Timing and Codes of Conduct*

Juan I. Block†

May 25, 2016

Abstract

In games where players can imperfectly observe an opponent's strategy, this paper examines how the timing of this information affects the equilibrium outcome set. The results show that for players to be able to obtain higher payoff in equilibrium, the stage at which players observe signals about their opponents' strategies must allow them to punish planned deviations severely but without imposing excessive punishments on the equilibrium path. A class of games in which players have an exit option, and as a consequence of executing that option the game ends, is shown to have a unique equilibrium outcome if players do not observe signals at the outset. This is in sharp contrast to finitely repeated games where signals observed arbitrarily late in the game can lead to efficiency.

Keywords: Folk theorem, self-referential game, commitment, exit game, repeated game.

JEL classification: C72, C73, D03.

*I am grateful to David Levine for his invaluable advice. I benefited from conversations with Mariagiovanna Baccara, Johannes Hörner, Anqi Li, John Nachbar, Paulo Natenzon, Ludovic Renou, Brian Rogers, Maher Said and Jonathan Weinstein. I acknowledge financial support from the Cambridge-INET Institute.

†Faculty of Economics, University of Cambridge. Email: jb2002@cam.ac.uk

1 Introduction

This paper studies situations in which players decide their strategies (codes of conduct) in a pre-play phase and are committed to them. Throughout the interaction, agents receive an informative signal about their opponent's strategy at a predetermined time. In addition to allowing players to commit, strategies also specify how players react to these signals. In industrial espionage, an entry firm would spy on the incumbent's response to market entries before expanding, and the entry firm could adjust the plan of actions after the incumbent launches a new product. Similar techniques are present in online click stream pricing which displays a price based on consumers' browsing history, and it may price a complementary product based on consumer's choice up to the checkout stage.

Standard models of behavior recognition typically suppose that information about strategies is revealed ex-ante.[1] One key insight of the existing literature is that a folk theorem for normal-form finite games holds. However, in dynamic games and the above applications this information may naturally arrive along the path of play. The model I propose accounts for this flexibility, and I show that for codes of conduct to expand the equilibrium outcome set, the value of adhering to the code of conduct needs to be greater than the planned deviation plus a punishment depending on the timing structure.

The framework I study is self-referential games (Block and Levine (2015)). The self-referential game is modeled as an extension of a two player multistage game where players observe the actions chosen at previous stages and may move simultaneously. Players also observe an upfront private signal at some stage of the game so an extended strategy is defined by public and private histories. In the self-referential game a strategy is then a code of conduct, which commits the player to an extended strategy and specifies one for his opponent. Players simultaneously choose a code of conduct and employ the corresponding extended strategy. Privately observed signals are drawn from a probability distribution and are determined by the code of conduct profile. This stochastic process captures both ideas that strategies are imperfectly observable and that players employing similar codes of conduct can recognize one another. Because the monitoring technology endows player with the ability to imperfectly distinguish whether rivals agree on the code of conduct, the basis to construct equilibria is simple trigger strategies.

In contrast to the case of pre-game signals studied by Block and Levine (2015), codes of conduct do not expand the equilibrium outcome set in splitting games despite players' ability to commit. In this class of finite-horizon games, players simultaneously decide whether to stay or exit from a joint partnership in every period, but once a player takes the exit option the game ends. Not taking such an option requires cooperation since late joint exit gives a higher payoff than early joint exit, whereas if in a period a player stays in the game and the opponent leaves, he incurs a large cost. The equilibrium set with outcomes where all players immediately exit is rendered unique by signals that are observed after the first period because the termination payoffs upon exit are increasing in the length of the game, but both the temptation deviation and the size of the punishment only

[1] See, for example, Tennenholtz (2004), Levine and Pesendorfer (2007), Kalai, Kalai, Lehrer, and Samet (2010), Kamada and Kandori (2011), and Block and Levine (2015).

depend on the initial period. A closely related class of games is preemption games, where players move alternately so that just one player can exit in each stage. In this class, a flexible timing structure may enlarge the equilibrium outcome set since the value of the exit option is increasing in the stage at which it is executed, creating an incentive to wait and delaying the planned deviation; while the punishment can be made sufficiently severe to deter such deviations—as long as the player terminating the game observes signals one active period preceding exit. However, codes of conduct do not require that there be signals at the outset as in Block and Levine (2015). The key difference is that players can inflict harsher punishments to potential deviations the sooner the signals can be observed because punishments are not costly to be carried over.

When players observe signals just before the start of the game, there could be costly on the equilibrium punishments in repeated games because the immediate penalty occurs in the first period and the punishment phase lasts the entire game, allowing agents to inflict severe punishments on the deviators. Unlike Block and Levine (2015), when relaxing the timing structure the punishment phase now starts early relative to when the opponent finds it optimal to deviate, ensuring players do choose the equilibrium code of conduct. Because punishments are triggered after a fixed threshold of remaining periods, the proportion of times punishments may occur on the equilibrium path becomes arbitrarily as the time horizon grows, indeed, any feasible and individually rational payoff can be approximately attained in equilibrium.

Finally, suppose that players receive signals only in the pre-play phase, thereby allowing them to punish any kind of potential deviation. It is shown that for any subgame perfect equilibrium of an infinite horizon game, there exists a Nash equilibrium of the self-referential truncation that coincides with such a subgame perfect equilibrium.

In the context of an evolutionary model Levine and Pesendorfer (2007) study self-referential games where players are pairwise matched to play a symmetric game and are able to identify behavior prior to play. They show that long-run stable strategies reward opponents that are likely to play similarly and punish otherwise. In a similar spirit, Bachi, Ghosh, and Neeman (2014) explore betrayal of intentions. Agents choose between a real and a deceptive strategy by incurring some cost, thereby showing a folk theorem for a sufficiently small cost of deception.

Codes of conduct have similar characteristics to conditional commitment devices, i.e. agents choose a device that conditions on other players' device. These devices support a folk theorem for complete information games (Tennenholtz (2004) and Kalai et al. (2010)) and for Bayesian games (Peters and Szentes (2012)). In particular, Tennenholtz (2004) develops a setting where players submit programs which take as input the other players' program and play on their behalf. At a more general level, Kalai et al. (2010) characterize the conditional commitment devices space. While they assume that these devices are perfectly observable, I analyze behavior in noisy environments. The current approach is also related to mechanism design problems in which agents have the ability to use self-referential contracts (Peters (2015) and Szentes (2015)).

This paper also contributes to the literature that studies the possibility of observing and revising strategies before the actual play of the game. For instance, Matsui (1989) considers infinitely repeated games in which players may observe opponents' strategy with small probability. He shows that any subgame perfect equilibrium payoff vector is Pareto efficient. More recently, Kamada and Kandori (2011) propose revision games in which agents stochastically revise actions, and prepared actions are mutually observable and implemented at a predetermined time. They find that the subgame perfect equilibrium set widens while Calcagno, Kamada, Lovo, and Sugaya (2014) argue that revision games narrow down the set of equilibrium payoffs in common and opposing interest games.[2] In contrast, in this paper strategies are imperfectly observable via signals with deterministic arrival time, and the time might be during the game.

