Spring 4-25-2014

Essays on Self-Referential Games

Juan Ignacio Block
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Economics

Dissertation Examination Committee:
David Levine, Chair
Mariagiovanna Baccara
John Nachbar
Paulo Natenzon
Brian Rogers
Maher Said

Essays on Self-Referential Games

by

Juan Ignacio Block

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2014
St. Louis, Missouri

Contents

Acknowledgments

First and foremost, I wish to thank my main advisor David Levine for being a true mentor, whose guidance and support helped me immensely to elaborate this dissertation. It has been a truly rewarding experience working with him. He has taught me, and he will continue doing so, to ponder about economic problems and be a scholar with academic integrity. In the future I hope that I could be able to envision potential paths and results as well as he can.

I would also like to thank John Nachbar. The joy and enthusiasm he has for teaching is contagious, and his rigorous approach to research has been a truly inspirational example to follow for me, and his insights always cut directly to the core of the matter. I am grateful to Paulo Natenzon, Mariagiovanna Baccara, Maher Said, Brian Rogers, Anqi Li, and Marcus Berliant for their constant support and valuable comments.

The Graduate School of Arts & Sciences, the Department of Economics and the Economic Theory Center provided generous funding during my entire doctoral studies.

A great supportive group of good friends have contributed to my personal and professional time at St. Louis. I am specially grateful to those with whom I went through my graduate studies experience: Jan Auerbach, Brian Bergfeld, Julieta Caunedo, Rohan Dutta, Hassan Faghani Dermi, David Kohn, Filippo Massari, Juan Passadore, Francisco Roch, Luis Silva, Mariano Somale, Faisal Sohal, and Emircan Yurdagul.

Last but certainly not least, I would like to dedicate this dissertation to my family and Pilar Alonso. I thank my mother Susana, whose unwavering love and support have sustained me during good and rough times. My father Juan for teaching me to push forward and never loose my sense of humor. My lovely sisters, Gisel and Solange, whose friendship and everyday encouragement have kept me grounded. In the past five years I consider myself very fortunate to have found Pilar. Most of all, I would like to express my heartfelt gratitude and admiration to Pilar, whose support, encouragement and patience made the last years an enjoyable journey.

Dedicated to my family and Pilar Alonso.

Chapter 1

Codes of Conduct, Private Information and Repeated Games

1.1 Introduction

The theory of repeated games has made enormous strides in penetrating the difficult but relevant setting in which players observe noisy signals of each other's play.[1] Unfortunately as our knowledge of equilibria in these games has expanded there is an increasing sense that the types of equilibria studied - involving as they do elaborately calibrated indifference - are difficult for players to play and unlikely to be observed in practice.

By way of contrast, if we give up on the notion of exact optimization, the theory of approximate equilibria in repeated games is simpler and generally more satisfactory than the theory of exact equilibrium. However, it is difficult to rationalize, for example, why a player who is aware that he has been lucky and his opponent has very favorable signals about his behavior, does not take advantage of this knowledge to behave badly. In fact, it is exactly

[1] See for example, [Sugaya(2011)] and references therein.

this type of small gain that approximate equilibrium constructions are based on. In addition, the abstract world of repeated games is not very like the world we inhabit. It is a world in which poker is a dull game because players can never guess that their opponent is bluffing from the expression on his face. It would also be surprising to have skilled interrogators who by asking a few pointed questions can tell whether a suspect is lying or telling the truth.

A class of games in which players have at least a chance of fathoming each others intentions whether through facial expressions or skilled interrogation - was introduced in [Levine and Pesendorfer(2007)]. The basic setup here is the [Levine and Pesendorfer(2007)] model generalized to allow for asymmetries. It utilizes the notion that players employ codes of conduct - complete specifications of how they and their opponents "should" play. Players also receive signals about what code of conduct their opponent may be using, while their own code of conduct enables them to respond to these signals. This is the "self-referential" nature of the games studied here. One key question addressed in this paper is how and when such self-referential codes of conduct make sense.

This paper views direct observation of opponents intentions and repetition of a game as complements rather than substitutes. Even if direct observation is unreliable, it may be enough to overcome the small ε's that arise when simple repeated game strategies are employed.

An effective code of conduct rewards players for using the same code of conduct, and punishes them for using a different code of conduct. Several examples explore such issues as when players in a repeated setting might get information about the past play of new partners from other players. Results of [Levine and Pesendorfer(2007)] about perfect discrimination are generalized to the asymmetric setting. General results about when approximate equilibria in a base game can be sustained as strict equilibria in the corresponding self-referential game are given. As an application a discounted strict Nash folk-like theorem for enforceable mutually punishable payoffs in repeated games with private monitoring is proven despite very limited ability to observe directly.

