Timing and Codes of Conduct

Juan I. Block*
Washington University in St. Louis

November, 2013

Abstract

In games where players can imperfectly observe an opponent's intentions, the time at which intentions can be discovered may have a significant impact on the equilibrium outcome set. When players infer intentions at the outset, I show that a folk theorem for finite horizon games holds, whereas if agents glean intentions afterwards the timing leads to different effects depending on the structure of the game. I identify two classes of games with antipodal results concerning the timing. In finitely repeated games with discounting, the folk theorem continues to apply regardless of the time at which intentions are observed, and whether the observation is synchronous or asynchronous. By contrast, the equilibrium outcome is unique in exit games, where players end the game endogenously.

KEYWORDS: Folk theorem, self-referential game, intentions, repeated game, exit game.

JEL CODES: C72, C73, D03.

*I am very grateful to my advisors, especially, David Levine for his invaluable insights and guidance; John Nachbar and Paulo Natenzon for their insightful comments. I thank Mariagiovanna Baccara, Anqi Li, Brian Rogers, Maher Said and Jonathan Weinstein for helpful discussions. I benefited from conversations with Drew Fudenberg, Johannes Hörner, Juan Passadore, Luciano Pomatto, Daniel Quint, Ludovic Renou, Aldo Rustichini, Marco Scarsini, Eran Shmaya, Mariano Somale and Tristan Tomala. Correspondence: Department of Economics, Washington University in St. Louis, MO 63130. Email: `juanblock@wustl.edu`

1 Introduction

In economic environments where agents have the ability to infer intentions, the time at which intentions can be discovered may have a substantial impact on the set of outcomes that can arise in equilibrium. When intentions can be inferred at the outset, a folk theorem for finite horizon games holds; however, when intentions are discovered later on the timing has different effects depending on environmental details. This paper identifies two classes of games with diametrically opposite results regarding such timing. In finitely repeated games, the folk theorem continues to hold even if players observe intentions in the last period of the game. In exit games, on the other hand, there exists a unique equilibrium outcome.

Many important situations in economics have agents capable of recognizing intentions. For example, in industrial espionage an entry firm would spy on the incumbent's response to market entries before expanding business to a new market (e.g., Airbus and Boeing developing the jumbo jet, Caruana and Einav (2008)). Likewise, in military conflicts armies spend resources to anticipate the enemy's battlefield plan (e.g., Solan and Yariv (2004) and Matsui (1989)). Another example is the Chairman of the Federal Reserve giving a public speech about a policy to be implemented and consumers predicting its time consistency. Recognition techniques are also present in online pricing strategies; by way of illustration, click stream pricing displays a price for the product depending on consumers' browsing history (e.g., Peters (2013)). Similarly when a security agency employs ex-ante verifications, such as random audits on passengers or internal audits in firms, it is attempting to recognize intentions. These situations are captured by self-referential games.

Standard models of intention recognition typically suppose that strategies can be revealed in the pre-play phase. Yet, in practice, agents partially observe intentions during the game such as Airbus forecasting Boeing's reaction to entries in the big jet segment after developing the A380 superjumbo. Likewise, in monetary policy consumers may adjust predictions about the annual speech in the third quarter. In click stream pricing, for instance, the seller may price a complimentary product based on consumer's choice up to the checkout stage. In contrast to all previous analyses, the model I propose extends self-referential games to address the realistic feature that agents might infer others' plans in the course of the interaction.

From the self-referential perspective, the ability to infer intentions is conceptually the result of conforming to a rule of behavior that might be too costly to change so that individuals are committed to this rule. In this context, lying is not fully costless (i.e. cheap talk, Crawford (2003)) in the sense that agents must imitate others' behavior to send the same signals as them. Moreover, agents may imperfectly identify opponents' rules of behavior because past play is observable, allowing for understanding intentions at some stage of the game. Alternatively, elements of strategies might exhibit themselves via communication or involuntary gestures. Such a situation could be Frank's (1987) example in which a sincere individual may blush whenever he lies. Consequently, blushing can be interpreted as evidence of potential dishonest behavior. As mentioned, there are costs associated with mimicking behavior, for instance, manipulating information or faking. In business, planning a feint would be costly because concealing the true state of a product's development might require continued investing in dead-end products.

Building on Block and Levine (2012), I model the self-referential game as an extension of a base game. The base game is a multistage game with observed actions in which players know the actions chosen at all previous stages and may move simultaneously in each stage, whereby strategies depend on public histories. Coupling with this game, the self-referential framework endows players with an upfront private signal at some stage. Accordingly, an *extended strategy* is defined by public and private histories. The self-referential game is played in two stages. At the first stage, players simultaneously choose a *code of conduct* which commits the player to an extended strategy, and specifies one for each of his opponents. Privately observed signals are drawn from an exogenous probability distribution that is determined by the code of conduct profile. This probability distribution is meant to capture both the idea that intentions are imperfectly observable and that codes of conduct might recognize one another. In the second stage, each agent employs the extended strategy according to his code of conduct.

In this model, players choose strategies that are indirectly conditioned on other players' as extended strategies consider the private signal which in turn depends on all players' choices—this is the self-referential property. These conditional commitment devices typically lead to the infinite-regress problem. Within this context, it means that a strategy depends on other player's strategy that is conditioned by the first strategy, and this inductive argument continues ad infinitum. Although such circularities are overcome because players triangulate this dependence through private signals, and the likelihood of these signals is determined by an exogenous probability distribution. As a matter of interpretation, codes of conduct should be thought of as social norms, to the extent that they provide a well-defined notion of agreement. Motivated by the vast evidence on reciprocal behavior, self-referential models focus on information structures that allow players to distinguish whether there is agreement.

As in the benchmark used in the literature, suppose that agents infer intentions only in the pre-play phase, thereby allowing them to punish any kind of intentions. 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. The key to construct equilibria is the probability of distinguishing whether rivals agree on the code of conduct. One implication of this result is that as the horizon grows long, the probability of detecting deviations approaches zero for any equilibrium strategy.

The main contribution of this paper is to identify two classes of games with starkly opposite predictions depending on the time at which intentions can be discovered. To begin with, I study finitely repeated games with discounting where the ability to observe intentions in the last round of the game suffices to prove a version of the folk theorem. Its proof hinges on patient players whose behavior is sensitive to changes in endpoint payoffs and on deviations that are likely to be detected. Under some regularity assumptions findings suggest that the sooner agents recognize intentions, the lower is the required probability of detecting deviations from the code of conduct in equilibrium.

Exit games, by contrast, is a class of games in which the equilibrium outcome set is immune to the possibility of inferring intentions later on. From this collection of games, I first consider splitting games where every player is able to terminate the game in each period, and the game also ends exogenously in the last round if none of them has exited. In this context, the equilibrium set with outcomes where everyone exits in the first period is rendered unique by intentions that can be discovered after the first stage. This is because the stakes in the beginning offset expected payoffs in any self-referential equilibrium exhibiting late exit profiles. Nonetheless, exit is delayed arbitrarily when intentions are inferred at the outset.

Another subclass of exit games is preemption games, where just one player—that may be active for many consecutive stages—can exit in each stage. When intentions are discovered in the next to the last active period, I show that there is a unique equilibrium outcome where the first mover exits immediately. The reason is that, although agents prefer to exit at late stages, leaving in the first active period entails enough payoffs without being punished. Yet, equilibria exhibiting delayed exit could be constructed as long as the player ending the game glean a rival's intentions one active period preceding exit, allowing him to punish deviations. Thus, players find it optimal to end the game conditional on the targeted exit profile.

The model extends to allow for asynchronous intention recognition, that is, players receive information about one another's code of conduct at different stages. In finitely repeated games, I find that a folk theorem applies since agents use signals in the last round of the game that help coordinate punishments and rewards. On the other hand, in splitting games the unique equilibrium outcome has all players exiting in the first period because sustaining exit profiles after that requires initial-period signals for all players. Finally, in preemption games there exists a unique equilibrium outcome under late signals; nevertheless, equilibria with exit profiles at late stages can be sustained as players alternate active periods complementing the asynchronous, but early, timing of signals.

1.1 Related Literature

Self-referential games were introduced by Levine and Pesendorfer (2007) in the context of an evolutionary model where players are pairwise matched to play a symmetric game. In their setting imitation of strategies is more likely than innovation, and identification of behavior prior to play is possible. They show that strategies that emerge in the long run are those that reward opponents that are likely to play similarly and punish opponents that are likely to behave differently. The self-referential framework was extended to multi-players asymmetric games by defining codes of conduct in Block and Levine (2012), who prove a folk theorem for repeated games with private monitoring. The key difference is that these models assume that behavior recognition occurs in the beginning of the game, while here it also happens during the game.

Codes of conduct have similar characteristics to conditional commitment devices. The self-referential framework is closely related to that of Tennenholtz (2004) and Kalai, Kalai, Lehrer, and Samet (2010).[1] In both papers, agents choose a commitment device that conditions on other players' commitment device. While they assume that these devices are perfectly observable, I analyze behavior that is conditioned by noisy information; although I also consider underlying games with complete information. See Peters and Szentes (2012) and Forges (2013) for the extension of Kalai et al. (2010)'s model to Bayesian games, and Peters and Troncoso-Valverde (2013) for a folk theorem in competing mechanism games in which agents employ this kind of devices.

This paper contributes to the literature that studies the possibility of observing strategies before the actual play of the game. For instance, Matsui (1989) considers two-player infinitely repeated games in which players may observe opponents' metagame strategy with small probability before the game starts, allowing revision of strategies. He shows that any subgame perfect equilibrium payoff vector is Pareto efficient. In contrast, in this paper information about rivals' code of conduct is imperfect and the class of games is much broader. In two-player normal form games, Solan and Yariv (2004) examine espionage games where one player can pay for a signal which delivers information about the other player's strategy. The main result says that the set of espionage equilibria coincides with the set of non-degenerate semi-correlated equilibrium distributions. While the espionage game is sequential and has only one-side spy, in my approach players simultaneously choose codes of conduct and receive a private signal with no explicit cost.

More recently, Kamada and Kandori (2011) study revision games in which the opportunity to revise actions arrive stochastically, 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 (2013) show that revision games narrow

[1] Tennenholtz (2004) develop a setting where players submit programs which take as input the other players' program and play on their behalf. A *program equilibrium* is constructed by programs that give an outcome action if they are syntactically identical and punish otherwise. Although, he did not describe the set of programs. At a more general level, Kalai et al. (2010) characterize the conditional commitment devices space. In self-referential games, this space is assumed to be common and very general.

down the set of equilibrium payoffs in common and opposing interest games. In the present model, strategies are imperfectly observable via signals with deterministic arrival time.

All these papers have information about strategies releasing at the outset, whereas I characterize how the size of the equilibrium outcome set gets determined by the time at which intentions are inferred.

The rest of the paper is organized as follows. In Section 2, I present the framework of the base and the self-referential game, and the information technology. Section 3 studies self-referential games in which recognition occurs only in advance. In section 4, I analyze finitely repeated games and characterize the equilibrium set in terms of the timing of signals. In Section 5, I examine exit games showing the implications of the timing on equilibrium behavior. Section 6 extends the analysis to asynchronous signals, contrasting asynchronicity results in each class of games with the equilibrium predictions obtained in the original setting. I conclude in Section 7. The Appendix collects the proofs.

2 The Model

I next outline the general framework. In section 2.1 I describe multistage games with observed actions. Section 2.2 presents the self-referential game, extending the setting in Block and Levine (2012) and allowing players to learn about opponents' intentions in the course of play.

2.1 Setup and Notation

I concentrate on multistage games with observed actions as defined in Fudenberg and Tirole (1991).[2] There are a set of players $\mathcal{I}$ with cardinality $|\mathcal{I}| = N$ and $T+1$ stages.[3]

Let $h^0 = \emptyset$ be the initial public history and $A_i(h^0) \ni a_i^0$ be the finite actions 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 $h^t = (a^0, a^1, \ldots, a^{t-1})$ whose length is $l(h^t)$. Player i chooses an action a_i^t from his finite actions set $A_i(h^t)$ at stage t with profile $a^t \in A(h^t)$, and $\{\overline{a}\}$ stands for the *no-decision action*. I write H^t for the set of all stage t public histories and $H := \bigcup_{t=0}^{\infty} H^t$ for the set of all public histories. Let Z be the set of terminal histories where h^{T+1} is finite if $T < \infty$, otherwise it is infinite, h^∞.

A (behavioral) strategy for player i is a map $\sigma_i : H \to \Delta A_i(h^t)$ where each $\Delta A_i(h^t)$ is endowed with the standard topology, and $\mathcal{S}_i^t := \Delta A_i(h^t)$ for notational convenience. Let

[2] This class of games is also known as multistage games with almost perfect information and perfect recall. Multistage games were generalized to *multistage situations* by Greenberg, Monderer, and Shitovitz (1996). Their framework applies to a broader class of social environments and allows for analysis to cases where, for example, the strategies tuples are not Cartesian product of the players' strategy sets.

[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 $X := \times_i X_i$ with typical element $x \in X$, and $X_{-i} := \times_{j \neq i} X_j$ with element x_{-i}.

$\Delta A(h^t)$ be the space of independent strategy profiles equipped with the product topology. The set Σ_i denotes pure strategies, and profiles are Σ with typical element σ_p. Write Ξ_i for behavioral strategies with profile Ξ.

The reward function for player i is $g_i : H \to \mathbb{R}$ where he receives payoff $g_i(h^t)$ following history h^t at stage $t-1$ that is discounted to stage $t-2$ by discount factor $\delta_i \in (0,1]$. Denote by A^∞ the set of possible outcomes with generic element a^∞. Specifically, the outcome path induced by σ_p is denoted by $a^\infty(\sigma_p)$. Player i's payoffs as a function of pure strategy profile, $u_i : \Sigma \to \mathbb{R}$, is

$$u_i(\sigma_p) = \sum_{t=0}^{\infty} \delta_i^t g_i(a^t(\sigma_p)).$$

I extend the domain of rewards to behavior strategies profile σ in the standard way denoting them by $u_i(\sigma)$. Finally, let Γ stand for the multistage game with observed actions.

2.2 The Self-Referential Game

In the self-referential game, the set of players is also $\mathcal{I}$. Every player i observes a signal y_i only in the beginning of stage τ_i, that belongs to the finite set Y_i with $|Y_i| \geq 2$.[4] The stage τ_i is deterministic and commonly known.

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 $t < \tau_i$, and $H_i^t \subset Y_i$ for all stages $t \geq \tau_i$. Let $H_i := \bigcup_{t=0}^{\infty} H_i^t$ 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 from public and private histories to actions, $s_i : H \times H_i \to \mathcal{S}_i^t$. Let $s \in S$ be a profile of extended strategies.[5]

The strategy of player i in the self-referential game, r^i, is called a code of conduct which is an $|\mathcal{I}| \times 1$ vector whose jth element corresponds to what player i assigns to player j's choice of extended strategies. Specifically, for any i and all $j \neq i$ the code of conduct r^i is a choice of $|\mathcal{I}|$ number of extended strategies, $r_j^i : H \times H_j \to \mathcal{S}_j^t$. I also refer to codes of conduct as *self-referential strategies*. Each player i is endowed with the common space of codes of conduct R_0 given by

$$R_0 := \left\{ r^i \mid r_j^i \in \mathcal{S}_j^{t\,H \times H_j} \text{and } \forall i,j \in \mathcal{I}, \forall h^t \in H, \forall h_j^t \in H_j, r_j^i(h^t, h_j^t) \in \mathcal{S}_j^t \right\},$$

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 the code of conduct r^i commits player i to an extended strategy and players participate in the

[4]The signal y_i parameterizes the information about intentions accumulated up to stage τ_i. The idea is that players accrue pieces of information during the game, and at some point they make use of them to evaluate adversaries' intentions.

[5]To avoid measure theoretic considerations, I assume that the set of strategies for player i, S_i is finite. Observe that finite mixed strategies are permitted, for example, rolling a finite n-dimensional dice.

self-referential game even if they pay no attention to signals.[6] A code of conduct vector is denoted by $r \in R$.

For each code of conduct profile $r \in R$, let $\pi(\cdot|r)$ be the probability distribution over signal profiles Y. I define the *intention monitoring structure* as 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 the code of conduct profile r. We say an intention monitoring structure (Y, π) is *stage-t timing* if signal profile $y \in Y$ is observed in the beginning of stage t, i.e. $\tau_i = t$ for all i.

Player i's expected payoffs in the self-referential game $U_i : R \to \mathbb{R}$ are

$$U_i(r) = \sum_{y \in Y} u_i(r_1^1(h, h_1(y_1)), \ldots, r_N^N(h, h_N(y_N)))\pi(y|r).$$

Let $\mathcal{G}(\Gamma) = \{\Gamma, Y, \pi, R\}$ represent the self-referential game. A vector of codes of conduct r^* is a Nash equilibrium of the self-referential game (or self-referential equilibrium) if for all players $i \in \mathcal{I}$ and codes of conduct $r^i \in R_0$, $U_i(r^*) \geq U_i(r^i, r^{*-i})$.

