_i) = 0$ *if* $r_i \in S_i \setminus \hat{S}_i$ *and* $D_i(\hat{p}, c'', r_i, t_i) > 0$ *if* $r_i \in \hat{S}_i$.

This lemma says that if there exists a probability distribution that is not pure on the boundary of the equilibrium set for some cost level (when this set is not already equal to the entire simplex), then for any higher cost level and any player, the difference between the expected payoff received by following the recommended strategy induced by this probability distribution (with strictly positive weight) and choosing any other strategy is strictly positive. This indicates that there is room for expansion under these conditions.

Proof. Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Suppose that assumptions in Lemma 2 hold. Pick any $c' \geq 0$ such that $\mathcal{P}(c') \neq \Delta(S)$ and pick any $i \in N$ and $c'' > c$. Also pick $\hat{p} \in \partial\mathcal{P}(c')$ and $\hat{S} \subseteq S$ such that $|\hat{S}| > 1$ and $\sum_{s_{-i} \in \hat{S}_{-i}} \hat{p}(s_i', s_{-i}) > 0$ if and only if $s' \in \hat{S}$. Then for any $r_i, t_i \in$

S_i, $D_i(\hat{p}, c'', r_i, t_i) = 0$ if $r_i \in S_i \setminus \hat{S}_i$ and $D_i(\hat{p}, c'', r_i, t_i) = D_i(\hat{p}, c', r_i, t_i) + (c'' - c')$ if $r_i \in \hat{S}_i$ by (2). Also, $\hat{p} \in \partial\mathcal{P}(c')$ implies $D_i(\hat{p}, c', r_i, t_i) \geq 0$. Since $c'' - c' > 0$, it follows that $D_i(\hat{p}, c_i'', r_i, t_i) > 0$ if $r_i \in \hat{S}_i$. ■

Proposition 1. *Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Suppose there exists $c' \geq 0$ is such that $\mathcal{P}(c') \neq \Delta(S)$. Then, for all $c'' > c'$ we have $\mathcal{P}(c'') \supset \mathcal{P}(c')$ if and only if $\partial\mathcal{P}(c') \setminus \mathcal{V}(\Delta(S)) \neq \emptyset$.*

Before presenting the proof, we would like to note that $\partial\mathcal{P}(c') \setminus \mathcal{V}(\Delta(S)) = \emptyset$ implies that $\partial\mathcal{P}(c')$ must be a singleton set (hence, so is $\mathcal{P}(c')$) and the unique equilibrium in $\partial\mathcal{P}(c')$ is a vertex of $\Delta(S)$, which means it is pure. To see this, take any $p, p' \in \partial\mathcal{P}(c')$ such that $p \neq p'$. Note that the convexity of $\mathcal{P}(c')$ implies that $\partial\mathcal{P}(c')$ is convex. So, $0.5p + 0.5p' \in \partial\mathcal{P}(c') \setminus \mathcal{V}(\Delta(S))$, contradicting that $\partial\mathcal{P}(c') \setminus \mathcal{V}(\Delta(S)) = \emptyset$. This means that for any normal-form game, the necessary and sufficient condition of Proposition 1 is violated at any $c' \geq 0$ if and only if $\partial\mathcal{P}(c')$ is a singleton set and the unique equilibrium in $\partial\mathcal{P}(c')$ is a vertex of $\Delta(S)$, i.e. there exists a unique equilibrium in pure strategies.

Proof. We will first prove the 'if part'. Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Suppose that the assumptions in Proposition 1 hold. Pick $c' \geq 0$ such that $\mathcal{P}(c') \neq \Delta(S)$ and pick c'' such that $c'' > c'$. Also pick $\hat{p} \in \partial\mathcal{P}(c')$ such that $\hat{p}$ is not in $\mathcal{V}(\Delta(S))$. Let $\hat{S}$ be the largest subset of S such that for any $i \in N$ we have $\sum_{s_{-i} \in \hat{S}_{-i}} \hat{p}(s_i', s_{-i}) > 0$ if $s' \in \hat{S}$. Clearly, $\hat{p}(s') = 0$ if $s' \in S \setminus \hat{S}$. Note that $\hat{S} \neq \emptyset$ since $\sum_{s \in S} \hat{p}(s) = 1$ and $|\hat{S}| > 1$ since $\hat{p}$ is not a vertex of $\Delta(S)$. The facts that $\mathcal{P}(c')$ and $\Delta(S)$ are convex, $\hat{p}$ is in $\partial\mathcal{P}(c')$, and $|\hat{S}|$ is larger than 1 together imply that there exists $p^* \in \Delta(S) \setminus \mathcal{P}(c')$ such that for any $i \in N$, we have $\sum_{s_{-i} \in \hat{S}_{-i}} p^*(s_i', s_{-i}) > 0$ if and only if $s' \in \hat{S}$ and the set $(\hat{p}, p^*)$, i.e., the interior of the arc connecting $\hat{p}$ to p^*, is nonempty and contained by $\Delta(S) \setminus \mathcal{P}(c')$. Pick such a p^*. Consider the sequences (α_k) and (q_k) such that for any positive integer k, we have $\alpha_k = (1/2)^k$ and $q_k = \alpha_k\, p^* + (1 - \alpha_k)\, \hat{p}$. Pick any $i \in N$ and $r_i, t_i \in S_i$. By construction,

