Costly Public Transfers in Repeated Cooperation under Imperfect Monitoring *

Mikhail Panov[†] Sergey Vorontsov[‡]

December 13, 2018

Abstract

We consider a two-player continuous-time repeated strategic interaction with imperfect monitoring of hidden production and the possibility that players can make public transfers between each other. Money transfers are costly: only a fraction $k < 1$ of the money sent is received by the recipient (the case $k = 0$ corresponds to pure money burning). We introduce the notion of self-enforcing public agreement which mimics the notion of pure-strategy public perfect equilibrium from the discrete time. For a fixed interest rate $r > 0$, we characterize the set of payoffs attainable in self-enforcing public agreements, as well as the dynamics in the efficient agreements. We show that adding the possibility of costly transfers increases the set of attainable payoffs, because it allows one to provide incentives to one player with less cost to the other player. We also show that costly transfers are used rarely and only after extreme histories when promised continuation payoffs hit players' individual rationality constraints.

*We are especially grateful to Michael Ostrovsky, Andrzej Skrzypacz, and Robert Wilson for the extensive guidance during the work on this project. We are also thankful to Dilip Abreu, Benjamin Brooks, Sebastian Di Tella, Elliot Lipnowski, Erik Madsen, Yuliy Sannikov, Takuo Sugaya, and Alexander Wolitzky for their thoughtful comments and suggestions. We also thank seminar participants at MIT/Harvard, New Economic School, Penn State, Stanford, University of New South Wales, University of Wisconsin Madison, University of Melbourne, University of Queensland, AEA 2017.

[†]New York University; panov@nyu.edu.

[‡]

1 Introduction.

We consider a two-player continuous-time repeated interaction with imperfect monitoring of hidden production and the possibility that players can make public transfers between each other. At each point in time, the players take hidden productive actions which are imperfectly observable through their effect on the drift of a public multidimensional Brownian signal. Besides hidden productive actions, the players are allowed to transfer money between each other. These transfers are instantaneously and commonly observable. The money transfers are costly: there is an exogenous retention parameter $k \in [0, 1)$. If at time t, a player sends to the opponent γ amount of money, the opponent immediately receives only $k\gamma$ with the remaining $(1 - k)\gamma$ being permanently lost. The limiting case $k = 1$ corresponds to perfect transfers. The motivation to study the case $k < 1$ is that in many important economic situations, perfect transfers may be infeasible (for example, in the context of cartels, they may be directly prohibited by the government). The case $k = 0$ corresponds to pure money burning. In the context of cartels, money burning can be implemented, for example, via open charity donations or any other expenditures, which are not beneficial directly to the stockholders of interacting firms. The intermediate case $k \in (0, 1)$ may be implemented in cartels, for example, when a firm or its subsidiary buys the final product from the competitor or its subsidiary (cf. Harrington and Skrzypacz (2007)).

In this continuous-time setting with both hidden and observable actions, we propose to study public agreements instead of extensive-form games. Our notion of *self-enforcing public agreements* then is designed to emulate pure-strategy public perfect equilibria (p-PPE) of discrete-time games.

Our main results are the following.

First, we provide the exact conditions on when a public agreement is self-enforcing. An agreement is self-enforcing if and only if it satisfies two separate conditions: One-Stage Deviation Principle in Hidden Actions and One-Stage Deviation Principle in Observable Actions. The One-Stage Deviation Principle in Hidden Actions condition is familiar in the literature. In fact, it is exactly the incentive compatibility condition from Sannikov (2007) and it does not contain any of the money-transfer part. The One-Stage Deviation Principle in Observable Actions condition is also quite familiar. It requires that essentially never, neither of the players will find it instantaneously profitable to publicly deviate in money transfers alone.

Second, we turn to the existence of optimal penal codes in our setting. The notion of optimal penal code was introduced by Abreu (1988). There, an optimal penal code is a tuple of p-SPNE's of a repeated game that deliver to each of the players their worst possible p-SPNE payoff. I.e., an optimal penal code implements the harshest possible subgame-perfect punishments for each of the players. For our second main results, we assume that minmaxing of each of the players may be *locally* enforced by the shift of promised continuation values (recall the notion of enforceability of an action profile from Fudenberg et al. (1994) and Sannikov (2007)). We show that if for each of the players, their stage-game minmaxing profile is enforceable (and under a few more technical restrictions), then there will exist a couple of self-enforcing public agreements that *globally* deliver the stage-game minmax payoffs to each of the players correspondingly. As any self-enforcing agreement must deliver to the players at least their stage-game minmax payoffs, these two agreements, indeed, implement the harshest punishments for the players.

Third, we provide the precise characterization of the set of payoffs attainable in self-enforcing public agreements. A pair of payoffs w is called individually rational if it lies above the players' minmax payoffs from the stage game. A subset S of the set of individually rational payoffs is called comprehensive if for any point $w \in S$, S also contains all individually rational payoffs that may be obtained from w by subtracting a positive linear combination of the money-transfer vectors $(1, -k)$ and $(-k, 1)$. For a subset of individually rational payoffs, the prefix ∂_+ denotes the part of the boundary of this subset which lies strictly above the minmax payoffs of both players. Finally, $\mathcal{N}$ denotes the convex hull of the p-NE payoffs of the stage game. Our third main result then states that for any $k \in [0, 1)$, any fixed interest rate $r > 0$, given that an optimal penal code exists, the set K of the payoffs attainable in self-enforcing agreements is precisely the largest bounded subset of the set of individually rational payoffs such that (1) K is comprehensive; (2) the boundary of K satisfies the optimality equation of Sannikov (2007) at any point $w \in \partial_+ K \backslash \mathcal{N}$; and (3) $\partial_+ K$ enters the minmax line of Player i, $i = 1, 2$, either at a p-NE payoff-pair of the stage game, or tangent to the corresponding money-transfer vector, $(1, -k)$ for Player 1 and $(-k, 1)$ for Player 2.

The rest of the paper is organized as follows. In Section 2, we briefly discuss our contributions to the existent literature. In Section 3, we introduce our model and give main definitions. In Section 4, we provide our main results. In Section 5, we discuss the dynamics in the efficient self-enforcing agreements.

2 Related Literature.

Our paper contributes to the existent literature in at least two different ways.

First, it continues the series of papers on repeated games with imperfect public monitoring started by Green and Porter (1984), Abreu et al. (1986), and Fudenberg et al. (1994). Our paper may be viewed as a continuation of Sannikov (2007) in which, besides the hidden productive actions, the players now also can use observable costly transfers. We show that the set K of payoffs attainable in self-enforcing public agreements in our setting subsumes the p-PPE payoff set from Sannikov (2007) and is typically larger. Also, unlike Sannikov (2007), the efficient payoffs in K can be achieved by outcomes of self-enforcing agreements such that, at least on the path of play, the promised continuation values always lie on the Pareto frontier of K. Thus, in our model, the issue on non-renegotiation proofness of optimal agreements may be less severe than in Sannikov (2007). Yet, our paper is related to the literature on relational contracts (such as Baker et al. (2002), Levin (2003), Rayo (2007)) and repeated games with perfect transfers (such as Fong and Surti (2009), Goldlucke and Kranz (2012),Goldlucke and Kranz (2013)). In fact, we may think that our paper bridges the gap between the papers on repeated games without transfers at all and repeated games with perfectly costless transfers by covering the intermediate case of costly transfers.

Second, our paper provides a methodological contribution towards modeling strategic interactions with observable actions in continuous time. The issues with treating observable actions in continuous time are discussed, for example, in Simon and Stinchcombe (1989) and Bergin and MacLeod (1993). Our main insight is that instead of trying to first define formally the whole extensive game and then searching for equilibria of the defined game, one can follow the approach similar to the one proposed in Abreu (1988): Define first the notion of an agreement for a given strategic situation as a collection of the initial outcome and all the outcomes specifying punishments after any finite sequence of observed deviations, then look for those agreements which are self-enforcing.

Finally, Hackbarth and Taub (2013) consider a continuous-time interaction with a hybrid of imperfectly observable productive actions and perfectly observable merger decisions by two colluding firms. The observability of the decision of when and how to merge introduces difficulties similar to those associated with observable transfers in our setting. However, to the best of our understanding, Hackbarth and Taub (2013) analyze the model in reduced form: they do not define deviating strategies for the players and they do not derive the corresponding incentive compatibility conditions. In contrast, in our paper, we focus precisely on these strategic aspects of the interaction.

3 The Model.

In this section, we introduce and discuss the main ingredients of our model.

3.1 Basic Setup.

Our model builds upon the model of two-player repeated games with imperfect monitoring in continuous time studied in Sannikov (2007).

Two players repeatedly interact in continuous time. At each point in time $t \in [0, \infty)$, Player i takes a productive action A_t^i from a finite set $\mathcal{A}^i$. These productive actions $A_t = (A_t^1, A_t^2)$ are imperfectly observable by their effect on the evolution of a d-dimensional public signal process X_t,

$$X_t = \int_0^t \mu(A_s)\, ds + Z_t,$$

where Z_t is a d-dimensional Brownian motion and $\mu : \mathcal{A}^1 \times \mathcal{A}^2 \to \mathbb{R}^d$ is a drift function. The arrival of public information is captured by an exogenously given filtration $\{\mathcal{F}_t\}$.

The new feature in our model is that besides the possibility of taking imperfectly observable productive actions, the players possess an exogenously given technology that allows them to publicly and verifiably transfer money between each other. Specifically, there is an exogenously given retention parameter $k \in [0, 1)$ characterizing how efficient these transfers can be. If at time t, Player i transfers amount $d\Gamma_t^i$ to the opponent, then Player $-i$ receives only $k \cdot d\Gamma_t^i$, with the remaining $(1-k) \cdot d\Gamma_t^i$ being permanently lost. Denote through Γ_t^i the cumulative process of transfers sent by Player i until time t.

Suppose that during the play of this interaction, the players take the profile of unobservable actions $(A_t^1, A_t^2)_{\{t \geqslant 0\}}$ and the profile of cumulative public transfers $(\Gamma_t^1, \Gamma_t^2)_{\{t \geqslant 0\}}$. In what follows, we will always restrict attention to such profiles that $(A_t^1, A_t^2)_{\{t \geqslant 0\}}$ are progressively measurable and $(\Gamma_t^1, \Gamma_t^2)_{\{t \geqslant 0\}}$ are weakly increasing and nonnegative RCLL-processes with respect to $\{\mathcal{F}_t\}$. Player i's random total discounted payoff under the play of this profile is

$$r \int_0^\infty e^{-rt} \big(c_i(A_t^i) dt + b_i(A_t^i) dX_t - d\Gamma_t^i + k d\Gamma_t^{-i} \big) - r\Gamma_0^i + rk\Gamma_0^{-i},$$

for some functions $c_i : \mathcal{A}^i \to \mathbb{R}$ and $b_i : \mathcal{A}^i \to \mathbb{R}^d$, where $r > 0$ denotes the common discount rate of the two players.

Denote

$$g_i(A_t) = c_i(A_t^i) + b_i(A_t^i)\mu(A_t)$$

Player i's continuation payoff expected at time t given the continuation profile (outcome) $(A, \Gamma)_{\{s \geqslant t\}} = (A_s^1, A_s^2, \Gamma_s^1, \Gamma_s^2)_{\{s \geqslant t\}}$ then can be written as

$$W_t^i(A, \Gamma) = E_t \Big[r \int_t^\infty e^{-r(s-t)} \big(g_i(A_s) - d\Gamma_s^i + k d\Gamma_s^{-i} \big) - r\Delta\Gamma_t^i + rk\Delta\Gamma_t^{-i} | A_s, s \geqslant t \Big],$$

where $\Delta\Gamma_t = \Gamma_t - \Gamma_{t-}$ if $t > 0$ and $\Delta\Gamma_0 = \Gamma_0$.

3.2 The Main Idea: the Abreu Approach.

Given an infinitely repeated discounted game with imperfect monitoring, one can be interested in finding the answers to the following questions:

> What are the pairs of expected payoffs attainable for the players in pure-strategy public perfect equilibria (p-PPE's)? What are the on-path dynamics in p-PPE's delivering the most efficient pairs of expected payoffs?