2 Setup

A two-player multistage game with observed actions (Fudenberg and Tirole (1991), Sect. 3.3) is characterized by a tuple $\Gamma = \langle \mathcal{I}, H, Z, (A_i)_{i\in\mathcal{I}} \rangle$, where $\mathcal{I} = \{1, 2\}$ is the set of players.[3] Each player i has a finite space of actions A_i. Let $h^0 = \emptyset$ be the initial public history and $a_i^0 \in A_i(h^0)$ be the finite action set available for player i at stage 0. The public history of play until stage t is defined recursively as a sequence of action profiles denoted by $h^t = (a^0, a^1, \ldots, a^{t-1})$. Player i chooses an action a_i^t from his finite action set $A_i(h^t)$ at stage t with profile $a^t \in A(h^t)$. Without loss of generality assume $A_i(h^t) \neq \emptyset$ for each h^t, player i is inactive if $|A_i(h^t)| = 1$, and active otherwise. Let H^t be the set of all stage t public histories and $H = \bigcup_{t=0}^{T} H^t$ be the set of all public histories. The set Z contains all terminal histories. A behavioral strategy for player i is a map $\sigma_i : H \to \Delta A_i(h^t)$ where $\mathcal{S}_i^t \equiv \Delta A_i(h^t)$ for notational convenience. Let the set Σ_i denote behavioral strategies with profile $\sigma \in \Sigma$.

The reward function for player i is $g_i : H \to \mathbb{R}$ where payoff $g_i(h^t)$ is awarded after history h^t at stage $t-1$ and players have discount factor $\delta \in (0, 1]$. Denote by A^{T+1} the set of possible outcomes with generic element a^{T+1}. The outcome path induced by pure strategy profile σ is denoted by $a^{T+1}(\sigma)$. Player i's payoffs as a function of pure strategy profile, $u_i : \Sigma \to \mathbb{R}$, is $u_i(\sigma) = \sum_{t=0}^{T} \delta^t g_i(a^t(\sigma))$. Payoffs are extended to behavioral strategy profile in the standard way and denoted by $u_i(\sigma)$. Finally, let Γ^∞ stand for the case $T = \infty$.

I present the self-referential game in Block and Levine (2015) with a flexible timing structure. The self-referential game $\mathcal{G}(\Gamma) = \langle \Gamma, \tau, Y, \pi, R \rangle$ couples with a multistage game with observed actions Γ. In the beginning of stage τ, which is deterministic and commonly known, every player i observes a signal y_i from a finite set Y_i.[4] Let H_i^t be the set of all stage t private histories of player i with element h_i^t. It follows that $H_i^t = \emptyset$ for all stages $t < \tau$, and $H_i^t \subset Y_i$ for all stages $t \geq \tau$. Let $H_i = \bigcup_{t=0}^{T} H_i^t$ denote the set of all private histories, and if $\overline{Y}_i \subset Y_i$ then $\overline{H}_i^t \subset \overline{Y}_i$ is accordingly defined. An *extended strategy* for player i is a map $s_i : H \times H_i \to \mathcal{S}_i^t$ and S_i is player i's finite extended strategy set. Let $s \in S$ be a profile of extended strategies.

In the self-referential game, a strategy for player i, r_i, is called a *code of conduct* which is a

[2] Lovo and Tomala (2015) demonstrate the existence of Markov perfect equilibria in stochastic revision games.

[3] For finite set X, let $\Delta(X)$ be the set of probability distributions on X. For list of sets $X_1, \ldots, X_N$, I write $X \equiv \times_i X_i$ with typical element $x \in X$, and $X_{-i} \equiv \times_{j\neq i} X_j$ with element x_{-i}.

[4] The set of signals Y_i parameterizes the information accumulated up to stage τ: players accrue pieces of information during the game, and at some point they make use of them to evaluate adversaries' strategies.

choice of two extended strategies, $s_j^i : H \times H_j \to \mathcal{S}_j^t$ denoting what player i assigns to player j's choice of extended strategies. For notational simplicity $s_i = s_i^i$. Each player i is endowed with the common space of codes of conduct R_0 given by

$$R_0 \equiv \left\{ r_i \middle| s_j^i \in \mathcal{S}_j^{t\,H \times H_j} \text{and } \forall i, j \in \mathcal{I}, \forall h^t \in H, \forall h_j^t \in H_j, s_j^i(h^t, h_j^t) \in \mathcal{S}_j^t \right\},$$

where $\mathcal{S}_j^{t\,H \times H_j}$ is the set of functions with domain $H \times H_j$ and range $\mathcal{S}_j^t$. Note well that r_i commits player i to an extended strategy. A code of conduct vector is denoted by $r \in R = \times_i R_0$.

For each code of conduct profile $r \in R$, let $\pi(\cdot|r)$ be the probability distribution over signal profiles Y. Define the *monitoring structure* (Y, π) to be the collection of probability distributions over private signal profiles $\{\pi(\cdot|r) \in \Delta(Y) : r \in R\}$. For each r, $\pi_i(\cdot|r)$ denotes the marginal distribution of $\pi(\cdot|r)$ over Y_i, that is, the probability that player i receives signal y_i under code of conduct profile r. Player i's expected payoff $U_i : R \to \mathbb{R}$ is $U_i(r) = \mathbb{E}_r \sum_{t=0}^{T} \delta^t g_i(a^t(s)) = \mathbb{E}_r u_i(s)$.

Each player i chooses simultaneously a code of conduct, $r_i \in R_0$, and then plays $s_i(h^t, h_i^t) \in \mathcal{S}_i^t$ for histories $h^t \in H$, $h_i^t \in H_i$. A vector of codes of conduct $r = (r_1, r_2)$ is a Nash equilibrium of $\mathcal{G}$ (or a self-referential equilibrium) if for all players i and all $r_i' \neq r_i$, $U_i(r) \geq U_i(r_i', r_{-i})$.

In order to distinguish the effect of the timing τ from that of the informativeness of the signals (Y, π), here the monitoring technology is the same as in Block and Levine (2015).[5] Specifically, the monitoring structure (Y, π) allows η-λ detection if for constants $\eta, \lambda \in [0, 1]$, for all players i there exist a $j \neq i$ and $\overline{Y}_j \subset Y_j$ such that for all $r \in R$, $y_j \in \overline{Y}_j$ and $\tilde{r}_i \neq r_i$, $\pi_j(y_j|\tilde{r}_i, r_{-i}) - \pi_j(y_j|r) \geq \eta$ and $\pi_j(y_j|r) \leq \lambda$. Throughout I focus on self-referential games that satisfy this property and $\eta > 0$.

3 Timing of Signals in Games with Exit Option

In exit games, $T < \infty$. For each player i, $A_i(h^t) \equiv \{F_i(h^t) \cup E_i(h^t)\}$ for $h^t \in H$, and $A_i(h^t) = \{\overline{a}\}$ when inactive. Action $\overline{a}$ stands for the *no decision* action. The subset $F_i(h^t)$ represents *forward* actions which are necessary for moving to the next stage but not sufficient. For any history h^t and any player i, $A_i(h^t) \neq \emptyset$ if $a_j^k \notin E_j(h^k)$ for all $k \leq t-1$ and all active players j. The other component, $E_i(h^t)$, represents *exit* actions where, in contrast to forward actions, the game ends if at least one player chooses one of these actions. If any active player i chooses $e_i^t \in E_i(h^t)$, it causes the game to end regardless of $a_{-i}^t \in A_{-i}(h^t)$. Note that $E_i(h^t) \cap F_i(h^t) = \emptyset$.

The last common feature is related to reward mappings: $g_i(h^t) = 0$ for all histories $h^t \in H$ such that $a_i^k \notin E_i(h^k)$ for all $i, k \leq t-1$. In addition, if some player j chooses e_j^t at stage t, then $g_i(h^k) = 0$ for all $k > t$ and all players i. If all players continue until and including the last stage T, the game ends and the players' payoffs are normalized to be zero. Players are assumed to only choose pure strategies.

[5] It allows players to discern whether rivals adhere to the same code of conduct. It can be interpreted as agents using simplified categorization of behavior, aiming a specific behavior while bundling all deviations into a single class.