The self-referential game $\mathcal{G}$ takes place in two stages. It starts with players choosing simultaneously code of conduct, $r^i \in R_0$. Then each player i chooses $r_i^i(h^t, h_i^t) \in \mathcal{S}_i^t$ for histories $h^t \in H$, $h_i^t \in H_i$, and observes private signal $y_i \in Y_i$ in the beginning of stage τ_i.

Finally, the detection technology allows players to discern whether rivals adhere to the same code of conduct (as in Block and Levine (2012)) and it plays an important role in developing the results in this paper. Specifically, we say that the self-referential game η-λ permits detection if for two constants $\eta, \lambda \in [0, 1]$, for all players i there exist some player $j \neq i$ and a subset of private signals $\overline{Y}_j \subset Y_j$ such that for all code of conduct profiles $r \in R$, any signal $\overline{y}_j \in \overline{Y}_j$ and each code of conduct $\tilde{r}^i \neq r^i$, it follows that $\pi_j(\overline{y}_j|\tilde{r}^i, r^{-i}) - \pi_j(\overline{y}_j|r) \geq \eta$ and $\pi_j(\overline{y}_j|r) \leq \lambda$.

In words, the lower bound η describes the minimum probability of detecting deviations from some code of conduct profile r, associating such deviations to signals in the set $\overline{Y}_i$. Although, players also observe this type of signals even if everyone follows the profile r leading us to interpret constant λ as the upper bound of the false positive probability. This technology suggests that agents use simplified categorization of intentions, aiming a specific behavior while bundling all deviations into a single class.[7] Note also that it gives the identity

[6] Kalai et al. (2010) and Forges (2013) define voluntary commitment devices so they allow the possibility of "not committing," while here players are committed to codes of conduct.

[7] One interpretation is that codes of conduct might be so complex that agents bundle intentions of opponents' behavior into analogy classes. These simplifications resembles Jehiel's (2005) analogy-based expectation equilibrium in which players partition histories into analogy classes and best-respond to beliefs that opponents' behavior is constant within each analogy class.

of the deviator but not the magnitude of the deviation.

3 Pre-Game Signals

In this section, I establish a connection between the equilibrium set in the infinite-horizon game and the set of self-referential equilibria in the finite-horizon version of the game. More precisely, I show that the set of outcomes that may arise in self-referential equilibrium of the finite horizon approximation of the infinite horizon game is equal to the set of equilibrium in the original game as long as players observe signals in the pre-play phase. A classic example is the infinitely repeated prisoner's dilemma game with patient enough players. Going from the plethora of equilibria in the infinite horizon game to the unique equilibrium of the finitely repeated game produces a discontinuity.

Let Γ^∞ be an infinite horizon multistage game with observed actions described in Section 2. Consider any finite stage τ, let Γ^τ represent the same game with time horizon truncated at τ. To approximate this game with its finite truncation we require players to be passive after stage τ by choosing the no-decision action $\overline{a}$ thereafter.[8] Of particular interest are games in which future payoffs are not relevant. 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

$$|u_i(\sigma) - u_i(\hat{\sigma})| < \varepsilon \quad \text{ if } \sigma^k = \hat{\sigma}^k \text{ for all } i \in \mathcal{I} \text{ and all } \sigma, \hat{\sigma} \in \Xi.$$

Examples of such games are repeated games with discounting and any finite horizon game. The set of games that fails continuity at infinity includes, for instance, repeated games with limit-average payoffs and alternating-offer bargaining games with no discounting.

The result of this section 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.

Theorem 1. *For $|\mathcal{I}| = 2$. Let Γ^∞ be continuous at infinity. Suppose that the self-referential game satisfies η-λ permit detection with $\lambda = 0$, and $\tau_i = 0$ for all i. For any subgame perfect equilibrium $\hat{\sigma}$ in Γ^∞ and τ-truncation Γ^τ, there exist a probability of detection $\eta_\tau > 0$ and a profile of codes of conduct r^τ such that for all $\eta \in [\eta_\tau, 1]$ in the self-referential equilibrium $r_i^i = \hat{\sigma}_{i,\tau}$ for all $h_i \notin \overline{H}_i, i \in \mathcal{I}$. Moreover, the probability of detection $\eta_\tau \to 0$ as $\tau \to \infty$.*

See Appendix A.1 for the proof.

[8] Apply the truncation of length k on σ to obtain the partial strategy σ^k for some $k \in \mathbb{Z}$ from $\sigma \in \Xi$.

The literature on the connection between infinite and finite horizon games has focused on perfect ε equilibria of the finite truncation Γ^{τ}.[9] In contrast, Theorem 1 presents a relationship between the perfect equilibrium set of Γ^{∞} and the *exact* equilibria of the self-referential game built on the truncation of the original game $\mathcal{G}(\Gamma^{\tau})$. We may interpret Theorem 1 as a lower hemi-continuity result, that is, exact equilibria of the self-referential game approach the limit point. A result of Fudenberg and Levine (1983, Theorem 3.3) guarantees that a subgame-perfect equilibrium in finite-action game exists.

Agents can only distinguish imperfectly whether opponents choose the same self-referential strategy because of η, λ detection technology. Consequently, a natural construction of self-referential equilibria uses *grim trigger* strategies, whereby each player rewards others unless he observes deviations from the code of conduct. The equilibrium code of conduct takes a simple form: in the case of evidence of agreement, agents play the truncated strategy $\hat{\sigma}_{\tau}$, while if signals indicate deviations from the code of conduct then players minmax opponents forever. On the equilibrium path players do not punish others following the code of conduct whenever $\lambda = 0$, but they do so to deviators. In fact, deviators are unlikely to be punished when the chance of detection is low. Therefore, there is a threshold level of detection η_{τ} below which the expected profit from deviation is high relative to the cost of punishment. When detection probability is above η_{τ}, the cost of punishment dominates and players adhere to this code of conduct.

The requirement of a detection probability $\eta_{\tau} > 0$ is weak because η_{τ} becomes arbitrarily small as $\tau \to \infty$. Consider the equilibrium code of conduct profile r^{τ} constructed above for a fixed truncation τ, and take a longer horizon. The difference between the gain to deviation and to adherence determines the critical detection probability η_{τ}. In the long run, this difference shrinks as the approximation improves ($\tau \to \infty$) due to continuity at infinity. Note that the hypothesis of the theorem cannot be strengthen to $\eta_{\tau} = 0$, namely players always detect disagreement about codes of conduct.[10]

4 Finitely Repeated Games

As mentioned, the effect of the timing of signals depends strongly on the underlying game. It is instructive then to do the analysis within specific classes of games, indeed I consider

[9] See, for instance, Radner (1981), Fudenberg and Levine (1986), Harris (1985a, 1985b) and Börgers (1989, 1991). A profile of strategies $\hat{\sigma}$ is an ε Nash equilibrium if $\forall i, \sigma_i, \varepsilon \geq 0$, $u_i(\hat{\sigma}) \geq u_i(\sigma_i, \hat{\sigma}_{-i}) - \varepsilon$. A strategy $\hat{\sigma}$ is a perfect ε-equilibrium if $\forall i, h, \sigma_i$, $u_i(\hat{\sigma}^h) \geq u_i(\sigma_i^h, \hat{\sigma}_{-i}^h) - \varepsilon$. The latter was defined as ex ante perfect ε equilibrium by Mailath, Postlewaite, and Samuelson (2005). They consider *contemporaneous* perfect ε equilibrium that evaluates best responses at the time of the deviation. For finite games and small epsilon, these two notions of equilibria coincide.

[10] The interpretation that the probability of detection goes to zero ($\eta \to 0$) is that signals become decreasingly informative about deviations from a profile r, as long as $\lambda = 0$.

two starkly opposite ones respecting the effect of the timing on the equilibrium outcome set. In this section, I study finitely repeated games with discounting where a folk theorem-like for one-shot games holds, getting approximately any feasible and individually rational payoff vector. In addition, I show a few different versions of a folk theorem for discounted repeated games even if intentions are observable in the last round of the game.

4.1 The Stage Game

Let Γ be the stage game. Each player $i \in \mathcal{I}$ has a finite actions set A_i with $|A_i| \geq 2$, and the profile of actions is $a \in A$. Reward functions are $g_i : A \to \mathbb{R}$. I write α_i for mixed actions for each player i with $\alpha_i \in \Delta(A_i)$, and I extend payoffs to mixed strategies in the standard manner $E_\alpha(g_i(a)) = g_i(\alpha)$. For each player i, I denote by $\underline{v}_i$ the (mixed strategies) minmax payoff of player i in the stage game as

$$\underline{v}_i \coloneqq \min_{\alpha_{-i} \in \Delta(A_{-i})} \max_{a_i \in A_i} g_i(a_i, \alpha_{-i}).$$

Then take action $\underline{\alpha}_{-i} \in \Delta(A_{-i})$ such that

$$\underline{v}_i \coloneqq \max_{a_i \in A_i} g_i(a_i, \underline{\alpha}_{-i}),$$

where $\underline{\alpha}_{-i}$ is the action profile that gives the minmax payoff to player i. Let

$$U \coloneqq \{(v_1, \ldots, v_N) : \exists a \in A, \forall i, g_i(a) = v_i\},$$
$$V \coloneqq co(U),$$
$$V^* \coloneqq int\left(\{(v_1, \ldots, v_N) \in V : \forall i \in \mathcal{I},\ v_i > \underline{v}_i\}\right).$$

V is the set of feasible payoff vectors and V^* is the set of feasible and strictly individually rational payoff vectors.[11] Players have access to a public randomization device which generates a public signal $\omega^t \in [0,1]$ uniformly distributed and independent across periods at the start of each period t, i.e. $(\omega^t)_{t \in \mathbb{N}}$ is an i.i.d. sequence. Thus, they may condition their actions on these signals.

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 one-shot game Γ, if some conditions on the self-referential information structure are satisfied.

Theorem 2. *Let $|\mathcal{I}| = 2$. Assume that the self-referential game η-λ permits detection with $\tau_i = 0$ for all i. For every feasible and strictly individually rational payoff vector $v \in V^*$*

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

in the stage game Γ*, there exist* $\eta_0 > 0$ *and* λ_0 *such that for all* $\eta \geq \eta_0, \lambda \leq \lambda_0$ *there is a self-referential equilibrium profile* r *where player* i*'s expected payoff is approximately* v_i *for each* i*.*

The proof may be found in Appendix A.2 along with the rest of the proofs corresponding to the results in this section.[12]

This is an approximate one-shot folk theorem due to the noisiness of signals. In particular, to establish a self-referential equilibrium with expected payoffs v, I show that we can only get close to v, but not arbitrarily close, when there are on-equilibrium punishments ($\lambda > 0$), and we must find a critical (small enough) λ_0 for supporting this equilibrium. It follows that the lower the λ, the closer are expected payoffs to v. Moreover, the threshold η_0 is determined by the condition that a player must not gain from deviating by choosing an alternative code of conduct. These two thresholds reflect a trade-off for players between the benefit from adhering to the code of conduct and thus the potential cost of either punishing some innocent opponent or being punished, and the benefit from deviating, thereby obtaining the immediate payoff and avoiding carrying out the punishment.

By having players submitting programs, Tennenholtz (2004) shows a similar result; however, programs use independent mixed strategies of the stage game Γ so it falls short of efficiency payoffs in some cases. A complete folk theorem is proved by Kalai et al. (2010). In their setting, players observe the choice of conditional commitment devices and use jointly controlled lotteries (a la Aumann and Maschler (1995)) to overcome the necessity of randomizing over feasible payoffs. Different from that paper, here agents can choose mixed strategies and use the public correlation device allowing us to dispense of controlled lotteries to obtain efficient payoffs. Unlike the complete folk theorem of Kalai et al. (2010), Theorem 2 is approximate since recognition is correct only probabilistically on the equilibrium path.

4.2 The Repeated Game

This subsection focuses on N-player discounted, finitely repeated games with perfect monitoring. Stages are referred to as periods. The finitely repeated game Γ^T is the T-fold repetition of the stage game Γ. In each period t players simultaneously choose actions $a_i \in A_i$, and after period T the game ends. Let $A^t := A^0 \times \cdots \times A^{t-1}$ be the t-fold Cartesian product of A, and by perfect monitoring the set of t-length public histories is $H^t = A^t$. A behavior strategy for player i is a map $\sigma_i : H \to \Delta(A_i)$. I omit public signals ω on the description of public history for conciseness.[13] Players discount future with common discount factor

[12] This result is the extension of Block and Levine (2012, Theorem 5.1) to imperfect identification of codes of conduct, and it is the non-evolutionary asymmetric version of Levine and Pesendorfer's (2007) result.

[13] In this case public history at period t would be $h^t = (a^0, \ldots, a^{t-1}, \omega^0, \ldots, \omega^t)$ and strategies would be measurable functions with respect to both past actions and the random variable ω.

$\delta \in (0,1)$. Given any strategy profile $\sigma_p \in \Sigma$, a path of play is induced $(a^t)_{t \leq T}$. Thus, the normalized payoffs for player i, $u_i : \Sigma \to \mathbb{R}$ can be written as

$$u_i(\sigma_p) = \frac{1-\delta}{1-\delta^{T+1}} \sum_{t=0}^{T} \delta^t g_i(a^t(\sigma_p)).$$

Finitely repeated games exhibit the so-called *unraveling property*—that is, players have incentive to choose a profitable action in the last period so any strategy other than repetition of a static Nash equilibrium unravels from the end to the beginning of the game. Roughly speaking, since the game ends after the final stage there is no room for retaliation. Within this class, agents' play is sensitive to the endpoint of the game. This leads us to conclude that last-period signals might be sufficient for the self-referential equilibrium set to span above the strictly individually rational payoff.

Next, I will show the main result of this section: A self-referential folk theorem for finitely repeated games with discounting.

Theorem 3. *For $|\mathcal{I}| = 2$. Consider a self-referential game that η-λ permits detection such that for any $t \leq T$, $\tau_i = t$ for all i. For all $v \in V^*$ and for any T-fold repetition of the stage game Γ^T, there exist a discount factor $\underline{\delta} < 1$ and parameters $\eta_\tau > 0$ and λ_τ such that for each $\delta \in (\underline{\delta}, 1), \eta \in [\eta_\tau, 1]$ and $\lambda \in [0, \lambda_\tau]$ there is a self-referential equilibrium profile of codes of conduct r^T so that player i's expected payoff is approximately v_i for all i.*

This theorem places no restrictions on the timing of signals. In other words, to construct a self-referential equilibrium that sustains approximately any $v \in V^*$ players may receive information about opponents' code of conduct in any period of the game, even in period T. The reason why this result is indifferent to the timing can be seen from the proof. Below, I outline a sketch of the argument.

The first step is to find a threshold $\underline{\delta}$, above which players find profitable deviations in the last round of the game for the case without self-referentiality. Then, consider a code of conduct such that grim-trigger strategies are used. Since there are on-equilibrium punishments ($\lambda > 0$), players would punish last-period deviations in period T, ensuring the lowest possible costs when these are triggered. Given these incentives, when signals arrive before the final stage players cannot infer opponents' realisation because this information will be used at the end. Next, the equilibrium code of conduct pins down the cutoffs η_τ and λ_τ that are determined by last-period behavior. This equilibrium code of conduct reinforces the fact that players' behavior is sensitive to payoff perturbations in the endpoint, thereby forgoing the unraveling logic discussed above.

Two features are worth noting in contrast to existing folk theorems for finitely repeated games. First, this result holds for any finite time horizon T and we do not need to find a sufficiently high threshold T^*. Second, it does not require multiple Nash equilibria of the stage game Γ to construct reward and punishment phases. These two conditions are necessary for the proof, for example, in Benoît and Krishna (1985) and Friedman (1985).[14] Unlike these papers, the equilibrium payoff vector cannot be arbitrarily close to $v \in V^*$, in fact, it hinges on the equilibrium thresholds η_τ and λ_τ.

The next corollary to Theorem 3 computes the approximation to payoff vector $v \in V^*$ in the self-referential game.

Corollary 1. *Consider the self-referential equilibrium code of conduct profile r^T in Theorem 3. All players i have expected payoffs given by*

$$U_i(r^T) = v_i - \frac{(1-\delta)\delta^T}{1-\delta^{T+1}}\mathcal{C} := v_i - \varepsilon_{(\delta,T)}\mathcal{C},$$

where the constant $\mathcal{C}$ depends on the stage-game payoffs and the parameter λ.