There could be at least two different approaches to finding payoffs and dynamics supportable in p-PPE's of *discrete-time* games. The first approach, frequently used in the literature, may be summarized as follows:

1. define first the full extensive-form game corresponding to a given repeated interaction: specify all the strategy profiles, the outcomes they induce, and the payoffs these outcomes deliver to the players;

2. find all the p-PPE's of the constructed game.

The alternative approach for finding p-PPE's outcomes may be similar to the one employed in Abreu (1988). We shall therefore refer to it as **the Abreu approach.** It can be summarized as follows:

1. View a potential p-PPE as the collection of recommended outcomes: the initial recommended outcome-path proposed by the p-PPE, as well as the collection of all those punishment continuation outcome-paths which follow any *finite* sequence of *observed* deviations from the previously recommended paths of play.

2. Specify the expected payoffs promised to the players by each of the outcomes from the collection assuming that nobody ever deviates from this outcome in the future;

3. Given the collection of recommended outcomes, a strategy for a player is defined as the rule specifying the sequence of *hidden actions* along the path of play of each of these outcomes, as well as the decisions of when and how to *observably* deviate from each of these outcomes;

4. define p-PPE as a self-enforcing collection of outcomes: after any finite history for each player, there is no continuation strategy that would yield them the expected payoff higher than the one promised by the current outcome.

Notably, for a given extensive game in discrete time, the set of p-PPE outcomes that can be found using the first approach is the same as the outcomes that can be found using the Abreu approach. In this paper, we are interested in modeling equilibrium outcomes for continuous-time strategic interactions with *observable* actions. It is well documented (Simon and Stinchcombe (1989), Bergin and MacLeod (1993)) that following the first approach for such strategic interactions gives rise to many difficulties (for instance, a strategy profile may not even define uniquely the associated outcome in the continuous-time game!). For this reason, we propose to follow the Abreu approach in this paper. Specifically, our plan will be:

> `Our Plan:` Instead of modeling the strategic interaction as a continuous-time game and then characterizing its p-PPE's, we use the Abreu approach. We first define the set of all public agreements, i.e., all of the collections of proposed initial and punishment outcomes following any finite sequence of observable deviations. We then characterize expected payoffs attainable for the players in self-enforcing public agreements, as well as the on-path dynamics in the Pareto-efficient self-enforcing public agreements.

3.3 Outcomes.

In this subsection, we describe outcomes, from which we eventually will build public agreements.

Within an agreement, an outcome Q is the description of the recommended path of play starting immediately after the last observed deviation from the previously effective outcome. In particular, Q contains a filtered probability space $(\Omega^Q, \mathcal{F}^Q, \{\mathcal{F}_t^Q\}_{t\geqslant 0}, \mathbf{P}^Q)$ capturing the arrival of public information after the last occurred public deviation. This information includes the evolution of the d-dimensional public signal X_t^Q and possibly the realizations of independent public randomizations. Further, Q specifies the profile of recommended hidden actions $\left(A^{1,Q}, A^{2,Q}\right)$ progressively measurable with respect to $\{\mathcal{F}_t^Q\}_{t\geqslant 0}$, and the recommended cumulative money-transfer processes

$(\Gamma^{1,Q}, \Gamma^{2,Q})$, which are weakly increasing nonnegative RCLL-processes adapted to $\{\mathcal{F}_t^Q\}_{t \geqslant 0}$. The measure $\mathbf{P}^Q$ agrees with the profile of recommended hidden actions in such a way that the process $X_t^Q - \int_0^t \mu(A_s^Q)\, ds$ is a standard d-dimensional Brownian motion under $\mathbf{P}^Q$.

More formally, the public information for outcome Q is constructed in the following way:

Definition (Public Information). *For an outcome Q, the public information $\mathcal{P}^Q$ is a filtered probability space $(\Omega^Q, \mathcal{F}^Q, \{\mathcal{F}_t^Q\}_{t \geqslant 0}, \boldsymbol{P}^Q)$, which is constructed as follows:*

1. *Take a filtered probability space $\mathcal{P}^0 = (\Omega^0, \mathcal{F}^0, \{\mathcal{F}_t^0\}_{t \geqslant 0}, \boldsymbol{P}^0)$ to be used for public randomization (take this space reach enough so that $\mathcal{F}^0$ includes realization of a random variable distributed $U[0,1]$).*

2. *Take a standard d-dimensional Brownian motion X_t on a filtered probability space $\mathcal{P}^X$.*

3. *Take the direct product $\mathcal{P} = (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \boldsymbol{P})$ of the above filtered probability spaces:*

$$\mathcal{P} = \mathcal{P}^0 \otimes \mathcal{P}^X.$$

4. *Set $\Omega^Q = \Omega$.*

5. *Take the profile $\left(A^{1,Q}, A^{2,Q}\right)$ of the recommended hidden actions in the outcome Q (which can be any progressively measurable process of hidden actions on $\mathcal{P}$).*

6. *Using Girsanov's theorem, construct measure $\boldsymbol{P}^Q$ on $(\Omega^Q, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0})$ so that $X_t^Q - \int_0^t \mu(A_s^Q)\, ds$ is a d-dimensional Brownian motion under $\boldsymbol{P}^Q$.*

7. *Finally, define $(\mathcal{F}^Q, \{\mathcal{F}_t^Q\}_{t \geqslant 0})$ as a right-continuous augmentation of $(\mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0})$ under $\boldsymbol{P}^Q$.*

We will say that public information $(\Omega^Q, \mathcal{F}^Q, \{\mathcal{F}_t^Q\}_{t \geqslant 0}, \mathbf{P}^Q)$ *agrees* with the profile of $\left(A^{1,Q}, A^{2,Q}\right)$ of hidden actions whenever this information is constructed using $\left(A^{1,Q}, A^{2,Q}\right)$.

Besides the recommended hidden actions, an outcome Q must also specify recommended money-transfer processes $(\Gamma^{1,Q}, \Gamma^{2,Q})$. We restrict $(\Gamma^{1,Q}, \Gamma^{2,Q})$ to be weakly increasing nonnegative RCCL-processes on $(\Omega^Q, \mathcal{F}^Q, \{\mathcal{F}_t^Q\}_{t \geqslant 0}, \mathbf{P}^Q)$. Further, we require that processes $(\Gamma^{1,Q}, \Gamma^{2,Q})$ be *M-nonmanipulable* for some $M > 0$ as defined below:

Definition (Well Bounded Process). *Given a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \boldsymbol{P})$, a weakly increasing nonnegative adapted RCLL-process Γ_t is said to be well bounded by $M > 0$ if for any finite $\{\mathcal{F}_t\}_{t \geqslant 0}$-stopping time T,*

$$E^{\boldsymbol{P}}\Big[\int_T^\infty e^{-r(s-T)} d\Gamma_s + \Delta\Gamma_T \Big| \mathcal{F}_T\Big] \leqslant M \quad \boldsymbol{P}\text{-a.s.}$$

Definition (M-Nonmanipulable Processes). *Given a public information* $(\Omega^Q, \mathcal{F}^Q, \{\mathcal{F}_t^Q\}_{t\geqslant 0}, \boldsymbol{P}^Q)$ *that agrees with the profile* $\left(A^{1,Q}, A^{2,Q}\right)$ *of hidden actions, a weakly increasing nonnegative adapted RCLL-process* Γ_t *is said to be* M*-nonmanipulable for some* $M > 0$ *if for each Player* i*,* $i = 1, 2$ *and for any progressively measurable process* $\tilde{A}^i$ *of hidden actions for Player* i*, the process* Γ_t *is well bounded by* M *under the measure* $\boldsymbol{P}(\tilde{A}^i, A^{-i,Q})$ *which is obtained from* $\boldsymbol{P}^Q$ *by changing* $A^{i,Q}$ *to* $\tilde{A}^i$*.*

M-nonmanipulability of Γ^{-i}, a money-transfer process for Player $-i$ recommended by outcome Q, guarantees that Player i would not be able to "jam" the public signal by changing their hidden actions in such a way so as to make the opponent to transfer them in expectation infinite amount of money.

We are now ready to introduce the formal definition of a public outcome, which will be the main building block in our construction of public agreements in the next subsection.

Definition (Outcome). *A public outcome* $Q = \{\mathcal{P}^Q, A^Q, \Gamma^Q\}$ *is a public information* $\mathcal{P}^Q$ *together with recommended processes of hidden actions* $\left(A^{1,Q}, A^{2,Q}\right)$ *and cumulative money transfers* $(\Gamma^{1,Q}, \Gamma^{2,Q})$ *such that*

1. $\left(A^{1,Q}, A^{2,Q}\right)$ *are progressively measurable and agree with* $\mathcal{P}^Q$*;*
2. $(\Gamma^{1,Q}, \Gamma^{2,Q})$ *are weakly increasing nonnegative adapted RCLL-processes which are* M*-nonnmanipulable for some* $M > 0$*;*

Note that we assume that whenever a certain outcome becomes effective during the play, we completely restart the clock: we set the current time back to $t = 0$ and we restart the public information anew.

3.4 Public Agreements.

Having introduced the concept of outcome in the previous subsection, we are ready to define public agreements, one of the main concepts in our paper. A public agreement is a collection of recommended public outcomes. It proposes to start with the initial outcome Q^0. It also specifies punishment outcomes suggesting the continuation play after any *finite* sequence of observed deviations. Below we introduce the important inertia restriction. Intuitively, inertia is the restriction on how frequently the players are allowed to *publicly* deviate from outcomes in an agreement. The inertia restriction will guarantee us that after essentially any finite history during the play of an agreement, there will be only finitely many observed deviations. In particular, this means that agreements will be well defined: an agreement will be recommending a well defined continuation play after essentially any finite history possible under the play of the agreement.

Within each outcome of an agreement, we restrict that players are only allowed to publicly deviate at times when they are prescribed to send positive transfers to the opponent, at permissible times of public deviations:

Definition (Permissible Time of Public Deviation). *Given an outcome $Q = \{\mathcal{P}^Q, A^Q, \Gamma^Q\}$, we say that an $\{\mathcal{F}_t^Q\}_{t\geqslant 0}$-stopping time T is a permissible time of public deviation for Player i if Player i is supposed to send positive amount of money at T. That is $T < \infty$ implies that $\Gamma_T^{i,Q}$ is increasing at T, or that $\Gamma_0^{i,Q} > 0$ and $T = 0$.*

An agreement contains the initial outcome and outcomes specifying punishments after observed public deviations. Fix small $\epsilon > 0$, the parameter of inertia. The following is the restriction on the punishment outcomes that can be employed in an agreement with the inertia parameter ϵ:

Inertia Restriction. *If $Q = \{\mathcal{P}^Q, A^Q, \Gamma^Q\}$ is a punishment outcome of an agreement with the inertia parameter $\epsilon > 0$, then Q must specify that in the beginning, no player sends positive transfers at least until the first time the public signal moves by ϵ or until ϵ amount of time elapses:*

$$\Gamma_{\tau-}^Q = (0,0), \text{ where } \tau = \min\{t : |X_t^Q| = \epsilon\} \wedge \epsilon.$$

We are now ready to provide the formal construction of a public agreement:

Definition (Public Agreement). *A public agreement $\mathcal{E}$ with the inertia parameter $\epsilon > 0$ is a collection of public outcomes which is constructed in the following steps:*

1. *$\mathcal{E}$ specifies the initial outcome Q^0;*
2. *given Q^0, the agreement $\mathcal{E}$ specifies all the punishment outcomes of level-1, the punishment outcomes after the first observed deviation by Player 1 or Player 2 for all permissible times of public deviations for these players in Q^0;*
3. *for each punishment outcome of level-1, Q^1, the agreement $\mathcal{E}$ specifies all the punishment outcomes of level-2 following Q^1, the punishment outcomes after the second observed deviation byd Player 1 or Player 2 for all permissible times of public deviations for these players in Q^1;*
4. *for each punishment outcome of level-2, Q^2, the agreement $\mathcal{E}$ specifies all the punishment outcomes of level-3 following Q^3, the punishment outcomes after the third observed deviation by Player 1 or Player 2 for all permissible times of public deviations for these players in Q^2;*
5. *and so on...*

Additionally, we require that there exists a uniform bound $M > 0$ such that for all of the outcomes in the agreement $\mathcal{E}$, the recommended money-transfer processes are M-nonmanipulable. The public informations from the outcomes in $\mathcal{E}$ are assumed to be independent of each other.