3.1 Deviation and Punishment at the Outset

Having established the general framework of exit games, this section studies splitting games. In splitting games Γ, players have discount factor $\delta \in (0,1)$. Let $|A_i(h^t)| \geq 2$ for each player i. Reward functions are additively separable in surplus share and costs, that is, $g_i(a^t) = w_i(a^t) - c_i(a^t), \forall a^t \in A(h^t)$. Player receives a surplus share $w_i(a^t)$ depending on his action and on opponent's. Similarly, each player incurs a cost of $c_i(a^t)$ by taking his share. For all i and $h^t \in H$ these preferences are characterized by

S.1 $w_i(e_i^t, \cdot) = \theta_i w^t > 0$ for $e_i^t \in E_i(h^t)$, and $w_i(f_i^t, \cdot) = 0$ for $f_i^t \in F_i(h^t)$;

S.2 $c_i(e_i^t, e_j^t) = \theta_i c > 0$ for $e_i^t \in E_i(h^t), e_j^t \in E_j(h^t)$, otherwise $c_i(\cdot) = 0$.

The surplus share w^t has present value w at any period t, player i's share weight is $\theta_i \in (0,1)$ where $\theta_i = 1 - \theta_j$, and the cost c is constant. The parameters w and c are assumed to satisfy $\lambda < \frac{w}{c} < \eta$. Intuitively, condition S.1 ensures that players would prefer to exit the game before his opponents rather than play a forward action. The constant w should be thought of as a steady state surplus. Condition S.2 in turn establishes that if all players decide to exit the game simultaneously in period t, they will pay a cost equal to $\delta^t \theta_i c$ in the beginning of the game. For each of the complement action profiles players pay zero. The constant c represents cost of reaching agreement. Together assumptions S.1 and S.2 imply that terminating the game late is a cooperative action, and hence the focus is on long time horizon T.

The next theorem says that if players receive signals about opponents' code of conduct in any period but period $t = 0$, there is, in fact, one self-referential equilibrium outcome in which every player immediately exits and so cannot attain higher payoffs.

Theorem 1. *Let $\tau \geq 1$. Then there exist a T^* and a unique self-referential equilibrium outcome in which all r have every player i choosing $s_i(h^t, h_i^t) = e_i^t$ for some $e_i^t \in E_i(h^t)$ and all $h^t \in H, h_i^t \in H_i$ and $T \geq T^*$.*

The surplus to be distributed when the partnership is jointly dissolved increases with the targeted time horizon T^*, but there is a uniform deviation in the sense that players find it optimal to exit in the first period irrespective of T^*. To induce the player to stay in the game, code of conduct must punish the player following a signal indicating deviation. This punishment could come in one of two forms: through an immediate penalty or from moving to a punishment phase that gives low expected payoff in the future. The problem here, unlike the model in Block and Levine (2015), is that both punishments are not feasible because the deviator terminates the game before they can be implemented. Notice that this means that the result holds even in the limit case of perfectly observed codes of conduct.

Theorem 2. *Suppose $\tau = 0$. Then there exist a T^* and a self-referential equilibrium r in which all players choose $s_i(h^T, h_i^T) = e_i^T$ for all $e_i^T \in E_i(h^T), h^T = (f^0, f^1, \ldots, f^{T-1}), h_i^T \notin \overline{H}_i$ and $T \geq T^*$.*

Despite the incentives to exit in the first period, the equilibrium code of conduct strikes a balance between the benefits from exiting the partnership with the opponent and the costs of dissolving the partnership at the end of the game. Because the opponent's code of conduct is observed imperfectly, however, the selected codes of conduct would call for punishments to be implemented on the equilibrium path but such punishments are not costly to be carried over. The payoff from joint exiting is increasing over the entire game, which generates incentives for cooperation, and leads to the trade-off between letting the value of the partnership grow and increasing the chances of being preempted by the opponent. But the payoff from exiting first is constant over time so players would exit in the first period. With the ability to be conditionally cooperative on the signals observed before the game starts, players can coordinate on dissolving the partnership in the last periods of the game thereby increasing their expected payoffs.

Example 1. Consider an elaboration of Mailath, Postlewaite, and Samuelson (2005)'s Example 2. Two players, 1 and 2, engage in a T finite horizon partnership and have discount factor $\delta \in (0,1)$. The partnership generates a payoff of $w_t = 2\delta^{-t}$ in period t. In each period t, they simultaneously choose to either *stay* or *exit*. The flow payoffs are as follows. If they both exit, each player obtains $\frac{1}{2}(2\delta^{-t} - 3)$. Whenever a single player decides to exit, that player obtains $\frac{1}{2}2\delta^{-t}$ and his opponent receives zero. As long as none of the players exits they receive zero payoffs. Once a player exits, the game ends, and players receive a continuation value of zero. Likewise, if players choose to stay in every period, they receive nothing.

In period τ, each player privately observes y_i in $\{\mathtt{G}, \mathtt{B}\}$. All the maps from histories to actions is $R_0 \equiv \left\{ r_i \mid s^i_j \in \{stay, exit\}^{H \times H_j} \text{ and } \forall i, j, \forall h^t \in H, \forall h^t_j \in H_j, s^i_j(h^t, h^t_j) \in \{stay, exit\} \right\}$. Assume $\pi(y|r) = \pi_1(y_1|r)\pi_2(y_2|r)$ for all $r \in R$ and $y \in \{\mathtt{G}, \mathtt{B}\}^2$. Let $\pi_i(y_i = \mathtt{B}|r_1, r_2) = q$ for $r_1 \neq r_2$, and $\pi_i(y_i = \mathtt{B}|r_1, r_2) = p$ for $r_1 = r_2$ where $p < 2/3 < q$. Consider r^*_i that satisfies

$$s^i_j(h^t, h^t_j) = \begin{cases} stay & \text{for } t = 0 \text{ or } h^k = (f^0, \ldots, f^{k-1}) \text{ and } h^k_j = \mathtt{G} \text{ for } 0 \leq k \leq T-1, \\ exit & \text{otherwise.} \end{cases}$$

Observe that every player receives an expected payoff of $-(1-p)(p + (1-p)3\delta^T) - \frac{3}{2}p^2\delta^\tau$ under r^*. For any $\tau \geq 1$ and T, player i may choose the code of conduct $\tilde{r}_i$ such that $\tilde{s}^i_j(h^t, h^t_j) = exit$ for all h^t, h^t_j, thereby obtaining an expected payoff of 1. Hence, in any self-referential equilibrium all players exit in the first period. Suppose now that $\tau = 0$ and that for all $\varepsilon > 0$ choose T^* such that $2 - 3\delta^T \geq \varepsilon$ for $T \geq T^*$. In this case, the gain to choosing $\tilde{r}_i$ instead of r^*_i is bounded by $(1-p)^2\varepsilon + p(1 - \frac{3}{2}p) - 1 + \frac{3}{2}q \leq 0$; therefore r^* is a self-referential equilibrium.