In words, the self-referential equilibrium payoff $U_i(r^T)$ is a perturbation of the targeted payoff v_i. This perturbation depends not only on the time horizon of the game T and level of patience δ, but also on rewards and punishments in the last round of the game captured by $\mathcal{C}$. The first component is on the equilibrium punishments ($\lambda \geq 0$). When these punishments are rare, λ is small, the expected payoff is close to v_i. The intuition is that adherence to code of conduct is "visible" among players which implies that the agent is unlikely to be punished if he follows r^T. Observe that agents dish out a relatively less costly on the equilibrium punishment whenever deviations are punished at the end of the game. The second component is $\varepsilon_{(\delta,T)}$. For fixed discount factor δ, as we take longer horizon of the finite game, the expected payoffs get closer to v_i, i.e. as $T \to \infty$, it follows that $\varepsilon_{(\delta,T)} \to 0$ and $U_i \to v_i$. Despite the improvement on the approximation to payoff vector $v \in V^*$, the probability of detection η does not approach zero ($\eta \not\to 0$) in the asymptotic limit of the time horizon $T \to \infty$.

Remark The result extends to time average payoffs. One interpretation is that it is continuous in the discount factor δ, that is, as $\delta \to 1$ the equilibrium payoff vector in $\mathcal{G}(\Gamma)$ remains close to V^*.

The result in Section 3 suggests that we may require a lower probability of detection if players acquire information in early periods of the game. To be consistent with Theorem 1, let us assume that $\lambda = 0$.[15] Indeed, I will show that any $v \in V^*$ is attainable in the

[14] For the extension to mixed strategies, Gossner (1995) uses sufficiently long horizon to build reward schemes. See also Smith (1995), Neyman (1999), and Miyahara and Sekiguchi (2013).

[15] Within this class of games it is easy to see that when punishments are triggered on the equilibrium path and close to the beginning of the game, the actual punishment for deviating from the code of conduct is still severe.

self-referential game if private signals are observed earlier than the last round of the game, and that the threshold on probability of detection is smaller than the one found in Theorem 3.

Proposition 1. *Let $|\mathcal{I}| = 2$. Suppose the self-referential game η-λ permits detection such that for all i, $\tau_i = T - k$ where $k \in \mathbb{N}, k > 1$ and $\lambda = 0$. For any $v \in V^*$ and any Γ^T, there exist $\underline{\delta} < 1$ and $\eta_{T-k} > 0$ such that for all $\delta \in (\underline{\delta}, 1), \eta \in [\eta_{T-k}, 1]$ every player i can obtain v_i in the self-referential equilibrium $\check{r}^T$, and $\eta_{T-k} \leq \eta_T$.*

The detection probability threshold η_{T-k} is relaxed relative to the threshold η_T found in Theorem 3 because players punish deviators in early periods without doing so on the equilibrium path. In the self-referential equilibrium, players adhere to a code of conduct profile $\check{r}^T$ that prescribes the minmaxing strategy in period $T-k$ whenever signals point to deviations, making punishments to such deviation more severe. Observe that threshold η_{T-k} is independent of the time horizon T. As long as private signals arrive k periods before the final round T, it is possible to construct these equilibria.

One might be interested in how early signals affect the approximation to payoffs in V^*. The actual computation of approximated payoff vector to v is presented in the following corollary.

Corollary 2. *Let $\check{r}^T$ be the self-referential equilibrium profile from Proposition 1. The approximate expected payoff of each player i is given by*

$$U_i(\check{r}^T) \approx v_i - \frac{(1-\delta)\delta^{T-k}}{1-\delta^{T+1}}\mathcal{C} := v_i - \varepsilon'_{\delta,T-k}\mathcal{C}.$$

This is a corollary to Proposition 1. The reason why the constant $\mathcal{C}$ adds to the perturbation of the expected payoffs is the same as in Corollary 1. Regarding the component $\varepsilon'_{(\delta,T-k)}$ observe that for a fixed time horizon T the sooner punishments are triggered, the larger would be the perturbation of expected payoffs. However, holding period k fixed, in the limit, as $T \to \infty$ the perturbation vanishes as in Theorem 3.

The findings of Theorem 1 and Proposition 1 can be combined to compute the speed of convergence of η_T. Recall that Theorem 1 says that $\eta_T \to 0$ as $T \to \infty$, thus, when signals arrive in the beginning of the game, the critical value of the detection probability under the equilibrium code of conduct profile converges to zero as the length of the truncation grows sufficiently large. Consequently I use the expression for η_T found in Proposition 1 and the corollary below follows this proposition.[16]

Corollary 3. *Consider stage-0 timing intention monitoring structure $\tau_i = 0$, the probability of detection $\eta_T > 0$ converges to zero at rate δ as $T \to \infty$. That is, η_T is $O(\delta^T)$.*

[16] If we consider average payoffs, η_T is $O(T^{-1})$.

Proof. By Proposition 1, we restrict attention to the highest probability of detection $\eta_T > 0$ which is

$$\max_{i\in\mathcal{I}} \eta_{i,T} = \frac{\delta^T(g_i(a_i, a_j^*) - g_i(a^*))}{g_i(\underline{\alpha}_j, \underline{\alpha}_i) - g_i(a_i, \underline{\alpha}_i) + \delta^T(g_i(a_i, a_j^*) - g_i(a^*))}.$$

For fixed $\delta > \underline{\delta}$, this goes to 0 at rate δ as $T \to \infty$. □

In summary the earlier players can detect deviations, the smaller the required probability of detection η that sustains the self-referential equilibrium but the more perturbed these payoffs would be.

5 Exit Games

Going further in the analysis of the signal timing, I explore a second class of games that represent situations where signals at the outset make a world of difference to the equilibrium outcome set but are redundant at later stages. In this section, I examine *exit games*.[17] The main feature of these games is that some player i can terminate the game at any stage t. After presenting a general framework, findings are developed by focusing on two subclasses: splitting and preemption games.

As shown in the previous section, signals have a considerable impact on equilibrium outcomes in finitely repeated games, regardless of the stage at which they arrive. In splitting games, on the other hand, equilibrium behavior is affected only if there are initial-period signals. More precisely, I show that there exists a unique equilibrium outcome when players cannot recognize an opponent's intentions in the beginning of the game, whereas if the recognition technology is available from the start of the game, we can arbitrarily delay when the game ends.

Preemption games lie between these two classes of games. Similarly as for finitely repeated games, results suggest that the set of equilibrium outcomes will be increased even though the intention monitoring happens relatively late in the game. As is the case for splitting games, the self-referential equilibrium that induces exit at early stages requires recognition possibilities ahead of these stages.

The General Framework

In the general environment, there is a set of players $\mathcal{I} \coloneqq \{1, \ldots, N\}$, and these players are involved for finitely T stages. Recall that public histories are defined recursively because players observe all previous interactions. Each player i accesses a finite actions set which is a bipartition, i.e. $A_i(h^t) \coloneqq \{\mathcal{F}_i(h^t) \cup \mathcal{E}_i(h^t)\}$ for all $h^t \in H$, $i \in \mathcal{I}$. The subset $\mathcal{F}_i(h^t)$

[17] These games are similar to *simple timing games*, see Fudenberg and Tirole (1991, Section 4.5).

represents the set of *forward* actions while not guaranteeing that the game continues are necessary for moving to the next stage. The other component, $\mathcal{E}_i(h^t)$, represents the set of *exit* actions; in contrast to forward actions, these actions are sufficient to end the game. Put differently, the game ends if there is only one player choosing exit actions. To lighten notation, let $\mathcal{F}_i(h^t) = \mathcal{F}_i^t$ and $\mathcal{E}_i(h^t) = \mathcal{E}_i^t$. By definition, the sets of forward and exit actions are disjoint, i.e. $\mathcal{E}_i^t \cap \mathcal{F}_i^t = \emptyset$. Idle players are allowed, that is, we may posit a player chooses action $\overline{a}$.

Whenever all active player i's choose forward actions at stage t, $f_i^t \in \mathcal{F}_i^t$, the game continues to the next stage $t+1$. Formally for any history h^t and any player i, $A_i(h^t) \neq \emptyset$ if $a_j^k \notin \mathcal{E}_j^k$ for all $k \leq t-1$ and all active players j. On the other hand, if any active player i plays an exit action $e_i^t \in \mathcal{E}_i^t$ for any $h^t \in H$, it causes the game to end regardless of the actions played by all the other players j.

The last common feature is related to reward mappings, i's payoff functions $g_i : H \to \mathbb{R}$. Until the game ends, each player receive no payoffs. Formally, $g_i(h^t) = 0$ for all histories $h^t \in H$ such that $a_i^k \notin \mathcal{E}_i^k$ for all $i \in \mathcal{I}, k \leq t-1$. In addition, once the game ends no further rewards are received. That is, if some player j chooses e_j^t at stage t, then $g_i(h^k) = 0$ for all $k > t$, any h^t and all players i. If all players continue until and including the last stage T, the game ends and the players' payoffs are zero. This is just a normalization, but all results are unchanged without it.

5.1 Splitting Games

The previous section established the general framework of exit games, including the structure of payoffs and the actions set. This subsection studies the first subclass of exit games which are defined as *splitting games*. These games capture situations in which agents take their surplus share that are conditioned on others, and they pay a positive cost to divide the surplus whenever other agents decide to take their portions. Henceforth, players have incentives to anticipate their rivals because this guarantees the surplus share without incurring the cost. Alternative, it could be interpreted as a partnership with exit (e.g., Chassang (2010)).

In this setup, stages are referred to as time periods. Players discount future payoffs using the constant discount factor $\delta_i \in (0,1)$ and none of them are idle. Reward functions $g_i : H \to \mathbb{R}$ for all players i are additively separable in surplus share and costs. These are represented by

$$g_i(a^t) = w_i(a^t) - c_i(a^t), \qquad \forall a^t \in A(h^t),$$

where the benefit function $w_i(a^t) \geq 0$ has present value at any period t whereas the cost function $c_i(a^t) \geq 0$ has period-t value. Players receive a share of this surplus $w_i(a^t)$ depending on his action and on opponents'. Similarly, each player incurs a cost of $c_i(a^t)$ by taking his share.

In particular, these preferences are characterized by the next set of assumptions. For all $i \in \mathcal{I}$, $h^t \in H$, $f_i^t \in \mathcal{F}_i^t$, $f_{-i}^t \in \mathcal{F}_{-i}^t$, $e_i^t \in \mathcal{E}_i^t$, and $e_{-i}^t \in \mathcal{E}_{-i}^t$ we have:

S.1 $w_i(e_i^t, \cdot) = w_i > 0$ with constant w_i, and $w_i(f_i^t, \cdot) = 0$;

S.2 $c_i(e_i^t, e_{-i}^t) = c_i > 0$ where c_i is constant, otherwise $c_i(\cdot) = 0$.

In words, condition S.1 ensures that players would prefer to exit the game before his opponents rather than play a forward action. The constant w_i 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 the constant cost c_i discounted by δ_i their time preference parameter, that is, $\delta_i^t c_i$. For each of the complement action profiles players pay nothing, $c_i = 0$. Think of the constant c_i as the cost of reaching agreement, deciding and proposing a voting rule to share the surplus, or as a one-time version of the transaction cost considered by Anderlini and Felli (2001). Together assumptions S.1 and S.2 imply that terminating the game late is a cooperative action.

The next theorem should be interpreted as an impossibility result. It says that if players comprehend cues 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.

Theorem 4. *Consider $|\mathcal{I}| = 2$ and any splitting game Γ. Suppose that the self-referential game η-λ permits detection with $\tau_i \geq 1$ for any i. Then for any η, λ there exists a unique self-referential equilibrium outcome in which all profiles of codes of conduct r^T have all players i choosing $r_i^i(h^0, h_i^0) = e_i^0$ for some exit action $e_i^0 \in \mathcal{E}_i^0$ and all $h_i^0 \in H_i$.*

The proof is included in Appendix A.3.[18]

The first point to make is that η-λ detection, players' ability to understand their adversaries' choice of strategies is irrelevant to this result. The driving force is precisely the timing of intention monitoring structure τ_i.

Here is a rough outline of the proof. Under assumption S.1 that players obtain their surplus fraction only if they choose an exit action, in any self-referential equilibrium they must simultaneously terminate the game. Consider now a profile of codes of conduct r' that prescribes exit in some period $t' > 0$. Against such profile, any player may choose an alternative code of conduct unilaterally exiting in period 0, thereby taking w_i without paying c_i, this is the optimal deviation because adherence to r' approximately gives an expected payoff of $w_i - \delta^{t'} c_i$. Such deviation is undetectable since information about intentions is acquired later than the initial period. Moreover, when the deviator exits, terminating the

[18] The reader may also find all the proofs of the results in this section in Appendix A.3.

game, opponents cannot punish such behavior by drawing on public history. This argument also applies to any period $t > 0$. Thus, I have shown that the self-referential game has a unique equilibrium in which the players exit in the first period and hence delayed exit is never reached.

This result leads us to conclude that signals are useless when either the game ends or signals arrive late. Late, in this sense, is relative to the point at which players wish to deviate from the code of conduct r. This important, but previously neglected, feature arises in self-referential games with different timing of informative signals.

As codes of conduct represent social norms, one would expect to see players agree upon exiting in periods $t > 0$. In fact, choosing an exit profile in the next period f^{t+1} Pareto dominates doing it in the current period f^t. A natural question is then, under what conditions agents will make these kinds of agreement? The following anti-impossibility theorem answers this question positively.

Theorem 5. *Let $|\mathcal{I}| = 2$ and Γ be a splitting game. Suppose the self-referential game η-λ permits detection with $\tau_i = 0$ for all i and any period $k \leq T$. Then there are $\eta_k > 0$ and λ_k such that for each $\eta \in [\eta_k, 1], \lambda \in [0, \lambda_k]$ there exists a self-referential equilibrium r^k in which all players i exit the game in period k, $r_i^i(h^k, h_i^k) = e_i^k$ for all $e_i^k \in \mathcal{E}_i^k, h_i^k \notin \overline{H}_i$.*

As with the intention monitoring structure $\tau_i > 0$, there is a self-referential equilibrium with immediate exit. However, contrary to Theorem 4, every exit profile can take place with a delay of an arbitrary number of periods. It corresponds to the fact that agents' ability to detect intentions of deviation occurs sufficiently in advance, thereby allowing players to punish deviations discussed above.

To see the intuition behind this result, recall that any r with exit in $t \geq 1$ must have the players simultaneously terminating the game in order to be a self-referential equilibrium. As a result, the optimal deviation against this code for any player i is to choose $e_i^0 \in \mathcal{E}^0$. When the possibility of detection happens in any period $t > 0$, these equilibria are unsustainable. If signals are observed in the first period $\tau_i = 0$, however, players may receive informative signals pointing to these kind of deviations. Turning to the construction of the self-referential equilibrium, consider a code of conduct r^k in which players agree to exit in period k. Then, it uses grim-trigger strategies in which a player chooses forward actions unless there is evidence of exit in any period $t \leq k$. Again as before, given the exit profile $e^k \in \mathcal{E}^k$, the parameters η_k and λ_k are chosen so as to provide each player with the right incentives to adhere to the code of conduct r^k.

One key observation is that the requirement that players receive signals at the outset $\tau_i = 0$ is independent of the particular exit profile e^k that is aimed to be sustained in r^k. The reason why this holds is that each player finds it optimal to exit in any period preceding period k, but then the reasoning works backwardly until the first period. Therefore, the deviation is characterized by exiting in the beginning regardless of period k. In addition, by inspecting the proof one may observe that the probability required for the equilibrium code of conduct decreases with the exit profiles sustained in late periods.

To conclude this subsection, I analyze welfare with respect to the results we have shown so far. Consider a parameterization of the information structure (Y, π) with $\tau_i = 0$ for each i and bounds η, λ such that at least one exit profile e^t in each period t is implementable by some code of conduct profile $r \in R$. Furthermore, the reward mappings $g_i(e^t) = w_i - \delta^t c_i$ are monotonically increasing in period $t \geq 0$ for all exit profiles $e^t \in \mathcal{E}^t$ since the discounted cost function $\delta^t c_i$ is decreasing in $t \geq 0$ and the benefits function w_i remains constant as it has period-0 value. Even though there are multiple self-referential equilibria, by monotonicity it is possible to have them Pareto-ranked: period-t exit profiles e^t are Pareto dominated by period-$(t+1)$ exit profiles e^{t+1} for any period t. This allows me to compute how much is lost if signals come late in the game, i.e. $\tau_i > 0$ for all i.

Players discovering adversaries' intentions after the game starts have a negative welfare effect; indeed, such agents obtain the lowest Pareto-ranked payoff. On the other hand, when the intention monitoring structure (Y, π) is such that players observe signals in the beginning, i.e. $\tau_i = 0$, there exist self-referential equilibria where the equilibrium expected payoff vector is greater than the lower bound of Pareto-ranked payoffs whenever $\tau_i \geq 1$. The next proposition summarises the welfare implications of the results about the timing of signals in splitting games.