it is true that for every integer $k \geq 1$ we have $q_k(s') = 0$ if $s' \in S \setminus \hat{S}$ and $\sum_{s_{-i} \in \hat{S}_{-i}} q_k(s'_i, s_{-i}) > 0$ if $s' \in \hat{S}$. Therefore, $D_i(q_k, c'', r_i, t_i) = 0$ if $r_i \in S_i \setminus \hat{S}_i$ and $D_i(q_k, c'', r_i, t_i) = D_i(q_k, c', r_i, t_i) + (c'' - c')$ if $r_i \in \hat{S}_i$ by (2). Moreover, $\lim_{k\to\infty} D_i(q_k, c'', r_i, t_i) = D_i(\hat{p}, c'', r_i, t_i)$. Since $c'' > c'$ and all assumptions in Lemma 2 are satisfied for $\hat{S}$ and $\hat{p}$, it is true that $D_i(\hat{p}, c'', r_i, t_i) = 0$ if $r_i \in S \setminus \hat{S}_i$ and $D_i(\hat{p}, c'', r_i, t_i) > 0$ if $r_i \in \hat{S}_i$. We have thus established that $\lim_{k\to\infty} D_i(q_k, c'', r_i, t_i) = 0$ if $r_i \in S_i \setminus \hat{S}_i$ and $\lim_{k\to\infty} D_i(q_k, c'', r_i, t_i) > 0$ if $r_i \in \hat{S}_i$. Since $D_i(p, c'', r_i, t_i)$ is continuous in p, there exists a positive integer $k_i(r_i, t_i)$ such that $D_i(q_k, c'', r_i, t_i) > 0$ for all $k \geq k_i(r_i, t_i)$ if $r_i \in \hat{S}_i$. Note that we can calculate $k_i(r_i, t_i)$ for any $i \in N$ and $r_i, t_i \in S_i$. Let $\bar{k} = \max_{i\in N} \max_{r_i,t_i\in S_i} k_i(r_i, t_i)$. Then, for any $i \in N$ and $r_i, t_i \in S_i$ it is true that $D_i(q_{\bar{k}}, c'', r_i, t_i) > 0$ if $r_i \in \hat{S}_i$ and $D_i(q_{\bar{k}}, c'', r_i, t_i) = 0$ if $r_i \in S \setminus \hat{S}_i$. Therefore, $q_{\bar{k}} \in \mathcal{P}(c'')$. Since $q_{\bar{k}} \in \Delta(S) \setminus \mathcal{P}(c')$, $\mathcal{P}(c'') \neq \mathcal{P}(c')$. Finally, since $c'' > c'$, Lemma 1 implies that $\mathcal{P}(c'') \supseteq \mathcal{P}(c')$. Therefore, $\mathcal{P}(c'') \supset \mathcal{P}(c')$, completing the proof of the 'if part'. Now we will prove the 'only if part'.

First, consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Suppose that there exists $c' \geq 0$ is such that $\mathcal{P}(c') \neq \Delta(S)$. Also suppose for a contradiction that $\partial\mathcal{P}(c') \setminus \mathcal{V}(\Delta(S)) = \emptyset$. This means that the necessary and sufficient condition of Proposition 1 is violated at any $c' \geq 0$ if and only if $\partial\mathcal{P}(c')$ is a singleton set and the unique equilibrium in $\partial\mathcal{P}(c')$ is a vertex of $\Delta(S)$. So, pick any $c' > 0$ such that $|\partial\mathcal{P}(c')| = 1$ and $\partial\mathcal{P}(c') \setminus \mathcal{V}(\Delta(S)) = \emptyset$. Let p^* be the unique distribution in $\mathcal{P}(c')$. Define for any $p \in \Delta(S) \setminus \{p^*\}$ and $i \in N$ the set $S_i(p, c') = \{(r_i, t_i) \in S_i \times S_i : \sum_{s_{-i}\in S_{-i}} p(r_i, s_{-i})[u_i(r_i, s_{-i}) - (u_i(t_i, s_{-i}) - c')] < 0\}$ and also the set of individuals $N(p, c') = \{i \in N : S_i(p, c') \neq \emptyset\}$ who find that (unilaterally) disobeying the recommendation of the mediator is strictly beneficial. For any $p \in \Delta(S) \setminus \{p^*\}$, we know that $N(p, c')$ is nonempty, since $p \notin \mathcal{P}(c')$. Let $k(c') = \max_{p\in\Delta(S)\setminus\{p^*\}} \max_{i\in N(p,c')} \max_{(r_i,t_i)\in S_i(p,c')} [u_i(r_i, s_{-i}) - (u_i(t_i, s_{-i}) - c')]$. Note that $k(c') < 0$ since $\Delta(S) \setminus \mathcal{P}(c') \neq \emptyset$. Pick any $c'' > 0$ such that $c' < c'' < -k(c')$. It follows that $i \in N(p, c'')$ if and only if $i \in N(p, c')$. So, for any $p \in \Delta(S) \setminus \{p^*\}$, $N(p, c'') \neq \emptyset$, implying that $\mathcal{P}(c'') = \{p^*\} = \mathcal{P}(c')$. Thus, it is not true that $\mathcal{P}(c'') \supset \mathcal{P}(c')$ for all

$c'' > c'$, completing the proof of the 'only if' part. ■

Proposition 1 states that if at any level of the cost of disobedience, c, there is any room for the set of CCE, $\mathcal{P}(c)$, to expand, then an increase in c can lead to an expansion if and only if the boundary of $\mathcal{P}(c)$ contains an equilibrium that is unpure, i.e., a non-vertex element of the probability simplex $\Delta(S)$. One can easily check that in Examples 1-3, the (necessary and) sufficient condition of Proposition 1 is satisfied, and therefore its prediction becomes true. On the other hand, Example 4 illustrates how this prediction fails when the necessary (and sufficient) condition of Proposition 1 does not hold. In that example, $\mathcal{P}(c') \neq \Delta(S)$ and thus there is room for $\mathcal{P}(c')$ to expand only if $c' < 1$. So, consider any $c' < 1$. We saw that an increase in the cost level from c' to c'' can be expansionary only if $c'' \geq 1$. This implies that for any cost increase ϵ that is smaller than $1 - c'$, the set of CCE is not larger when the cost of disobedience is $c'' = c' + \epsilon$ than when it is c'.

Proposition 1 also indicates when costly mediation leads to a coarser set of CCE than costless mediation does.

Corollary 1. *Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. If $\mathcal{P}(0) \neq \Delta(S)$ and G has a Nash equilibrium that is not a vertex of $\Delta(S)$, then $\mathcal{P}(c) \supset \mathcal{P}(0)$ for any $c > 0$.*