The notion of self-referential strategies has a long history of discovery and rediscovery. The earliest notion we are aware of appears in [Howard(1988)] who discusses computer programs that play based on reading each other. This idea was formalized in [Tennenholtz(2004)] that developed the concept of program equilibrium. In the context of two-player games [Kalai et al.(2010)Kalai, Kalai, Lehrer, and Samet] developed conditional devices that play on behalf of players and condition on their opponent's conditional device. Using these conditional devices as a source of commitment they proved a folk like theorem for one-shot games. [Levine and Pesendorfer(2007)] were primarily interested in self-referential games as a simple alternative to repeated games that exhibit many of the same features. For example they showed that in a two player symmetric setting if players can accurately determine whether or not their opponent is using the same strategy as they are then a type of folk theorem holds. The simple structure of static self-referential games made it possible to answer questions about which of many equilibria have long-run stability properties in an evolutionary setting. These questions were impractical to study in repeated games.

1.2 The Model

Consider a finite *base game* $\Gamma = \{I, \{S_i, u_i\}_{i \in I}\}$ with set of players I. Player i has finitely many strategies S_i. Note that we allow mixed strategies, there are only a finite number of them: for example, only a six-sided dice is available. Notice also that we assume implicitly either a finite horizon, or a very small subset of strategies in an infinite horizon – for example, finite automata with an upper bound on the number of states. We denote by $S = \times_i S_i$

the corresponding profile of strategies. Utility of player i is given by the payoff function $u_i : S \to \mathbb{R}$ with the usual notation $u_j(s_j, s_{-j})$ for the strategy profile of all players but player j.

The *self-referential game* is defined upon any base game Γ. The set of players is I as in the base game. Each player i has a finite set of private signals Y_i with element y_i and profile $y \in Y = \times_i Y_i$. We assume that the cardinality of the set of signals is $|Y_i| \geq 2$ for each player i. The strategy of player i in the self-referential game is an $|I| \times 1$-vector whose element corresponding to jth is a mapping from player j's set of private signals to player i's set of strategies in the base game, i.e. $r^i_j : Y_j \to S_j$. We call this strategy *code of conduct* and we denote it by r^i where $r^i = (r^1_i, \ldots, r^i_N)$. The common abstract space of codes of conduct is defined as the set of all maps from player's private signals set into player's strategy set, $R_0 := \{r^i \mid r^i_j : Z_j \to S_j \text{ for all } i, j\}$. We write $r \in R$ for the profile of codes of conduct.

For each profile $r \in R$, let $\pi(\cdot|r)$ be the probability distribution over profiles Y. The collection of probability distributions over profile of private signals is denoted by $\{\pi(\cdot|r)|r \in R\}$. Let $\pi_i(\cdot|r)$ be the marginal probability distribution of $\pi(\cdot|r)$ over profile Y. That is, $\pi_i(y_i|r)$ is the probability that player i observes y_i if players have chosen profile of codes of conduct r.

Notice that codes of conduct play two roles. First, they determine how players play as a function of the signals they receive: that is, a player who has chosen the code of conduct r^i and who observes the signal y_i plays $r^i_i(y_i)$. Second, codes of conduct influence the signals y_j players j receive about each others' intentions through the probability distribution π.

In the self-referential game all players i simultaneously choose codes of conduct $r^i \in R_0$. If the profile of codes of conduct is $r \in R$ the corresponding expected utility of player i is given by

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

A Nash equilibrium of the self-referential game is a profile of codes of conduct r such that for all players i and any alternative code of conduct $\tilde{r}^i$, it follows that $U_i(r) \geq U_i(\tilde{r}^i, r^{-i})$.

Once we have defined formally the self-referential game, let us sketch its timing. Before playing the base game and observing any signals, players simultaneously choose codes of conduct r^i. Afterwards, each player i privately observes his signal y_i generated by the probability distribution $\pi(y|r)$ for $y \in Y$. Finally, players execute their codes of conduct and play the base game $r_i^i(y_i) = s_i$ for $s_i \in S_i$. When choosing a code of conduct r^i, players commit to adhere to it.

1.3 Examples

Let the self-referential game consist of a binary space of signals $Y_i = \{0, 1\}$ for each player i, where 0 may be interpreted as we are both using the same code of conduct and 1 may be interpreted as we are both using different codes of conduct. This interpretation would be consistent with the probability distribution of signals profile where private signals are conditionally independent $\pi(y|r) = \pi_i(y_i|r)\pi_j(y_j|r)$ and for all players i, j with $q \geq p$ let

In words, if the two players employ different codes of conduct they are more likely to receive the signal 1. Finally, the space of codes of conduct R_0 we will take to be all pairs of maps (r^1, r^2) where $r^i : \{0, 1\} \to S_1 \times S_2$. We maintain this self-referential game throughout the next two examples.