Proposition 2. *Suppose that* Γ *is a splitting game.*

(i) If the intention monitoring structure (Y, π) *allows* $\tau_i = t$ *for any* $t \geq 1$ *and all* i*, then the unique outcome of any self-referential equilibrium gives the worst Pareto-ranked payoff vector;*

(ii) If the intention monitoring structure (Y, π) *satisfies* $\tau_i = 0$ *for each* i*, then for any period* $k \leq T$*,* $\eta \geq \eta_k$ *and* $\lambda \leq \lambda_k$*, there exists a self-referential equilibrium with payoff vector greater than the worst Pareto-ranked payoff vector.*

5.2 Preemption Games

The analysis of the previous sections suggest that at one extreme, in finitely repeated games with discounting, the timing of signals is irrelevant to construct self-referential equilibria, obtaining a few versions of the folk theorem. At the other extreme, in splitting games, I find that informative signals must arrive at the outset to have some impact on the equilibrium outcomes set of the self-referential game. Identifying a class of games that fits between these two, perhaps, is of particular interest to better understand the timing of signals. In this subsection, I examine *preemption games*. This class captures situations in which players alternate veto power to terminate the game. For example, the two most influential parties in the Congress, or a wage-bargaining between a company and a labor union, both may alternate such veto power. Furthermore, it includes the well-studied centipede game.

The preemption game is typically modeled as follows, it runs from stage $t = 0$ to the odd finite stage T. There is a set of two players and each player i does not discount between stages, i.e. $\delta_i = 1$. At each stage t, there is only one player i *active*. The game starts with player 1 moving and ends with player 2 choosing an action.

Let $\iota : H \setminus Z \to \mathcal{I}$ be the player function. This function assigns to each nonterminal history $h \in H \setminus Z$ a player i. I write $\iota(h^t) = i$ for the case in which player i makes a choice from $A_i(h^t)$ after history h^t in stage t. To make the analysis interesting, there must be a minimum level of alternation between players. More specifically, there is no terminal history $h \in Z$ in the game where for some stage $k \in \mathbb{N}$, for all $t < k$ $\iota(h^t) = 1$ and for all $t \geq k$ $\iota(h^t) = 2$. To track the number of identity changes along the path I define the function $\phi : \mathbb{N}_0 \to \mathbb{N}_0$ which is given by

$$\phi(n) := \left\{ \max\ l(h^t) \mid h^t \in H \text{ such that } \sum_{k=0}^{T} \mathbb{1}_{\{\iota(h^k) \neq \iota(h^{k+1})\}} = n \right\},$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.[19] The function ϕ simply returns the stage at which there are exactly n alternations between players. Let $\overline{n}$ be the maximum number of shifts in the game. Given any stage k and n number of shifts, I refer to k_{-n} as the stage from which we observe n number of shifts up to stage k and it is computed as $k_{-n} = \phi(\phi^{-1}(k) - n)$. Analogously, I define $k_{+n} = \phi(\phi^{-1}(k)) + n)$ to be the stage at which n shifts have occurred after stage k.

The reward mappings $g_i : H \to \mathbb{R}$ are required to fulfill the following conditions. For all players $i, j \in \mathcal{I}$ and any pair of stages k, t with $k > t$

P.1 $g_i(e_i^t, \overline{a}) < g_i(e_i^k, \overline{a})$ for all $e_i^t \in \mathcal{E}_i^t, e_i^k \in \mathcal{E}_i^k$;

P.2 $g_i(e_i^t, \overline{a}) > g_i(\overline{a}, e_j^k)$ for all $e_i^t \in \mathcal{E}_i^t, e_j^k \in \mathcal{E}_j^k$.

The first condition P.1 guarantees that whenever players are active, they prefer to exit the game at later stages of the game. This implies that if player i is active between stage t and stage t', i.e. $\iota(h^k) = i$ for $t \leq k \leq t'$, then his choice of ending the game in any of these stages k before stage t' is strictly dominated by the choice of ending it at stage t'.[20]

[19] I write $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

[20] For instance, consider the bargaining between some group of firms and a labor union. In general, the labor union bargains with no flexibility until the period of mandatory conciliation by law. This idea also carries through to *inter-active period* decisions, basically, agents have incentive to terminate the game at later stages.

Nevertheless, players face a trade-off between waiting to the 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 stage he is active, but in order to reach it he will go through an inactive stage. The transition between active stages is threatened by having the opponent exiting the game.

The analysis of preemption games starts by showing that the unique self-referential equilibrium outcome exhibits player 1 finishing the game at the end of his first active period if signals arrive in the penultimate active period.

Theorem 6. *Let* Γ *be a preemption game. For any stage* $k \geq \phi(\overline{n})$*, assume that the self-referential game* η-λ *permits detection and allows* $\tau_i \geq k_{-1}$ *for any* i*. Then for any pair of parameters* η, λ *there is a unique equilibrium outcome for any code of conduct profile* r^k *such that* $r_i^i(h^t, h_i^t) = e_i^t$ *for some* $e_i^t \in \mathcal{E}_i^t$ *and all* $h_i^t \in H_i$ *where* $t = \phi(1) - 1$ *for player 1 and* $t = \phi(2) - 1$ *for player 2.*

As in the case of splitting games, signals arriving late play a key role in establishing this impossibility result, except that now they must arrive after a critical stage that is not necessarily the first stage. The threshold $\tau_i \geq k_{-1}$ (with $k \geq \phi(\overline{n})$) is necessarily greater than the threshold $\tau_i \geq 1$, in Theorem 4 because the critical stage $\phi(\overline{n})$ is at least three by the minimum level of alternations assumed, implying that $\tau_i \geq 2$. Note that stage k_{-1} might be at the very end of the game as is defined by alternations. Thus, if the game lasts for long horizon T and there are one-period alternations, the critical value k_{-1} will be $T - 1$ as the parameter satisfies $\phi(\overline{n}) = T$.

To build some intuition for this result, consider without loss $k = \phi(\overline{n})$ and $\tau_i = k_{-1}$. Suppose for a moment that the code of conduct profile stipulates exit in stage, say, $k - k'$ for some $k' < k$. By assumption P.2, the active player in stage k'_{-1} deviates from this code of conduct because he can always terminate the game without being detected as informative signals are observed afterwards. Accordingly, the candidate profile of codes of conduct r^k must have players exiting after stage k_{-1}. Suppose that it requires exit in period k without loss, noting that player 2 terminates the game. As before, player 1 wishes to exit at stage k_{-1} by assumption P.2, while player 2 observes informative signals after this profitable deviation. Player 2 would then find it beneficial to choose an exit action at stage k_{-2}. But player 1's response to this behavior would be to exit at stage k_{-3}. Proceeding in this way, it turns out that with timing of signals $\tau_i \geq k_{-1}$, there is a unique self-referential equilibrium. In this equilibrium, player 1 exits in his first active period terminating the game at stage $\phi(1) - 1$, whereas player 2 chooses exit actions at the last stage when active $\phi(1)$.

In the same sense as in splitting games, the uniqueness of the self-referential equilibrium outcomes set is independent of the accuracy of signals captured by parameters η and λ.

Precisely, the timing of signals impedes the construction of equilibria that exhibit delayed exit.

As discussed above, when $\tau_i = k_{-1}$ for any stage $k \geq \phi(\overline{n})$ there is a unique equilibrium outcome in which the game terminates in the first active period. In what follows, I explore under what conditions there also exist equilibria with delayed exit. From the previous result two insights emerge, first the lower bound on the fixed stage $\phi(\overline{n})$ does not allow players to use signals at stage k_{-1} so as to exit later than such stage. The second insight is that the timing of signals τ_i depends only on this lower bound.

The following result characterizes self-referential equilibria in which the game terminates after the first active period only insofar as signals are observed early in the game.

Theorem 7. *Consider a preemption game* Γ. *For any stage* $k \geq \phi(2)$, *suppose that the self-referential game* η-λ *permits detection and provides* $\tau_i \leq k_{-2}$ *for all* i. *Then there exist an* $\eta_k > 0$, *a* λ_k *and a code of conduct profile* r^k *such that for all* $\eta \in [\eta_k, 1]$, $\lambda \in [0, \lambda_k]$ *in the self-referential equilibrium player* $i = \iota(h^k)$ *chooses an exit action at stage* k, *i.e.* $r_i^i(h^k, h_i^k) = e_i^k$ *for some* $e_i^k \in \mathcal{E}_i^k$ *and any* $h_i^k \notin \overline{H}_i$.

Condition $k \geq \phi(2)$ is a mild restriction, requiring that the targeted code of conduct profile is sufficiently rich so that players can punish intentions of exit before stage k. This is because any stage below this threshold where the game terminates cannot be a self-referential equilibrium. To see this, suppose that $k < \phi(2)$, say player 2 is active and the code of conduct profile dictates exit there. If that is the case, player 1 wishes to exit in his first active period by condition P.2, meaning that this is the only equilibrium outcome since player 2 cannot punish this behavior. Clearly, the same argument applies when player 1 is active.

To illustrate the role of the timing of signals in this result, I briefly describe the proof. Fix a stage $k \geq \phi(2)$, and suppose that player ι is active. Once again, the constructed codes of conduct must use grim-trigger strategies because of the detection technology assumed here, stating exit at stage k. First, consider player $j \neq \iota$. By assumption P.2, he has incentives to deviate at stage k_{-1}, but not at earlier stages by assumption P.1. Correspondingly, player ι punishes these intentions, whenever there is evidence of such behavior, at stage k_{-2} which is possible as $\tau_i \leq k_{-2}$. At the same time, player ι may actually exit in the last stage of his active period containing stage k motivated by condition P.1; therefore, player j will punish these intentions of behavior at stage k_{-1} provided that signals point out deviations. Under the proposed codes of conduct, the incentives to adhere are aligned through the parameterization of η_k and λ_k as before. Observe that there is still an equilibrium in which player 1 concludes the game in his first active period.

As in splitting games, early informative signals permit one to construct self-referential equilibria with delayed exit. However, contrary to such class of games, the required stage at which agents observe signals need not be the initial stage. In contrast to finitely repeated games, where the point at which signals are observed is completely irrelevant, and splitting games, in which signals must arrive at the outset in order to construct nontrivial equilibrium, in the present environment the timing of signals depends only on the stage at which players wish to terminate the game. The difference hinges on the fact that players can asynchronously terminate the game and they may remain active for more than one stage.

Finally, consider the assumptions in Theorem 7. Because the timing of signals suffices to construct self-referential equilibrium with exit after the fixed stage k, we obtain the following result.

Corollary 4. *Given stage k, for each stage $t \geq k$ there is self-referential equilibrium where player $i = \iota(h^t)$ plays $r_i^i(h^t, h_i^t) = e_i^t$ for some $e_i^t \in \mathcal{E}_i^t$ and each $h_i^t \notin \overline{H}_i$.*

Proof. This follows by noting that the detection probability η_k from Theorem 7 allows us to construct a self-referential equilibrium in which both players adhere to the code of conduct r^i that dictates for player $j = \iota(h^t)$ the strategy $r_j^i(h^t, h_j^t) = e_j^t$ for some $e_j^t \in \mathcal{E}_i^t$ and any $h_j^t \notin \overline{H}_i$. □

6 Asynchronous Intention Monitoring

So far, I have restricted the analysis to information structures of $\mathcal{G}(\Gamma)$ in which all players observe signals in the same stage, that is, $\tau_i = t'$ for all i and any stage t'. In many applications, however, agents could have heterogenous abilities to recognize others' rules of behavior. For instance, an entry firm may have better information about pricing strategies than the incumbent. Similarly, consider an underlying game in which one player moves in early stages and therefore the others might observe his behavior acquiring information before such player. This section studies an information structure of self-referential games that allows for heterogeneity in the timing at which players receive private signals, maintaining the assumption that the timing of signals is deterministic and commonly known. To facilitate the analysis I focus on two-player discounted, finitely repeated games and exit games.

The description of the self-referential game is exactly as in Section 2.2. Although, the key difference is the intention monitoring structure so I redefine it here. We say an intention monitoring structure (Y, π) is *stage-(t, t') timing* if for two stages t, t' with $t \neq t'$, for any code of conduct profile $r \in R$, $\tau_1 = t$ and $\tau_2 = t'$. This definition says that players 1 and 2 receive signals at stages t and t', respectively.

To begin with, I analyze finitely repeated games with discounting, continue with splitting games and conclude with preemption games. In particular, the interest is to compare all previous results with synchronous signals relative to self-referential games with asynchronous monitoring information structure.

6.1 Finitely Repeated Games

In what follows, I will show that all feasible and strictly individually rational payoffs can be approximated in the self-referential game with asynchronicity as for δ big enough. In fact, the key qualitative property of self-referential equilibrium in repeated games with synchronicity, that players can deter deviations regardless of the period in which they observe informative signals, extends to the case of asynchronicity provided that players are sufficiently patient.

Proposition 3. *Suppose that the self-referential game η-λ permits detection and endows players with $\tau_1, \tau_2 \leq T$. For any $v \in V^*$ and Γ^T, there exist $\underline{\delta} < 1$, $\eta_T > 0$ and λ_T such that if $\delta \geq \underline{\delta}$, $\eta \geq \eta_T$ and $\lambda \leq \lambda_T$ then there is a self-referential equilibrium r^T where each player i gets an expected payoffs approximately equal to v_i.*

The proof is in Appendix A.4.[21]

Thus, the self-referential folk theorem is independent of asynchronicity. The reason why asynchronous timing does not affect previous results is that finitely repeated games with sufficiently patient players are sensitive to the endpoint. In other words, since players are patient enough they find it optimal to deviate in the last round of the game, implying that for any asynchronous timing each player has received his private signal by that time, and then agents simultaneously use this information. The proof follows closely the same structure as the proof of Theorem 3.

Recall that Theorem 3 points out that players observing signals in period $T - k$ could punish in such period whenever there is evidence of potential deviations. With asynchronous signals this logic applies as well. Although, for this profile to be a self-referential equilibrium the punishment stage must be $T - k = \max(t, t')$. Otherwise, the agent receiving signals late could make an inference about his opponent's realisation.

6.2 Splitting Games

In this class of games, the fact that each player can terminate the game in any period, together with the properties of reward functions, were identified as the reason why initial-period signals are required to construct self-referential equilibria that exhibit late exit profiles. Indeed, every player finds it optimal to exit the game in period zero irrespective of the code of conduct profile, meaning that each player must observe informative signals at the outset. With asynchronicity, on the other hand, at least one player receives information about intentions in period $t \geq 1$, that in turn cannot punish the other player intending to exit in the first period. It is clear then that there is no timing so that the self-referential equilibrium outcomes set is not unique.

[21] The rest of the proofs related to this section are also in Appendix A.4.

Proposition 4. *Consider any splitting game* Γ*. Assume that the self-referential game* η*-*λ *permits detection where any* $\tau_1, \tau_2 \leq T$*. Then for all* η, λ *there exists a unique self-referential equilibrium outcome where for any* r *each player* i *conforms to* $r_i^i(h^0, h_i^0) = e_i^0$ *for all* $h_i^0 \in H_i$ *and some* $e_i^0 \in \mathcal{E}_i^0$*.*

This result contrasts sharply with the results found in Section 5.2, especially, Theorem 7 that entails conditions such that the construction of equilibrium in which late exit profiles is possible. As discussed above, in splitting games every player must observe signals at the outset $\tau_i = 0$ for a self-referential equilibrium to sustain delayed exit, i.e. $e^t \in \mathcal{E}^t$ for all periods $t \geq 1$. When there is asynchronicity, such delayed exit profiles are not feasible, thereby leading to immediate exit. This case is of special interest, because it shows that once we relax the assumption of synchronous intention monitoring, the consequences in terms of welfare can be quite severe. The impact of heterogeneity in timing of signals on welfare is characterized by predicting the worst Pareto-ranked payoff vector as the unique equilibrium outcome.

6.3 Preemption Games

Now, I conclude the analysis of asynchronous monitoring structure by revisiting preemption games. Similar to the case of synchronous timing, I find that there exists a unique equilibrium outcome in which player 1 exits the game in his first active period while under less restrictive conditions. In particular, only player 2 may receive signals sufficiently late in the course of play. As already pointed out, on the other hand, the existence of alternation between players to exit the game, where such decisions depend only on the targeted code of conduct profile, may help to construct self-referential equilibrium with late exit profiles.

From this discussion, I state the next result which parallels Theorem 6 obtained for synchronous monitoring structure.