3.2 Late Deviation and Early Punishment

In this section, I examine preemption games in which players alternate veto power to terminate the game. The preemption game Γ runs from stage $t = 0$ to the odd finite stage T. Player 1 moves first and the game ends with player 2 choosing an action. There is no discounting $\delta = 1$. At each stage t, there is only one player active. Let $\iota : H \setminus Z \to \mathcal{I}$ be the player function that assigns to each history $h \in H \setminus Z$ a player i. That is, player $\iota(h^t) = i$ makes a choice from $A_i(h^t)$ after history h^t in stage t. The function $\phi : \mathbb{N} \to \mathbb{N}$ returns the stage k at which there are n alternations, that is, $\phi(n) = k$. Let $\overline{n}$ be the maximum number of shifts in Γ, and assume $\overline{n} \geq 3$. For all $i \neq j$ and any stage t the reward mappings are required to fulfill two conditions:

P.1 $g_i(e_i^t, \overline{a}) < g_i(e_i^{t+k}, \overline{a})$ for $e_i^t \in E_i(h^t), e_i^{t+k} \in E_i(h^{t+k})$;

P.2 $g_i(e_i^t, \overline{a}) > g_i(\overline{a}, e_j^{t+k})$, $g_i(\overline{a}, e_j^t) < g_i(\overline{a}, e_j^{t+k})$ for $e_i^t \in E_i(h^t)$, $e_j^t \in E_j(h^t)$, $e_j^{t+k} \in E_j(h^{t+k})$.

The first condition P.1 guarantees that if player $i = \iota(h^k)$ is active for $t \leq k \leq t'$, then his choice of ending the game in any stage k before stage t' is strictly dominated by the choice of ending it at stage t'. Nevertheless, players face a trade-off between waiting to the next active period and terminating the game in the current active stage determined by condition P.2. Any player prefers to choose an exit action in the next active stage, but in order to reach it, he will go through an inactive stage which is threatened by having the opponent exiting the game.

The next result establishes that the unique self-referential equilibrium outcome exhibits player 1 finishing the game at the end of his first active period if signals are observed in the penultimate active period.

Theorem 3. *Assume that $\tau \geq \phi(\overline{n} - 1)$. Then there exist a unique self-referential equilibrium outcome in which all r have every player i choosing $s_i(h^k, h_i^k) = e_i^k$ for some $e_i^k \in E_i(h^k)$ and all $h^k \in H, h_i^k \in H_i$ and $k = \phi(i) - 1$ for player i.*

Each player faces the tradeoff between continuing and stopping when they are active, but the player finds it optimal to preempt the opponent depending on the stage of the game. If players were to choose a code of conduct characterized by exiting at some specific stage, the inactive player at such stage would have the incentives to stop during his preceding active period so that code of conduct must implement a punishment during the active period before that. As long as $\tau \geq \phi(\overline{n} - 1)$, such punishments are not feasible since a player that has exited terminates the game. In contrast to Block and Levine (2015), this signal timing pushes each player to exit too early in equilibrium, as players do not observe signals about the opponent's code of conduct while the game continues. In equilibrium, each player therefore exits at the end of their first active period.

The following result characterizes a self-referential equilibrium in which the game is terminated in the last stage only insofar as signals are observed relatively early to that stage.

Theorem 4. *Let $\tau \leq \phi(\overline{n} - 1) - 1$, then there exist a T^* and a self-referential equilibrium r where all players choose $s_i(h^t, h_i^t) = f_i^t$ for all $f_i^t \in F_i(h^t)$, $h^t \in H^t$ in which all active players have chosen a forward action, $h_i^t \notin \overline{H}_i$, for all $t \leq T - 1$, and player 2 exits in period $T \geq T^*$.*

The equilibrium code of conduct is constructed by finding the optimal stopping time T^* in which player 2 terminates the game. Given $\overline{n}$, a larger threshold T^* increases the active period of player 2. At T^*, player 2 would always stay in the game until the final active period since the expected payoff is higher at T^* than at the end of her penultimate active period. However, player 1 values the option of not ending the game less than his opponent, so that player 1 has the incentives to exit in his last active period. Since the timing satisfies $\tau \geq \phi(\overline{n}-1)$, the equilibrium code of conduct calls player 2 for punishing such planned deviation during her penultimate active period discouraging player 1 from exiting the game. Thus, players obtain higher expected payoffs in equilibrium because both the deviation and the punishment are not uniformly determined by initial-stage incentives.

4 Timing of Signals in Repeated Games

In the stage game each finite action set A_i is independent of history. Let $\alpha_i \in \Delta(A_i)$ be a mixed strategy, and payoffs are extended to α_i accordingly. Denote player i's minmax payoff in the stage game as $\underline{v}_i \equiv \min_{\alpha_j \in \Delta(A_j)} \max_{a_i \in A_i} g_i(a_i, \alpha_j)$. Then take action $\underline{\alpha}_j \in \Delta(A_j)$ such that $\underline{v}_i \equiv \max_{a_i \in A_i} g_i(a_i, \underline{\alpha}_j)$, where $\underline{\alpha}_j$ is the action that gives the minmax payoff to player i. Let $V \equiv \text{co}\,\{(v_1, v_2) : \exists a \in A, \forall i,\ g_i(a) = v_i\}$ and $V^* \equiv \text{int}\,\{(v_1, v_2) \in V : \forall i,\ v_i > \underline{v}_i\}$. V is the set of feasible payoff vectors and V^* is the set of feasible and strictly individually rational payoff vectors.[6] There is a public randomization device.

The T finitely repeated game with perfect monitoring Γ is the T-repetition of the stage game. In each period t players simultaneously choose $a_i \in A_i$, and after period T the game ends. Let A^t be the t-fold Cartesian product of A, then $H^t = A^t$. Given any σ, the average payoffs for player i, can be written as $u_i(\sigma) = \frac{1}{T}\sum_{t=1}^{T} g_i(a^t(\sigma))$.

The next result states that the set of self-referential equilibria payoffs approximately coincides with the set of feasible and strictly individually rational payoffs of the stage game. Similar folk theorems have been established (see e.g. Benoît and Krishna (1987), González-Díaz (2006), and Renou (2009)).

Theorem 5. *For any $v \in V^*$ and all $\varepsilon > 0$, there exist a T^* and a self-referential equilibrium r so that each player i's expected payoff satisfies $|U_i(r) - v_i| \leq \varepsilon$ for any $T \geq T^*$ and $\tau \leq T - T^*$.*

In equilibrium, trigger codes of conduct require players to punish the opponent based on the observed signal. Such punishment consists of an immediate penalty plus an additional punishment phase until the end of the game. Since signals are noisy there could be punishments on the equilibrium path, and potentially those punishments could be very costly for the punisher with the timing of Block and Levine (2015).[7] However, players find it optimal to deviate within some fixed number of stages before the game ends, and it, in turn, implies that the cost of punishing can be small by allowing a flexible timing structure. Note that the value of equilibrium code of conduct increases with the time horizon T, whereas the punishment phase remains fixed. Therefore, as T grows large the expected payoff of players is very close to v_i and the non-punishment phase virtually consumes the entire game.

[6] co denotes the convex-hull operator and int$X$ stands for the topological interior of a set X.

[7] On the equilibrium path punishments are not a problem with perfectly observed conditional commitment devices assumed in Kalai et al. (2010) and Tennenholtz (2004).

	C	D
C	$5, 5$	$0, 6$
D	$6, 0$	$1, 1$

Figure 1: Prisoner's dilemma.