Proof. We suppose that the normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$ is such that $\mathcal{P}(0) \neq \Delta(S)$ and G has a Nash equilibrium σ that is in $\Delta(S) \setminus \mathcal{V}(\Delta(S))$. Pick any $c > 0$. If $\sigma \in \partial\mathcal{P}(0)$, then $\sigma \in \partial\mathcal{P}(0) \setminus \mathcal{V}(\Delta(S))$, and by Proposition 1 we obtain $\mathcal{P}(c) \supset \mathcal{P}(0)$. Now suppose that $\sigma \notin \partial\mathcal{P}(0)$. Pick any $p \in \partial\Delta(S)$ such that $p \notin \mathcal{V}(\Delta(S))$. Since $\mathcal{P}(0) \subseteq \Delta(S)$, it is true that $\sigma \notin \partial\Delta(S)$, implying $\sigma \neq p$. Define $q(\alpha) = \alpha\sigma + (1 - \alpha)p$ for any $\alpha \in (0, 1)$. Since $\mathcal{P}(0)$ is bounded and $\mathcal{P}(0) \subseteq \Delta(S)$, there exists some $\hat{\alpha} \in (0, 1)$ such that $q(\hat{\alpha}) \in \partial\mathcal{P}(0)$. Also, since $\{\sigma, p\} \cap \mathcal{V}(\Delta(S)) = \emptyset$, it is true that $q(\hat{\alpha}) \notin \mathcal{V}(\Delta(S))$, implying $q(\hat{\alpha}) \in \partial\mathcal{P}(0) \setminus \mathcal{V}(\Delta(S))$. Since $\partial\mathcal{P}(0) \setminus \mathcal{V}(\Delta(S)) \neq \emptyset$, again we have $\mathcal{P}(c) \supset \mathcal{P}(0)$ by Proposition 1. ■

The above corollary to Proposition 1 says that if any normal-form game has an unpure (yet, not necessarily totally mixed) Nash equilibrium, then the set of CCE is always strictly coarser when disobedience is costly than when it is not. If, on the other hand, a normal-form game has only one equilibrium in pure strategies, then the set of CEE does not strictly expand unless the cost is sufficiently high. Thus, we can say that for any finite normal-form game that has strictly dominant strategy equilibrium or that is dominance solvable, adding small non-zero cost to disobedience does not affect the set of CCE. The next result shows that in any normal-form game every probability distribution becomes a CCE, as expected, when disobedience becomes sufficiently costly.

Proposition 2. *For any normal-form game* $G = [N, \{S_i\}, \{u_i(.)\}]$*, there exists* $\bar{c} \geq 0$ *such that* $\mathcal{P}(c) = \Delta(S)$ *if and only if* $c \geq \bar{c}$*.*

Proof. Pick any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Let $\bar{c} = \max_{i \in N} \max_{r_i, t_i \in S_i} \max_{s_{-i} \in S_{-i}} [u_i(r_i, s_{-i}) - u_i(t_i, s_{-i})]$. Then, equation (2) implies that for any $p \in \Delta(S)$, $c \geq \bar{c}$, $i \in N$, and $r_i, t_i \in S_i$ we have $D_i(p, c, r_i, t_i) \geq 0$. Thus, $\mathcal{P}(c) = \Delta(S)$ if $c \geq \bar{c}$. To prove the 'only if' part, first assume that $\bar{c}$ defined above is equal to zero. Since $c < 0$ is not possible, it is true that $\mathcal{P}(c) = \Delta(S)$ only if $c \geq 0$. Now, suppose $\bar{c} > 0$. Then, pick any $c \in [0, \bar{c})$. The definition of $\bar{c}$ implies that there exist $i \in N$, $r_i, t_i \in S_i$, and $s_{-i} \in S_{-i}$ such that $u_i(r_i, s_{-i}) - u_i(t_i, s_{-i}) + c < 0$. Pick any such $i \in N$, $r_i, t_i \in S_i$, and $s_{-i} \in S_{-i}$. Let $p \in \Delta(S)$ be such that $p(r_i, s_{-i}) = 1$ and $p(s) = 0$ for all $s \in S \setminus \{(r_i, s_{-i})\}$. Then, equation (2) implies that $D_i(p, c, r_i, t_i) = u_i(r_i, s_{-i}) - u_i(t_i, s_{-i}) + c < 0$, implying that $p \notin \mathcal{P}(c)$. Therefore, $\mathcal{P}(c) \neq \Delta(S)$ if $c \in [0, \bar{c})$, completing the proof. ■

Note that in Examples 1 and 3 we have $\mathcal{P}(c) = \Delta(S)$ if $c \geq 2$ and in Examples 2 and 4 we have $\mathcal{P}(c) = \Delta(S)$ if $c \geq 1$.

3.2 Welfare Effects

Now we will study how the social welfare in any mediated normal-form game can be affected by the change in, as well as the presence/absence of, the cost of disobedience. Given any disobedience cost $c \geq 0$, we suppose that the mediator has the task of implementing a CCE that maximizes the sum of the expected utilities of all players. Since the solution to this maximization problem may not be unique, we define the set of (socially) optimal CCE as given by

$$SO(\mathcal{P}(c)) = \left\{ p \in \mathcal{P}(c) : \sum_{i \in N} E[u_i|p] \geq \sum_{i \in N} E[u_i|p'] \text{ for all } p' \in \mathcal{P}(c) \right\}, \quad (3)$$

where

$$E[u_i|p'] = \sum_{s \in S} p'(s) u_i(s_i, s_{-i}). \quad (4)$$

Note that all probability distributions in $SO(\mathcal{P}(c))$ must lead to the same expected utility sum for the players. For simplicity, we will denote this sum by $E[u_i|SO(\mathcal{P}(c))]$, by slightly abusing the notation. Also, we will use $SO(\Delta(S))$ to denote the set of probability distributions in $\Delta(S)$ that maximize the sum of the expected utilities of all players.

Proposition 3. *Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Pick $c', c'' \geq 0$. It is true that*

(i) if $c'' > c'$, then $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c''))] \geq \sum_{i \in N} E[u_i|SO(\mathcal{P}(c'))]$, and

(ii) if $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c''))] > \sum_{i \in N} E[u_i|SO(\mathcal{P}(c'))]$, then $c'' > c$.