Pure public strategies for the players are defined only against a given public agreement in the following way:

Definition (Pure Public Strategy). *Given a public agreement $\mathcal{E}$ with the inertia parameter $\epsilon > 0$, a pure public strategy σ for Player i is a collection of separate rules σ^Q prescribing Player i's behavior in each of the outcomes Q from $\mathcal{E}$. Each σ^Q consist of:*

1. $A^{i,Q,\sigma}$*, a process of hidden actions for Player* i *progressively measurable with respect to the public filtration* $\{\mathcal{F}_t^Q\}_{t\geqslant 0}$*;*

2. An $\{\mathcal{F}_t^Q\}_{t\geqslant 0}$*-stopping time* $T^{i,Q,\sigma}$ *prescribing the moment when Player* i *announces the public deviation from* Q*. The stopping time* $T^{i,Q,\sigma}$ *is restricted to be a permissible time of public deviation for Player* i*.*

$S^i(\mathcal{E})$ *denotes the set of all pure public strategies for Player* i *given an agreement* $\mathcal{E}$*.*

During the play of an agreement, there is always exactly one currently effective outcome recommending the continuation play to the players. The outcome remains effective until the first time T at which either of the players (possibly both of them) publicly deviates. A public deviation at time $T(\omega)$ causes instantaneous hold on money transfers, i.e., sets $\Delta\Gamma_T^1(\omega) = \Delta\Gamma_T^2(\omega) = 0$. Also, the deviation makes the continuation play switch to the new effective outcome, the corresponding punishment outcome prescribed in the agreement.

Suppose that given an agreement $\mathcal{E}$ with the inertia parameter ϵ, the players decide to play a profile of pure public strategies (σ^1, σ^2). Notice that because of the inertia restriction and the fact that public deviations are only permissible at times when the deviating player is supposed to be sending positive amount of money, for any finite time $t > 0$, "with probability 1," there will be only finitely many public deviations observed by time t. Indeed, if there is a finite history before time t, such that the players have deviated infinitely many times, then infinitely many times along this history, the public deviations became possible by an ϵ-jump of the currently effective public signals X^Q. But there exist $c > 0$, such that for any outcome Q and any hidden action profile of the players, ϵ-jump of X^Q before time ϵ happens with probability less than $1 - c$. As we treat public signals independently for different outcomes, the probability that infinitely many such jumps happened before time t then is at most $(1-c)^\infty = 0$. Therefore, $\mathcal{E}$, indeed, correctly determines the proposed continuation play for almost all finite public histories arising from the play of any pure public strategy profile.

Finally, the inertia restriction together with the restriction on when public deviations are permitted implies that we only focus our attention on agreements in which:

1. any deviations in money transfers are ignored at times when on path, the deviating player is supposed to send 0 amount of money;

2. the only deviations in money transfers that are considered for the players are to deviate and send 0 amount of money;

3. once a player decides to deviate and send 0 amount of money, they do so immediately and for at least some discrete amount of time.

In discrete-time analogue of our model, restrictions 1 and 2 can be shown to be without loss of generality (cf. Abreu (1988)) and restriction 3 is built in the discreteness of the periods itself. In our continuos-time setting, we instead impose these restrictions directly.

3.5 Promised Continuation Values.

Having introduced the concept of public agreement in the previous subsection, we can now specify the continuation values promised under the play of a public agreement. As usual, these continuation values are computed assuming that nobody further deviates from the currently proposed outcome.

Suppose that the players are playing against an agreement $\mathcal{E}$ and after some history of play find themselves in the situation when some outcome Q is effective (either initial, or punishment). Within this outcome Q, we can define the process of promised continuation values as the discounted sum of future stage-game payoffs and net money transfers evaluated at time $t \geqslant 0$ similarly to how it is done in Sannikov (2007). Specifically, at time t after the start of the outcome Q, Player i's promised continuation value under Q is

$$W_t^{i,Q} = E_t^{\mathbf{P}^Q}\Big[r\int_t^{\infty} e^{-r(s-t)}\big(g_i(A_s^Q)ds - d\Gamma_s^{i,Q} + k\,d\Gamma_s^{-i,Q}\big) - r\Delta\Gamma_t^{i,Q} + rk\Delta\Gamma_t^{-i,Q}\Big|\mathcal{F}_t^Q\Big].$$

Note that the boundedness of the stage game payoffs together with the the well boundedness of the money-transfer processes ensures that we can always find a bounded modification for the process $W_t^{i,Q}$. Note that $W_t^{i,Q}$ is a random variable for each t and we shall not attach any game-theoretic meaning to it, we will only use $W_t^{i,Q}$ throughout our derivations. The only continuation value to which we actually attach a game-theoretic meaning and do interpret it as the value from the outcome as assessed by the player, will be $W^{i,Q}$, the unconditional expectation of $W_0^{i,Q}$ computed at the very beginning of Q:

$$W^{i,Q} := E^{\mathbf{P}^Q}\Big[W_0^{i,Q}\Big].$$

Given an agreement $\mathcal{E}$ with the initial outcome Q^0, define *the expected payoff* $W^{i,\mathcal{E}}$ *promised by* $\mathcal{E}$ to Player i as

$$W^{i,\mathcal{E}} := W^{i,Q_0}.$$

The following is the straightforward adaptation of Proposition 1 from Sannikov (2007) to our setting:

Proposition 1. *(Representation and Promise Keeping) A bounded stochastic process W_t^i is the process of promised continuation values $W_t^{i,Q}$ of Player i under an outcome Q if and only if there exist processes $\beta^{i,Q} = (\beta^{i1,Q}, ..., \beta^{id,Q})$ in $\mathcal{L}^*$ of the space $\mathcal{P}^Q$ and a martingale $\tilde{\epsilon}^{i,Q}$ on $\mathcal{P}^Q$ orthogonal to X^Q with $\tilde{\epsilon}_0^{i,Q} = 0$ such that for all $t > 0$, W_t^i satisfies*

$$W_t^i = W_0^i + r\int_0^t \big(W_s^i - g_i(A_s^Q)\big)ds + r\Big(\Gamma_0^{i,Q} + \int_0^t d\Gamma_s^{i,Q}\Big) - r\Big(k\Gamma_0^{-i,Q} + \int_0^t kd\Gamma_s^{-i,Q}\Big) +$$

$$+ r\int_0^t \beta_s^{i,Q}\big(dX_s^Q - \mu(A_s^Q)ds\big) + \tilde{\epsilon}_t^{i,Q}. \quad (1)$$

Proof. The proof is essentially the same as the proof of Proposition 1 from Sannikov (2007). □

The shorthand form for representation (1) is

$$dW_t^{i,Q} = r\big(W_t^{i,Q} - g_i(A_t^Q)\big)dt + rd\Gamma_t^{i,Q} - rkd\Gamma_t^{-i,Q} + \beta_t^{i,Q}\big(dX_t^Q - \mu(A_t^Q)dt\big) + d\tilde{\epsilon}_t^{i,Q}. \quad (2)$$

Comparing to Sannikov (2007), the new terms in equation (2) are $rd\Gamma_t^{i,Q}$ and $(-rkd\Gamma_t^{-i,Q})$. Intuitively, if at time t, a player is supposed to send G dollars to the opponent, then their promised continuation value at the very next moment must go up by rG so as to keep them equally happy. At the same time, the opponent's continuation value must go down by rkG so as to reflect the receipt of the transfer.The new terms capture precisely this intuition.

3.6 The Value of a Strategy.

Our next task is to define the value of a strategy for a player. Suppose the players are playing under some agreement $\mathcal{E}$. Take Player i and a pure public strategy σ for them. What can be the value from following σ evaluated *at the beginning* of some outcome Q from $\mathcal{E}$?

Suppose that σ does not prescribe any public deviation from Q, i.e., the stopping time of the deviation $T^{i,Q} = +\infty$ everywhere. Naturally, we can assign the continuation value for σ at the beginning of Q by computing the sum of expected discounted payoffs along Q:

$$V(\sigma, Q) := E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big[r\int_0^\infty e^{-rs}\big(g_i(A_s^{i,Q,\sigma}, A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\, d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big],$$

where $\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})$ is the measure induced in $\mathcal{P}^Q$ if the players take the profile of hidden actions $(A^{i,Q,\sigma}, A^{-i,Q})$.

Suppose now that starting at Q, σ prescribes at most one further public deviation. Denote by $\tilde{Q}\big(T, \omega\big)$ the punishment outcome specified by the agreement if Player i publicly deviates from Q in state ω at time T. Naturally, we can assign the continuation value of σ after this deviation as $V\big(\sigma, \tilde{Q}(T, \omega)\big)$. What about the value of σ evaluated at the beginning of Q? Naively, we might want to write it down as

$$V(\sigma, Q) = E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big[r \int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma}, A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\, d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] +$$
$$+ E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big[e^{-rT^{i,Q}} \Big(V\big(\sigma, \tilde{Q}(T^{i,Q}, \omega)\big) + r\Delta\Gamma_{T^{i,Q}}^{i,Q} - rk\Delta\Gamma_{T^{i,Q}}^{-i,Q}\Big)\Big]. \quad (3)$$

Unfortunately, the second term in the above expression generally is not well defined because $V\big(\sigma, \tilde{Q}(T^{i,Q}, \omega)\big)$ is not necessary a random variable! Instead of assigning the precise value to $V(\sigma, Q)$, let us assign the upper bound for this value, $V^*(\sigma, Q)$, and the lower bound for this value, $V_*(\sigma, Q)$, by using correspondingly the upper and lower integrals for the second term in (3). Formally,

$$V^*(\sigma, Q) := E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big[r \int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma}, A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\, d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] +$$
$$+ \Big(E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big)^*\Big[e^{-rT^{i,Q}} \Big(V\big(\sigma, \tilde{Q}(T^{i,Q}, \omega)\big) + r\Delta\Gamma_{T^{i,Q}}^{i,Q} - rk\Delta\Gamma_{T^{i,Q}}^{-i,Q}\Big)\Big]$$

and

$$V_*(\sigma, Q) := E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big[r \int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma}, A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\, d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] +$$
$$+ \Big(E^{\mathbf{P}(A^{i,Q,\sigma}, A^{-i,Q})}\Big)_*\Big[e^{-rT^{i,Q}} \Big(V\big(\sigma, \tilde{Q}(T^{i,Q}, \omega)\big) + r\Delta\Gamma_{T^{i,Q}}^{i,Q} - rk\Delta\Gamma_{T^{i,Q}}^{-i,Q}\Big)\Big],$$

where $\Big(E^{\mathbf{P}}\Big)^*$ and $\Big(E^{\mathbf{P}}\Big)_*$ denote the upper and the lower integrals. [1]

Next, for any strategy σ prescribing only finitely many observable deviations, we can define the upper and lower bounds on its value recursively as

[1] For any function $f : \Omega \to \mathbb{R}$, define
$\Big(E^{\mathbf{P}}\Big)^*(f) := \inf_{\substack{g \text{ is a random variable} \\ \forall \omega \in \Omega, g(\omega) \geqslant f(\omega)}} E^{\mathbf{P}}(g)$ and $\Big(E^{\mathbf{P}}\Big)_*(f) := \sup_{\substack{g \text{ is a random variable} \\ \forall \omega \in \Omega, g(\omega) \leqslant f(\omega)}} E^{\mathbf{P}}(g)$.