Proposition 5. *Suppose that* Γ *is a preemption game, and for each stage* $k \geq \phi(\overline{n})$ *suppose that the self-referential game* η*-*λ *permits detection and allows* $\tau_2 \geq k_{-1}$ *and any* τ_1*. Then, for any pair* η, λ *there is a unique self-referential equilibrium outcome such that for all* r *each player* i *chooses* $r_i^i(h^t, h_i^t) = e_i^t$ *for some* $e_i^t \in \mathcal{E}_1^t$*, all* $h_i^t \notin \overline{H}_i$*,* $t = \phi(1) - 1$ *for player 1 and* $t = \phi(2) - 1$ *for player 2.*

The key observation concerns player 2's timing of signals. If he observes signals in his last active period ($\tau_2 \geq k_{-1}$), then there is a unique equilibrium outcome in which player 1 leaves the game in the first active period ($\phi(1) - 1$). In contrast to the analogous result in the synchronous case, this requires only that player 2 receives signals late enough in the game. The intuition is simple. Suppose that we try to sustain a self-referential equilibrium exhibiting exit at any stage $k \geq \phi(\overline{n})$. Consider first player 2 who cannot observe any informative cues about his opponent's behavior until his last active period. As player 1 benefits from exiting later on (by assumption P.1), he finds it optimal to leave the game at stage k_{-1} so that he preempts player 2. Then, player 2 chooses to end the game before this stage knowing that player 1 will not wait until that period, and because he prefers to be the one terminating the game rather than player 1 by condition P.2. Consequently, this logic applies to any stage $k \geq \phi(\overline{n})$, therefore, both players will exit in their first active periods. At the end, player 1 exits terminating the game in the last stage of his first active period. As was the case when signals arrive late relative to the exit profile, η-λ detection technology—the precision of information—is completely irrelevant.

As highlighted before, the optimal exit of players is conditioned on the aimed code of conduct. Thus, one might be interested in knowing whether there exist information structures (Y, π) allowing agents to adhere to self-referential equilibrium codes of conduct with delayed exit.

Proposition 6. *Pick any stage $k \geq \phi(2)$ in any preemption game Γ with $\iota(h^k) = \iota$. Assume that the self-referential game η-λ permits detection with $\tau_i \leq k_{-1}$ and $\tau_\iota \leq k_{-2}$ for $i \neq \iota$. Then, there are $\eta_k > 0$ and λ such that for all $\eta \geq \eta_k$, $\lambda \leq \lambda_k$ in the self-referential equilibrium r^k player ι chooses $r^\iota_\iota(h^k, h^k_\iota) = e^k_\iota$ for some $e^k_\iota \in \mathcal{E}^k_\iota$ and for each $h^k_\iota \notin \overline{H}_\iota$.*

This proposition says that given some stage k (for $k \geq \phi(2)$) if the self-referential game allows players to observe signals early relative to the stage k, and these signals are sufficiently informative, then there is a self-referential equilibrium where the game ends at this particular stage.

The lower bound $\phi(2)$ guarantees that both players are able to punish potential deviations from any code of conduct. Recall that for any stage $k < \phi(2)$, none of the stages within player 2's first active period could be sustained as a self-referential equilibrium exit profile, and that there exists a unique equilibrium outcome.

To see why asynchronous signals do not affect the equilibrium outcomes set, resulting in a unique prediction as in the case of splitting games, consider a code of conduct profile r^k with exit at stage $k \geq \phi(2)$. Suppose that ι ends the game with $j \neq \iota$. By assumption P.1, player j and player ι find it optimal to exit at stage k_{-1} and $(k_{+1} - 1)$, respectively. The timing of signals satisfies $\tau_j \leq k_{-1}$ and $\tau_\iota \leq k_{-2}$, by implication, such deviations are punishable. What is more important is that because players cannot terminate the game simultaneously—there are alternations on active periods between players—the proposed codes of conduct could be a self-referential equilibrium, as long as signals arrive at different time but allow each player to punish intentions of deviation with accuracy. It remains to find parameters η_k and λ_k that balance incentives so that agents prefer to adhere to r^k rather than to exit before stage k.

A further implication of Proposition 6 is that we can sustain $e^t \in \mathcal{E}^t$ for $t \geq k$ in $\mathcal{G}(\Gamma)$ for some parameters η_t and λ_t.

Corollary 5. *Consider stage k, for any stage $t \geq k$ there exists a self-referential equilibrium in which player ι chooses $r_\iota^t(h^t, h_\iota^t) = e_\iota^t$ for some $e_\iota^k \in \mathcal{E}_\iota^k$ and all $h_\iota^k \notin \overline{H}_\iota$.*

7 Concluding Remarks

In this paper, I have developed a model that allows agents to learn about opponents' intentions not only at the outset, but also in the course of the game. This paper characterizes how the time at which intentions are inferred shapes the size of the equilibrium outcome set, which in turn crucially depends on the underlying game. Because of this dependence, by focusing on games with perfect information the role of the signal timing is clearly identified. In particular, I provide a characterization, for certain classes of games, in terms of the relation between the number of equilibria in the infinite horizon game and the number of equilibria in the finite horizon version of the game.

As in the benchmark recognition technology model, I found a significant impact of pregame signals on equilibrium outcomes. In particular, for generic games I established a connection between infinite horizon equilibria and self-referential equilibria of the finite truncation.

A couple of principles emerge from the families of games studied here. First, sustaining the proposed code of conduct profile as a self-referential equilibrium hinges on agents' ability to anticipate deviations. Of course, the timing of signals must allow players to observe these signals before the actual deviation, only insofar as observing intentions requires this per se. More importantly, informative signals may arrive sufficiently in advance for punishment to be severe, providing agents with incentives to adhere promptly to the code of conduct. Second, the noisiness of the recognition technology implies that there exist on-equilibrium punishments that might be very costly to players. Henceforth, even when agents observe intentions early they might delay punishments to avoid these on-equilibrium costs.

There are three extensions that will be part of future research. Throughout the analysis, I assumed that the time of arrival is deterministic and commonly known. It would be interesting to examine what happens when the arrival of signals is stochastic, for example, it could follow a Bernoulli process. This assumption seems to be natural when a firm, perhaps, is uncertain whether its opponents have received information about the stage of a developing product.

The methods developed in this paper can be applied to study other settings, for instance, repeated games with imperfect public monitoring. In that case, although there might be a tension between public signals and signals from codes of conduct, these two sources of information would complement each other. The decision to trigger punishments may depend on sufficient statistics based on public history and on the period in which signals arrive.

Finally, I have considered a recognition technology—η-λ permit detection, which captures the idea of reciprocal behavior—that allows us to construct simple self-referential equilibria that uses grim-trigger codes of conduct. With a more general information structure (Y, π), one could allow a richer set of detection possibilities, for instance, codes of conduct that recognize other codes of conduct as long as they provide the same outcome in the game.

References

Anderlini, L., and Felli, L. (2001). Costly bargaining and renegotiation. *Econometrica*, *69*(2), 377–411.

Aumann, R., and Maschler, M. (1995). *Repeated games with incomplete information*. Cambridge, MA: The MIT Press.

Benoît, J.-P., and Krishna, V. (1985). Finitely repeated games. *Econometrica*, *53*(4), 905-22.

Block, J. I., and Levine, D. K. (2012). *Codes of conduct, private information and repeated games* (Working paper). Washington University in St. Louis.

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

Börgers, T. (1991). Upper hemicontinuity of the correspondence of subgame-perfect equilibrium outcomes. *Journal of Mathematical Economics*, *20*(1), 89–106.

Calcagno, R., Kamada, Y., Lovo, S., and Sugaya, T. (2013). Asynchronicity and coordination in common and opposing interest games. *Theoretical Economics*. (forthcoming)

Caruana, G., and Einav, L. (2008). A theory of endogenous commitment. *Review of Economic Studies*, *75*(1), 99–116.

Chassang, S. (2010). Fear of miscoordination and the robustness of cooperation in dynamic global games with exit. *Econometrica*, *78*(3), 973–1006.

Crawford, V. P. (2003). Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, *93*(1), 133–149.

Forges, F. (2013). A folk theorem for bayesian games with commitment. *Games and Economic Behavior*, *78*(0), 64–71.

Frank, R. H. (1987). If homo economicus could choose his own utility function, would he want one with a conscience? *American Economic Review*, *77*(4), 593–604.

Friedman, J. W. (1985). Cooperative equilibria in finite horizon noncooperative supergames. *Journal of Economic Theory*, *35*(2), 390–398.

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.

Gossner, O. (1995). The folk theorem for finitely repeated games with mixed strategies. *International Journal of Game Theory*, *24*(1), 95–107.

Greenberg, J., Monderer, D., and Shitovitz, B. (1996). Multistage situations. *Econometrica*, *64*(6), 1415–1437.

Harris, C. (1985a). A characterisation of the perfect equilibria of infinite horizon games. *Journal of Economic Theory*, *37*(1), 99–125.

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

Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory*, *123*, 81–104.

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.

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.

Miyahara, Y., and Sekiguchi, T. (2013). Finitely repeated games with monitoring options. *Journal of Economic Theory*, *148*(5), 1929–1952.

Neyman, A. (1999). Cooperation in repeated games when the number of stages is not commonly known. *Econometrica*, *67*(1), 45–64.

Peters, M. (2013). *Reciprocal contracting* (Mimeo). University of British Columbia.

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

Peters, M., and Troncoso-Valverde, C. (2013). A folk theorem for competing mechanisms. *Journal of Economic Theory*, *148*(3), 953–973.

Radner, R. (1981). Monitoring cooperative agreements in a repeated principal-agent relationship. *Econometrica*, *49*(5), 1127–1148.

Smith, L. (1995). Necessary and sufficient conditions for the perfect finite horizon folk theorem. *Econometrica*, *63*(2), 425–430.

Solan, E., and Yariv, L. (2004). Games with espionage. *Games and Economic Behavior*, *47*(1), 172–199.

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

Appendix A Proofs

A.1 Proof of Theorem 1

Before proving Theorem 1, I need to state some notation that is used in the proof. The magnitude of payoffs after stage τ could be measured by the greatest variation in payoffs due to events after stage τ for any player $i \in \mathcal{I}$:

$$\zeta^{\tau} := \{\sup |u_i(\sigma) - u_i(\hat{\sigma})| \;\mid\; i \in \mathcal{I},\ \sigma, \hat{\sigma} \in \Xi \text{ such that } \sigma_{\tau} = \hat{\sigma}_{\tau}\}.$$

The constant ζ^{τ} describes how much weight we put on payoffs at the tail of the game. Continuity at infinity implies that $\lim_{\tau \to \infty} \zeta^{\tau} = 0$.[22] Lastly, I define the minmax payoff of player $i \in \mathcal{I}$ in (mixed strategies) in the τ-truncation Γ^{τ} as

$$\underline{u}_{i,\tau} := \min_{\sigma_{-i,\tau}|H^{\tau} \in \Xi_{-i}} \max_{\sigma_{i,p,\tau}|H^{\tau} \in \Sigma_i} u_i(\sigma_{i,p,\tau}, \sigma_{-i,\tau}) \text{ for histories } H^{\tau} \subset H,$$

and I write $\underline{\sigma}_{-i,\tau}$ to denote the minmax profile against player i, and let $\underline{\sigma}_{i,p,\tau}$ be the best respond to $\underline{\sigma}_{-i,\tau}$ by player i. For any σ_{-i}, denote by $BR_i(\sigma_{-i}) = \operatorname{argmax}_{\sigma_i \in \mathcal{E}_i} u_i(\sigma_i, \sigma_{-i})$ the set of best responses to σ_{-i} of player i.

To construct the truncation choose an arbitrary strategy σ in the infinite horizon game Γ^{∞}. In this case it is convenient to work with the strategy $\overline{\sigma}$ which is the constant repetition of the no-decision action $\overline{a}$. Then, embed the strategy σ_{τ} in the truncation version Γ^{τ} into the infinite horizon game Γ^{∞} by concatenating the strategy σ_{τ} with the strategy $\overline{\sigma}$. The strategy σ_{τ} states the plan of play in all stages up to and including stage τ, and that players follow $\overline{\sigma}$ in subsequent stages $t > \tau$. I will evaluate the limit of the τ truncation of the game as the truncation grows, $\tau \to \infty$. Since the action space is finite it is sufficient to work with the product topology.[23]

Proof. Suppose that $\tau_i = 0$ for all players i. Fix any subgame perfect equilibrium $\hat{\sigma} \in \Xi$ of the infinite horizon game Γ^{∞}. Suppose we take a τ-truncation of this game, Γ^{τ}. If the

[22] To obtain continuity at infinity it suffices to assume that players discount and rewards are bounded functions, i.e. there exists some constant $\mathcal{C}$ such that $\max_{a^t} |g_i(a^t)| < \mathcal{C}$ for all $i \in \mathcal{I}$.

[23] 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 $h \in H$.

truncated strategy profile $\hat{\sigma}_\tau$ turns out to be an equilibrium of Γ^τ, then the profile of codes of conduct $\hat{r}^\tau \in R$ chosen by all players i is $\hat{r}^i_j(h^t, h^t_j) = \hat{\sigma}^t_{j,\tau}(h^t)$ for all $j \in \mathcal{I}, h^t \in H$ and $h^t_j \in H_j$. It follows immediately that it would form a self-referential equilibrium.
On the other hand, suppose that $\hat{\sigma}_\tau$ is not an equilibrium of the truncated game Γ^τ. Let $\sigma_{i,\tau}$ be the optimal deviation of player i from $\hat{\sigma}_\tau$, that is, $\sigma_{i,\tau} \in BR_i(\hat{\sigma}_{-i,\tau})$. Pick the profile of codes of conduct $\hat{r}^\tau \in R$ which prescribes for all $i, j \in \mathcal{I}$:

$$\hat{r}^i_j(h^t, h^t_j) := \begin{cases} \hat{\sigma}^t_{j,\tau}(h^t) & \text{for all } h^t \in H^t, h^t_j \notin \overline{H}^t_j, \\ \underline{\sigma}^t_{-i,\tau} & \text{otherwise.} \end{cases}$$

If all players choose $\hat{r}^i$, player i gets an expected payoff equal to $U_i(\hat{r}) = u_i(\hat{\sigma}_\tau)$. If not, suppose that player i's choice involves some code of conduct $\tilde{r}^i$ such that $\tilde{r}^i_i(h^t, h^t_i) = \sigma^t_{i,\tau}(h^t)$ for all $y^t_i \in Y^t_i$ and for any $j \neq i$ it says $\tilde{r}^i_j = \hat{r}^i_j$. Let the highest payoffs associated to $\tilde{r}^i$ for player i be $\overline{W}_i \geq U_i(\tilde{r}^i, r^j)$, and it is given by

$$\overline{W}_i = u_i(\sigma_{i,\tau}, \hat{\sigma}_{i,\tau}) + \eta_{\tau,i}(u_i(\sigma_\tau, \underline{\sigma}_{-i,\tau}) - u_i(\sigma_{i,\tau}, \hat{\sigma}_{i,\tau})).$$

From this, adherence to $\hat{r}^\tau$ requires that $U_i(\hat{r}) \geq \overline{W}_i$, namely, for each player $i \in \mathcal{I}$:

$$\eta_{\tau,i} = \frac{u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\hat{\sigma}_\tau)}{u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\sigma_\tau, \underline{\sigma}_{-i,\tau})}.$$

We take the maximum of these probabilities of detection, thus, $\eta_\tau := \max_{i \in \mathcal{I}} \eta_{\tau,i}$ and this constitutes the lower threshold such that players find it optimal to adhere to code of conduct profile $\hat{r}$. Whenever $\eta \geq \eta_\tau$ the proposed code of conduct profile $\hat{r}$ is a Nash equilibrium of the self-referential game defined on the truncated game Γ^τ.
Finally, for this τ-truncation Γ^τ we may find an upper bound on the probability of detection η_τ. Recall that $\hat{\sigma}$ is a subgame perfect equilibrium of the infinite horizon game Γ^∞. Thus, for any player i we may bound the numerator of the last expression as follows

$$\begin{aligned} u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\hat{\sigma}_\tau) &\leq u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\hat{\sigma}_\tau) + u_i(\hat{\sigma}) - u_i(\sigma_i, \hat{\sigma}_{-i}) \\ &\leq 2\zeta^\tau \end{aligned} \tag{1}$$

and working similarly on the denominator we find the bound

$$\begin{aligned} u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\sigma_\tau, \underline{\sigma}_{-i,\tau}) &\leq u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\sigma_\tau, \underline{\sigma}_{-i,\tau}) + u_i(\hat{\sigma}) - u_i(\sigma_i, \hat{\sigma}_{-i}) \\ &\leq 2\zeta^\tau + u_i(\hat{\sigma}_\tau) - u_i(\sigma_\tau, \underline{\sigma}_{-i,\tau}) \end{aligned} \tag{2}$$

Hence, combining expressions (1) and (2) we get

$$\eta_\tau \leq \frac{2\zeta^\tau}{2\zeta^\tau + u_i(\hat{\sigma}_\tau) - u_i(\sigma_\tau, \underline{\sigma}_{-i,\tau})}$$

By continuity at infinity for all $\varepsilon > 0$ we can find a sufficiently long τ-truncation $\tau^* \in \mathbb{N}$ such that for all $\tau > \tau^*$, $|u_i(\sigma^\infty) - u_i(\hat{\sigma}^\infty)| < \varepsilon/2$ where $\sigma_\tau = \hat{\sigma}_\tau$, then $\eta_\tau < \varepsilon$. □

A.2 Proofs for Section 4

Proof of Theorem 2. First, note that timing must be $\tau_i = 0$ for any i. Then, fix any feasible and strictly individually rational payoff vector $v \in V^*$. Suppose that $v = g(a^*)$ for some profile of actions $a^* \in A$, and let $M := \max_{a \in A, i \in \mathcal{I}} g_i(a)$ and $m := \min_{a \in A, i \in \mathcal{I}} g_i(a)$ be the maximum and minimum possible payoffs for any player $i \in \mathcal{I}$.[24] For if the profile of actions $a^* \in A$ is a Nash equilibrium of the stage game, then the code of conduct vector $\hat{r}$ would require that all players $i \in \mathcal{I}$ select $\hat{r}^i \in R_0$ such that for all $i, j \in \mathcal{I}$, $\hat{r}^i_j(h^0, h^0_j) = a^*_j$ for all private histories $h^0_j \in H_j$. Notice that the self-referential strategy calls for the static Nash equilibrium strategy.