Proposition 1. *Suppose $\lambda = 0$. For any $v \in V^*$ there exist a T^* and a self-referential equilibrium r where every player i obtains v_i for any $T \geq T^*$ and $\tau \leq T - T^*$.*

Example 2. To illustrate the flexibility of the signal timing in the self-referential approach consider the example studied in Block and Levine (2015). Specifically, Γ is a T finitely repeated prisoner's dilemma in which each agent chooses between cooperate (C) or defect (D) in each period. Figure 1 represents players' stage game payoffs. Note that (D, D) is the unique Nash equilibrium of the finitely repeated game. Embed this game with a self-referential information structure which consists of a tuple $\langle \tau, Y, \pi, R \rangle$. Suppose that each player privately observes a signal from the set $Y_i = \{\mathtt{G}, \mathtt{B}\}$. Let R_0 be given by all the maps from histories to actions

$$R_0 \equiv \left\{ r_i \mid s^i_j \in \{C, D\}^{H \times H_j} \text{ and } \forall i, j, \forall h^t \in H, \forall h^t_j \in H_j, s^i_j(h^t, h^t_j) \in \{C, D\} \right\}.$$

The signal structure satisfies $\pi(y|r) = \pi_1(y_1|r)\pi_2(y_2|r)$ for all $r \in R$ and $y \in \{\mathtt{G}, \mathtt{B}\}^2$. Let the marginal probability distribution for each i be such that $\pi_i(y_i = \mathtt{B}|r_1, r_2) = q$ for $r_1 \neq r_2$, and $\pi_i(y_i = \mathtt{B}|r_1, r_2) = 0$ for $r_1 = r_2$. Signal $\mathtt{B}$ is more likely when players disagree on behavior. Given this self-referential game, Block and Levine (2015) constructed an equilibrium where cooperation could arise when $\tau = 1$. Suppose r^*_i satisfies

$$s^i_j(h^t, h^t_j) = \begin{cases} C & \text{for } t = 1 \text{ or } h^k = (a^{*1}, \dots, a^{*k-1}) \text{ and } h^k_j = \mathtt{G} \text{ for } 1 \leq k \leq t-1, \\ D & \text{otherwise.} \end{cases}$$

Hence, r^* yields an expected payoff of 5 for each player. Player i may choose an alternative code of conduct $\tilde{r}_i$ such that $\tilde{s}^i_j(h^t, h^t_j) = C$ if $t = 1$ or $h^k = (a^{*1}, \dots, a^{*k-1})$ and $h^k_j \neq \mathtt{B}$ for $2 \leq k \leq t-1$ and $t < T$, and $\tilde{s}^i(h^T, h^T_i) = D$ for any h^T, h^T_i, yielding an expected payoff of $\frac{1}{T}(5T + 1 - q(4T + 2))$. Given the proposed strategy profile r^*, nobody wants to deviate if $q \geq \frac{1}{4T+2}$. The idea of this construction is that the play in the prisoner's dilemma game depends upon what players have observed about their opponents' strategy. Note that, in the beginning of the game, players receive information about a potential defection that would occur in the last period. Suppose that $\tau = T$ and that players choose r^*_i. Again, players' expected payoff is 5, however, $\tilde{r}_i$ now yields $\frac{1}{T}(5T + 1 - q5)$. Yet, r^* is a self-referential equilibrium if $q \geq \frac{1}{5}$. For any $v \in V^*$, Proposition 1 shows that signals may be observed sufficiently early in the game.

5 A Partial Characterization with Pre-Game Signals

The perfect equilibrium set was characterized by Radner (1980), and Fudenberg and Levine (1986) for infinite horizon games where future events are uniformly less important.[8] These characterizations are crucial to reconcile the difference between the number of equilibria in an infinite horizon game and in its finite horizon version. Similar to these papers I work with a finite approximation to the infinite horizon game but the focus is on self-referential equilibria in this approximation instead of ε-Nash equilibria. Agents are endowed with pre-game signals and this in turn helps to develop a general treatment, encompassing the large class of multistage games with observed actions.

Let $\mathcal{E}(\Gamma^\infty)$ denote the set of subgame perfect equilibrium in Γ^∞. For any $T < \infty$, Γ^T represents the T-truncation of Γ^∞ with time horizon truncated at T, where we require players to be inactive after T. Apply the truncation of length T on σ to obtain the partial strategy σ_T from σ. Of particular interest are games in which distant future events are uniformly less important, for instance, repeated games with discounting and any finite horizon game. Formally, an infinite horizon game Γ^∞ is said to be continuous at infinity (Fudenberg and Levine (1983)) if for any $\varepsilon > 0$ there exists some $k < \infty$ such that $|u_i(\sigma) - u_i(\hat{\sigma})| < \varepsilon$ if $\sigma_k = \hat{\sigma}_k$ for all i and all $\sigma, \hat{\sigma} \in \Sigma$. A result of Fudenberg and Levine (1983, Theorem 3.3) guarantees that a subgame perfect equilibrium in finite action game exists.

The partial characterization holds for all multistage games with observed actions that are continuous at infinity, and it says that we can reconstruct any subgame perfect equilibrium of the infinite horizon game in its self-referential finite truncation if players are likely to detect disagreement on codes of conduct in the beginning of the game (as in Block and Levine (2015)).

Theorem 6. *Let Γ^∞ be continuous at infinity and take any Γ^T. Suppose that $\lambda = 0$ and $\tau = 0$. For any $\sigma \in \mathcal{E}(\Gamma^\infty)$, there exist a T^* and a self-referential equilibrium r such that $s_i = \sigma_{i,T}$ for all i and any $T \geq T^*$.*

Existing results on the connection between infinite and finite horizon games has focused on perfect ε equilibria of the finite truncation Γ^T.[9] In contrast, Theorem 6 presents a relationship between $\mathcal{E}(\Gamma^\infty)$ and the exact equilibria of $\mathcal{G}(\Gamma^T)$. It can be also interpreted as a lower hemi-continuity result to the extent that exact equilibria of the self-referential game approach the limit point.

6 Conclusion

This paper has developed a model that allows players to receive information about opponents' strategy along the path of play and highlights the role of this information in characterizing the equilibrium outcome set. The results provide new insights as to when the ability to commit conditional on the opponent's strategy, encoded in the self-referential property of codes of conduct, may lead to an increase in equilibrium payoffs.

[8] See also Harris (1985), Börgers (1989), and Mailath et al. (2005).

[9] A strategy profile $\hat{\sigma}$ is a perfect ε-equilibrium if it is an ε-Nash equilibrium after any history h^t.

Given a candidate equilibrium code of conduct, the key interplay is between when players find it optimal to deviate and the ability to punish such planned deviations sufficiently in advance. Since the environment considered here is noisy, there is also a trade-off between early and late punishments based on signals. It has been shown that signals at the outset are important to construct punishment phases because it allows players to impose harsher punishments on potential deviator, especially if the deviator may terminate the game. On the other hand, delaying the time at which players observe signals can reduce costly on the equilibrium punishments.

References

Bachi, B., Ghosh, S., and Neeman, Z. (2014). Communication and deception in 2-player games. (Mimeo)

Benoît, J.-P., and Krishna, V. (1987). Nash equilibria of finitely repeated games. *International Journal of Game Theory*, *16*(3), 197-204.

Block, J. I., and Levine, D. K. (2015). Codes of conduct, private information and repeated games. *International Journal of Game Theory*, forthcoming.

Börgers, T. (1989). Perfect equilibrium histories of finite and infinite horizon games. *Journal of Economic Theory*, *47*(1), 218–227.

Calcagno, R., Kamada, Y., Lovo, S., and Sugaya, T. (2014). Asynchronicity and coordination in common and opposing interest games. *Theoretical Economics*, *9*(2), 409-434.

Fudenberg, D., and Levine, D. K. (1983). Subgame-perfect equilibria of finite- and infinite-horizon games. *Journal of Economic Theory*, *31*(2), 251–268.

Fudenberg, D., and Levine, D. K. (1986). Limit games and limit equilibria. *Journal of Economic Theory*, *38*(2), 261–279.

Fudenberg, D., and Tirole, J. (1991). *Game theory.* Cambridge, MA: The MIT Press.

González-Díaz, J. (2006). Finitely repeated games: A generalized nash folk theorem. *Games and Economic Behavior*, *55*(1), 100–111.

Harris, C. (1985). Existence and characterization of perfect equilibrium in games of perfect information. *Econometrica*, *53*(3), 613–28.