Naturally, $\Big(E^{\mathbf{P}}\Big)^*(f) \geqslant \Big(E^{\mathbf{P}}\Big)_*(f)$. Also, $\Big(E^{\mathbf{P}}\Big)^*(f) = \Big(E^{\mathbf{P}}\Big)_*(f) \in (-\infty, +\infty)$ if and only if f is measurable and integrable, in which case $E^{\mathbf{P}}(f) = \Big(E^{\mathbf{P}}\Big)^*(f) = \Big(E^{\mathbf{P}}\Big)_*(f)$.

$$V^*(\sigma,Q) := E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big[r\int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma},A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\,d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] +$$
$$+\Big(E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big)^*\Big[e^{-rT^{i,Q}}\ \Big(V^*\big(\sigma,\tilde{Q}(T^{i,Q},\omega)\big) + r\Delta\Gamma_{T^{i,Q}}^{i,Q} - rk\Delta\Gamma_{T^{i,Q}}^{-i,Q}\Big)\Big]$$

and

$$V_*(\sigma,Q) := E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big[r\int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma},A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\,d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] +$$
$$+\Big(E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big)_*\Big[e^{-rT^{i,Q}}\ \Big(V_*\big(\sigma,\tilde{Q}(T^{i,Q},\omega)\big) + r\Delta\Gamma_{T^{i,Q}}^{i,Q} - rk\Delta\Gamma_{T^{i,Q}}^{-i,Q}\Big)\Big].$$

Finally, for a strategy prescribing arbitrary many observable deviations, define the upper and lower bounds on its value as

$$V^*(\sigma,Q) := \limsup_{N\to\infty} V^*(\sigma_N,Q),$$

$$V_*(\sigma,Q) := \liminf_{N\to\infty} V_*(\sigma_N,Q),$$

where σ_N is the N-th truncation of σ. I.e., σ_N coincides with σ until the N-th public deviation by Player i, and it follows the actions recommended by the agreement ever after.

This last step is a crucial one and needs to be justified! Indeed, as $N\to\infty$, because of the inertia restriction on how frequently the players can deviate, the strategies σ and σ^N become different either in the event with vanishingly small probability, or after the time horizon that is going to ∞. As the payoffs in the stage game are bounded and all the money-transfer processes are uniformly M-nonmanipulable for some $M>0$, this difference can be effectively ignored in the limit. For a similar reason, lim sup and lim inf in the above definitions can be replaced by the usual limits.

4 Main Results.

In this section, we show and discuss the three main results of our paper: the characterization of self-enforcing agreements through the one-stage deviation principle; the existence of optimal penal codes; and the characterization of the set of payoffs attainable in self-enforcing agreements.

4.1 Self-Enforcing Public Agreements.

The main concept in our paper is that of *self-enforcing public agreement.* It is formally defined as follows:

Definition (Self-Enforcing Public Agreement). *A public agreement $\mathcal{E}$ is called self-enforcing if for each of its outcomes $Q \in \mathcal{E}$, no player can find a pure public strategy with the upper bound on the value higher than the promised continuation value as evaluated at the beginning of Q,*

$$\forall Q \in \mathcal{E}, \quad \forall i = 1, 2 \quad, \forall \sigma \in S^i(\mathcal{E}), \quad V^*(\sigma, Q) \leqslant W^{i,Q}.$$

The following measurability restriction imposes a technical condition on selecting different punishment outcomes in public agreements:

Definition (Measurable Public Agreement). *A public agreement $\mathcal{E}$ is called measurable if for any outcome $Q \in \mathcal{E}$, any Player i, and any permissible time of public deviation T for Player i , the promised continuation value $W^{i,\tilde{Q}(T)}$ from the resulting punishment is an $\mathcal{F}_T^Q$-random variable.*

Recall the representation (1) of promised continuation values given in Proposition 1. The following theorem is the first main result in our paper.

Theorem 1 (One-Stage Deviation Principle). *Let $\mathcal{E}$ be a public agreement. Consider the following restrictions:*

1. *(One-Stage Deviation in Hidden Actions)*

   For each outcome $Q \in \mathcal{E}$ and for any $T \in (0, \infty)$, the inequalities

   $$\forall i = 1, 2, \quad \forall a_i' \in \mathcal{A}^i, \quad g_i(A_t^Q) + \beta_t^i \mu(A_t^Q) \geqslant g_i(a_i', A_t^{-i,Q}) + \beta_t^i \mu(a_i', A_t^{-i,Q})$$

   are satisfied $\boldsymbol{P}^Q \otimes \lambda[0,T]$-almost surely on $\Omega^Q \times [0,T]$, where $\lambda[0,T]$ is the standard Lebesgue measure on $[0,T]$.

2. *(One-Stage Deviation in Observable Actions)*

   For each outcome $Q \in \mathcal{E}$ and for any $\{\mathcal{F}_t^Q\}_{t \geqslant 0}$-stopping time T that is a permissible time of pubic deviation for Player i, the instantaneous gains from disrupting the money transfers and going to the punishment outcome $\tilde{Q}(T, \omega)$ are nonpositive $\boldsymbol{P}^Q$-almost surely,

   $$W_T^{i,Q} \geqslant W^{i,\tilde{Q}(T,\omega)} + r\Delta\Gamma_T^{i,Q} - rk\Delta\Gamma_T^{-i,Q} \qquad \boldsymbol{P}^Q\text{-a.s.}$$

Then:

- *(Sufficiency) If $\mathcal{E}$ satisfies restrictions 1 and 2, it is self-enforcing.*
- *(Necessity of 1) If $\mathcal{E}$ does not satisfy restriction 1, it is not self-enforcing. Moreover, there exists an outcome $Q \in \mathcal{E}$ and a strategy σ for some Player i such that*

  $$V^*(\sigma, Q) = V_*(\sigma, Q) > W^{i,Q}.$$

- *(Necessity of 2) If $\mathcal{E}$ does not satisfy restriction 2, it is not self-enforcing. Moreover, if $\mathcal{E}$ is measurtable, then there exists an outcome $Q \in \mathcal{E}$ and a strategy σ for some Player i such that*

$$V^*(\sigma, Q) = V_*(\sigma, Q) > W^{i,Q}.$$

Proof. See Appendix A. □

Theorem 1 provides the precise characterization of the necessary and sufficient conditions for a public agreement to be self-enforcing: a public agreement is self-enforcing if and only if it satisfies the One-Stage Deviation Principle in both hidden and observable actions.

Recall that in our definition, we required that for an agreement to be self-enforcing, there should be no deviating strategy for either of the players with the upper bound on the value, rather than the exact expected value, higher than the continuation value promised by the agreement. This may seem too restrictive. Fortunately, Theorem 1 also establishes that for measurable agreements, this is without loss: if a measurable agreement is not self-enforcing, then there exists a deviating strategy for some player that is a strictly profitable deviation in the sense of usual expected values.

4.2 Optimal Penal Codes.

In this subsection, we turn to the problem of constructing punishments in self-enforcing agreements. Abreu (1988) proves the existence of optimal penal codes for his discrete-time setting. There, an optimal penal code is a pair of punishment outcomes Q^1 and Q^2 which punish observable deviations by Player 1 and Player 2 correspondingly such that using them alone, one can construct two p-SPNE's, $\mathcal{E}^1$ and $\mathcal{E}^2$, delivering the worst possible p-SPNE payoffs to Player 1 and to Player 2 correspondingly. In this subsection, we prove an analogous result for self-enforcing public agreements in our continuous-time setting under some additional restrictions.

Denote by $K(\epsilon)$ the set of payoffs attainable in self-enforcing public agreements with the inertia parameter ϵ. The next lemma shows that the sets $K(\epsilon)$ are decreasing in ϵ.

Lemma 1 (Monotonicity)**.** *For any $\epsilon_1 > \epsilon_2 > 0$,*

$$K(\epsilon_1) \subseteq K(\epsilon_2).$$

Proof. See Appendix B.1. □

Consider the stage game G in hidden actions played by the players in our setting. The set of players is $N = \{1, 2\}$, the set of actions for Player i is $\mathcal{A}_i$, the payoff functions are g_i:

$$G = \big\{N, (\mathcal{A}_i)_{i \in N}, (g_i)_{i \in N}\big\}.$$

Denote by $\underline{v}_i$ the pure-strategy minmax payoff of Player i in G:

$$\underline{v}_i = \min_{a_{-i} \in \mathcal{A}_{-i}} \max_{a_i \in \mathcal{A}_i} g_i(a_i, a_{-i}).$$

A corresponding profile of pure actions delivering Player i their mixmax payoff is called a profile minmaxing Player i. The minmax line for Player i is the straight line in the space of players' payoffs $(w_1, w_2) \in \mathbb{R}^2$ which is given by the equation $w_i = \underline{v}_i$. The pure-strategy minmax payoff of Player i can be interpreted as Player i's individual rationality constraint against any pure action of the opponent in the stage game. For our repeated setting, we still can interpret it as Player i's individual rationality constraint, the per-period average expected payoff that can be guaranteed by Player i against any process of pure hidden actions and money transfers of the opponent. To guarantee this payoff, a player should simply never transfer any money to the opponent and always keep playing the myopic best-response hidden action from the stage game against the current hidden action of the opponent. The following lemma establishes that any self-enforcing agreement must deliver to both players individually rational payoffs.

Lemma 2 (Individual Rationality)**.** *For any $\epsilon > 0$, any self-enforcing agreement $\mathcal{E}$ with the inertia parameter ϵ delivers to each Player i the expected payoff at least as high as their minmax payoff in the stage game G,*

$$W^{i,\mathcal{E}} \geqslant \underline{v}_i.$$

Moreover, if Q is an outcome in a self-enforcing agreement, then for any stopping time τ, not necessary a time of permissible public deviation, and for any Player i,

$$W_\tau^{i,Q} \geqslant \underline{v}_i + r\Delta\Gamma_\tau^{i,Q} - rk\Delta\Gamma_\tau^{-i,Q} \qquad \boldsymbol{P}^Q\text{-a.s.}$$

Proof. See Appendix B.2. □

Lemma 2 establishes that the worst payoffs deliverable to the players in self-enforcing agreements must be at least their static minmax payoffs. One can ask whether there exist the worst self-enforcing agreements for each of the players that deliver them precisely this lower bound. It turns out that we can answer this in the affirmative provided several additional assumptions are satisfied. Specifically, in the remainder of the paper, we assume that the following condition is satisfied:

Condition C. *There exists $\epsilon_0 > 0$ and $(w_1, w_2) \in K(\epsilon_0)$ such that $w_1 > \underline{v}_1$ and $w_2 > \underline{v}_2$.*

Condition C is guaranteed to be satisfied if there is a p-NE of the stage game with higher than minmax payoffs, or if the set of p-PPE's from Sannikov (2007) has nonempty interior.