1.3.1 The Prisoner's Dilemma

We will study a prisoner's dilemma game. The actions are denoted C for cooperate and D for defect, and the payoffs are given in the table below. We focus on pure strategies in this base game $S_i = \{C, D\}$ for all players i.

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

One equilibrium profile of codes of conduct $r \in R$ is simply to ignore the signal and defect, that is, all players $i = 1, 2$ choose strategy $r^i \in R_0$ such that $r^i_i(y_i) = D$ for any signal y_i, and pick any map for their opponent $r^i_j(y_j) = s_j$ for all y_j, s_j. This is a Nash equilibrium of the self-referential game exactly as in the strict Nash equilibrium of the base game, and each player gets 1.

Let us investigate the possibility of sustaining cooperation through self-referentiality. In particular, we consider the code of conduct $\hat{r}^i$ that chooses C if the signal 0 is received, and chooses D if the signal 1 is received. Each player i adheres to the code of conduct $\hat{r}^i$ where for all players i, j it prescribes

$$\hat{r}^i_j(y_j) := \begin{cases} C & \text{for } y_j = 0, \\ D & \text{otherwise.} \end{cases}$$

If both players adhere to the code of conduct $\hat{r}^i$, they receive an expected utility of $U_i(\hat{r}) = 5 - 4p$. A player who chooses instead to always defect, thus $\tilde{r}^i_i(y_i) = D$ for all y_i and $\tilde{r}^i_j(y_j) = s_j$ for any s_j, y_j, gets $U_i(\tilde{r}^i, \hat{r}^j) = 6 - 5q$, and does worse by always cooperating.

The code of conduct profile $\hat{r}$ is a strict Nash equilibrium of the self-referential game when $q \geq 4/5 + 1/5p$. This says, in effect, that the signal must be informative enough.

1.3.2 The Repeated Prisoner's Dilemma

We now consider the prisoner's dilemma repeated twice and we use the sum of payoffs between the two periods. Consider the following code of conduct r^i.

$$r^i_j(y_j) := \begin{cases} CC & \text{if } y_j = 0, \\ DD & \text{if } y_j = 1. \end{cases}$$

Since play is not conditioned on what the other player does in the first period, the optimal alternative code of conduct $\tilde{r}^i$ against this code r^i is given by $\tilde{r}^i_i(y_i) = DD$ for any y_i and any kind of mapping $\tilde{r}^i_j(y_j) = s_j$ for any $s_j \in \{C, D\}^2$ and y_j. Thereafter, the analysis is the same as in the one-period case.

Next, we wish to examine whether it might nevertheless be possible to have cooperation in the two period game when the marginal probability distribution satisfies $q < 1/5 + 4/5p$. For simplicity of the exposition we analyze the case in which $p = 0$. We consider the code of conduct $\hat{r}^i$ that prescribes for all players i, j

$$r^i_j(y_j) := \begin{cases} tit\text{-}for\text{-}tat & \text{if } y_j = 0, \\ DD & \text{if } y_j = 1. \end{cases}$$

In other words, following the good signal 0 players play tit-for-tat, following the bad signal 1 players defect in both periods. If both players adhere to the code of conduct $\hat{r}^i$, they get expected payoffs equal to $U_i(\hat{r}) = 10$.

There are two alternative codes of conduct of interest $\tilde{r}^i, \check{r}^i$: to defect in both periods, or to cooperate in the first period then defect in the second period.

Consider first the code of conduct $\tilde{r}^i$ that says $\tilde{r}^i_i(y_i) = DD$ for any y_i, and any map $\tilde{r}^i_j(y_j) = s_j$ for all y_j, s_j. A player who chooses $\tilde{r}^i$ has a $1-q$ chance of getting 6, and a q chance of getting 1 in the first period, while he gets 1 in the second period for sure. Thus, the expected utility is $U_i(\tilde{r}^i, \hat{r}^j) = 7 - 5q$. Since this is less than 10 for any q, it is never optimal the choice of this alternative code of conduct $\tilde{r}^i$. Next, suppose the code of conduct $\check{r}^i$ characterized by $\check{r}^i_i(y_i) = CD$ for any y_i, and $\check{r}^i_j(y_j) = s_j$ for any y_j and s_j. Player i gets expected payoff of $U_i(\check{r}^i, \hat{r}^j) = 11 - 10q$. From these results we can work out that our code of conduct $\hat{r}^i$ would be chosen over $\check{r}^i$ when the probability distribution satisfies $q > 1/10$. If this condition holds, the profile of codes of conduct $\hat{r}$ is a Nash equilibrium of the self-referential game. By comparison in the one-period game we require $q > 1/5$ so for $1/10 \leq q < 1/5$ we can sustain cooperation in the two period game, but not in the one-period game. By violating the code of conduct with the intention to deviate in the final period the deviator risks being found out and punished when cooperating in the first period. Notice also that the code of conduct $\hat{r}^i$ is a strict self-referential Nash equilibrium except in the boundary case when $q = 1/10$.