Kalai, A. T., Kalai, E., Lehrer, E., and Samet, D. (2010). A commitment folk theorem. *Games and Economic Behavior*, *69*(1), 127–137.

Kamada, Y., and Kandori, M. (2011). *Revision games* (Mimeo).

Levine, D. K., and Pesendorfer, W. (2007). The evolution of cooperation through imitation. *Games and Economic Behavior*, *58*(2), 293–315.

Lovo, S., and Tomala, T. (2015). *Markov perfect equilibria in stochastic revision games* (Mimeo).

Mailath, G. J., Postlewaite, A., and Samuelson, L. (2005). Contemporaneous perfect epsilon-equilibria. *Games and Economic Behavior*, *53*(1), 126–140.

Matsui, A. (1989). Information leakage forces cooperation. *Games and Economic Behavior*, *1*(1), 94–115.

Peters, M. (2015). Reciprocal contracting. *Journal of Economic Theory*, *158, Part A*(0), 102-126.

Peters, M., and Szentes, B. (2012). Definable and contractible contracts. *Econometrica*, *80*(1), 363–411.

Radner, R. (1980). Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. *Journal of Economic Theory*, *22*(2), 136-154.

Renou, L. (2009). Commitment games. *Games and Economic Behavior*, *66*(1), 488–505.

Szentes, B. (2015). Contractible contracts in common agency problems. *Review of Economic Studies*, *82*(1), 391-422.

Tennenholtz, M. (2004). Program equilibrium. *Games and Economic Behavior*, *49*(2), 363–373.

Appendix A Proofs

Proof of Theorem 1. By way of contradiction, fix an arbitrary period $\tau = \hat{t}$ with $1 \leq \hat{t} \leq \uptau$ and suppose players aim to exit in period $\uptau$. Then consider $\hat{r}_i$ characterized by $\hat{s}^i_j(h^t, h^t_j) = f^t_j \in F_j(h^t)$ for $t = 0$ or $h^t = (f^0, f^1, \ldots, f^{t-1})$ and $h^t_j \notin \overline{H}^t_j$ for $0 \leq t \leq \uptau$, otherwise, $\hat{s}^i_j(h^t, h^t_j) = e^t_j \in E_j(h^t)$. The lowest expected payoff for any i, $\underline{W}_i(\hat{r})$, from choosing $\hat{r}_i$ gives

$$\underline{W}_i(\hat{r}) = (1-\lambda)^2\theta_i(w - \delta^{\uptau}c) + \lambda\theta_i(w - \lambda c\delta^{\hat{t}}).$$

Let $\tilde{r}_i$ be the optimal code of conduct against $\hat{r}$. By S.1 and S.2, $\tilde{s}_i(h^t, h^t_i) = e^t_i$ for all $h^t \in H, h^t_i \in H^t_i$ and some $e^t_i \in E_i(h^t)$; for all players $j \neq i$, it says $\tilde{r}_j = \hat{r}_j$. Player i obtains $\theta_i w$ by choosing $\tilde{r}_i$. Pick T^* such that $w - \delta^{T^*}c \geq \varepsilon$ for $\varepsilon > 0$. Observe that all alternative codes of conduct will require deviation in the first period of the game since

$$\begin{aligned}\underline{W}_i(\hat{r}) - \theta_i w &= (1-\lambda)^2\theta_i(w - \delta^{T^*}c) + \lambda\theta_i(w - \lambda c\delta^{\hat{t}}) - \theta_i w \\ &= (1-\lambda)\theta_i(-\delta^{T^*}c - \lambda(w - \delta^{T^*}c)) - \lambda^2\theta_i c\delta^{\hat{t}} \leq -(1-\lambda)\theta_i(\delta^{T^*}c + \lambda\varepsilon) - \lambda^2\theta_i c\delta^{\hat{t}} \leq 0.\end{aligned}$$

Hence, pick $\tilde{r}_i$ for all i, such that $\tilde{s}^i_j(h^t, h^t_j) = e^t_j$ for all $h^t \in H, h^t_j \in H^t_j$ and for some $e^t_j \in E_j(h^t)$. In the self-referential equilibrium, each player i gets $\theta_i(w - c)$ for $T \geq T^*$. Any deviation from $\tilde{r}$ gives expected payoffs of 0. This profile constitutes a self-referential equilibrium with the unique outcome where all players exit in period $t = 0$. □

Proof of Theorem 2. Pick some period $\uptau \in \mathbb{N}$, and assume $\tau = 0$. Let $\hat{r}_i$ be $\hat{s}^i_j(h^t, h^t_j) = f^t_j \in F_j(h^t)$ for $t = 0$ or $h^t = (f^0, f^1, \ldots, f^{t-1})$ and $h^t_j \notin \overline{H}^t_j$ for $0 \leq t \leq \uptau - 1$, $\hat{s}^i_j(h^t, h^t_j) = e^t_j \in E_j(h^t)$ otherwise. The lowest expected payoff associated to $\hat{r}$ for player i is

$$\underline{W}_i(\hat{r}) = (1-\lambda)^2\theta_i(w - \delta^{\uptau}c) + \lambda\theta_i(w - \lambda c).$$

Given assumptions S.1 and S.2, consider $\tilde{r}_i$ such that $\tilde{s}_i(h^t, h^t_i) = e^t_i$ for all $h^t \in H, h^t_i \in H^t_i$ and some $e^t_i \in E_i(h^t)$; and $\tilde{r}_j = \hat{r}_j$ for all $j \neq i$. It gives an expected payoff of at most $\overline{W}_i(\tilde{r}_i, \hat{r}_j)$ for player i:

$$\overline{W}_i(\tilde{r}_i, \hat{r}_j) = \theta_i w - (\pi_j(\overline{y}_j|\hat{r}) + \eta)\theta_i c.$$

For all $\varepsilon > 0$, there is a T^* such that $w - \delta^{T^*} c \geq \varepsilon$. Note that $T^* > 0$ since $w < c$. Combining the last two expressions, we have

$$\begin{aligned} \underline{W}_i(\hat{r}) - \overline{W}_i(\tilde{r}_i, \hat{r}_j) &= (1-\lambda)^2\theta_i(w - \delta^{T^*}c) + \lambda\theta_i(w - \lambda c) - \theta_i w + (\pi_j(\overline{y}_j|\hat{r}) + \eta)\theta_i c \\ &\geq (1-\lambda)^2\theta_i\varepsilon + \lambda\theta_i(w - \lambda c) - \theta_i(w - \eta c) > 0. \end{aligned}$$

where the last inequality follows from $\lambda < \frac{w}{c} < \eta$. This shows that for all $T \geq T^*$, $\hat{r}$ is a self-referential equilibrium. □

Proof of Theorem 3. Fix a period $\tau = \hat{t}$ with $\phi(\overline{n} - 1) \leq \hat{t} \leq T$ and suppose players aim to have player $\iota(h^\tau)$ exit in period τ. Then consider $\hat{r}_i$ characterized by $\hat{s}^i_j(h^t, h^t_j) = f^t_j \in F_j(h^t)$ for $t = 0$ or h^t where all players have played a forward action, and $h^t_j \notin \overline{H}^t_j$ for $0 \leq t \leq \tau$, otherwise, $\hat{s}^i_j(h^t, h^t_j) = e^t_j \in E_j(h^t)$. The expected payoff for player $\iota(h^\tau) \equiv \iota$ is $g_\iota(e^\tau_\iota, \overline{a})$, and for player $i \neq \iota$ is $g_i(\overline{a}, e^\tau_\iota)$. Let k be the last stage player i is active before entering the active period in which player ι is active and exits in stage τ. Clearly, player i would choose $\tilde{r}_i$, the optimal code of conduct against $\hat{r}$. By P.1 and P.2, $\tilde{s}_i(h^t, h^t_i) = f^t_i \in F_i(h^t)$ for $t = 0$ or h^t where all players have played a forward action, and $h^t_i \notin \overline{H}^t_i$ for $0 \leq t \leq k-1$, otherwise, $\tilde{s}_i(h^t, h^t_i) = e^t_i \in E_i(h^t)$; for all players $j \neq i$, it says $\tilde{r}_j = \hat{r}_j$. Player i would get then $g_i(e^k_i, \overline{a}) > g_i(\overline{a}, e^\tau_\iota)$ by choosing $\tilde{r}_i$. Since $\hat{t}$ was arbitrary, in any self-referential equilibrium, each player i would exit at the end of their first active period. □