Contrary, suppose that the action profile $a^* \in A$ is not a Nash equilibrium of the stage game. Consider the profile of codes of conduct $\hat{r} \in R$: For all $i, j \in \mathcal{I}$,

$$\hat{r}^i_j(h^0, h^0_j) := \begin{cases} a^*_j & \text{if } h^0_j \notin \overline{H}_j, \\ \underline{\alpha}_{-i} & h^0_j \in \overline{H}_j. \end{cases}$$

It remains to show that this profile of codes of conduct $\hat{r}$ forms a self-referential equilibrium for a sufficiently high probability of detection η_T. For some profile $r \in R$, let $\underline{W}_i(r)$ be the lowest expected payoffs for any player i given profile $r \in R$. Suppose that all players adhere to the profile of codes of conduct $\hat{r}$, then $\underline{W}_i(\hat{r})$ is given by the following expression

$$\underline{W}_i(\hat{r}) = g_i(a^*) + (1-(1-\lambda)^2)(g_i(\underline{\alpha}_{-j}, a^*_j) - g_i(a^*) - (M-m)) + \pi_j(\overline{y}_j|\hat{r})(g_i(a^*_i, \underline{\alpha}_{-i}) - g_i(a^*)).$$

Consider instead that player i chooses an alternative code of conduct $\tilde{r}^i \in R_0, \tilde{r}^i \neq \hat{r}^i$ that says for any private history $h^t_i \in H_i$, player i chooses some action $a_i \in A_i$ where $a_i \in \text{argmax}_{\tilde{a}_i \in A_i} g_i(\tilde{a}_i, a^*_j) \geq g_i(a^*)$, and for the rest of the players $j \neq i$ it states $\tilde{r}^i_j = \hat{r}^i_j$. Given that $j \in \mathcal{I} \setminus \{i\}$ adhere to the code of conduct $\hat{r}$, then the highest expected payoff for

[24] The same proof works for the case of mixed strategies, α.

player i, denoted by $\overline{W}_i(\tilde{r}^i)$ is

$$\overline{W}_i(\tilde{r}^i) = g_i(a_i, a_j^*) + (1 - (1 - \lambda)^2)(g_i(a_i, a_j^*) - g_i(\underline{\alpha}_{-j}, a_j^*) + (M - m)) \\ + (\pi_j(\overline{y}_j|\hat{r}) + \eta)(g_i(a_i^*, \underline{\alpha}_{-i}) - g_i(a^*) + g_i(\underline{\alpha}_{-j}, a_j^*) - g_i(a_i, a_j^*)).$$

In order to have any player i adhering to the profile of codes of conduct $\hat{r}$, it requires that $\underline{W}_i(\hat{r}) \geq \overline{W}_i(\tilde{r}^i)$, namely, $\underline{W}_i(\hat{r}) - \overline{W}_i(\tilde{r}^i) \geq \varepsilon$ for some $\varepsilon > 0$. Then, players find it optimal to follow the profile of codes of conduct $\hat{r}$ if the probability of detecting deviations from code of conduct profile $\eta_{i,T}$ satisfies the following condition

$$\eta_{i,0}\kappa_1 \geq g_i(a_i, a_j^*) - g_i(a^*) + 2\lambda\kappa_2,$$

where $\kappa_1 = g_i(a_i^*, \underline{\alpha}_{-i}) - g_i(a^*) + g_i(\underline{\alpha}_{-j}, a_j^*) - g_i(a_i, a_j^*)$ and $\kappa_2 = g_i(a_i, a_j^*) + g_i(a^*) - 2g_i(\underline{\alpha}_{-j}, a_j^*) + 2(M - m)$. Take the highest probability of detection among players so that the last condition holds for all players i. Let $\eta_0 := \max_{i \in \mathcal{I}} \eta_{i,0}$. Given this probability, we pin down λ_0; for any $\varepsilon > 0$

$$\lambda_0 := \frac{1 + g_i(a^*) - g_i(a_i, a_j^*)}{2\kappa_2} - \varepsilon.$$

It follows that if the probability of detection is high enough, that is $\eta \geq \eta_0$, and on-equilibrium punishments are not too costly $\lambda \leq \lambda_0$, then the profile of codes of conduct $\hat{r}$ is a self-referential equilibrium. □

I will make use of the following piece of notation to prove the next result. Let $\sigma_i^{h^t} := \{\sigma_i \mid \sigma_i^k = a_i^k \text{ with } a_i^t \in h^k, \forall k \leq t\}$ be the strategy σ_i modified in the information sets of player i preceding h^t so that it prescribes the pure actions that induce the history h^t, with profile $\sigma^{h^t} = (\sigma_i^{h^t}, \sigma_{-i}^{h^t})$. Observe that multistage games with observed actions have unique set of actions from σ^{h^t} at each of the previous information set. For any history $h^t \in H$, let $\sigma_{i,p}|h^t$ denote the continuation strategy prescribed by $\sigma_{i,p}$ after history h^t and let $\sigma_{i,p}|H^t$ denote the restriction of $\sigma_{i,p}$ to the subset of histories $H^t \subset H$. For strategy profile σ_p I write $\sigma_p|h^t$ and $\sigma_p|H^t$, respectively. Likewise for behavioral strategies. Let $u_i(\sigma_p|h^t)$ be the continuation payoff to player i induced by the strategy profile $\sigma_p \in \Sigma$ conditional on h^t being reached:

$$u_i(\sigma_p|h^t) = \sum_{t'=t}^{\infty} \delta_i^{t'} g_i(a^{t'}(\sigma_p^{h^t}))$$

Because actions are observed, the strategy profile σ^{h^t} determines a unique history of length

$l(h^t)$ in which we reach h^t. Thus payoffs can be written as, for all players $i \in \mathcal{I}$

$$u_i(\sigma^{h^t}) = \sum_{t'=0}^{t-1} \delta_i^{t'} g_i(a^{t'}(\sigma^{h^t})) + \delta_i^t u_i(\sigma^{h^t}|h^t)$$

Proof of Theorem 3. From Theorem 1, signals are more useful the earlier they arrive. It is then sufficient to consider $\tau_i = T$ for all i. Start by picking any feasible and strictly individually rational payoff vector $v \in V^*$. Again, assume that $v = g(a^*)$ for some profile of actions $a^* \in A$.[25] First, if the profile of actions $a^* \in A$ is a Nash equilibrium of the stage game then the profile of codes of conduct $\hat{r}$ would require that all players $i \in \mathcal{I}$ select $\hat{r}^i \in R_0$ such that for all $i, j \in \mathcal{I}$, $\hat{r}^i_j(h^t, h^t_j) = a^*_j$ for all histories $h^t \in H, h^t_j \in H^t_j$. It follows $\hat{r}$ would be a self-referential equilibrium.

Otherwise, we begin with the construction of the trigger strategy denoted by $\hat{\sigma}_{i,T}$. This strategy is defined as, for all players, $i \in \mathcal{I}$

$$\hat{\sigma}_{i,T}(h^t) := \begin{cases} a^*_i & \text{if } t = 0 \text{ or } h^s = a^{*s} \text{ for } 0 \leq s \leq t-1, \\ \underline{\alpha}_{-j} & \text{otherwise.} \end{cases}$$

Since the profile of strategies $\hat{\sigma}_T$ is not a subgame-perfect equilibrium of the finitely repeated game Γ^T by backward induction argument, there exists at least one profitable one-shot deviation. It is enough to study the case in which there are two of the kind. We define two profitable one-shot deviations denoted by $\sigma_{i,T}$ and $\sigma'_{i,T}$ for each player $i \in \mathcal{I}$, any two actions $a_i, a'_i \in A_i$ and public history $h^T \in H$:

$$\sigma_{i,T}(h^s) := \begin{cases} \hat{\sigma}_{i,T}(h^s) & \text{if } h^s \neq h^T, \\ a_i & \text{if } h^s = h^T, \end{cases} \quad \text{and} \quad \sigma'_{i,T}(h^s) := \begin{cases} \hat{\sigma}_{i,T}(h^s) & \text{if } h^s \neq h^{T-1}, \\ a'_i & \text{if } h^s = h^{T-1}. \end{cases}$$

That is, strategies $\sigma_{i,T}$ and $\sigma'_{i,T}$ have different timing of deviation and potentially different deviation actions. We pick the threshold discount factor $\underline{\delta} \in (0, 1)$ so that for all players $i \in \mathcal{I}$

$$u_i(\sigma_{i,T}, \hat{\sigma}_{-i,T}) \geq u_i(\sigma'_{i,T}, \hat{\sigma}_{-i,T})$$

It is sufficient to have $\underline{\delta}$ satisfying

$$\underline{\delta} = \frac{M - \min_{i \in \mathcal{I}} \underline{v}_i}{\min_{i \in \mathcal{I}} \underline{v}_i - m}$$

Given $\underline{\delta}$, each player $i \in \mathcal{I}$ may only find a profitable one-shot deviation at the final round of the finitely repeated game Γ^T. We may restrict attention to all discount factors δ such

[25] The same argument applies to mixed strategies.

that $\delta \in (\underline{\delta}, 1)$. We write $\underline{\sigma}_i^{j,t}$ for the strategy of player i that is the minmax strategy against player $j \in \mathcal{I} \setminus \{i\}$ at period $t \geq 0$ (constant repetition of the minmax strategy $\underline{\alpha}_{-j}$) with profile $\underline{\sigma}^t = (\underline{\sigma}_i^{-i,t}, \underline{\sigma}_{-i}^{i,t})$. We now proceed to construct the profile of codes of conduct $\hat{r} \in R$ such that for all $i, j \in \mathcal{I}$:

$$\hat{r}_j^i(h^t, h_j^t) := \begin{cases} \hat{\sigma}_{j,T}^t(h^t) & \text{if } h_j^t \notin \overline{H}_j, \text{ for all } t \geq 0, \\ \underline{\sigma}_j^{i,t}(h^t) & \text{if } h_j^t \in \overline{H}_j, \text{ for } t = T. \end{cases}$$

We claim that this profile of codes of conduct $\hat{r}$ forms a self-referential equilibrium for a sufficiently high probability of detection η_T. Given the choice of δ the optimal deviation for any player $i \in \mathcal{I}$ from this profile $\hat{r}$ is the strategy $\sigma_{i,T}$ we defined above. Observe that the strategy $\sigma_{i,T}$ only differs from $\hat{\sigma}^i$ after period $T-1$. Let $\hat{h}^T = (a^{*0}, \ldots, a^{*T-1})$ be the T-length history induced by strategy profile $(\sigma_{i,T}, \hat{\sigma}_{-i,T})$ which in turn is also induced by the strategy profile $\hat{\sigma}_T$. For history $\hat{h}^T$ if all players adhere to the profile of codes of conduct $\hat{r}$, the least expected payoff for any player $i \in \mathcal{I}$ by adhering is

$$\underline{W}_i(\hat{r}) = \frac{1-\delta}{1-\delta^{T+1}} \Bigg[\sum_{t=0}^{T-1} \delta^t g_i(a^t(\hat{\sigma}^{\hat{h}^T})) + \delta^T \big(g_i(a^*) + (1-(1-\lambda)^2)(g_i(\underline{\alpha}_{-j}, a_j^*) - g_i(a^*) - (M-m)) \\ + \pi_j(\overline{y}_j | r)(g_i(a_i^*, \underline{\alpha}_{-i}) - g_i(a^*)) \big) \Bigg]$$

Suppose instead that player i chooses the code of conduct $\tilde{r}^i \in R_0$ where he plays $\tilde{r}_i^i(h^t, h_i^t) = \sigma_{i,T}^t(h^t)$ for any private history $h_i^t \in H_i^t$, and $\hat{r}_j^i = \tilde{r}_j^i$ for all $j \in \mathcal{I} \setminus \{i\}$. Given that $j \in \mathcal{I} \setminus \{i\}$ adhere to the code of conduct $\hat{r}^T$, then the highest expected payoff for player i is

$$\overline{W}_i(\tilde{r}^i, \hat{r}^j) = \frac{1-\delta}{1-\delta^{T+1}} \Bigg[\sum_{t=0}^{T-1} \delta^t g_i(a^t(\sigma_i^{\hat{h}^T}, \hat{\sigma}_{-i}^{\hat{h}^T})) + \delta^T \big(g_i(a_i, a_j^*) + (1-(1-\lambda)^2)(g_i(a_i, a_j^*) \\ - g_i(\underline{\alpha}_{-j}, a_j^*) + (M-m)) + (\pi_j(\overline{y}_j | r) + \eta)(g_i(a_i^*, \underline{\alpha}_{-i}) - g_i(a^*) \\ + g_i(\underline{\alpha}_{-j}, a_j^*) - g_i(a_i, a_j^*)) \big) \Bigg]$$

We now find $\eta_{i,T}$ as the minimum probability of detection that the self-referential game must satisfy to deter player i choosing the code of conduct $\tilde{r}^i$, i.e. $\underline{W}_i(\hat{r}) \geq \overline{W}_i(\tilde{r}^i, \hat{r}^j)$. Thus,

$$\hat{\eta}_{i,T} \kappa_1 \geq g_i(a_i, a_j^*) - g_i(a^*) + 2\lambda \kappa_2$$

where $\kappa_1 = g_i(a_i^*, \underline{\alpha}_{-i}) - g_i(a^*) + g_i(\underline{\alpha}_{-j}, a_j^*) - g_i(a_i, a_j^*)$ and $\kappa_2 = g_i(a_i, a_j^*) + g_i(a^*) -$

$2g_i(\underline{\alpha}_{-j}, a_j^*) + 2(M-m)$. Set $\eta_T := \max_{i\in\mathcal{I}} \eta_{i,T}$, and to pin down λ_T, let this bound satisfies $\lambda_T := \frac{1+g_i(a^*)-g_i(a_i,a_j^*)}{2\kappa_2} - \varepsilon$ for some $\varepsilon > 0$. By construction, the profile of codes of conduct $\hat{r}^T$ is a self-referential equilibrium. Moreover, the expected payoffs for any player $i \in \mathcal{I}$ under $\hat{r}$ is at least

$$\underline{W}_i(\hat{r}) = v_i - \frac{(1-\delta)\delta^T}{1-\delta^{T+1}}\big((1-(1-\lambda)^2)(g_i(\underline{\alpha}_{-j}, a_j^*) - g_i(a^*) - (M-m)) + \eta(g_i(a_i^*, \underline{\alpha}_{-i}) - g_i(a^*))\big)$$

□

Proof of Proposition 1. Fix any $v \in V^*$, again assume that $v_i = g_i(a^*)$ for all $i \in \mathcal{I}$ for some $a^* \in A$. If $a^* \in A$ is an equilibrium the argument follows from using the profile of codes of conduct that ignores signals, that is, for all players $i, j \in \mathcal{I}$ the code of conduct says $\breve{r}_j^i(h^t, h_j^t) = a_j^*$ for any history $h^t \in H, h_j^t \in H_j$. Otherwise, construct strategies σ_T and $\hat{\sigma}_T$ as in the proof Theorem 3 from which we pick $\underline{\delta}$ and consider $\delta \geq \underline{\delta}$. The proposed profile of codes of conduct $\breve{r} \in R$ is such that for all players $i, j \in \mathcal{I}$ and all public histories $h^t \in H$

$$\breve{r}_j^i(h^t, h_j^t) := \begin{cases} \hat{\sigma}_{j,T}^t(h^t) & \text{if } h_j^t \notin \overline{H}_j^t, \text{ for all } t \geq 0, \\ \underline{\sigma}_j^{i,t}(h^t) & \text{if } h_j^t \in \overline{H}_j^t, \text{ for all } t \geq T-k. \end{cases}$$