Recall the definitions of enforceable action profiles and enforceability along hyperplanes from Fudenberg et al. (1994) and Sannikov (2007):

Definition. *A $2 \times d$ matrix*

$$B = \begin{bmatrix} \beta^1 \\ \beta^2 \end{bmatrix} = \begin{bmatrix} \beta^{11} & \dots & \beta^{1d} \\ \beta^{21} & \dots & \beta^{2d} \end{bmatrix}$$

enforces action profile $a \in \mathcal{A}$ *if for* $i = 1, 2$,

$$\forall a_i' \in \mathcal{A}^i, \quad g_i(a) + \beta^i \mu(a) \geqslant g_i(a_i', a_{-i}) + \beta^i \mu(a_i', a_{-i}).$$

An action profile $a \in \mathcal{A}$ *is enforceable if there exists some matrix* B *that enforces it.*

Definition. *A vector of volatilities* $\phi \in \mathbb{R}^d$ *enforces action profile* $a \in \mathcal{A}$ *along vector* $\boldsymbol{T} = (t_1, t_2)$ *if the matrix*

$$B = \boldsymbol{T}^\top \phi = \begin{bmatrix} t_1\phi_1 & \dots & t_1\phi_d \\ t_2\phi_1 & \dots & t_2\phi_d \end{bmatrix}$$

enforces a*. Of all vectors* ϕ *that enforce* a *along* $\boldsymbol{T}$*, let* $\phi(a, \boldsymbol{T})$ *be the one with the smallest length.*

Consider the following assumptions:

Assumption 1. *All action profiles* $(a_1, a_2) \in \mathcal{A}^1 \times \mathcal{A}^2$ *of the stage game are pairwise identifiable, i.e., the spans of the* $d \times (|\mathcal{A}^1| - 1)$ *matrix* $M_1(a)$ *with columns* $\mu(a_1', a_2) - \mu(a)$*,* $a_1' \neq a_1$ *and the* $d \times (|\mathcal{A}^2| - 1)$ *matrix* $M_2(a)$ *with columns* $\mu(a_1, a_2') - \mu(a)$*,* $a_2' \neq a_2$ *intersect only at the origin.*

Assumption 2. *Either*

1. *For all* $i = 1, 2$ *and* $a_i \in \mathcal{A}^i$*, the static best response to* a_i *is unique or*

2. *For all* $a \in \mathcal{A}$*, the spans of* $M_1(a)$ *and* $M_2(a)$ *are orthogonal.*

Assumption 3. *For each of the players, at least one of the pure-action profiles minmaxing them is a Nash equilibrium of the stage game or is enforceable.*

Assumptions 1 and 2 are precisely the assumptions used in Sannikov (2007). In particular, these assumptions guarantee that an enforceable action profile is enforceable along all regular vectors. Moreover, an enforceable action profile a is enforceable along vector $\mathbf{T}$ with $t_i = 0$ if and only if a_i is a best response to a_{-i} in the stage game G. Assumption 3 is a new and the most crucial one. It requires that at least locally, one can provide incentives via shift in promised continuation values to each of the players to minmax their opponents. Note that this restriction is much weaker than to directly require that minmaxing can be incentivized forever.

Consider any two punishment outcomes Q^1 and Q^2 satisfying the inertia restriction for some inertia parameter $\epsilon > 0$. Define two agreements $\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$ constructed from Q^1 and Q^2 as follows:

- $\mathcal{E}^1(Q^1, Q^2)$ proposes Q^1 as the initial outcome and then at any point of public deviation, proposes to start the punishment outcome Q^i if this deviation was caused solely by Player i;

- $\mathcal{E}^2(Q^1, Q^2)$ proposes Q^2 as the initial outcome and then at any point of public deviation, proposes to start the punishment outcome Q^i if this deviation was caused by solely Player i;

- in both $\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$, if a public deviation is caused by both players simultaneously, the prescribed punishment is Q^1.

Notice that for any two punishment outcomes Q^1 and Q^2, the agreements $\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$ are measurable.

The following theorem establishes the existence of optimal penal codes in our setting and is our second main result.

Theorem 2 (Optimal Penal Code)**.** *Suppose that Condition C and Assumptions 1, 2, and 3 are satisfied. Then, there exists $\bar{\epsilon} > 0$ and public outcomes Q^1 and Q^2 such that for any $\epsilon \in (0, \bar{\epsilon})$,*

1. *Q^1 and Q^2 are punishment outcomes with the inertia parameter ϵ;*
2. *$\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$ are self-enforcing public agreements;*
3. *$\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$ deliver the minmax payoffs to Player 1 and Player 2 correspondingly,*

$$\forall i = 1, 2, \quad W^{i, \mathcal{E}^i(Q^1, Q^2)} = \underline{v}_i.$$

Proof. See Appendix B.4 for a constructive proof. □

4.3 Characterization of the Payoff Set.

In this subsection, we provide the characterization of $K(\epsilon)$, the set of payoffs attainable in self-enforcing public agreements for sufficiently small inertia parameters ϵ. In the previous subsection, we showed that $K(\epsilon)$ consists of individually rational payoffs.

Definition (Comprehensive Set)**.** *A subset S of the set of individually rational payoffs is called comprehensive if for any payoff pair $(w_1, w_2) \in S$, S also contains all individually rational payoffs obtainable from (w_1, w_2) by subtracting a positive linear combination of money-transfer vectors $(1, -k)$ and $(-k, 1)$. I.e., if S contains all $(w_1', w_2') \geqslant (\underline{v}_1, \underline{v}_2)$ that can be represented as $(w_1, w_2) - \alpha_1(1, -k) - \alpha_2(-k, 1)$ for some $\alpha_1 \geqslant 0$ and $\alpha_2 \geqslant 0$.*

Lemma 3 (Comprehension)**.** *Under Condition C and Assumptions 1, 2, and 3, there exists $\bar{\epsilon} > 0$, such that for any $\epsilon \in (0, \bar{\epsilon})$, the set $K(\epsilon)$ is comprehensive.*

Proof. Take the initial outcome of some self-enforcing agreement $\mathcal{E}$ with payoffs $(w_1, w_2) \in K(\epsilon)$. For an individually rational payoff pair $(w_1', w_2') = (w_1, w_2) - \alpha_1(1, -k) - \alpha_2(-k, 1)$ construct the outcome which requires initially Player 1 to send α_1 amount of money and Player 2 to send α_2 with the retention parameter k, and then implements the initial outcome of $\mathcal{E}$. Support this outcome by the optimal penal code from Theorem 2. This agreement will also be self-enforcing with uniformly non-manipulable outcomes. □

Lemma 4 (Stabilization)**.** *Under Condition C and Assumptions 1, 2, and 3, there exists $\bar{\epsilon} > 0$ such that for all $\epsilon \in (0, \bar{\epsilon})$, the set $K(\epsilon)$ is the same.*

Proof. By Theorem 1, the set of outcomes supportable in self-enforcing agreements is the same for all inertia parameters for which there exists an optimal penal code. The rest follows from Theorem 2. □

Lemma 5 (Convexity)**.** *For any $\epsilon > 0$, the set $K(\epsilon)$ is convex.*

Proof. By the standard argument of convexification through the initial public randomization. □

Lemma 6 (Inclusion)**.** *For any $\epsilon > 0$, the set $K(\epsilon)$ contains the set of p-PPE-payoffs from Sannikov (2007).*

Proof. Take any p-PPE from Sannikov (2007) resulting in the outcome Q. Construct the agreement $\mathcal{E}$ which specifies Q as the initially proposed outcome. As Q prescribes no positive transfers, there are no possible public deviations for the players. Hence, $\mathcal{E}$ does not need to specify punishment outcomes. Conditions of Theorem 1 then simplify to the incentive compatibility conditions from Proposition 2 in Sannikov (2007), which are supposed to be satisfied for Q. Also, as there are no punishment outcomes, the inertia restriction is vacuous. Q.E.D. □

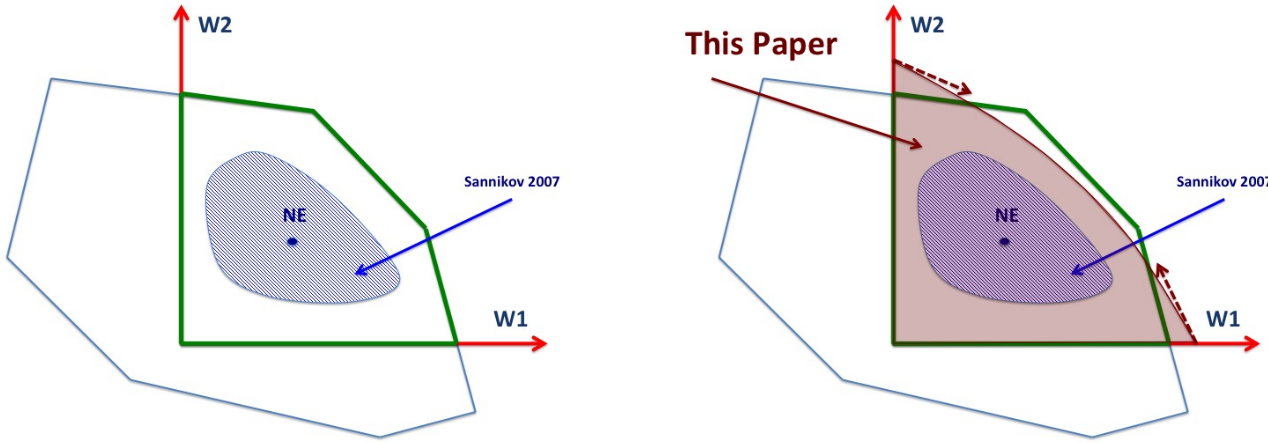
Denote by $\partial_+ K(\epsilon)$ the part of the boundary of $K(\epsilon)$ which lies strictly above the minmax lines of the players. Take any point $w = (w_1, w_2) \in \partial_+ K(\epsilon)$. Let $\mathbf{T}(w)$ and $\mathbf{N}(w)$ denote the unit tangent and outward normal vectors for $\partial_+ K(\epsilon)$ at w. As $K(\epsilon)$ is convex, these vectors are uniquely defined for all but at most countably many points of $\partial_+ K(\epsilon)$. Let $\kappa(w)$ be the curvature of $\partial_+ K(\epsilon)$ at w. Recall that $\phi(a, \mathbf{T})$ denotes the vector of volatilities that enforces an action profile a along vector T and has the smallest length. If a is not enforceable along $\mathbf{T}$, set $|\phi(a, \mathbf{T})| = \infty$. Also, let $\mathcal{A}^N$ be the set of pure Nash equilibria of the stage game G and $\mathcal{N}$ be the convex hull of the payoffs from $\mathcal{A}^N$. The following equation is the *optimality equation* from Sannikov (2007):

$$\kappa(w) = \max \left\{ 0; \max_{a \in (\mathcal{A}^1 \times \mathcal{A}^2) \setminus \mathcal{A}^N} \frac{2\mathbf{N}(w)(g(a) - w)}{r|\phi(a, \mathbf{T}(w))|^2} \right\} \tag{4}$$

Sannikov (2007) shows that in his setting, the boundary of the set of p-PPE payoffs satisfies the optimality equation at each point outside of $\mathcal{N}$. The following lemma is the analogous result for our model:

Lemma 7 (Optimality Equation)**.** *Under Condition C and Assumptions 1, 2, and 3, for any $\epsilon > 0$, at all points which are not in $\mathcal{N}$, $\partial_+ K(\epsilon)$ satisfies the optimality equation (4). Moreover, for each $i = 1, 2$, $\partial_+ K(\epsilon)$ enters the minmax line of Player i either at a payoff pair corresponding to a p-NE of the stage game, or tangent to the corresponding money-transfer vector, $(1, -k)$ for Player 1 and $(-k, 1)$ for Player 2.*

Proof. The proof is similar to the proof of Proposition 5 in Sannikov (2007). In fact, the proof that the curvature of $\partial_+ K(\epsilon)$ can not be smaller than the one prescribed by the optimality equation



(a) the p-PPE payoff set $\mathcal{S}$ from Sannikov (2007). (b) $K(\epsilon)$ in our setting.

Figure 1: Payoffs sets.

is almost exactly the same. The proof that the curvature can not be greater than the one in the optimality equation, i.e., "the escape argument", differs in our setting by the introduction of pushes of continuation values caused by money transfers. However, as $K(\epsilon)$ is comprehensive, at any point along $\partial_+ K(\epsilon)$, the outward normal vector is positively correlated with the money-transfer pushes. Thus, these pushes can only make the escape argument to be more compelling. See Appendix C.1 for the formal argument.

□

The above lemmata are summarized in the theorem below, which is our third main result.