Proof. Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. To prove part (i), pick any $c', c'' \geq 0$ such that $c'' > c'$. Lemma 1 implies $\mathcal{P}(c'') \supseteq \mathcal{P}(c')$. Then, equations (3) and (4) imply that $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c''))] \geq \sum_{i \in N} E[u_i|SO(\mathcal{P}(c'))]$. To prove part (ii), pick any $c', c'' \geq 0$ and assume that $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c''))] > \sum_{i \in N} E[u_i|SO(\mathcal{P}(c'))]$. Then, it cannot be true that $SO(\mathcal{P}(c')) \supseteq SO(\mathcal{P}(c''))$. It follows from equations (3) and (4)

that $\mathcal{P}(c') \supseteq \mathcal{P}(c'')$ cannot be true. Consequently, Lemma 1 implies that $c' \geq c''$ cannot be true, implying that $c'' > c$, which completes the proof. ■

The above proposition asserts that the total payoffs of players at an optimal CCE cannot be lower whenever the cost of disobedience becomes higher in the mediated game. Moreover, a cost change cannot increase the total payoffs of players at an optimal CCE unless it is positive. However, whether the total payoffs increase, when the cost of disobedience does so, depends on the payoff structure of the game. To illustrate this point, let us first consider the Matching Pennies Game in Example 1. Note that for any $p \in \Delta(S)$, one can calculate that $\sum_{i \in N} E[u_i|p] = 0$, implying $SO(\mathcal{P}(c)) = \mathcal{P}(c)$ for any $c \geq 0$, i.e., any probability distribution in $\mathcal{P}(c)$ is optimal. Moreover, $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c))] = 0$ for any $c \geq 0$, implying that the optimal value of the expected social welfare is independent from the cost of disobedience. As another example, let us now consider the Prisoners' Dilemma Game in Example 4, where the payoffs are not zero-sum. We can easily calculate $SO(\mathcal{P}(c))$ for any $c \geq 0$ as follows:

$$SO(\mathcal{P}(c)) = \begin{cases} \{(0,0,0,1)\} & \text{if } c < 1, \\ \{(1,0,0,0)\} & \text{if } c \geq 1. \end{cases}$$

Clearly, for any cost levels c and c'' such that $1 > c'' > c' \geq 0$, we have $\sum_{i \in N} E[u_i|\ SO(\mathcal{P}(c''))] = \sum_{i \in N} E[u_i|SO(\mathcal{P}(c'))] = 2$. On the other hand, for any $c' < 1$, we can always find $c'' \geq 1$ such that $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c''))] = 4 > 2 = \sum_{i \in N} E[u_i|SO(\mathcal{P}(c'))]$. Note that the singleton set $\{(1,0,0,0)\}$ is incidentally equal to $SO(\Delta(S))$, the set of probability distributions that maximize the total payoffs (of two players) in $\Delta(S)$. These observations can be generalized in the following result.

Proposition 4. *Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Suppose $c' \geq 0$ is such that $\sum_{i \in N} E[u_i|SO(\Delta(S))] > \sum_{i \in N} E[u_i|\ SO(\mathcal{P}(c'))]$. Then, there exists $c'' > c'$ such that $\sum_{i \in N} E[u_i|SO(\mathcal{P}(c''))] > \sum_{i \in N} E[u_i|\ SO(\mathcal{P}(c'))]$.*

Proof. Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$. Suppose there exists $c' \geq 0$ is such that $\sum_{i \in N} E[u_i | SO(\Delta(S))] > \sum_{i \in N} E[u_i | SO(\mathcal{P}(c'))]$. Pick any such c'. Then, equations (3) and (4) imply that $\Delta(S) \supset \mathcal{P}(c')$. Proposition 2 implies that there exists some $c'' \geq 0$ such that $\mathcal{P}(c'') = \Delta(S)$, implying $SO(\mathcal{P}(c'')) = SO(\Delta(S))$. Hence, $\sum_{i \in N} E[u_i | SO(\mathcal{P}(c''))] > \sum_{i \in N} E[u_i | SO(\mathcal{P}(c'))]$. Then, Proposition 3(ii) implies that $c'' > c'$, which completes the proof. ■

Proposition 4 implies that if in any normal-form game the cost of disobedience is at such a level that the total expected welfare obtained by the players when they obey to play according to the recommendations of the mediator implementing an optimal CCE is below the maximum total expected welfare the players can ever obtain from this game, then the mediator can increase the total expected welfare of the players by increasing the cost of disobedience to a sufficiently high level. As a matter of fact, this is exactly the case in Examples 2 and 4. For instance, consider Example 2. Let $p = (p_{11}, p_{12}, p_{21}, p_{22})$ for any $p \in \Delta(S)$. One can easily check that $SO(\Delta(S)) = \{(0, 0, 1, 0)\}$ and $\sum_{i \in N} E[u_i | SO(\Delta(S))] = 8$. Also,

$$SO(\mathcal{P}(c))] = \begin{cases} \left\{\left(\frac{1-c}{3-c}, 0, \frac{1+c}{3-c}, \frac{1-c}{3-c}\right)\right\} & \text{if } c \in [0, 1) \\ \{(0, 0, 1, 0)\} & \text{if } c \geq 1 \end{cases}$$

and

$$\sum_{i \in N} E[u_i | SO(\mathcal{P}(c))] = \begin{cases} \frac{20-4c}{3-c} & \text{if } c \in [0, 1) \\ 8 & \text{if } c \geq 1. \end{cases}$$

Note that the sum $\sum_{i \in N} E[u_i | SO(\mathcal{P}(c))]$ is increasing in c over $(0, 1)$ and $\lim_{c \to 1^-} \sum_{i \in N} E[u_i | SO(\mathcal{P}(c))] = 8 = \sum_{i \in N} E[u_i | SO(\Delta(S))]$. So, for any $c' \in [0, 1)$ and any $c'' \geq 1$ in Example 2, the claim of Proposition 4 becomes true. What is more interesting is that this claim even becomes true for any $c' \in [0, 1)$ and any $c'' > c'$. That is, the mediator can raise the total expected welfare of the players at an optimal CCE by increasing the cost of disobedience even by an arbitrarily small amount as long as this cost is sufficiently low (i.e., it is below 1).

Example 2 reveals that there are normal-form games in which the social benefit of mediation increases with the cost of disobedience, suggesting that mediation in these games performs better when disobedience is more costly. In fact, we can formally quantify the performance of mediation under costly disobedience by extending the measures in Ashlagi et al (2008) proposed for costless mediation. Given any normal-form game, we say that *the value of mediation*, $m(c)$, at cost level c is equal to the ratio between the total payoff obtained in any optimal CCE at cost level c and the maximal total Nash equilibrium payoff obtained in the absence of any mediation. Also, we say that *the value of enforcement*, $e(c)$, is the ratio between the total payoff obtained in any optimal CCE at cost level c and the maximal total payoff in the normal-form game. Given these definitions, Proposition 3 implies the following result.

Corollary 2. *Consider any normal-form game* $G = [N, \{S_i\}, \{u_i(.)\}]$. *Pick* $c', c'' \geq 0$. *It is true that*

(i) $m(c'') \geq m(c')$ *if and only if* $e(c'') \geq e(c')$,

(ii) *if* $c'' > c'$, *then* $m(c'') \geq m(c')$,

(iii) *if* $m(c'') > m(c')$, *then* $c'' > c$.