Proof of Theorem 4. Let $\tau \leq \phi(\overline{n} - 1) - 1$. Suppose that $\hat{r}_i$ prescribes $\hat{s}^i_1(h^t, h^t_1) = f^t_1 \in F_1(h^t)$ for $t = 0$, or $h^t \in H^t$ where all players have played a forward action, and $h^t_1 \notin \overline{H}^t_1$ for $0 \leq t \leq T$, or $h^t_1 \in \overline{H}^t_1$ for $0 \leq t < \phi(\overline{n}) - 1$; and in any other case $\hat{s}_1(h^t, h^t_1) = e^t_1 \in E_1(h^t)$. For player 2, it prescribes $\hat{s}^i_2(h^t, h^t_2) = f^t_2 \in F_2(h^t)$ for $t = 0$, or $h^t \in H^t$ where all players have played a forward action, and $h^t_2 \notin \overline{H}^t_2$ for $0 \leq t \leq T - 1$, or $h^t_2 \in \overline{H}^t_2$ for $0 \leq t < \phi(\overline{n} - 1) - 1$; and in any other case $\hat{s}_2(h^t, h^t_2) = e^t_2 \in E_2(h^t)$. Let $k = \phi(\overline{n}) - 1$, $k' = \phi(\overline{n} - 1) - 1$. The lowest expected payoff for player 2, $\underline{W}_2(\hat{r})$, from choosing $\hat{r}_i$ gives

$$\underline{W}_2(\hat{r}) = (1-\lambda)((1-\lambda)g_2(e^T_2, \overline{a}) - \lambda g_2(\overline{a}, e^k_1)) + \lambda g_2(e^{k'}_2, \overline{a}).$$

Since player 2 ends the game, combined with assumptions P.1 and P.2, player 2 has incentives to choose the alternative code of conduct $\tilde{r}_2$ where $\tilde{s}_2(h^t, h^t_2) = f^t_2 \in F_2(h^t)$ for $t = 0$, or $h^t \in H^t$ where all players have played a forward action for $0 \leq t \leq \phi(\overline{n} - 1) - 1$; and in any other case $\tilde{s}_2(h^t, h^t_2) = e^t_2 \in E_2(h^t)$; and $\tilde{r}_1 = \hat{r}_1$. This code of conduct gives $g_2(e^{k'}_2, \overline{a})$ to player 2, next choose T_2 so that $g_2(e^{T_2}_2, \overline{a}) \geq (1/(1-\lambda))(\lambda g_2(\overline{a}, e^k_1) - g_2(e^{k'}_2, \overline{a}))$. The lowest expected payoff for player 1 is

$$\underline{W}_1(\hat{r}) = (1-\lambda)((1-\lambda)g_1(\overline{a}, e^T_2) + \lambda g_1(e^k_1, \overline{a})) + \lambda g_1(\overline{a}, e^{k'}_2).$$

Assumptions P.1 and P.2, player 1 would choose the alternative code of conduct $\tilde{r}_1$ where $\tilde{s}_1(h^t, h^t_1) = f^t_1 \in F_1(h^t)$ for $t = 0$, or $h^t \in H^t$ where all players have played a forward action for $0 \leq t \leq \phi(\overline{n}) - 1$; and in any other case $\tilde{s}_1(h^t, h^t_1) = e^t_1 \in E_1(h^t)$; and $\tilde{r}_2 = \hat{r}_2$. It gives an expected payoff of at most

$\overline{W}_1(\tilde{r}_1, \hat{r}_2)$ for player 1:

$$\overline{W}_1(\tilde{r}_1, \hat{r}_2) = g_1(e_1^k, \overline{a}) - (\pi_j(\overline{y}_j|\hat{r}) + \eta)(g_1(e_1^k, \overline{a}) - g_1(\overline{a}, e_2^{k'})).$$

so choose T_1 be such that $g_1(\overline{a}, e_2^{T_1}) \geq (1/(1-\lambda))(\lambda g_1(\overline{a}, e_2^k) - g_1(e_1^k, \overline{a}))$. Let $T^* = \max\{T_1, T_2\}$. For all $T \geq T^*$ $\hat{r}$ is a self-referential equilibrium. □

Proof of Theorem 5. Start by picking any $v \in V^*$ and assume that $v_i = g_i(a^*)$ for some $a^* \in A$. Consider the code of conduct r_i such that $s_j^i(h^t, h_j^t) = a_j^*$ if $t = 1$ or $h^k = (a^{*1}, a^{*2}, \ldots, a^{*k-1})$ and $h_j^t \notin \overline{H}_j^t$ for $1 \leq k \leq t-1$, and $s_j^i(h^t, h_j^t) = \underline{\alpha}_j$ otherwise. Suppose that there are $\tau + 1$ periods left after $y_i \in Y_i$ is observed by each player i.

If player i chooses r_i, the least expected payoff he obtains is

$$\frac{T-\tau-1}{T} g_i(a^*) + \frac{1}{T}\left[(1-\lambda)^2 g_i(a^*) + \lambda(1-\lambda)\underline{v}_i + \lambda \min_a g_i(a)\right] + \frac{\tau}{T}\left[\underline{v}_i + (1-\lambda)^2(g_i(a^*) - \underline{v}_i)\right].$$

Next, player i chooses $\tilde{r}_i \neq r_i$. It prescribes $\tilde{s}_j^i(h^t, h_j^t) = a_j^*$ if $t = 1$ or $h^k = (a^{*1}, a^{*2}, \ldots, a^{*k-1})$ and $h_j^t \notin \overline{H}_j^t$ for $1 \leq k \leq t-1$ and $t < t'$, and $\tilde{s}^i(h^{t'}, h_i^{t'}) = a_i$ for some $t' \leq T - \tau$ with $a_i \neq a_i^*$. Let $t \leq T - \tau$ be the period in which player i deviates and chooses some alternative action. The highest expected payoff for player i is

$$\frac{t-1}{T} g_i(a^*) + \frac{1}{T}\left[(\eta + \pi_j(\overline{y}_j|r))\underline{v}_i + (1 - \eta - \pi_j(\overline{y}_j|r)) \max_a g_i(a)\right] + \frac{T-t}{T}\underline{v}_i.$$

Note that this expression has a maximum for $t = T - \tau$. Now compute the number of periods τ_i that satisfies the following inequality

$$\begin{aligned}\tau_i(1-\lambda)^2(g_i(a^*) - \underline{v}_i) \geq \max_a g_i(a) - g_i(a^*) - (\eta + \pi_j(\overline{y}_j|r))(\max_a g_i(a) - \underline{v}_i)\\ + (1-\eta)\eta(g_i(a^*) - \underline{v}_i) + \eta(g_i(a^*) - \min_a g_i(a)),\end{aligned}$$