Theorem 3 (Payoff-Set Characterization). *Under condition C and Assumptions 1, 2, and 3, for any $k \in [0,1)$, there exists $\bar{\epsilon} > 0$ such that for $\epsilon \in (0, \bar{\epsilon})$, the set $K(\epsilon)$ is the largest bounded closed set that satisfies the following properties:*

1. *$K(\epsilon)$ is a convex and comprehensive subset of the set of individually rational payoffs;*

2. *at all points outside of $\mathcal{N}$, $\partial_+ K(\epsilon)$ satisfies the optimality equation (4) and for each $i = 1, 2$, $\partial_+ K(\epsilon)$ enters the minmax line of Player i either at a payoff pair corresponding to a p-NE of the stage game, or tangent to the corresponding money-transfer vector, $(1, -k)$ for Player 1 and $(-k, 1)$ for Player 2.*

Proof. See Appendix C.2. □

Figure 1 compares schematically the set of p-PPE payoffs $\mathcal{S}$ from Sannikov (2007) (Figure 1a) and the corresponding set $K(\epsilon)$ from our setting (Figure 1b). The blue polygon on both pictures corresponds to the boundary of the convex hull of the stage-game payoffs; the red lines are the players' minmax lines; the green polygon is the boundary of the set $\mathcal{V}^*$, the set of individually rational and feasible-without-transfers payoffs. The blue solid shape in Figure 1a is the set $\mathcal{S}$, the

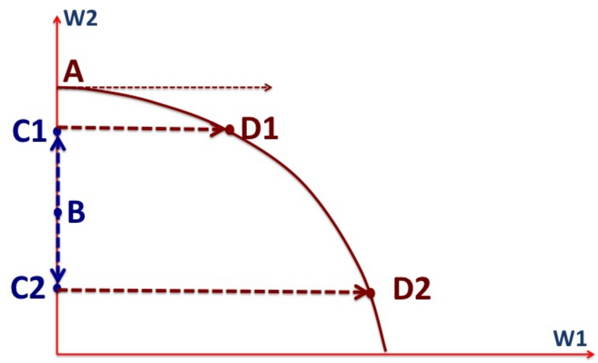


Figure 2: Punishing a deviation from Player 1, $k = 0$.

red solid shape in Figure 1b is the corresponding set $K(\epsilon)$. Note that $\mathcal{S}$ typically does not reach the players' minmax lines unless there are p-NE payoffs on them. Also, $\mathcal{S}$ must lie inside of $\mathcal{V}^*$. In contrast, in our setting, the set $K(\epsilon)$ reaches both minmax lines as long as the conditions of Theorem 2 are satisfied. Also, $K(\epsilon)$ may extend outside of $\mathcal{V}^*$ as the feasible set is larger when money transfers are available. The upper part of the boundary of $K(\epsilon)$, $\partial_+ K(\epsilon)$ is smooth at all points outside of $\mathcal{N}$ and enters the minmax lines of the players either at p-NE payoffs, or parallel to the money transfer vectors (the red dashed vectors in Figure 1b). Finally, $\partial_+ K(\epsilon)$ will typically have strictly positive curvature outside of $\mathcal{N}$. The only exception that may be is that $\partial_+ K(\epsilon)$ contains a segment of a straight line which starts at a player's mixmax line, ends at a p-NE payoff-pair, and is parallel to the corresponding money-transfer vector.

5 Efficient Self-Enforcing Public Agreements.

Let us now discuss the dynamics in efficient self-enforcing public agreements in our setting. Figure 2 depicts schematically a typical path of continuation values along the initial outcome and punishments in an efficient self-enforcing agreement in the case of pure money burning, $k = 0$ (The picture for the case $0 < k < 1$ will look similar).

Unless there is a p-NE payoff-pair on $\partial_+ K(\epsilon)$, an agreement that delivers the payoffs $w \in \partial_+ K(\epsilon)$ starts with $W_0 = w$ and supports players' incentives by the shift of promised continuation values along $\partial_+ K(\epsilon)$ without costly transfers. The recommended profiles of hidden actions and volatilities of continuation values are determined by the optimality equation (4). This continues until the promised continuation values hit the minmax line (the individual rationality constraint) of either of the players. For example, point A in Figure 2 is the point at which the continuation values hit the minmax line of Player 1. At point A, the agreement introduces the reflective boundary for the process of promised continuation values following the SDE from Proposition 1. To implement this reflective boundary, the agreement asks Player 1 to burn (transfer in case $0 < k < 1$) money so as the cumulative amount of money burnt, Γ_t^1, exactly matches the compensating process of the reflected continuation values. In particular, money burning will be happening in infinitesimal installments and only after extreme histories, when it will be no longer possible to support incentives by the shift of promised continuation values without violating the individual rationality constraint of Player 1.

Suppose that at point A, Player 1 announces a public deviation. In that case, the agreement will go to the stage of punishing Player 1. This can be done using the construction from Theorem 2. An alternative punishment is shown in Figure 2. The punishment starts by moving the continuation values to point B. This will upset the promises made to Player 2, but this is fine since Player 2 is not the deviating player. The punishment outcome then proposes to support minmaxing Player 1 by moving the promised continuation values along the minmax line of Player 1 until they hit either C_1, or C_2. At C_1, Player 1 is asked to burn a positive amount of money so as to jump to D_1. Similarly, at C_2, Player 1 is asked to burn a positive amount of money so as to jump to D_2. The punishment outcome then is concatenated with the initial outcomes of the efficient agreements that deliver D_1, or D_2 correspondingly.

There are two more things to be said about the efficient self-enforcing agreements in our setting. First, an efficient p-PPE in Sannikov (2007) typically is supported by the evolution of promised continuation values that will eventually drive them into the area Pareto dominated by other p-PPE payoffs. This may raise concerns regarding the renegotiation proofness of such p-PPE's. In contrast, in our setting, the promised continuation values of an efficient agreement on the path of play will always stay on $\partial_+ K(\epsilon)$, the Pareto frontier of $K(\epsilon)$. Thus, the renegotiation-proofness concerns are less severe in our setting. Of course, punishing observable deviations in our case still requires that the continuation values plunge into the Pareto-dominated areas. However, the "depth" of such plunges may be made arbitrary small with the inertia parameters $\epsilon \to 0$. Second, the dynamics in the efficient agreements in the case of costly transfers are in sharp contrast with the dynamics in efficient equilibria of repeated games with perfect transfers (such as Levin (2003), Goldlucke and Kranz (2012)). With perfect transfers, the timing of the transfers itself is not important and so it may be efficient to use them frequently (for example, at the end of every period). With costly transfers, it is efficient to postpone their use for as long as possible. Thus, costly transfers are used rarely, only when the individual rationality constraint of either of the players become binding.

A Proof of Theorem 1.

`Sufficiency.`

Take any agreement $\mathcal{E}$. Suppose that it satisfies both restrictions of the One-Stage Deviation Principle. We shall prove that $\mathcal{E}$ is self-enforcing. Indeed, consider any strategy σ for any Player i. As the upper bound on value from σ is the limsup of upper bounds on values of its finite truncations, it is sufficient for us to check that $V^*(\sigma, Q) \leqslant W^{i,Q}$ for any $Q \in \mathcal{E}$ and for σ that prescribes only finitely many public deviations. We do so by induction in L, the number of public deviations prescribed by σ.

Base: $L = 0$. Suppose σ does not prescribe any public deviation. Then $V^*(\sigma, Q) \leqslant W^{i,Q}$ may be proven using the One-Stage Deviation in Hidden Actions restriction only. In fact, the proof essentially repeats the proof of Proposition 2 from Sannikov (2007). (The money-transfer processes cancel out when we evaluate the effect on Player i payoff caused by the change in hidden-action profile!)

Induction: Suppose that $V^*(\sigma, Q) \leqslant W^{i,Q}$ for any σ prescribing less that $L = l$ observable deviations. Prove for σ that prescribes $L = l$ observable deviations. Take any outcome $Q \in \mathcal{E}$. Recall the definition of $V_*(\sigma, Q)$:

$$V^*(\sigma,Q) = E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big[r\int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma},A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\,d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] + \\ + \Big(E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big)^*\Big[e^{-rT^{i,Q}(\omega)}\ \Big(V_*\big(\sigma,\tilde{Q}(T^{i,Q}(\omega))\big) + r\Delta\Gamma^{i,Q}_{T^{i,Q}(\omega)} - rk\Delta\Gamma^{-i,Q}_{T^{i,Q}(\omega)}\Big)\Big]$$

But, starting from the beginning of the punishment $\tilde{Q}\big(T^{i,Q}\big)$ that follows immediately after Player i deviates from Q at $T^{i,Q}$, σ prescribes at most $l-1$ observable deviations. By the induction hypothesis then, $V_*\big(\sigma, \tilde{Q}(T^{i,Q}(\omega))\big) \leqslant W^{i,\tilde{Q}(T^{i,Q}(\omega))}$. Therefore,

$$V^*(\sigma,Q) \leqslant E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big[r\int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma},A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\,d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] + \\ + \Big(E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big)^*\Big[e^{-rT^{i,Q}(\omega)}\ \Big(W^{i,\tilde{Q}(T^{i,Q}(\omega))} + r\Delta\Gamma^{i,Q}_{T^{i,Q}(\omega)} - rk\Delta\Gamma^{-i,Q}_{T^{i,Q}(\omega)}\Big)\Big]$$

Applying the One-Stage Deviation in Observable Actions restriction to the outcome Q and the stopping time $T^{i,Q}$, we then get that moreover

$$V^*(\sigma,Q) \leqslant E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big[r\int_0^{T^{i,Q}} e^{-rs}\big(g_i(A_s^{i,Q,\sigma},A_s^{-i,Q})ds - d\Gamma_s^{i,Q} + k\,d\Gamma_s^{-i,Q}\big) - r\Gamma_0^{i,Q} + rk\Gamma_0^{-i,Q}\Big] + \\ + E^{\mathbf{P}(A^{i,Q,\sigma},A^{-i,Q})}\Big[e^{-rT^{i,Q}} W^{i,Q}_{T^{i,Q}}\Big]$$

But the RHS of the above inequality is the value evaluated at the beginning of Q of the strategy that prescribes to follow σ until the moment of the first observable deviation and then instead of announcing this deviation, to abide to Q. By the base of induction, this value is weakly below $W^{i,Q}$. Thus, $V^*(\sigma, Q) \leqslant W^{i,Q}$.

`Necessity of 1.`

Suppose, $\mathcal{E}$ fails the One-Stage Deviation in Hidden Actions restriction. Then a profitable deviation σ can be constructed by deviating only in hidden actions within a given outcome Q, similarly to how it can be done in Sannikov (2007). Moreover, the value of this deviating strategy can be computed with the usual integrals so that $V^*(\sigma, Q) = V_*(\sigma, Q) > W^{i,Q}$.

`Necessity of 2.`

Suppose, $\mathcal{E}$ fails the One-Stage Deviation in Observable Actions restriction. Suppose the restriction fails for some outcome $Q \in \mathcal{E}$, Player i, and a stopping time T. Consider the set $B = \{\omega \in \Omega^Q : W_T^{i,Q} < W^{i,\tilde{Q}(T)} + r\Delta\Gamma_T^{i,Q} - rk\Delta\Gamma_T^{-i,Q}\}$. On B, consider the function $f(\omega) = e^{-rT(\omega)}\Big(W^{i,\tilde{Q}(T)} + r\Delta\Gamma_T^{i,Q} - rk\Delta\Gamma_T^{-i,Q} - W_T^{i,Q}\Delta\Gamma_T^{-i,Q}\Big)$, the discounted instantaneous gains from publicly deviating at time $T(\omega)$. Then, $f(\omega) > 0$ on B and B is not $\mathbf{P}^Q$-measure zero set. Set $f(\omega) = 0$ outside of B. Then $\exists\delta_1 > 0$, $\exists\delta_2 > 0$ such that any measurable set C that contains the set $B(\delta_1) = \{\omega \in \Omega^Q : f(\omega) > \delta_1\}$ has $\mathbf{P}^Q$-measure of at least δ_2. Therefore, for any measurable function $g(\omega)$ such that $g(\omega) \geqslant f(\omega)$, we will have $E^{\mathbf{P}^Q}(g) > \delta_1 \cdot \delta_2$. Define the stopping time $\hat{T} = T \cdot \mathbb{1}_B + \infty \cdot \mathbb{1}_{\Omega^Q \setminus B}$. Consider the strategy σ for Player i that prescribes no deviations in hidden actions and just one observable deviation from Q at $\hat{T}$. Clearly, this strategy will be a profitable deviation with $V^*(\sigma, Q) \geqslant W^{i,Q} + \delta_1\delta_2$. The first part is proven.