Proof. Directly follows from Proposition 3 and the definitions of $m(.)$ and $e(.)$. ■

Corollary 2 says that when the cost of disobedience c changes, $m(c)$ and $e(c)$ always move in the same direction and they are always non-decreasing. Whether they can be increasing at some (Lebesgue) measurable interval of cost values depends on the structure of game. To see this, note that for the normal-form game in Example 2, one can calculate that the value of mediation and the value of enforcement are given by

$$m(c) = \begin{cases} \dfrac{10-2c}{9-3c} & \text{if } c < 1 \\ \dfrac{4}{3} & \text{if } c \geq 1 \end{cases}$$

and

$$e(c) = \begin{cases} \dfrac{5-c}{6-2c} & \text{if } c < 1 \\ 1 & \text{if } c \geq 1 \end{cases}$$

respectively. One may check that both $m(c)$ and $e(c)$ are increasing over the interval $(0, 1)$. In particular, we note that $m(0) = 10/9$ and $m(c) = 12/9$ for any $c \geq 1$. We note that $m(c)$ is always above 1, implying that mediation is always beneficial for the society. Also, we note that $e(0) = 5/6$, $e(c) < 1$ for any $c < 1$, and $e(c) = 1$ for any $c \geq 1$. The full enforcement is attained if and only if $c \geq 1$.

On the other hand, for the Prisoners' Dilemma Game in Example 4, the value of mediation and the value of enforcement are given by

$$m(c) = \begin{cases} 1 & \text{if } c < 1 \\ 2 & \text{if } c \geq 1 \end{cases}$$

and

$$e(c) = \begin{cases} \dfrac{1}{2} & \text{if } c < 1 \\ 1 & \text{if } c \geq 1 \end{cases}$$

respectively. We observe that both $m(c)$ and $e(c)$ are constant for $c < 1$ and $c > 1$ and they both jump to a higher value at $c = 1$. Mediation is not beneficial (the value of mediation is not higher than 1) unless the cost of disobedience is sufficiently large, i.e., $c \geq 1$. Likewise, if c is less than 1, mediation cannot enforce the players to any outcome that is not obtained as a Nash equilibrium in the absence of mediation (let alone the outcome with the maximal social welfare). Full enforcement is attained only if $c \geq 1$, while

very small changes in the value of c can increase enforcement only if it is below, but also sufficiently close to, 1.

Clearly, there exist normal-form games (like the one in Example 2) in which the payoff obtained by each player at an optimal CCE changes as the cost of disobedience is varied. In these games, any rule (whether it is optimal or not) used by the mediator to select an equilibrium from the set of possible CCE would accordingly induce for each player a preference (or an expected utility function) over the possible values of disobedience cost. These preferences of players may result in strategic issues which we will investigate in the next section.

3.3 An Extension for Future Research: Cost-Selection Game under Mediation

Here we will introduce, and briefly study, an extension for future research. Consider a situation where each player in the mediated game is asked, before the game starts, to non-cooperatively select (and announce to the mediator) the cost of disobedience that s/he has to bear in case s/he disobeys to any recommendation made by the mediator. Suppose that before observing the cost chosen by any player i, the mediator announces an equilibrium rule that specifies a CCE for each possible cost profile reported by the society. Given this rule, we assume that each player will choose his/her disobedience cost given his/her conjectures about the choices of the others.

So, consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$ that players face in the absence of any mediation. Let c_i denote the cost of disobedience (punishment fee) player i non-cooperatively chooses and announces to the mediator, and let $C_i = [0, \bar{c}_i]$ denote for each player i the set of all admissible cost reports, where $\bar{c}_i = \max_{r_i,t_i \in S_i} \max_{s_{-i} \in S_{-i}} [u_i(r_i, s_{-i}) - u_i(t_i, s_{-i})]$. Define $C = \times_{i \in N} C_i$, with $c \in C$ denoting the cost profile for the set of all players. As we said earlier, before the players choose their costs of disobedience, the mediator announces a CCE rule f. Formally, f is a CCE rule if $f : C \to \Delta(S)$ is a function such that $f(c) \in \mathcal{P}(c)$ for all $c \in C$. Clearly, given any CCE rule f, the utility function $u_i(.)$ of each player i over the set

of strategies S in a given normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$ induces a utility function $u_i^{f,G}(.)$ over the set of strategies C such that $u_i^{f,G}(c) = E[u_i|p]$ for any $c \in C$, where $p = f(c)$ and $E[u_i|p] = \sum_{s \in S} p(s)u_i(s_i, s_{-i})$. Let $G^f = [N, \{C_i\}, \{u_i^{f,G}(.)\}]$ denote the cost-selection game obtained from G under the rule f. Given a normal-form game G and a CCE rule f, we define the best response correspondence of player $i \in N$ in the cost-selection game G^f as $b_i^f : C_{-i} \to C_i$ that assigns to each profile c_{-i} in C_{-i} the set

$$b_i^f(c_{-i}) = \{c_i \in C_i : u_i^{f,G}(c_i, c_{-i}) \geq u_i^{f,G}(c_i', c_{-i}) \text{ for all } c_i' \in C_i\}.$$

We say that a cost profile $c^* \in C$ is a Nash equilibrium of the cost-selection game G^f if $c_i^* \in b_i^f(c_{-i}^*)$ for all $i \in N$. Since for each $i \in N$, the set $C_i \subset \mathbb{R}$ is a nonempty, convex, and compact subset of an Euclidean space, we know by the works of Debreu (1952), Glicksberg (1952), and Fan (1952) that (whenever f is single-valued) a pure-strategy Nash equilibrium of G^f exists if $u_i^{f,G}(c)$ is continuous in c and quasiconcave in c_i.

Below, we will extend the game in Example 2, borrowed from Aumann (1974), to a cost-selection game to gain some insights.