It is interesting also to see what happens in the $T < \infty$ period repeated prisoner's dilemma game with no discounting. Let us consider the time-average payoff. Consider the code of conduct $\hat{r}^i$ that says that both players should play the grim-strategy on the good signal, and always defect on the bad signal. We write D_T for the $T \times 1$-vector of all D entries

$$r^i_j(y_j) := \begin{cases} \textit{tit-for-tat} & \text{if } y_j = 0, \\ D_T & \text{if } y_j = 1. \end{cases}$$

This gives a payoff of $U_i(\hat{r}) = 5$. The optimal alternative code of conduct $\tilde{r}^i$ against this code of conduct is to play the grim-strategy until the final period, then defect. Formally, $\tilde{r}^i_i(y_i) = (C_{T-1}, D)$ for all y_i with C_{T-1} representing a $T - 1 \times 1$-vector of all C entries, and for the opponent it says $\tilde{r}^i_j(y_j) = s_j$ for all y_j and $s_j \in \{C, D\}^T$. The expected payoff would be $U_i(\tilde{r}^i, \hat{r}^j) = 1/T\,[(1 - q)(5T + 1) + q(T - 1)]$. Hence it is optimal to adhere to the code of conduct $\hat{r}^i$ when $q \geq 1/(4T + 2)$. The salient fact is that as $T \to \infty$ only a very tiny probability of "getting caught" is needed to sustain cooperation.

1.4 Are Self-Referential Games Relevant?

There are three issues to address. First: to what extent is it possible for strategies to recognize one another and make use of that information? Second: what is the proper extension of self-referential strategies from two player symmetric games to general games? Third: does this model capture some aspect of reality? Since there is scarcely reason to discuss whether the model captures an aspect of reality if it is impossible to implement self-referential strategies, we address each of these issues in turn.

1.4.1 Codes of Conduct as Computer Algorithms

A simple physical model of strategies is to imagine that players play by submitting computer programs to play on their behalf. In this setting, computer programs work as follows. Fix a signal profile space Y and break the program into two parts, one of which generates Y based on analyzing the programs, the other of which maps the signal profiles Y into the strategy profile space S. The programs are self-referential in the sense that they also receive as input the program of the other player.

A fairly well-known result is the impossibility of running an algorithm in which we are able to read the opponents program and best respond to it. On the other hand, it is still possible to write down a computer program that gives one response if both programs are the same, and gives an alternative response if different. One sequence of commands may execute a particular strategy. This raises the question whether we are able to compare two programs having the same function while written differently. The answer is no. The last comparison must be based on actual code, and it cannot be made on function of program. For a more detailed discussion, see [Levine and Szentes(2006)].

Specifically, we assume that there is a finite language L of computer statements, and a finite limit l on the length of a program. The (finite) space of computer programs P consists of all sequences in L of length less than or equal to l. Each program $p^i \in P$ produces outputs which have the form of a map $p^i : P \times P \to \{1, 2, \ldots, \infty\} \times S$. The interpretation is that $p^i(p^1, p^2) = (\nu^i, s^i)$ produces the result s^i after ν^i steps. In case $\nu^i = \infty$, the program does not halt. Notice that depending on the language L these programs can be either Turing machines or finite state machines. A "self-referential strategy" consists of a pair consisting of a "default strategy profile" and a program $r^i = (\bar{s}^i, p^i)$, where $\bar{s}^i \in S$. After players submit their program p^1, p^2, each program p^i is given itself and the program submitted by the opposing player p^{-i} as inputs. All programs are halted after an upper limit of $\bar{\nu}$ steps. If $p^i(p^i, p^{-i}) = (\nu^i, s^i)$ and $\nu^i \leq \bar{\nu}$, that is, the program halted in time, we then define the mapping $r^i(p^1, p^2) = s^i$, otherwise $r^i(p^1, p^2) = \bar{s}^i$. To map this to a self-referential game, we take the signal space to be $Y = S$. Then the probability distribution of signal profiles is $\pi(y|r) = 1$ if $y_i = r^i(p^i, p^{-i})$ for all players $i = 1, 2$, and $\pi(y|r) = 0$ otherwise. In this context, self-referential signals are defined as

Definition 1.4.1. *The strategy space S is self-referential with respect to the deadline $\bar{\nu}$ if for every pair of actions $\bar{a}, \underline{a}$ there exists a strategy $s = (d, p)$ such that*

$$p(\tilde{d}, \tilde{p}) := \begin{cases} \bar{\nu}, \bar{a} & \text{if } \tilde{d} = d, \tilde{p} = p, \\ \nu, \underline{a} & \text{otherwise.} \end{cases}$$