Suppose further that $\mathcal{E}$ is measurable. Suppose the restriction fails for some outcome $Q \in \mathcal{E}$, Player i and a stopping time T. Consider the set $B = \{\omega \in \Omega^Q : W_T^{i,Q} < W^{i,\tilde{Q}(T)} + r\Delta\Gamma_T^{i,Q} - rk\Delta\Gamma_T^{-i,Q}\}$. By measurability of $\mathcal{E}$, B is an $\mathcal{F}_T^Q$-measurable event. By the failure of the One-Stage Deviation in Observable Actions restriction, $Pr^{\mathbf{P}^Q}(B) > 0$. Define the stopping time $\hat{T} = T \cdot \mathbb{1}_B + \infty \cdot \mathbb{1}_{\Omega^Q \setminus B}$. Consider the strategy σ for Player i that prescribes no deviations in hidden actions and just one observable deviation from Q at $\hat{T}$. Clearly, this strategy will be a profitable deviation with $V^*(\sigma, Q) = V_*(\sigma, Q) > W^{i,Q}$.

B Proof of Theorem 2.

B.1 Proof of Lemma 1.

Take any point $(w_1, w_2) \in K(\epsilon_1)$. This point can be achieved as expected-payoff pair in some self-enforcing agreement $\mathcal{E}$ with the inertia parameter ϵ_1. But then, $\mathcal{E}$ is also a self-enforcing agreement with the inertia parameter ϵ_2 because it satisfies the conditions of Theorem 1 and because the ϵ_2-inertia restriction is less restrictive than the ϵ_1-inertia for $\epsilon_2 < \epsilon_1$. Thus, $(w_1, w_2) \in K(\epsilon_2)$.

B.2 Proof of Lemma 2.

Clearly, the second statement in the formulation of Lemma 2 implies the first one. So it is sufficient to show that the second statement is correct. Suppose on the contrary, that there is a public outcome Q in a self-enforcing agreement $\mathcal{E}$, a stopping time τ in Q, and Player i such that $W_\tau^{i,Q} \geqslant \underline{v}_i + r\Delta\Gamma_\tau^{i,Q} - rk\Delta\Gamma_\tau^{-i,Q}$ is violated on the event $A \in \mathcal{F}_\tau^Q$ of positive positive probability. Consider the following deviating strategy for Player i: follow the plan of actions and transfers suggested in Q on the event "not A"; on the event A, follow the proposed plans until τ and then switch to "dropping out from the cooperation", i.e., start always playing a hidden action that is a myopic best response to the current action of the opponent and always announce to refuse to send positive transfers if asked by the agreement. Notice, that the switch to the dropping-out regime happens only after time τ, at which point Player i will know whether A has happened or not. Thus, so described strategy for Player i is indeed a well-defined public strategy. Yet, this strategy will be a strictly profitable deviation, which contradicts the assumption that $\mathcal{E}$ is self-enforcing. Q.E.D.

B.3 Concatenation of Outcomes.

In this subsection, we show how having two public outcomes Q^α and Q^β and a stopping time τ^α in outcome Q^α, one can construct the concatenated outcome $Con(Q^\alpha, Q^\beta, \tau^\alpha)$ which corresponds to the play of Q^A in the beginning until the time hits τ^α and then switches to the beginning of Q^B.

Suppose we are given two outcomes $Q^\alpha = \{\mathcal{P}^{Q^\alpha}, A^{Q^\alpha}, \Gamma^{Q^\alpha}\}$ and $Q^\beta = \{\mathcal{P}^{Q^\beta}, A^{Q^\beta}, \Gamma^{Q^\beta}\}$. Suppose τ^α is a stopping time in $\mathcal{P}^{Q^\alpha}$ at which the play should switch from Q^α to Q^β. Let us construct the concatenated outcome $Con(Q^\alpha, Q^\beta, \tau^\alpha) = \{\mathcal{P}, \mathcal{A}, \Gamma\}$:

- The state-space Ω for the concatenated outcome is the direct product of the state-spaces of Q^α and Q^β, i.e., $\Omega = \{\omega = (\omega_1, \omega_2) | \omega_1 \in \Omega^{Q^\alpha}, \omega_2 \in \Omega^{Q^\beta}\}$.
- The probability measure $\mathbf{P}$ for the concatenated outcome is the direct product $\mathbf{P} = \mathbf{P}^{Q^\alpha} \otimes \mathbf{P}^{Q^\beta}$.
- The moment of switch τ corresponds to τ^α, i.e., $\tau(\omega_1, \omega_2) = \tau^\alpha(\omega_1)$.
- The public filtration $(\mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0})$ is the following:
  - $\mathcal{F} = \sigma(\mathcal{F}^{Q^\alpha} \otimes \mathcal{F}^{Q^\beta})$;
  - $\mathcal{F}_t$ consists of all those events $A \in \mathcal{F}$ such that for any $0 \leqslant s_1 \leqslant s_2 \leqslant t$, the event $A \cap \{s_1 \leqslant \tau \leqslant s_2\}$ belongs to the σ-algebra $\sigma(\mathcal{F}^{Q^\alpha}_{s_2} \otimes \mathcal{F}^{Q^\beta}_{t-s_1})$ and the event $A \cap \{\tau > t\}$ belongs to $\mathcal{F}^{Q^\alpha}_t \otimes \{\varnothing, \Omega^{Q^\beta}\}$.
- The public signal X_t is $X_t(\omega_1, \omega_2) = X^{Q^\alpha}_t(\omega_1) \cdot \mathbb{1}_{\tau \geqslant t} + \left(X^{Q^\alpha}_\tau(\omega_1) + X^{Q^\beta}_{t-\tau}(\omega_2)\right) \cdot \mathbb{1}_{\tau < t}$.
- The recommended hidden actions A_t are $A_t(\omega_1, \omega_2) = A^{Q^\alpha}_t(\omega_1) \cdot \mathbb{1}_{\tau < t} + A^{Q^\beta}_{t-\tau}(\omega_2) \cdot \mathbb{1}_{\tau \geqslant t}$.
- The recommended cumulative money transfers Γ_t are $\Gamma_t = \Gamma^{Q^\alpha}_t(\omega_1) \cdot \mathbb{1}_{\tau < t} + \left(\Gamma^{Q^\alpha}_\tau(\omega_1) + \Gamma^{Q^\beta}_{t-\tau}(\omega_2)\right) \cdot \mathbb{1}_{\tau \geqslant t}$. Note that the switch happens only after the transfers recommended at time τ^α in Q^α are processed.

Note that in this construction, $X_t - \int_0^t \mu(A_s) ds$ is a d-dimensional Brownian motion under $\mathbf{P}$, the processes A_t and k are progressively measurable for $(\mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0})$, and Γ_t are adapted to $(\mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0})$ weakly increasing nonnegative RCLL-processes. Thus, $Con(Q^\alpha, Q^\beta, \tau^\alpha)$ is a public outcome.

B.4 Proof of Theorem 2.

By Condition C, there exists a self-enforcing agreement $\mathcal{E}$ delivering payoffs (w_1, w_2) with $w_1 > \underline{v}_1$ and $w_2 > \underline{v}_2$. Let Q^0 be the initial outcome of $\mathcal{E}$. By Lemma 2, we can find a modification for the processes of hidden actions (A^1_t, A^2_t) and the processes of promised continuation values (W^{1,Q^0}_t, W^{2,Q^0}) such that the One-Stage Deviation in Hidden Actions and the restriction $\forall i = 1, 2, W^{i,Q^0}_t \geqslant \underline{v}_i + r\Delta\Gamma^{i,Q^0}_t - rk\Delta\Gamma^{-i,Q^0}_t$ are satisfied always.

We shall now construct the required Q^1 and Q^2.

Let us construct Q^1, the harshest punishment for Player 1:

Suppose first that the minmaxing profile $a = (a_1, a_2)$ for Player 1 is a Nash Equilibrium of the stage game. Then set Q^1 to be the public outcome corresponding to the play of (a_1, a_2) forever with zero volatility of promised continuation values, $\big(W^{1,Q^1}_t, W^{2,Q^1}_t\big) = \big(g_1(a), g_2(a)\big)$.

Suppose now that the profile $a = (a_1, a_2)$ minmaxing Player 1 is not a Nash equilibrium, but only enforceable. Then the construction is the following. Set $L_1 = (\underline{v}_1, w_2 + k(w_1 - \underline{v}_1))$ and $L_2 = \big(\underline{v}_1, \underline{v}_2 + k(w_1 - \underline{v}_1) + k^2(w_2 - \underline{v}_2)\big)$. As $0 \leqslant k < 1$, L_1 is strictly above L_2. For the beginning of the outcome, call it Q, set $(W^1_0, W^2_0) = \frac{L_1 + L_2}{2}$. Recommend the players to play the minmaxing profile (a_1, a_2) enforcing it along the vector $(0, 1)$ and requiring no money transfers. I.e., set the matrix $B = (0, 1)^\top \phi(a, (0, 1))$. Construct (W^1_t, W^2_t) as a continuous weak solution to the system of SDE's

$$dW^i_t = r(W^i_t - g_i(a))dt + B^i(dX_t - \mu(a)dt).$$

Notice that for this solution, we will always have constant $W^1_t = \underline{v}_1$. Thus, the solution (W^1_t, W^2_t) moves along the minmax line of Player 1. Consider the stopping time τ, the first time when (W^1_t, W^2_t) hits either L_1 or L_2. At L_1, require that the players send transfers $(w_1 - \underline{v}_1, 0)$, at L_2, require that they send transfers $\big(w_1 - \underline{v}_1 + k(w_2 - \underline{v}_2), w_2 - \underline{v}_2\big)$. Then stop the outcome Q and start playing the outcome Q^0. Define Q^1 as the concatenated outcome, $Q^1 = Con(Q, Q^0, \tau)$. Set the process of promised continuation values in the concatenated outcome as the concatenation of the processes of promised continuation values from Q and Q^0. Clearly, these processes will satisfy representation (1) and so they are indeed the processes of promised continuation values for Q^1. Moreover, Q^1 satisfies the One-Stage Deviation in Hidden Action restriction. Indeed, in the beginning, it is satisfied by enforceability of (a_1, a_2), and after the switch it is satisfied for Q^0. Yet, Q^1 delivers the worst possible payoff to Player 1, $W^{1,Q^1} = \underline{v}_1$.

The outcome Q^2 for punishing Player 2 is constructed analogously.

Next, Q^1 does not require any transfers until the hitting time τ. As the incentives until τ are enforced by the constant matrix of volatilities, there exists $\epsilon_1 > 0$ such that Q^1 is a punishment outcome for any inertia parameter $\epsilon \in (0, \epsilon_1)$. Similarly, here exists $\epsilon_2 > 0$ such that Q^2 is a punishment outcome for any inertia parameter $\epsilon \in (0, \epsilon_2)$. Set the required $\bar{\epsilon} = \min\{\epsilon_1, \epsilon_2\}$.

Finally, the agreements $\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$ satisfy the One-Stage Deviation in Observable Action restriction. Indeed, by construction, the processes of continuation values plus the current transfers always stay above the minmax payoffs $(\underline{v}_1, \underline{v}_2)$, exactly the payoffs promised to the players in case either of them deviates. Therefore, $\mathcal{E}^1(Q^1, Q^2)$ and $\mathcal{E}^2(Q^1, Q^2)$ are self-enforcing. Q.E.D.

C Proof of Theorem 3.

C.1 Proof of Lemma 7.