Example 5. Consider the normal-form game in Example 2. Suppose that players can non-cooperatively choose their costs of disobedience. For any $p \in \Delta(S)$, let $p = (p_{11}, p_{12}, p_{21}, p_{22})$ where $p_{11} = p((U, L))$, $p_{12} = p((U, R))$, $p_{21} = p((D, L))$, and $p_{22} = p((D, R))$. One can easily check that for any cost profile $c = (c_1, c_2) \in C$, the set of CCE can be calculated as

$$\mathcal{P}(c) = \left\{ p \in \Delta(S) : \begin{array}{ll} (1 + c_1)p_{11} \geq (1 - c_1)p_{12}, & (1 + c_1)p_{22} \geq (1 - c_1)p_{21}, \\ (1 + c_2)p_{11} \geq (1 - c_2)p_{21}, & (1 + c_2)p_{22} \geq (1 - c_2)p_{12}. \end{array} \right\}$$

Clearly, for any $c \in C$ such that $c \neq (0, 0)$, we have $\mathcal{P}(c) \supset \mathcal{P}((0, 0))$. More generally, for any $c', c'' \in [0, 1]^2$ such that $c'' \geq c'$ with $c_i'' > c_i'$ for some $i \in N$, one can easily check that $\mathcal{P}(c'') \supset \mathcal{P}(c')$.

Now, suppose that the mediator announces a CCE rule f such that $f(c) \in SO(\mathcal{P}(c))$ for any $c \in [0, 1]$. One can easily calculate that $\sum_{i \in N} E[U_i|p] =$

$6 + 2p_{21}$ for any $p \in \Delta(S)$, implying that

$$SO(\mathcal{P}(c))] = \begin{cases} \left\{\left(\frac{(1+c_1)(1-c_2)}{3+c_1+c_2-c_1c_2}, 0, \frac{(1+c_1)(1+c_2)}{3+c_1+c_2-c_1c_2}, \frac{(1-c_1)(1+c_2)}{3+c_1+c_2-c_1c_2}\right)\right\} & \text{if } c_1, c_2 \in [0,1) \\ \{(\frac{1-c_2}{2}, 0, \frac{1+c_2}{2}, 0)\} & \text{if } c_1 = 1 \text{ and } c_2 \in [0,1) \\ \{(0, 0, \frac{1+c_1}{2}, \frac{1-c_1}{2})\} & \text{if } c_1 \in [0,1) \text{ and } c_2 = 1 \\ \{(0,0,1,0)\} & \text{if } c_1 = 1 \text{ and } c_2 = 1. \end{cases}$$

Since $SO(\mathcal{P}(c))$ is always a singleton set, $f(c)$ is uniquely determined. Noting that $E[U_1|p] = 5p_{11} + 4p_{21} + p_{22}$ and $E[U_2|p] = p_{11} + 4p_{21} + 5p_{22}$ for any $p \in \Delta(S)$, one can easily calculate that

$$E[U_1|f(c)] = \begin{cases} \frac{10+8c_1-2c_1c_2}{3+c_1+c_2-c_1c_2} & \text{if } c_1 \in [0,1) \text{ and } c_2 \in [0,1) \\ \frac{9-c_2}{2} & \text{if } c_1 = 1 \text{ and } c_2 \in [0,1) \\ \frac{5+3c_1}{2} & \text{if } c_1 \in [0,1) \text{ and } c_2 = 1 \\ 4 & \text{if } c_1 = 1 \text{ and } c_2 = 1 \end{cases}$$

and

$$E[U_2|f(c)] = \begin{cases} \frac{10+8c_2-2c_1c_2}{3+c_1+c_2-c_1c_2} & \text{if } c_1 \in [0,1) \text{ and } c_2 \in [0,1) \\ \frac{9-c_1}{2} & \text{if } c_1 \in [0,1) \text{ and } c_2 = 1 \\ \frac{5+3c_2}{2} & \text{if } c_1 = 1 \text{ and } c_2 \in [0,1) \\ 4 & \text{if } c_1 = 1 \text{ and } c_2 = 1. \end{cases}$$

Thus, we have constructed the cost-selection game $G^f = [N, \{C_i\}, \{u_i^{f,G}(.)\}]$. For this game, we can calculate the best-response correspondences as $b_1^f(c_2) = \{1\}$ for any $c_2 \in C_2$ and $b_2^f(c_1) = \{1\}$ for any $c_1 \in C_1$. Clearly, the cost profile $(0,0)$ is not a Nash equilibrium of G^f. It is true that if player $j \in \{1,2\}$ chooses his/her cost at $c_j = 0$, player $i \neq j$ can secure an expected utility of $9/2$ by choosing $c_i = 1$. As a matter of fact, (c_1, c_2) is never Nash equilibrium when $c_1 \in [0,1)$ or $c_2 \in [0,1)$. The unique Nash equilibrium of G^f arises at

$(c_1, c_2) = (1, 1)$, at which both players obtain an expected utility of 4. ■

Example 5 reveals the following.

Remark 1. *There exist cost-selection games where selecting the cost of disobedience at zero level (or below some positive level) is a strictly dominated strategy for each player. In such games, it is to the interest of each player to voluntarily commit to pay some positive penalty fee to the mediator in case of disobedience.*

Inspecting the expected utilities in Example 5 closely, we can observe that for each $i \in N$, $E[U_i|f(c)]$ is strictly increasing in the own cost level c_i (for any level of the opponent's cost c_j) below some threshold. This observation leads us to note down another simple remark below. Let $\mathbf{0}$ denote the zero cost profile c where $c_i = 0$ for all $i \in N$.

Remark 2. *Consider any normal-form game $G = [N, \{S_i\}, \{u_i(.)\}]$ and any CCE rule f. If there exists some $i \in N$ such that $E[U_i|f(c)]$ is increasing in c_i around $\mathbf{0}$, then $c^* = \mathbf{0}$ is not a Nash equilibrium of the game $G^f = [N, \{C_i\}, \{u_i^{f,G}(.)\}]$.*

The next example shows that the games in Remark 1 or Remark 2 are not universal. There are also games where none of the players finds it beneficial to voluntarily commit to pay positive penalty fees to the mediator in case of disobedience.