Perhaps the easiest way to provide convincing proof that there are self-referential strategy spaces is to provide a simple example of a strategy that satisfies the properties of definition 4.1. We consider the trading game with common action space $A = \{0, 1\}$. The default action is $\bar{s}^i = 0$. The computer language is the Windows command language; the listing is given below:

```
@echo off
if "0" EQU "\%3" goto sameactions
echo 0
goto finish
:sameactions
echo n | comp \%2 \%4
if \%errorlevel\% EQU 0 goto cooperate
echo 0
goto finish
:cooperate
echo 1
```

```
:finish
```

This program runs from the Windows command line, and takes as inputs four arguments: a digit describing the “own” default action, a “own” filename, an opponent default action and an opponent filename. If the opponent default action is 0, and the opponent program is identical to the listing above, the program generates as its final output the number 1; otherwise it generates the number 0. The point is, since it has access to sequence of its own instructions, it simply compares them to the sequence of opponents program instructions to see if they are the same or not. [Howard(1988)] uses this idea to analyze the case in which two computer programs play a prisoners dilemma game. [Tennenholtz(2004)] developed the concept of program equilibrium which works similarly to the Nash equilibrium of the self-referential game within this context.

When one of the players knows that the other player has a relatively bounded memory size and is not a skilled programmer, he might exploit this. Writing a program that seems to be the same but actually is different requires large memory and clever computational skills. In addition, the use of these computer programs may work just as well as secret handshake: they will be visible to each other, but not to less sophisticated programs. For example, if a portion of a program is not visible to a naïve opponent, a clever programmer could fill it with a particular meaningless sequence of code that is never executed but that serves only to identify the program to a sophisticated opponent.

People playing games do not often do so by submitting computer programs. From an evolutionary perspective, genes serve to an important extent as computer programs governing how we behave. Modern anthropological research, such as [Tooby and Cosmides(1996)], emphasizes the extent to which social organization is influenced by the ability and inability to detect genetic differences using cues ranging from appearance to smell. These studies suggest emotional programs create involuntary emitted signals that may be forecast of the mind state.

1.4.2 From Two Players to Many

In a two player symmetric game, a player can compare himself to his opponent and tell whether they both follow the same strategy or not. When we extend the environment and allow more players in different roles such a simple comparison is no longer possible. Our notion of a code of conduct is intended to capture what it means to "be the same" in a more general setting.

In a multiplayer multi-role game a code of conduct may be interpreted as the specification of how all players are supposed to play. This definition is convenient with asymmetry: players can compare themselves to their opponents by determining if they have the same expectations for how players in different roles (including their own) should play. Applying this interpretation we can characterize two players agreeing about how all players should behave as "adhering" to the same code of conduct. Players do not only agree about how they would behave, but also about how third parties would behave. With two player games, one would expect to observe situations in which players are able to recognize opponents' kind same or different. Signals that allow you to understand your opponents intentions might be as simple as information systems or might be as the form of micro facial expressions. In the context of player games, agents are endowed with the possibility of discovering everyone elses intentions. This is motivated by the fact that humans are able to interpret signals from other people. We assume that players recognize intentions irrespectively of the number of players in the game.

1.4.3 Do People Use Self-Referential Strategies?

Here is one possible motivation for this recognition technology. Strategies govern the behavior of agents over many matches possibly randomly pairwise matches. Players are committed to a particular strategy because it is too costly to change behavior in any particular match. Suppose a player could observe the past interactions of an upcoming opponent. It might be difficult on this basis to form an exact prediction of how that opponent would behave during their own upcoming match. However, it would be considerably easier to determine if that opponent conformed to a particular rule for example a player might be able to tell with a reasonable degree of reliability whether that opponent was following the same or a different strategy than he was employing himself. Moreover, portions of strategies might be directly observable. For instance, an individual who rarely lies may blush whenever he is dishonest. Seeing an opponent blush would indicate that he would be unlikely to be dishonest in future interactions. (This example is due to [Frank(1987)].)

1.5 Perfect Information

Throughout this section we assume that signals in the self-referential game are perfectly revealing, i.e. player directly observes the code of conduct chosen by his opponents.

Consider any base game $\Gamma = \{I, \{S_i, u_i\}_{i \in I}\}$, which is not necessarily a repeated game. The self-referential game consists of a finite set of signal spaces Y_i for each i, a code of conduct space R_0, and probability distribution over profile of signals given by $\pi(\cdot|r)$. To analyze this case we assume that each player $i \in I$ is able to detect all opponents that do not adhere to the same code of conduct. We say that the self-referential game *permits detection* if for each player i and all players $j \in I \setminus \{i\}$ there exists a set of private signals $\bar{Y}_j^i \subset Y_j$ such that for any profile of codes of conduct $r \in R$, and any $\tilde{r}^i \neq r^i$ we have $\pi_j(\bar{Y}_j^i|\tilde{r}^i, r^{-i}) = 1$ and $\pi_j(\bar{Y}_j^i) = 0$. Intuitively, this says that if player i deviates every other player j detects this deviation with certainty.