We obtain the probability of detection $\breve{\eta}_T$ by following the proof of Theorem 3. (I omit the calculation of the probability of detection η_T in the analogous result to Theorem 3 when $\lambda = 0$.) With these probabilities in hand, we show that

$$\begin{aligned}
\breve{\eta}_T := \max_{i\in\mathcal{I}} \breve{\eta}_{i,T} &= \frac{u_i(\sigma_i^{\hat{h}^{T-k}}, \hat{\sigma}_{-i}^{\hat{h}^{T-k}} | \hat{h}^{T-k}) - u_i(\hat{\sigma}^{\hat{h}^{T-k}} | \hat{h}^{T-k})}{u_i(\sigma_i^{\hat{h}^{T-k}}, \hat{\sigma}_{-i}^{\hat{h}^{T-k}} | \hat{h}^{T-k}) - u_i(\sigma_i^{\hat{h}^{T-k}}, \underline{\sigma}_{-i}^{i,\hat{h}^{T-k}} | \hat{h}^{T-k})} \\
&\leq \frac{u_i(\sigma_i^{\hat{h}^{T-k}}, \hat{\sigma}_{-i}^{\hat{h}^{T-k}} | \hat{h}^{T-k}) - u_i(\hat{\sigma}^{\hat{h}^{T-k}} | \hat{h}^{T-k}))}{u_i(\sigma_i^{\hat{h}^{T}}, \hat{\sigma}_{-i}^{\hat{h}^{T}} | \hat{h}^{T}) - u_i(\sigma_i^{\hat{h}^{T}}, \underline{\sigma}_{-i}^{i,\hat{h}^{T}} | \hat{h}^{T})} \\
&= \frac{u_i(\sigma_i^{\hat{h}^{T}}, \hat{\sigma}_{-i}^{\hat{h}^{T}} | \hat{h}^{T}) - u_i(\hat{\sigma}^{\hat{h}^{T}} | \hat{h}^{T}))}{u_i(\sigma_i^{\hat{h}^{T}}, \hat{\sigma}_{-i}^{\hat{h}^{T}} | \hat{h}^{T}) - u_i(\sigma_i^{\hat{h}^{T}}, \underline{\sigma}_{-i}^{i,\hat{h}^{T}} | \hat{h}^{T})} = \hat{\eta}_{T,i} \\
&\leq \max_{i\in\mathcal{I}} \hat{\eta}_{T,i} := \hat{\eta}_T
\end{aligned}$$

The first inequality follows from the fact that the profile of strategies $(\sigma_i^{\hat{h}^{T-k}}, \hat{\sigma}_{-i}^{\hat{h}^{T-k}})$ differs from the profile of strategies $(\sigma_i^{\hat{h}^{T}}, \hat{\sigma}_{-i}^{\hat{h}^{T}})$ after period $T-1$ for a given player i. Moreover, the punishment profile $(\sigma_i^{\hat{h}^{T-k}}, \underline{\sigma}_{-i}^{i,\hat{h}^{T-k}})$ triggered in period $T-k$ would be harsher than punishment profile triggered in the last period $(\sigma_i^{\hat{h}^{T}}, \underline{\sigma}_{-i}^{i,\hat{h}^{T}})$. The equality after this inequality follows from the construction in which players find it optimal to deviate in the last period of the repeated game. Finally, for all players $i \in \mathcal{I}$ the least expected payoff by adhering to

$\breve{r}$ is given by $U_i(\breve{r}) = v_i$. □

A.3 Proofs for Section 5

Proof of Theorem 4. First note that for splitting games the best feasible stage-t timing intention monitoring structure of the self-referential game $\mathcal{G}(\Gamma)$ is $\tau_i = 1$ for all i, because players have strong incentives to exit early, and by Proposition 1, earlier signals allows agents to construct broader codes of conduct as self-referential equilibria. We select the profile of codes of conduct $\hat{r} \in R$ so that for all players $i, j \in \mathcal{I}$

$$\hat{r}^i_j(h^t, h^t_j) := \begin{cases} s^t_j(h^t, h^t_j) & \text{if } h^t_j \notin \overline{H}^t_j, \text{ for all } t \geq 0, \\ \underline{s}^t_j(h^t, h^t_j) & \text{if } h^t_j \in \overline{H}^t_j, \text{ for all } t \geq 1. \end{cases}$$

Fix some period $1 \leq \hat{t} \leq T$. For all players i, the strategy $s_i \in S_i$ is given by $s^t_i(h^t, h^t_i) = f^t_i$ for all $t < \hat{t}, h^t \in H$ and some forward action $f^t_i \in \mathcal{F}^t_i$; and for all $t \geq \hat{t}$ $s^t_i(h^t, h^t_i) = e^t_i$ for some exit action $e^t_i \in \mathcal{E}^t_i$ and $h^t \in H$. The strategy $\underline{s}_i \in S_i$ is given by $\underline{s}^0_i(h^0) = f^0_i$ for $f^0_i \in \mathcal{F}^0_i$, and $\underline{s}^t_i(h^t, h^t_i) = e^t_i$ for all $t \geq 1, e^t_i \in \mathcal{E}^t_i, h^t \in H$. The lowest expected payoff for any player $i \in \mathcal{I}$ from adherence to this code of conduct profile gives

$$\underline{W}_i(\hat{r}) = w_i - \delta^{\hat{t}}_i c_i - (1 - (1 - \lambda)^2)\delta^{\hat{t}}_i c_i - \pi_j(\overline{y}_j | \hat{r})(w_i - \delta^{\hat{t}}_i c_i).$$

Alternatively, let $\tilde{r}^i$ be the optimal code of conduct against $\hat{r}$. By assumptions S.1 and S.2, this code of conduct calls for immediate deviation in period $\hat{t} - 1$. Formally, $\tilde{r}^i_i(h^t, h^t_i) = \tilde{s}^t_i$ for all histories $h^t_i \in H^t_i$ and some strategy $\tilde{s}_i \in S_i$ with $\tilde{s}_i \neq s_i$. The strategy $\tilde{s}_i$ unravels as follows, $\tilde{s}^t_i(h^t, h^t_i) = f^t_i$ for all $t < \hat{t} - 1, h^t \in H, h^t_i \in H^t_i$ and $f^t_i \in \mathcal{F}^t_i$; and $\tilde{s}^t_i(h^t, h^t_i) = e^t_i$ for any $h^t \in H, h^t_i \in H^t_i$ and $e^t_i \in \mathcal{E}^t_i$ for all $t \geq \hat{t} - 1$. For all players $j \neq i$, it says $\tilde{r}^i_j = \hat{r}^i_j$.
This gives an expected payoff of $U_i(\tilde{r}^i, \hat{r}^j) = w_i$ which is higher than $\underline{W}_i(\hat{r})$ the lowest expected payoff under the profile $\hat{r}$. In fact, if $\lambda = 0$, the expected payoff for player i is $U_i(\hat{r}) = w_i - \delta^{\hat{t}}_i c_i$. This is for any arbitrary period $\hat{t}$. Observe that all alternative codes of conduct will require deviation in the first period of the game. In particular, we are left only with codes of conduct which ignore signals arriving at any period $t \geq 1$. For instance, pick the profile of codes of conduct $\tilde{r}$ that is characterized by the following behavior: for all $i, j \in \mathcal{I}$, such that $\tilde{r}^i_j(h^t, h^t_j) = \tilde{s}^t_j(h^t)$ for all $h^t_j \in H^t_j, t \geq 0$ and for some $\tilde{s}_i \in S_i$. The strategy $\tilde{s}_i$ states that $\tilde{s}^t_i(h^t, h^t_i) = e^t_i$ for all $t \geq 0$ with $e^t_i \in \mathcal{E}^t_i$. In equilibrium, each player gets $U_i(\tilde{r}) = w_i - c_i$. Any deviation from this profile of codes of conduct gives expected payoffs of 0. This profile constitutes a self-referential equilibrium with the unique outcome where all players i exit in period $t = 0$. □

Proof of Theorem 5. From Theorem 4, for any r such that $r^i_i(\dot{}) = e^t_i$ in any period $t > 0$, each player i finds it optimal to choose an alternative code of conduct, playing some exit action $e_i^0 \in \mathcal{E}_i^0$, guaranteeing himself w_i irrespective of what his opponents do. This is because (Y, π) satisfies $\tau_i = t$ for some $t > 0$ and for all i. However, here (Y, π) provides $\tau_i = 0$ for any i. Pick some period $k \in \mathbb{N}$ with $0 \leq k \leq T$. Let us focus on the profile of codes of conduct $\hat{r} \in R$ so that for all players $i, j \in \mathcal{I}$

$$\hat{r}_j^i(h^t, h_j^t) := \begin{cases} s_j^t(h^t, h_j^t) & \text{if } h_j^t \notin \overline{H}_j^t, \\ \underline{s}_j^t(h^t, h_j^t) & \text{if } h_j^t \in \overline{H}_j^t. \end{cases}$$

In strategy $s_i \in S_i$, player i chooses $s_i^t(h^t, h_i^t) = f_i^t$ for some $f_i^t \in \mathcal{F}_i^t$, any $h^t \in H$ and all periods $t \leq k-1$; and $s_i^t(h^t, h_i^t) = e_i^t$ for some $e_i^t \in \mathcal{E}_i^t$, history $h^t \in H$ and all periods $t \geq k$. In addition, the strategy $\underline{s}_i \in S_i$ is described as $\underline{s}_i^t(h^t, h_i^t) = e_i^t$ for some $e_i^t \in \mathcal{E}_i^t$, $h^t \in H$ and all $t \geq 0$. The lowest expected payoffs associated to this profile are

$$\underline{W}_i(\hat{r}) = w_i - \delta_i^k c_i + (1 - (1-\lambda)^2)\delta_i^k c_i - \pi_j(\overline{y}_j | \hat{r})(w_i - \delta_i^k c_i).$$

One alternative optimal code of conduct could be $\tilde{r}^i$ such that $\tilde{r}_i^i(h^t, h_i^t) = \tilde{s}_i^t(h^t, h_i^t)$ for all $h_i^t \in H_i^t$, some strategy $\tilde{s}_i \in S_i$, and $\tilde{r}_j^i = \hat{r}_j^i$ for all $j \neq i$. Here, the strategy $\tilde{s}_i \in S_i$ requires $\tilde{s}_i^t(h^t, h_i^t) = f_i^t$ for some $f_i^t \in \mathcal{F}_i^t$, $h^t \in H$ and all $t < k-1$; and $\tilde{s}_i^t(h^t, h_i^t) = e_i^t$ for some $e_i^t \in \mathcal{E}_i^t$, $h^t \in H$ and all $t \geq k-1$. Thus, it gives an expected payoff of at most $\overline{W}_i$ for player i.

$$\overline{W}_i(\tilde{r}^i) = w_i - (\pi_j(\overline{y}_j | \hat{r}) + \eta)(w_i - \delta_i^k c_i).$$

Thus, for $\varepsilon > 0$ we must have $\underline{W}_i(\hat{r}) \geq \overline{W}_i(\tilde{r}^i) + \varepsilon$. Working in the same line as in the proof of Theorem 4, we find the required probability of detection $\eta_k := \max_{i \in \mathcal{I}} \eta_{i,k}$ to sustain this profile of codes of conduct $\hat{r}$ where each $\eta_{i,k}$ is given by $\eta_{i,k} = \delta_i^k c_i (1-\lambda)^2 / (w_i - \delta_i^k c_i)$. Then, take λ_k such that it satisfies $(1-\lambda_k)^2 \delta c_i \leq (w_i - \delta c_i)(1-\varepsilon)$ for some $\varepsilon > 0$. □

Proof of Theorem 6. It is sufficient to take stage $k = \phi(\overline{n})$. Note that player 2 makes a choice at this stage, i.e. $\iota(h^k) = 2$. Suppose that players adhere to the code-of-profile profile $r \in R$, where for players i prescribes

$$r_j^i(h^t, h_j^t) := \begin{cases} s_j^t(h^t, h_j^t) & \text{if } h_j^t \notin \overline{H}_j^t, \\ \underline{s}_j^t(h^t, h_j^t) & \text{otherwise.} \end{cases}$$

where for player 1, the strategy $s_1 \in S_1$ requires $s_1^t(h^t, h_1^t) = f_1^t$ for all $t \geq 0$, and the strategy $\underline{s}_1 \in S_1$ is given by $\underline{s}_1^t(h^t, h_1^t) = f_1^t$ for all $t < k_{-1}$ and $\underline{s}_1^t(h^t, h_1^t) = e_1^t$ for all $t \geq k_{-1}$. On the other hand, for player 2 his strategy $s_2 = \underline{s}_2$ with $s_2, \underline{s}_2 \in S_2$ says that $s_2^t(h^t, h_2^t) = f_2^t$ for all $0 \leq t < T$, and $s_2^T(h^T, h_2^T) = e_2^T$. It is clear that player 2 finds it optimal to adhere to this code of conduct, it is the highest possible payoff. However, player 2 will detect deviations from the equilibrium code of conduct while he is inactive at stage k_{-1}—that is, $\iota(h^{k_{-1}}) = 2$. Player 1 could deviate from this code at stage k_{-1} and get $g_1(e_1^{k_{-1}}, \overline{a})$. By adhering to the code of conduct, player 1 obtains $g_1(\overline{a}, e_2^T)$. By assumption S.2, player 1 finds it optimal not to adhere. The same argument goes through any stage $k < \phi(\overline{n})$. If that it is the case, consider a profile r where for some stage t player $\iota(h^t)$ chooses an exit action $e^t_{\iota(h^t)}$ and $f^k_{\iota(h^t)}$ for any $k \neq t$, moreover, player $j \neq \iota(h^t)$ plays f_j^k for all k. But again player j would be better off by exiting at stage t_{-1}, i.e. $g_j(e_j^{t_{-1}}, \overline{a}) > g_j(\overline{a}, e^t_{\iota(h^t)})$. This implies players exit whenever they are active and that the code of conduct profile $r \in R$ such that all players i choose exit actions, i.e. $r_i^i(h^t, h_i^t) = e_i^t$ for all $e_i^t \in \mathcal{E}_i^t$, $h_i^t \notin \overline{H}_i^t$, $h^t \in H$ where $t = \phi(1) - 1$ for player 1, and $t = \phi(1)$ for player 2 is the unique self-referential equilibrium. □

Proof of Theorem 7. Pick any stage $k \in \mathbb{N}$ such that $k \geq \phi(2)$ and in which the game ends will end in equilibrium. It suffices to check the case $\tau_i = k_{-2}$ for each i. Suppose that the code of conduct profile $\hat{r} \in R$ where all players i choose according to

$$r_j^i(h^t, h_j^t) := \begin{cases} s_j^t(h^t, h_j^t) & \text{if } h_j^t \notin \overline{H}_j^t, \\ \underline{s}_j^t(h^t, h_j^t) & \text{otherwise.} \end{cases}$$

Let $\iota = \iota(h^k)$ be the active player that ends the game at stage k. Player ι's strategies satisfy for all stages $t < k$, $s_\iota^t(h^t, h_\iota^t) = f_\iota^t$ for all public histories $h^t \in H$, private histories $h_\iota^t \notin \overline{H}_\iota$, and forward actions $f_\iota^t \in \mathcal{F}_\iota^t$; and for all stages such that $t \geq k$ it requires $s_\iota^t(h^t, h_\iota^t) = e_\iota^t$ for all $h^t \in H$, $h_\iota^t \notin \overline{H}_\iota$ and exit actions $e_\iota^t \in \mathcal{E}_\iota^t$. For the punishment strategy $\underline{s}_\iota^t$, for all stages $t < k_{-2} - 1$ it would be $\underline{s}_\iota^t(h^t, h_\iota^t) = f_\iota^t$ for all $h^t \in H$, $h_\iota^t \in \overline{H}_\iota$ and $f_\iota^t \in \mathcal{F}_\iota^t$; and for all $t \geq k_{-2} - 1$ it says $\underline{s}_\iota^t(h^t, h_\iota^t) = e_\iota^t$ for all $h^t \in H$, $h_\iota^t \in \overline{H}_\iota$ and $e_\iota^t \in \mathcal{E}_\iota^t$. On the other hand, for inactive player $i \neq \iota(h^k)$ at stage k, her strategy is as follows. For all stages $t \geq 0$, the strategy s_i requires $s_i^t(h^t, h_i^t) = f_i^t$ for all $h^t \in H$, $h_i^t \notin \overline{H}_i$ and $f_i^t \in \mathcal{F}_i^t$. The punishment strategy $\underline{s}_i^t$ says that for all stages $t < k_{-1} - 1$ we have that $\underline{s}_i^t(h^t, h_i^t) = f_i^t$ for all $h^t \in H$, $h_i^t \in \overline{H}_i$ and $f_i^t \in \mathcal{F}_i^t$; and for all $t \geq k_{-1} - 1$ it must be the case that $\underline{s}_i^t(h^t, h_j^t) = e_j^t$ for all $h^t \in H$, $h_i^t \in \overline{H}_i^t$ and $e_i^j \in \mathcal{E}_i^t$.