It ensures that player i has no profitable alternative code of conduct. Let $T^* = \max_i \tau_i$, and notice that for all $T \geq T^*$, r is a self-referential equilibrium. It follows that equilibrium candidate code of conduct profile r is feasible for $\tau \leq T - T^*$. Since T^* is independent of the length of the game, take T large enough so that $|U_i(r) - v_i| \leq \varepsilon$ for any $\varepsilon > 0$. □

Proof of Proposition 1. Fix any $v \in V^*$, and assume that $v_i = g_i(a^*)$ for all i for some $a^* \in A$. Let r_i be $s_j^i(h^t, h_j^t) = a_j^*$ if $t = 1$ or $h^k = (a^{*1}, a^{*2}, \ldots, a^{*k-1})$ and $h_j^t \notin \overline{H}_j^t$ for $1 \leq k \leq t-1$, and $s_j^i(h^t, h_j^t) = \underline{\alpha}_j$ otherwise. When everyone chooses r_i, players' expected payoff is v_i. Assume that $\tau + 1$ periods remain after $y_i \in Y_i$ is observed for all i. Suppose that player i chooses $\tilde{r}_i \neq r_i$, where $\tilde{s}_j^i(h^t, h_j^t) = a_j^*$ if $t = 1$ or $h^k = (a^{*1}, a^{*2}, \ldots, a^{*k-1})$ and $h_j^t \notin \overline{H}_j^t$ for $1 \leq k \leq t-1$ and $t < t'$, and $\tilde{s}^i(h^{t'}, h_i^{t'}) = a_i$ for some $t' \leq T - \tau$ with $a_i \neq a_i^*$. Let $t \leq T - \tau$ be the period in which player i

deviates and chooses some alternative action. The highest expected payoff for player i is

$$\frac{t-1}{T}g_i(a^*) + \frac{1}{T}\left[\eta\underline{v}_i + (1-\eta)\max_a g_i(a)\right] + \frac{T-t}{T}\underline{v}_i.$$

This expression is maximized when $t = T - \tau$, and now find τ_i that solves

$$\tau_i(g_i(a^*) - \underline{v}_i) \geq \max_a g_i(a) - g_i(a^*) - \eta(\max_a g_i(a) - \underline{v}_i).$$

Define $T^* = \max_i \tau_i$, and note that for $T \geq T^*$ and $\tau \leq T - T^*$, r is a self-referential equilibrium with payoff v_i for each player i. $\square$

Notation. The constant $\zeta^T \equiv \{\sup |u_i(\sigma) - u_i(\hat{\sigma})| : \forall i, \forall \sigma, \hat{\sigma} \in \Sigma \text{ s. t. } \sigma_T = \hat{\sigma}_T\}$ is greatest variation in payoffs due to events after T for any i. Continuity at infinity implies that $\lim_{T\to\infty} \zeta^T = 0$.

For any $h^t \in H$, let $\sigma_i|h^t$ denote the continuation strategy prescribed by σ_i after history h^t and let $\sigma_i|H^t$ denote the restriction of σ_i to $H^t \subset H$. For $\sigma \in \Sigma$, $\sigma|h^t$ and $\sigma|H^t$, respectively. The minmax payoff of player i in the T-truncation Γ^T is given by $\underline{u}_{i,T} \equiv \min_{\sigma_{j,T}|H^T} \max_{\sigma_{i,T}|H^T} u_i(\sigma_{i,T}, \sigma_{j,T})$, where $\underline{\sigma}_{j,T}$ denotes the minmax profile against player i. For σ_{-i}, $BR_i(\sigma_{-i})$ is the set of best responses to σ_{-i} by i.

To construct the truncation choose any σ in Γ^∞. Let $\overline{\sigma}$ be the strategy which is the constant repetition of the no-decision action $\overline{a}$. Then, embed σ_T in Γ^T into Γ^∞ by concatenating σ_T with $\overline{\sigma}$: players follow σ_T in all stages up to and including stage T, and $\overline{\sigma}$ at $t > T$. I will evaluate the limit of Γ^T as $T \to \infty$. Since A_i is finite it is sufficient to work with the product topology.[10]

Proof of Theorem 6. Suppose that $\tau = 0$. Fix any $\hat{\sigma} \in \mathcal{E}(\Gamma^\infty)$. Take a T-truncation of this game, Γ^T. If the truncated $\hat{\sigma}_T$ turns out to be an equilibrium of Γ^T, then pick $\hat{r} \in R$ such that $\hat{s}^i_j(h^t, h^t_j) = \hat{\sigma}_{j,T}(h^t)$ for all $i, j, h^t \in H$ and $h^t_j \in H_j$. It follows $\hat{r}$ would form a self-referential equilibrium. If not, let $\sigma_{i,T} \in BR_i(\hat{\sigma}_{-i,T})$. Pick $\hat{r} \in R$ which prescribes for all i, j:

$$\hat{s}^i_j(h^t, h^t_j) = \begin{cases} \hat{\sigma}_{j,T}(h^t) & \text{for all } h^t \in H^t, h^t_j \notin \overline{H}^t_j, \\ \underline{\sigma}_{j,T}(h^t) & \text{otherwise.} \end{cases}$$

If all i choose $\hat{r}_i$, they get $U_i(\hat{r}) = u_i(\hat{\sigma}_T)$. If not, suppose that player i's choice involves some $\tilde{r}_i$ such that $\tilde{s}_i(h^t, h^t_i) = \sigma_{i,T}(h^t)$ for all $h^t_i \in H^t_i$ and $\tilde{s}^i_j = \hat{s}^i_j$ for any $j \neq i$. Let the highest payoffs associated to $\tilde{r}_i$ for player i be $\overline{W}_i = u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \eta(u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i)$. Each player i may choose $\hat{r}_i$ if $U_i(\hat{r}) \geq \overline{W}_i$. Rewrite this inequality as $u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - u_i(\hat{\sigma}_T) - \eta(u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i) \leq 0$. For any player i find bounds as follows

$$u_i(\sigma_{i,T}, \hat{\sigma}_{-i,T}) - u_i(\hat{\sigma}_T) \leq u_i(\sigma_{i,T}, \hat{\sigma}_{-i,T}) - u_i(\hat{\sigma}_T) + u_i(\hat{\sigma}) - u_i(\sigma_i, \hat{\sigma}_{-i}) \leq 2\zeta^T,$$

[10] A sequence $\{\sigma_{i,n}\}_{n\in\mathbb{N}}$ converges to σ_i in the product topology if and only if $\sigma_{i,n}(h) \to \sigma_i(h)$ for any $h \in H$.

and

$$u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i \leq u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i + u_i(\hat{\sigma}) - u_i(\sigma_i, \hat{\sigma}_{-i}) \leq 2\zeta^T + u_i(\hat{\sigma}_T) - \underline{u}_i.$$

Set $\varepsilon \equiv \frac{\eta}{1-\eta} \min_i (u_i(\hat{\sigma}_T) - \underline{u}_i)$. Observe that since the game is continuous at infinity, one can take T^* such that $\zeta^{T^*} \leq \frac{\varepsilon}{2}$. Now, for all $T \geq T^*$ and for all i we have

$$\begin{aligned} \overline{W}_i - U_i(\hat{r}) &= u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - u_i(\hat{\sigma}_T) - \eta(u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i) \\ &\leq \varepsilon - \eta(\varepsilon + (u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i) = (1-\eta)\varepsilon - \eta(u_i(\sigma_{i,T}, \hat{\sigma}_{j,T}) - \underline{u}_i) \\ &= \min_i (u_i(\hat{\sigma}_T) - \underline{u}_i) - (u_i(\hat{\sigma}_T) - \underline{u}_i) \leq 0. \end{aligned}$$

□