We define the (possibly mixed) minmax strategy of the base game against player i by $\underline{s}_{-i}^i$ as the argument of $\min_{s_{-i} \in S_{-i}} \max_{s_i \in S_i} u_i(s_i, s_{-i})$. Let $\underline{s}_i = u_i(\underline{s}_i^i, \underline{s}_{-i}^i)$ be the smallest payoff that his opponent can keep player i below and $\underline{s}_i^i$ denote i's best response to $\underline{s}_{-i}^i$.

Our first result in the perfect information case is similar to [Levine and Pesendorfer(2007)] with the difference that we consider asymmetry and more than two players:

Theorem 1.5.1. *If $v_i = u_i(s) \geq \underline{u}_i$ for all players i with strategy profile $s \in S$, then there exists a profile of codes of conduct $r \in R$ such that $(v_1, \dots, v_N)$ is a Nash equilibrium payoff of the self-referential version of the game.*

Proof. Take any profile $s \in S$ such that for any player $i \in I$, $u_i(s) \geq \underline{u}_i$. Consider the code of conduct $\hat{r}^i \in R_0$ that prescribes

$$\hat{r}_j^i(y_j) := \begin{cases} s_j & \text{if } y_j \in Y_j \setminus \bar{Y}_j^i, \\ \underline{s}_j^i & \text{if } y_j \in \bar{Y}_j^i. \end{cases}$$

If all players choose this code of conduct, any player i would get $U_i(\hat{r}) = u_i(s)$. Contrary, if player i adheres to some $\tilde{r}^i$ so that $\tilde{r}_i^i(y_i) = \tilde{s}_i$ for all $y_i \in Y_i$ and any $\tilde{s}_j$; and $\tilde{r}_j^i(y_j) = s_j$

for all y_j, s_j, he gets $U_i(\tilde{r}^i, \hat{r}^{-i}) = \underline{u}_i$. It follows then that $\hat{r}$ is a Nash equilibrium of the self-referential game. □

We turn now to the case with more than two players but we allow the possibility that only some people receive these perfectly revealing signals. We define the notion of this group by saying that a self-referential game is *locally perfectly informative* if there exists a proper nonempty subset of players $l \subset I$ such that all player k's contained in that set its self-referential structure satisfies permit detection. Formally, the self-referential game is said to be *locally perfectly informative* if there is a subset of players $l \in I$ and for all players $k \in l$ there exists a set of private signals $\bar{Y}_j^i \subset Y_j$ such that for any profile of codes of conduct $r \in R$, for each player i, and any $\tilde{r}^i \neq r^i$ we have $\pi_j(\bar{Y}_j^i|\tilde{r}^i, r^{-i}) = 1$ and $\pi_j(\bar{Y}_j^i|r) = 0$. All players $i \in I \setminus l$ observe a trivial signal $y_i = \emptyset$.

In this specification, there is a possible scenario in which one of the players who receive the perfectly revealing signal needs some other player to punish the deviator. To provide some sort of communication we assume cheap talk after receiving private signals Y_i and before playing the base game Γ.

The game has the following timing. First, players select their code of conduct r^i. Second, after signals are received $y_i \in Y_i$, players make announcements on violation of the code of conduct. Finally, agents play the base game. Whether this cheap talk communication turns out to be useful or not is one of the goals of this exercise.

Each player adheres to a code of conduct, $r^i \in R_0$. After receiving private signals $y_i \in Y_i$ according to $\pi_i(\cdot|r)$, players send cheap talk signals defined as *announcements* taken from a finite set $\tilde{y}_i \in \tilde{Y}_0$, and a profile of announcements is defined as $\tilde{y} \in \tilde{Y} := \times_i \tilde{Y}_0$. We allow for the possibility of not sending a message, i.e. $\{\emptyset\} \in \tilde{Y}_0$. The space of announcements is common for all players. A *message* from player i is a map $m_i : Y_i \to \tilde{Y}_0$ with message profile denoted by $m = (m_1, \ldots, m_N)$. Once announcements have been made, players are engaged to play Γ. A strategy for player i is the decision about a strategy $s_i \in S_i$ to take and a message m_i to send, that is, $s'_i = (m_i, s_i)$ with $s'_i \in \tilde{Y}_0 \times S_i$.