The proof is similar to the proof of Proposition 5 from Sannikov (2007). Indeed, notice first that the following adaptation of Proposition 3 from Sannikov (2007) applies to our case:

Proposition 3′. *Suppose that a curve $\mathcal{C}$ satisfies the optimality equation (4). Suppose further that $\mathcal{C}$ has endpoints which are attainable as payoffs in self-enforcing agreements with the inertia parameter ϵ. Then any point in $\mathcal{C}$ is attainable as a payoff-pair in a self-enforcing public agreement with the inertia parameter ϵ, i.e., $\mathcal{C} \subseteq K(\epsilon)$.*

Proof. The construction in the proof is similar to the one used in the proof of Theorem 2. If the curve $\mathcal{C}$ has positive curvature, then the idea is to take any point $w \in \mathcal{C}$ and to construct the beginning of the outcome by supporting incentives without money-transfers, solely by the drift-diffusion of the promised continuation values along $\mathcal{C}$ until the values hit either of the endpoints of $\mathcal{C}$ (exactly how it is done in Sannikov (2007)). After that, use the concatenation with the initial outcomes of the corresponding self-enforcing agreements. Theorem 1 then will insure that so constructed outcome will be supportable in a self-enforcing agreement and will deliver to the players payoffs w. If $\mathcal{C}$ is a segment of a straight line, then w can be obtained as a public randomization between the agreements corresponding to the end points of $\mathcal{C}$. And so w will indeed be in $K(\epsilon)$. □

Then notice that the following variant of Lemma 8 from Sannikov (2007) is valid in our setting:

Lemma 8′. *Consider a point $w \in \partial_+ K(\epsilon) \backslash \mathcal{N}$ with the outward normal vector $\mathbf{N}$. Then the curve $\mathcal{C}$, which solves the optimality equation (4) with initial conditions $(w, \mathbf{N})$ does not enter the interior of $K(\epsilon)$.*

Proof. The proof uses Proposition 3′ and is otherwise the same as the proof of Lemma 8 in Sannikov (2007). □

Thus, indeed, the curvature of $\partial_+ K(\epsilon)$ can not be smaller than the one prescribed by the optimality equation (4). To prove that the curvature of $\partial_+ K(\epsilon)$ can not be greater than the one in the optimality equation (4), we use the following adaptation of Lemma 6′ from Hashimoto (2010):

Lemma 6″. *It is impossible for a solution $\mathcal{C}'$ of the optimality equation (4) with endpoints v_L and v_H to satisfy the following properties simultaneously:*

1. *There is a unit vector $\hat{\mathbf{N}}$ such that $\forall x > 0$, $v_L + x\hat{\mathbf{N}} \notin K(\epsilon)$ and $v_H + x\hat{\mathbf{N}} \notin K(\epsilon)$.*

2. *For all $w \in \mathcal{C}'$ with an outward unit normal $\mathbf{N}(w)$ for $\mathcal{C}'$ at w, we have*

$$\max_{v_N \in \mathcal{N}} \mathbf{N}(w) v_N < \mathbf{N}(w) w.$$

3. *$\mathcal{C}'$ "cuts through" $K(\epsilon)$, that is, there exists a point $v \in \mathcal{C}'$ such that $W_0 = v + x\hat{\mathbf{N}} \in K(\epsilon)$ for some $x > 0$.*

4. $\inf_{w \in \mathcal{C}'} \hat{\mathbf{N}} \mathbf{N}(w)^\top > 0$*, where $\mathbf{N}(w)$ is the outward normal vector for $\mathcal{C}'$ at w.*

5. *$\hat{\mathbf{N}}$ is positively correlated with the money-transfer vectors, $\hat{\mathbf{N}} \cdot (1, -k) \geqslant 0$ and $\hat{\mathbf{N}} \cdot (-k, 1) \geqslant 0$.*

Proof. The proof almost exactly repeats the proof from Hashimoto (2010). The only difference now is that with money transfers, the RHS of the Ito formula in footnote 2 of Hashimoto (2010) will have an extra term, $P_t = \int_0^t (1, -k) \cdot \hat{\mathbf{N}} \; d\Gamma_s^1 + \int_0^t (-k, 1) \cdot \hat{\mathbf{N}} \; d\Gamma_s^2 + (1, -k) \cdot \hat{\mathbf{N}} \; \Delta\Gamma_0^1 + (-k, 1) \cdot \hat{\mathbf{N}} \; \Delta\Gamma_0^2$. But since $\hat{\mathbf{N}}$ is positively correlated with both $(1, -k)$ and $(-k, 1)$, the term P_t is positive. Therefore, equation (6) from Hashimoto (2010) still applies in our case and the rest of their proof works. □

To finish the proof Lemma 7, take ϵ small enough that the optimal penal codes exist. Take any point $w \in \partial_+ K(\epsilon) \backslash \mathcal{N}$. Set $\hat{\mathbf{N}}$ to be any outward unit-normal vector for $\partial_+ K(\epsilon)$ at w. By Lemma 3, the set $K(\epsilon)$ is comprehensive, and so $\hat{\mathbf{N}}$ is positively correlated with both $(1, -k)$ and $(-k, 1)$. If the curvature of $\partial_+ K(\epsilon)$ at w is greater than the one prescribed by the optimality equation, or if $\partial_+ K(\epsilon)$ has a kink at w, then apply Lemma 6″ for w, $\hat{\mathbf{N}}$ and a solution $\mathcal{C}'$ which starts inside of $K(\epsilon)$ with the initial normal $\hat{\mathbf{N}}$ very close to w. This will lead to a contradiction. Therefore, the curvature of $\partial_+ K(\epsilon)$ at w must indeed be given by the optimality equation.

Finally, suppose $\partial_+ K(\epsilon)$ enters the minmax line for Player 1 at point w outside of $\mathcal{N}$. We need to show then that $\partial_+ K(\epsilon)$ is tangent to $(1, -k)$ at w. Indeed, as $K(\epsilon)$ is comprehensive, the slope of $\partial_+ K(\epsilon)$ at w must be at least as steep as $-\frac{1}{k}$. But if that slop is even steeper, then we can apply Lemma 6″ for w, $\hat{\mathbf{N}} = (\frac{k}{\sqrt{1+k^2}}, \frac{1}{\sqrt{1+k^2}})$, and a solution starting inside $K(\epsilon)$ in the vicinity of w, which would yield a contradiction. Similarly, $\partial_+ K(\epsilon)$ must enter he minmax line for Player 2 either at a point from $\mathcal{N}$, or tangent to $(-k, 1)$. To finish the proof, it remains to notice that any point from $\mathcal{N}$ on the boundary of $K(\epsilon)$ must correspond to the payoffs of some p-NE. Q.E.D.

C.2 Proof of Theorem 3.

By Lemmata 3, 4, 5, 6, and 7, we already know that the set $K(\epsilon)$ must satisfy properties 1 and 2 from Theorem 3. It remains to show the converse, if K is a bounded set satisfying properties 1 and 2, then $cl(K) \subseteq K(\epsilon)$.

Indeed, take any $w \in \partial_+ K$. Let us construct an outcome Q^0 that will satisfy the One-Stage deviation restriction in Hidden Actions and deliver to the players the payoffs equal to w. There could be three different cases.

Case 1: $w \in \mathcal{N}$. Then take Q^0 as the initial public randomization among p-NE's of the stage game that would yield w followed by the infinite repetition of the corresponding realized p-NE without money transfers. Collusion u Case 2: $w \in \partial_+ K \backslash \mathcal{N}$ and the curvature of $\partial_+ K$ is strictly positive at w. Then start Q^0 as a weak solution to representation (1) with $W_0 = w$ that moves along the curve $\mathcal{C}$, which is the solution to the optimality equation 4 with the initial condition (w, $\mathbf{N}(w)$). The underlying action profile is going to be determined as the maximizer in the optimality equation. As the volatility along $\mathcal{C} \backslash \mathcal{N}$ is uniformly bounded away from 0, the curve $\mathcal{C}$ eventually hits either a payoff from $\mathcal{N}$, or the minmax of either of the players. In the former case, concatenate the play with the subsequent randomization and indefinite play of the realized p-NE. In the later, when $\mathcal{C}$ hits the minmax line of Player i at point v, introduce the money-transfers from Player i made with the retention parameter k such that they coincide with the pushing process of W_t on $\mathcal{C}$ with the reflection boundary at v. So constructed money-transfer processes will be M-nonmanipulable for some $M > 0$. Indeed, if the reflexion happens on one side of $\mathcal{C}$, then the rate of growth of the transfers as time $t \to \infty$ is that of order $\sqrt{t}$. If the reflexion happens on both ends, the rate of growth is of order t. As there are only finitely many hidden action profiles and as the volatility of W_t is uniformly bounded on $\mathcal{C}$, there will exist a constant $C > 0$, such that for any $t > 0$, any manipulations with the drift of the public signal can not increase either of the cumulative transfers by more than Ct. Since the interest rate $r > 0$, the money-transfers processes indeed will be nonmanipulable. Thus, Q^0 will be a required public outcome. Support Q^0 by the optimal penal code from Theorem 2. This will give as a self-enforcing agreement with the payoffs w.

Case 3: $w \in \partial_+ K \backslash \mathcal{N}$, but the curvature of $\partial_+ K$ at w is 0. Then the solution to the optimality equation with the initial condition (w, $\mathbf{N}(w)$) is a straight line. As $K(\epsilon)$ is bounded, this solution has to stop somewhere. If both of the endpoints are in $\mathcal{N}$, then w can be obtained by initial public randomization of between the agreements corresponding to this two endpoints. If one of the endpoints is on minmax line on Player i, while another is in $\mathcal{N}$, then w can be obtained in agreement which first asks Player i to transfer positive amount of money so as to jump to the endpoint in $\mathcal{N}$, and then follows with the agreement corresponding to this end point. Finally, as $k \neq 1$, it is not possible for a straight solution $\mathcal{C}$ to enter both minmax lines while at the same time being parallel to $(1, -k)$ and $(-k, 1)$.

Now, if w is an end point of $\partial_+ K$ which is not in $\mathcal{N}$, we can construct an agreement by using the money transfers to push away from the minmax lines similarly to how it is done for $w \in \partial_+ K$.

Finally, any point that may be obtained from $cl(\partial_+ K)$ by subtracting the money-transfer vectors will also belong to $K(\epsilon)$ by Lemma 3.

Thus, $cl(K) \subseteq K(\epsilon)$. Q.E.D.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica 56*(2), 383–396.

Abreu, D., D. Pearce, and E. Stacchetti (1986). Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory 39*, 251–269.

Baker, G., R. Gibbons, and K. Murphy (2002). Relational contracts and the theory of the firm. *Quarterly Journal of Economics 2*, 39–84.

Bergin, J. and W. B. MacLeod (1993). Continuous time repeated games. *International Economic Review 34*(1), 21–37.

Fong, Y.-F. and J. Surti (2009). On the optimal degree of cooperation in the repeated prisoner's dilemma with side payments. *Games and Economic Behavior 67*(1), 277–291.

Fudenberg, D., D. Levine, and E. Maskin (1994). The folk theorem with imperfect public information. *Econometrica 62*(5), 997–1039.

Goldlucke, S. and S. Kranz (2012). Infinitely repeated games with public monitoring and transfers. *Journal of Economics Theory 147*(3), 1191–1221.

Goldlucke, S. and S. Kranz (2013). Renegotiation-proof relational contracts. *Games and Economic Behavior 80*, 157–178.

Green, E. J. and R. H. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica 52*(1), 87–100.

Hackbarth, D. and B. Taub (2013). Does the dearth of mergers mean more competition? *Working Paper*.

Harrington, J. E. and A. Skrzypacz (2007). Collusion under monitoring of sales. *RAND Journal of Economics 38*(2), 314–331.

Hashimoto, T. (2010). Corrigendum to "games with imperfectly observable actions in continuous time". *Econometrica 78*(3), 1155–1159.

Levin, J. (2003). Relational incentive contracts. *American Economic Review 93*(3), 835–857.

Rayo, L. (2007). Relational incentives and moral hazard in teams. *Review of Economic Studies 74*(3), 937–963.

Sannikov, Y. (2007). Games with imperfectly observable actions in continuous time. *Econometrica 75*(5), 1285–1329.

Simon, L. K. and M. J. Stinchcombe (1989). Extensive form games in continuous time: Pure strategies. *Econometrica 57*(5), 1171–1214.