Example 6. Suppose that the players in Example 4 can non-cooperatively choose their costs of disobedience in the mediated game of Prisoners' Dilemma. For any $p \in \Delta(S)$, let $p = (p_{11}, p_{12}, p_{21}, p_{22})$ where $p_{11} = p((C, C))$, $p_{12} = p((C, D))$, $p_{21} = p((D, C))$, and $p_{22} = p((D, D))$. For any cost profile, $c = (c_1, c_2) \in [0, 1]^2$, the set of CCE can be calculated as

$$\mathcal{P}(c) = \{p \in \Delta(S) : (1 - c_1)(p_{11} + p_{12}) \leq 0 \text{ and } (1 - c_2)(p_{11} + p_{21}) \leq 0\}.$$

It is easy to see that

$$\mathcal{P}(c)=\begin{cases}\{(0,0,0,1)\} & \text{if } c_1,c_2\in[0,1)\\ \{p\in\Delta(S): p_{11}=p_{12}=0,\ p_{21}+p_{22}=1\} & \text{if } c_1\in[0,1) \text{ and } c_2=1\\ \{p\in\Delta(S): p_{11}=p_{21}=0,\ p_{12}+p_{22}=1\} & \text{if } c_1=1 \text{ and } c_2\in[0,1)\\ \Delta(S) & \text{if } c_1=1 \text{ and } c_2=1.\end{cases}$$

Suppose that the mediator announces, before s/he observes the cost report of players, a CCE rule f such that $f(c)\in SO(\mathcal{P}(c))$ for any $c\in[0,1]^2$. One can easily check that $\sum_{i\in N}E[U_i|p]=4p_{11}+3p_{12}+3p_{21}+2p_{22}=3+p_{11}-p_{22}$ for any $p\in\Delta(S)$, implying that

$$SO(\mathcal{P}(c))=\begin{cases}\{(0,0,0,1)\} & \text{if } c_1\in[0,1) \text{ and } c_2\in[0,1)\\ \{(0,0,1,0)\} & \text{if } c_1\in[0,1) \text{ and } c_2=1\\ \{(0,1,0,0)\} & \text{if } c_1=1 \text{ and } c_2\in[0,1)\\ \{(1,0,0,0)\} & \text{if } c_1=1 \text{ and } c_2=1.\end{cases}$$

Since $SO(\mathcal{P}(c))$ is always a singleton set, $f(c)$ is uniquely determined. Noting that $E[U_1|p]=2p_{11}+3p_{21}+p_{22}$ and $E[U_2|p]=2p_{11}+3p_{12}+p_{22}$ for any $p\in\Delta(S)$, one can easily calculate that

$$E[U_1|f(c)]=\begin{cases}1 & \text{if } c_1\in[0,1) \text{ and } c_2\in[0,1)\\ 3 & \text{if } c_1\in[0,1) \text{ and } c_2=1\\ 0 & \text{if } c_1=1 \text{ and } c_2\in[0,1)\\ 2 & \text{if } c_1=1 \text{ and } c_2=1\end{cases}$$

and

$$E[U_2|f(c)]=\begin{cases}1 & \text{if } c_1\in[0,1) \text{ and } c_2\in[0,1)\\ 0 & \text{if } c_1\in[0,1) \text{ and } c_2=1\\ 3 & \text{if } c_1=1 \text{ and } c_2\in[0,1)\\ 2 & \text{if } c_1=1 \text{ and } c_2=1.\end{cases}$$

Thus, we have constructed the extended game $G^f=[N,\{C_i\},\{u_i^{f,G}(.)\}]$ where $C_i=[0,1]$ for any $i\in N$. Note that for this game, the payoffs can be

simply represented as follows:

	$c_2 = 1$	$c_2 \in [0,1)$
$c_1 = 1$	$2, 2$	$0, 3$
$c_1 \in [0,1)$	$3, 0$	$1, 1$

We can calculate the best-response correspondences for G^f as $b_1^f(c_2) = [0,1)$ for any $c_2 \in [0,1]$ and $b_2^f(c_1) = [0,1)$ for any $c_1 \in [0,1]$. Clearly, the cost profile $(0,0)$ is a Nash equilibrium of the extended game G^f. In general, it is true that any $c \in [0,1]^2$ is a Nash equilibrium of G^f if and only if $c_1 \in [0,1)$ and $c_2 \in [0,1)$. ■

Example 6 shows that the players that face a situation of Prisoners' Dilemma (between cooperation and defection, say, in reporting that they are guilty) when the game played is not mediated, remain to face a similar dilemma (between cooperation and defection in reporting that their costs of disobedience are not less than 1) also when their game is mediated if they non-cooperatively select the penalty fees they commit to pay in case they disobey the mediator's recommendations. Interestingly, any cost profile $c \in [0,1]^2$ that is arbitrarily close to, but smaller than, $(1,1)$ is a Nash equilibrium of the cost selection game in Example 6, while it yields a payoff of 1 to each player. On the other hand, each player could obtain a payoff of 2 if the mediator were to interfere and slightly increase the cost of disobedience for each player to a level equal to 1. It seems that the discontinuities in the value of mediation and the value of enforcement at the cost level $(1,1)$ –that we already calculated for the Prisoners' Dilemma Game in Section 3.2– create a strategic barrier for the players that cannot be overcome non-cooperatively. Example 6 also suggests the following.

Remark 3. *There exist cost-selection games where selecting the cost of disobedience below some positive level, hence at zero level, is a weakly dominant strategy for each player. In any Nash equilibrium of these games the payoff of any player is equal to what s/he obtains when each player reports his/her*

cost of disobedience as zero.

The discontinuity of the expected utlity functions in Example 6 illustrates that their continuity is, as already known, not essential but just sufficient for the existence of a pre-strategy Nash equilibrium in G^f. As a matter of fact, $E[U_i|f(c)]$ is constant for each $i \in N$ and higher in the own cost level c_i below the threshold of 1 (but constant and lower in the opponent's cost level c_j below the same threshold). This makes zero cost profile weakly dominant. However, if the players were to choose a cost level for their opponents (not for themselves) non-cooperatively, then we would get zero cost profile strictly dominated as in the previous example. Indeed, one may simply extend this observation to predict that in dominance solvable games with a strictly dominant strategy profile, in order to achieve efficiency, players should be enforced to select the disobedience costs of their opponents. In this regard, Examples 5 and 6 together suggest that given any normal-form game a mediator who has the capacity to calculate the expected utility of each player at all admissible cost profiles can profitably investigate whether the socially efficient outcome can be attained through a correlated equilibrium of a mediated game when, prior to this game, the disobedience cost of each player will be non-cooperatively selected by some player, not necessarily himself/herself.