Potentially, players may respond to such announcements. If player j announces player i has violated the code of conduct profile and this was the only announcement, all players might play the prescribed action required to implement punishment to player i. Moreover player i may try to take advantage of this information structure by announcing somebody else has violated the code of conduct. Therefore, two players could be pointing to each other and it is not possible to tell who actually deviated. We rule out this mutually implication by assuming that the self-referential game *strongly permits detection* meaning that the notion of permitting detection is not reciprocal. We say that the self-referential game *strongly permits detection* if for all pairs of players $i, j \in l$ there exists a set of private signals $\bar{Y}_j^i \subset Y_j$ such that for any profile of codes of conduct $r \in R$, and any $\tilde{r}^i \neq r^i$ we have $\pi_j(\bar{Y}_j^i | \tilde{r}^i, r^{-i}) = 1$ and $\pi_j(\bar{Y}_j^i) = 0$, but for any $\tilde{r}^j \neq r^j$ it holds that $\pi_j(y_j | \tilde{r}^i, r^{-i}) = \pi_j(y_j | r)$ for all $y_j \in Y_j$.

Theorem 1.5.2. *Let $v_i = u_i(s) \geq \underline{u}_i$ for all $i \in I, s \in S$. If the self-referential game strongly permits detection and is locally perfectly informative with cheap talk, then there is a code of conduct profile $r \in R$ such that the payoff vector $v = (v_1, \ldots, v_N)$ is a Nash equilibrium of the self-referential game.*

Proof. Fix an arbitrary profile of strategies $s \in S$ such that $u_i(s) \geq \underline{u}_i$ for all $i \in I$. Given the profile of strategies, we begin with constructing the profile of codes of conduct, $\hat{r}$, that implements $u_i(s)$ for all i.

Let $\hat{r}^i$ be the code of conduct that prescribes for any player $i \in I$ the following: For any player $i \in l$ if he is able to unilaterally punish player j, $\hat{r}^i_i(y_i) = (\emptyset, \underline{s}^j_i)$, for any $y_i \in \bar{Y}^j_i$. Alternatively, if player i needs player k's to punish j, he chooses $\hat{r}^i_i(y_i) = (\tilde{y}^j_i, \underline{s}^j_i)$ for any $y_i \in \bar{Y}^j_i$, and for some specific $\tilde{y}^j_i \in \tilde{Y}_0$. If he does not receive any signal from the set $\bar{Y}^j_i$ but the message $\tilde{y}^j_i \in \tilde{Y}_0$ to punish player j, play $\hat{r}^i_i(y_i) = (\emptyset, \underline{s}^j_i)$ where $y_i \in Y_i \setminus \bar{Y}^j_i$. In any other case, $\hat{r}^i_i(y_i) = (\emptyset, s_i)$ for all $y_i \in Y_i \setminus \bar{Y}^j_i$. For all players $i \in I \setminus l$, they choose $\hat{r}^i_i(\emptyset) = s_i$, and $\hat{r}^i_i(\emptyset) = \underline{s}^j_i$ for any $\tilde{y}^j_k \in \tilde{Y}_0$ announced by any player $k \in l$. We now proceed to show that this code of conduct implements the payoff proposed. If players adhere to $\hat{r}^i$ their payoffs are $U_i(\hat{r}) = u_i(s)$ for all i.

The relevant alternative code of conduct $\tilde{r}^i \in R_0$ for players $i \in l$ is to announce somebody else has deviated $\tilde{y}^k_i \in \tilde{Y}_0$ and to not play the prescribed strategy by choosing $\tilde{s}_i \in S_i$. There are two possible cases to consider here. First, player i announces that a player k who cannot observe i's play has deviated. Since there is some other player $k' \neq i, k$ points out player i's deviation and i is correspondingly punished, player i's expected payoffs are equal to $U_i(\tilde{r}^i, \hat{r}^{-i}) = \underline{u}_i$. More interestingly, suppose player i accuses player k who does observe i's play and who accuses i. By strong detection, detection is unidirectional hence everyone knows that while k observes i's play, i does not observe k's play, hence i should be punished and obtains $\underline{u}_i$. From this we can conclude that the code of conduct $\hat{r}$ is a Nash equilibrium of the self-referential game. □

In this theorem we require the strong version of "permits detection." The reason for this is simple: if players respond only to unique announcements then a player can foil the system by violating the equilibrium code of conduct and announcing also that another player has violated it. At worst when he is detected there will be two such announcements. This is a fairly common strategy in criminal proceedings: try to obscure guilt by blaming everyone else.

However if the game strongly permits detection then we can specify that when two players announce violations and one points to the other, then the one who has no information is punished. What strong detection says in a sense is that there are neutral witnesses people who observe wrong-doing but who cannot be credibly accused of wrong-doing by the wrong-doer.

1.6 Approximate Equilibria

In this section, we ask to what extent a small probability of detecting deviations from a code of conduct can be used to sustain approximate equilibria of the base game as strict equilibria of the self-referential game.