Next, we find a sufficiently high probability of detection. To do so, the optimal alternative code of conduct for player ι is the following. For player i, $\tilde{r}_i^\iota(h^t, h_i^t) = \hat{r}_i^\iota$ and $\tilde{r}_\iota^\iota(h^t, h_\iota^t) = s_\iota$ where $s_\iota \in S_\iota$ is as follows, for all $h^t \in H$ and $h_\iota^t \in H_\iota$, for any $t \leq k$, $\tilde{s}_\iota^\iota(h^t, h_\iota^t) = f_\iota^t$ for $f_\iota^t \in \mathcal{F}_\iota^t$; and for all $t \geq k_{+1} - 1$, $\tilde{s}_\iota^\iota(h^t, h_\iota^t) = e_\iota^t$ for $e_\iota^t \in \mathcal{E}_\iota^t$. The lowest expected payoffs by adhering to $\hat{r}$ for player ι is

$$\underline{W}_\iota(\hat{r}) = g_\iota(e_\iota^k, \overline{a}) + (1 - (1-\lambda)^2)(g_\iota(e_\iota^{k-1}, \overline{a}) - g_\iota(e_\iota^k, \overline{a})) + \pi_i(\overline{y}_i|\hat{r})(g_\iota(\overline{a}, e_i^{k-1}) - g_\iota(e_\iota^k, \overline{a}))$$

If $\tilde{r}^\iota$ is chosen instead,

$$\overline{W}_\iota(\tilde{r}^\iota) = g_\iota(e_\iota^{k_{+1}-1}, \overline{a}) - (\pi_i(\overline{y}_i|\hat{r}) + \eta)(g_\iota(e_\iota^{k_{+1}-1}, \overline{a}) - g_\iota(\overline{a}, e_i^{k_{-1}}))$$

For ι, it means that $\eta_{\iota,k}$ satisfies $\underline{W}_\iota(\hat{r}) - \overline{W}_\iota(\tilde{r}^\iota) \geq 0$

$$\begin{aligned} g_\iota(e_\iota^k, \overline{a}) + (1-(1-\lambda)^2)(g_\iota(e_\iota^{k_{-1}}, \overline{a}) - g_\iota(e_\iota^k, \overline{a})) + \pi_i(\overline{y}_i|\hat{r})(g_\iota(\overline{a}, e_i^{k_{-1}}) - g_\iota(e_\iota^k, \overline{a})) \\ - g_\iota(e_\iota^{k_{+1}-1}, \overline{a}) + (\pi_i(\overline{y}_i|\hat{r}) + \eta_{\iota,k})(g_\iota(e_\iota^{k_{+1}-1}, \overline{a}) - g_\iota(\overline{a}, e_i^{k_{-1}})) \geq 0 \end{aligned}$$

Then, $\lambda_{\iota,k}$, for any $\varepsilon > 0$ is given by

$$\begin{aligned} g_\iota(e_\iota^{k_{+1}-1}, \overline{a}) - g_\iota(e_\iota^k, \overline{a}) - (1-(1-\lambda)^2)(g_\iota(e_\iota^{k_{-1}}, \overline{a}) - g_\iota(e_\iota^k, \overline{a})) - \pi_i(\overline{y}_i|\hat{r})(g_\iota(\overline{a}, e_i^{k_{-1}}) - g_\iota(e_\iota^{k_{+1}-1}, \overline{a})) \\ \leq ((1-\varepsilon) + \pi_i(\overline{y}_i|\hat{r}))(g_\iota(e_\iota^{k_{+1}-1}, \overline{a}) - g_\iota(\overline{a}, e_i^{k_{-1}})) \end{aligned}$$

The optimal code of conduct for player i, in this case, is $\tilde{r}_i^i(h^t, h_i^t) = \tilde{s}_i^t$ with for all $t < k_{-1}$, $\tilde{s}_i^t(h^t, h_i^t) = f_i^t$, for any $h^t \in H$, $h_i^t \in H_i^t$ and $f_i^t \in \mathcal{F}_i^t$; and for all $t \geq k_{-1}$, $\tilde{s}_i^t(h^t, h_i^t) = e_i^t$ for $h^t \in H$, $h_i^t \in H_i^t$ and $e_i^t \in \mathcal{E}_i^t$. For player i, $\underline{W}_i(\hat{r}) - \overline{W}_i(\tilde{r}^i) \geq 0$

$$\eta_{i,k}(g_i(e_i^k, \overline{a}) - g_i(e_i^{k_{-1}}, \overline{a})) = g_i(\overline{a}, e_\iota^{k_{-2}}) - g_i(e_i^k, \overline{a}) - (1-(1-\lambda)^2)(g_i(e_i^{k_{-1}}, \overline{a}) - g_i(e_i^k, \overline{a}))$$

Thus, $\lambda_{i,k}$ must satisfy for $\varepsilon > 0$

$$g_i(\overline{a}, e_\iota^{k_{-2}}) - g_i(e_i^k, \overline{a}) - (1-(1-\lambda)^2)(g_i(e_i^{k_{-1}}, \overline{a}) - g_i(e_i^k, \overline{a})) \leq (1-\varepsilon)(g_i(e_i^k, \overline{a}) - g_i(e_i^{k_{-1}}, \overline{a}))$$

Taking both $\eta_k \coloneqq \max_i \eta_{i,k}$ and $\lambda_k \coloneqq \max_i \lambda_{i,k}$, the profile $\hat{r}$ forms a self-referential equilibrium. □

A.4 Proofs for Section 6

Proof of Proposition 3. Fix any $\tau_1, \tau_2 \leq T$ where by asynchronicity $\tau_1 \neq \tau_2$. Let $v \in V^*$. Assume $v_i = g_i(a^*)$ for $a^* \in A$. If $a^* \in A$ is a Nash equilibrium of Γ, then the code of conduct for all i, j, the code of conduct $\hat{r}_j^i(h^t, h_j^t) = a_j^*$ for all $h^t \in H, h_j^t \in H_j$ forms a self-referential equilibrium for any η, λ. Otherwise, focus on $\delta > \underline{\delta}$ such that players only deviate in T. Then we can mimic the the proof approach used for Theorem 3, it follows r is a self-referential equilibrium here as well. □

Proof of Proposition 4. Suppose that $\tau_1, \tau_2 \leq T$, $\tau_1 \neq \tau_2$ and say $\tau_1 < \tau_2$ provided by (Y, π).[26] Pick some $k \in \mathbb{N}, 0 < k \leq T$ such that $k \geq \tau_1, \tau_2$. Observe that if $k \leq \tau_1$, since the game

[26]The argument is invariant to permutation.

ends in period k, player 2 receives her signal too late to materialise any punishment. Let r^i be the code of conduct $\forall i, j$

$$r^i_j(h^t, h^t_j) := \begin{cases} s^t_j(h^t, h^t_j) & \text{if } h^t_j \notin \overline{H}^t_j, \text{ for all } t \geq 0, \\ \underline{s}^t_j(h^t, h^t_j) & \text{if } h^t_j \in \overline{H}^t_j, \text{ for all } t \geq 0. \end{cases}$$

By triggering punishments in period τ_2 none of the players can infer opponent's y_i. Let $s_i \in S_i$ be for all periods $t \leq k$, $s_i(h^t, h^t_i) = f^t_i$ for all $h^t \in H$, $h^t_i \notin \overline{H}_i$ and $f^t_i \in \mathcal{F}^t_i$, and for all $t \geq k$ we have $s_i(h^t, h^t_i) = e^t_i$ for any $h^t \in H$, $h^t_i \notin \overline{H}_i$ and $e^t_i \in \mathcal{E}^t_i$. For $\underline{s}_i \in S_i$, for all $t < \tau_2$, $\underline{s}_i(h^t, h^t_i) = f^t_i$ for $h^t \in H$, $h^t_i \in \overline{H}_i$ and $f^t_i \in \mathcal{F}^t_i$, and for all $t \geq \tau_2$ it says $\underline{s}_i(h^t, h^t_i) = e^t_i$ for $h^t \in H$, $h^t_i \in \overline{H}_i$ and $e^t_i \in \mathcal{E}^t_i$. Then, the lower bound in expected payoffs is given by

$$\underline{W}_i(r) = w_i - \delta^k_i c_i - (1 - (1 - \lambda)^2)\delta^k_i c_i - \pi_j(\overline{y}_j | r)(w_i - \delta^k_i c_i).$$

The punishment under this timing must occur in period τ_2 so that player 2 does not infer player 1 receive a signal in $Y_1 \setminus \overline{Y}_1$ if he continues playing. For player 1, it is clear that the alternative code of conduct $\tilde{r}^1$ stating that $\tilde{r}^1_1 = s_1$ where $s_1(h^t, h^t_1) = e^t_1$ for all $h^t \in H$, $h^t_1 \in H_1$, $t \geq \tau_1$ and $e^t_1 \in \mathcal{E}^t_1$, and for his opponent $\tilde{r}^1_2 = r^1_2$ gives higher payoffs as it delivers a payoff of w_1. Moreover, player 2 finds it optimal to choose the alternative code of conduct $\hat{r}^2$ where $\hat{r}^2_2 = s_2$ where $s_2(h^t, h^t_2) = e^t_2$ for all $h^t \in H$, $h^t_2 \in H_2$, $t \geq \tau_1 - 1$ and $e^t_2 \in \mathcal{E}^t_2$, and for $t < \tau_1 - 1$ simply $s_2(h^t, h^t_2) = f^t_2$, for all $h^t \in H$, $h^t_2 \in H_2$ and $f^t_2 \in \mathcal{F}^t_2$. Finally for player 1 it requires $\hat{r}^2_1 = r^2_1$. This gives a payoff of w_2. For both players is optimal to not adhere to r^i. Moreover, the same logic applies to any timing even for the case $\tau_1 = 0$ and $\tau_2 = 1$. Suppose that we aim to have an exit profile after period 1 (or even in period 1). Player 1 can always exit in period $t = 0$ taking his surplus w_1 without paying the cost because player 2 receives her signal about player 1's intentions to exit when this actually already happened while the game ended so no punishment is possible. By choosing the code of conduct r^i such that $r^i_j = s_j$ with $s_j(h^t h^t_j) = e^t_j$ for all t, $h^t \in H$, $h^t_j \in H^t_j$ and any $e^t_j \in \mathcal{E}^t_j$. It follows that the only equilibrium outcome in the self-referential game exhibits all players exiting in period $t = 0$. □

Proof of Proposition 5. Suppose without loss of generality that $k = \phi(\overline{n})$. Recall that the timing is such that $\tau_2 = k_{-1}$. Pick τ_1, say, $\tau_1 = t'$ with $0 \leq t' \leq k_{-2}$. Then we aim to construct a code of conduct featuring exit at stage k. To do this, consider

$$r^i_j(h^t, h^t_j) := \begin{cases} s^t_j(h^t, h^t_j) & \text{if } h^t_j \notin \overline{H}^t_j, \text{ for all } t, \\ \underline{s}^t_j(h^t, h^t_j) & \text{if } h^t_j \in \overline{H}^t_j, \text{ for all } t. \end{cases}$$

The code of conduct states for player 2, thus, $r^2_2 = s_2$ for $s_2 \in S_2$ where for all $t \geq k$,

$s_2(h^t, h_2^t) = f_2^t$ for all $h^t \in H$, $h_2^t \in H_2$, and $f_2^t \in \mathcal{F}_2^t$, while for $t \geq k$ it requires $s_2(h^t, h_2^t) = e_2^t$ for all $h^t \in H$, $h_2^t \in H_2$, and $e_2^t \in \mathcal{E}_2^t$. In addition, $\underline{s}_2 = s_2$. On the other hand, for player 1 it says for all $t \geq 0$, $s_1(h^t, h_1^t) = f_1^t$ for all $h^t \in H$, $h_1^t \notin \overline{H}_1$, and $f_1^t \in \mathcal{F}_1^t$, whereas for any $t \leq k_{-1}$, $\underline{s}_1(h^t, h_1^t) = f_1^t$, for all $h^t \in H$, $h_1^t \in \overline{H}_1$, and $f_1^t \in \mathcal{F}_1^t$; and for $t \geq k_{-1}$, $\underline{s}_1(h^t, h_1^t) = e_1^t$, for all $h^t \in H$, $h_1^t \in \overline{H}_1$, and $e_1^t \in \mathcal{E}_1^t$.
Again, player 2 has no incentives to deviate by condition P.1 and P.2 on reward mappings. It remains to check player 1. Consider the optimal deviation to this code of conduct r, denoted by $\tilde{r}^1$ such that $\tilde{r}_2^1 = r_2^1$. For player 1, $\tilde{r}_1^1 = \tilde{s}_1$ so that for all $t \leq \phi(\overline{n}) - 1$, this strategy is $\tilde{s}_1(h^t, h_1^t) = f_1^t$, for all $h^t \in H$, $h_1^t \in \overline{H}_1$, and $f_1^t \in \mathcal{F}_1^t$. For all $t \geq \phi(\overline{n}) - 1$, $\tilde{s}_1(h^t, h_1^t) = e_1^t$, for all $h^t \in H$, $h_1^t \in H_1$, and $e_1^t \in \mathcal{E}_1^t$. This gives a lower bound of at least $g_1(e_1^{\phi(\overline{n})-1}, \overline{a}) > g_1(\overline{a}, e_2^k)$ by condition (ii). Trying to sustain exit actions before will imply that player 2 is not able to observe signals sufficiently in advance. This argument can be applied to any other exit profile. Henceforth, the unique equilibrium outcome is player 1 leaving the game at stage $\phi(1) - 1$, and player 2 exiting at $\phi(\overline{n})$. □

Proof of Proposition 6. First, the requirement $k \geq \phi(2)$ ensures that both players have a incentive to continue beyond their first active period, determined by condition P.2 on reward mappings. Pick a stage $k \in \mathbb{N}$ such that $\phi(2) \leq k \leq T$. We aim to construct a code of conduct where player $\iota(h^k)$ exits the game in stage k as he is active. For notational convenience, write $\iota(h^k) = \iota$. Recall that $\tau_\iota \leq k_{-2}$ and $\tau_i \leq k_{-1}$. The proposed code of conduct is r^i requires for all i, j

$$r_j^i(h^t, h_j^t) := \begin{cases} s_j^t(h^t, h_j^t) & \text{if } h_j^t \notin \overline{H}_j^t, \\ \underline{s}_j^t(h^t, h_j^t) & \text{otherwise.} \end{cases}$$

Therefore, the strategy for player ι is for all $t < k$, $s_\iota^t(h^t, h_\iota^t) = f_\iota^t$ for all $h^t \in H$, $h_\iota^t \notin \overline{H}_\iota$, and $f_\iota^t \in \mathcal{F}_\iota^t$; and for all $t \geq k$, $s_\iota^t(h^t, h_\iota^t) = e_\iota^t$ for all $h^t \in H$, $h_\iota^t \notin \overline{H}_\iota$, and $e_\iota^t \in \mathcal{E}_\iota^t$. Further, the punishment strategy is such that for all $t < k_{-2}$, $\underline{s}_\iota^t(h^t, h_\iota^t) = f_\iota^t$ for all $h^t \in H$, $h_\iota^t \in \overline{H}_\iota$, and $f_\iota^t \in \mathcal{F}_\iota^t$, moreover, for all $t \geq k_{-2}$, $\underline{s}_\iota^t(h^t, h_\iota^t) = e_\iota^t$ for all $h^t \in H$, $h_\iota^t \in \overline{H}_\iota$, and $e_\iota^t \in \mathcal{E}_\iota^t$. It remains to state the strategies for player i with $i \neq \iota$. For player i, for each $t < k$, $s_i^t(h^t, h_i^t) = f_i^t$ for all $h^t \in H$, $h_i^t \notin \overline{H}_i$, and $f_i^t \in \mathcal{F}_i^t$; and for all $t \geq k$, $s_i^t(h^t, h_i^t) = e_i^t$ for all $h^t \in H$, $h_i^t \notin \overline{H}_i$, and $e_i^t \in \mathcal{E}_i^t$. Similar to player ι but with different timing the punishment strategy is characterized by: for all $t < k_{-1}$, $\underline{s}_i^t(h^t, h_i^t) = f_i^t$ for all $h^t \in H$, $h_i^t \in \overline{H}_i$, and $f_i^t \in \mathcal{F}_i^t$, moreover, for all $t \geq k_{-1}$, $\underline{s}_i^t(h^t, h_i^t) = e_i^t$ for all $h^t \in H$, $h_i^t \in \overline{H}_i$, and $e_i^t \in \mathcal{E}_i^t$. By parameterizing η_k and λ_k as in Theorem 7, the proposed code of conduct profile forms a self-referential equilibrium. □