4 Conclusions

In this paper, we have extended the notion of correlated equilibrium in normal-form games to a notion of *costly correlated equilibrium* (CCE) by allowing players to involuntarily or voluntarily bear a cost whenever they disobey recommendations of the mediator. In case the cost of disobedience is involuntary and common for all players, we have showed that the set of CCE at any cost level expands (whenever there is a room for it) if and only if the boundary of this set contains an unpure equilibrium, i.e., a non-vertex element of the probability simplex associated with the normal-form game. We have also showed that if the payoffs of a normal-form game and the cost of

disobedience are such that the total expected payoff of the players (the social welfare) at an optimal CCE is lower than the maximal total expected payoff they can ever get in the game, then the mediator can increase the social welfare by raising the cost of disobedience to a sufficiently high level. We have also discussed how our model can be extended, for a profitable investigation by future research, to a setting where the cost of disobedience is strategically chosen by each player. In more detail, we have considered a cost-selection game (prior to every mediated normal-form game) in which each player non-cooperatively chooses his/her cost after the mediator announces a CCE rule that specifies an optimal CCE for each possible cost profile of the players. We have showed that there exist cost-selection games in which choosing the cost of disobedience at zero level (or below some positive level) is a strictly dominated strategy for each player as well as games this strategy becomes weakly dominant for each player.

Future research may also extend the notion of costly correlated equilibrium in a number of directions. For example, using the generalization of correlated equilibrium by Hart and Schmeidler (1989) for infinite games, the notion of CCE can be extended to any game with infinitely many strategies. In particular, one can investigate the implication of disobedience costs in potential games –a special class of infinite strategy games, introduced by Monderer and Shapley (1994) and first studied within the context of correlated equilibrium by Neyman (1997). Also, given the two well-known, and generally unrelated, generalizations of correlated equilibrium in normal-form games, namely the weak (coarse) correlated equilibrium (also known as the simple extension) of Moulin and Vial (1978) and the soft correlated equilibrium of Forgó (2010), one can respectively define and study the weak costly correlated equilibrium (WCCE) and the soft costly correlated equilibrium (SCCE) taking the cost of disobedience into consideration. Another line of research can integrate the cost of disobedience to a refinement of correlated equilibrium known as "acceptable correlated equilibrium" introduced by Myerson (1986) as the analogue of trembling-hand perfection.

Finally, one can experimentally investigate whether in mediated normal-

form games the presence of a cost of disobedience for each player increases the performance of the recommendations of the mediator in overcoming coordination problems.

References

Anbarci, N., Feltovich, N. and Gurdal, M. Y. (2018). Payoff Inequity Reduces the Effectiveness of Correlated-Equilibrium Recommendations. *European Economic Review* **108**, 172-190.

Aumann, R. J. (1974). Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics* **1**, 67-95.

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

Bone, J., Drouvelis, M. and Ray, I. (2012). Coordination in 2x2 Games by Following Recommendations from Correlated Equilibria . *University of Birmingham Economics Discussion Paper,* **12–04**.

Cason, T. N. and Sharma, T. (2006). Recommended Play and Correlated Equilibria: an Experimental Study. *Economic Theory* **33**, 11-27.

Debreu, D. (1952). A Social Equilibrium Existence Theorem. *Proceedings of the National Academy of Sciences* **38**, 886-893.

Duffy, J. and Feltovich, N. (2010). Correlated Equilibria, Good and Bad: An Experimental Study. *International Economic Review* **51**, 701–721.

Fan, K. (1952). Fixed Point and Minimax Theorems in Locally Convex Topological Linear Spaces. *Proceedings of the National Academy of Sciences* **38**, 121-126.

Forgó, F. (2010). A Generalization of Correlated Equilibrium: A New Protocol. *Mathematical Social Sciences* **60**, 186-190.

Georgalos, K., Ray, I. and Sen Gupta, S. (2020). Nash versus Coarse Correlation. *Experimental Economics,* forthcoming.

Gerard-Varet, L. A. and Moulin, H. (1978). Correlation and Duopoly. *Journal of Economic Theory* **19**, 123-149.

Glicksberg, I. L. (1952). A Further Generalization of the Kakutani Fixed Point Theorem with Application to Nash Equilibrium Points. *Proceedings of the National Academy of Sciences* **38**, 170-174.

Hart, S. (2005). Adaptive Heuristics. *Econometrica* **73**, 1401-1430.

Hart, S. and Mas-Collell, A. (2000). A Simple Adaptive Procedure Leading to Correlated Equilibrium. *Econometrica* **68**, 1127-1150.

Hart, S. and Schmeidler D. (1989). Existence of Correlated Equilibria. *Mathematics of Operations Research* **14**, 18-25.

Liu, L. (1996). Correlated Equilibrium of Cournot Oligopoly Competition. *Journal of Economic Theory* **68**, 544-548.

Monderer, D. and Shapley L. S. (1996). Potential Games. *Games and Economic Behavior* **14**, 124-143.

Moreno, D. and Wooders, J. (1998). An Experimental Study of Communication and Coordination in Noncooperative Games. *Games and Economic Behavior* **24**, 47–76.

Moulin, H., Ray, I. and Sen Gupta, S. (2014). Improving Nash by Coarse Correlation. *Journal of Economic Theory* **150**, 852-865.

Moulin, H. and Vial J.-P. (1978). Strategically Zero-sum Games: The Class of Games Whose Completely Mixed Equilibria Cannot be Improved Upon. *International Journal of Game Theory* **7**, 201-221.

Myerson, R. (1986). Acceptable and Predominant Correlated Equilibria. *International Journal of Game Theory* **15**, 133-154.

Neyman, A. (1997). Correlated Equilibrium and Potential Games. *International Journal of Game Theory* **26**, 223-227.

Ray, I. and Sen Gupta S. (2013). Coarse Correlated Equilibria in Linear Duopoly Games. *International Journal of Game Theory* **42**, 541-562.

Rosenthal, R. W. (1974). Correlated Equilibria in Some Classes of Two-Person Games. *International Journal of Game Theory* **3**, 119-128.

Ui, T. (2008). Correlated Equilibrium and Concave Games. *International Journal of Game Theory* **37**, 1-13.

Vanderschraaf, P. (1995). Endogenous Correlated Equilibria in Noncooperative Games. *Theory and Decision* **38**, 61-84.

Yi, S.S. (1997). On the Existence of a Unique Correlated Equilibrium in Cournot Oligopoly. *Economics Letters* **54**, 235-239.

Young, H.P. (2004). *Strategic Learning and Its Limits.* Oxford University Press.