We relax the detection technology assumption from last section and allow for noisy signals instead of perfectly revealing signals. To this end, we define the notion of imperfect identification of self-referential strategies. The self-referential game is said to E, D *permit detection* where constants E, D satisfy $E, D \in [0, 1]$ and $E + D \leq 1$ if for every player $i \in I$ there exists some player j and a nonempty set $\bar{Y}_j \subset Y_j$ such that for any profile code of conduct $r \in R$, any signal $\bar{y}_j \in \bar{Y}_j$, and any $\tilde{r}^i \neq r^i, \tilde{r}^i \in R_0$ we have $\pi_j(\bar{y}_j|\tilde{r}^i, r^{-i}) \geq D$ and $\pi_j(\bar{y}_j|r) \leq E$. We view D as the probability of detection, that is, how likely it is that player j observes intentions of deviating from the other player i. You can think of E as the probability of someone who is being falsely accused of cheating when he behaves honestly.

We assume that in the base game all players have access to N individual randomizing devices each of which has an independent probability ε_R of an outcome we call *punishment*.

Depending on the private signals in the self-referential game players decide whether to ignore or not these outcomes.

For any $\varepsilon \geq 0$, the strategy profile $s \in S$ is an ε-Nash equilibrium if for all players i and for any strategy $\tilde{s}_i \neq s_i$ it holds that $u_i(s_i, s_{-i}) - u_i(\tilde{s}_i, s_{-i}) \geq \varepsilon$. Suppose that the strategy profile s^0 is an ε_0-Nash equilibrium giving utility $u_i(s^0)$ for each player i in the base game. For any strategy profile $s \in S$ and strategies s^i_j played by player j for any pair of players $i, j \in I$ with $i \neq j$ suppose that $s^i = (s^i_j, s_{-j})$ are ε_1-Nash equilibria. Define for each player i, $P_i = u_i(s^0) - u_i(s^i)$. We assume that $P_i \geq \underline{P} \geq 0$ and for some $\varepsilon_p \geq 0$ that $|u_j(s^i) - u_j(s^0)| \leq \varepsilon_p$. The number P_i stands for player i's loss when punished and the punishment must be at least $\underline{P}$. Think of ε_p as a measure of the closeness of s^i to s^0, that is, a measure of how far the punishment equilibria are from the original equilibrium. Since the base game is finite, we can denote by $\underline{u}, \bar{u}$ be the lowest and highest payoffs to any player in the game. Define two parameters ε and K

$$\varepsilon = \varepsilon_0 + (N + \bar{u} - \underline{u})(\varepsilon_1 + \varepsilon_p)E,$$

$$K = \max\{(N + \bar{u} - \underline{u}) \left[3N^4(1 + \bar{u} - \underline{u})\right], N \left[N^4(\bar{u} - \underline{u} + 1)\right] (\bar{u} - \underline{u})\}.$$

Observe that K depends only on the number of players in the game, and the highest and lowest possible utility and not, for example, the size of the strategy spaces or other details of the game.

Theorem 1.6.1. *Suppose $(D(\underline{P} - \varepsilon_1))^2 > 4K\varepsilon$. Then there exists an ε_R and a strict Nash equilibrium code of conduct $\hat{r} \in R$ such that for all players $i \in I$,*

$$\left|u_i(s^0) - U_i(\hat{r})\right| \leq \varepsilon + D(\underline{P} - \varepsilon_1) - \sqrt{(D(\underline{P} - \varepsilon_1))^2 > 4K\varepsilon}.$$

The proof, which can be found in the Appendix, is simply a computation. The key point is that if $\underline{P} > \varepsilon_1$ then small enough ε implies a strict Nash equilibrium of the self-referential game giving players very nearly what they get at the approximate equilibrium. To better understand what the theorem says let us answer the following question: When is ε small? Not surprisingly we must have ε_0 small. In addition we must either have E small or both ε_p and ε_1 small.

Recall that we are holding D the chance of being "caught" fixed. Here E measures how frequently we must punish if nobody deviates. The quantities ε_p and ε_1 measure how costly the punishment is and how credible it is respectively. That is, if ε_p is large players who carry out punishments stand to lose quite a lot compared to sticking at s^0, while if ε_1 is large players have a lot of incentive to deviate from the punishments. These two forces together make any code of conduct hard to adhere by players. But, we are able to overcome this issue by exploiting the E, D possibility of detection. At the extreme case, when $E = 0$, the parameter ε turns out to be ε_0. The smaller the ε_0, the closer our strict equilibrium to the approximate equilibrium in the base game.

Holding E fixed is more problematic, because in a general game it is not clear how we can choose punishments s_j^i that have little cost to the punishers and are also credible. Given an ε_0-Nash equilibrium s^0 we might expect to be able to find nearby approximate equilibrium $\bar{s}^i$ which punishes player i, but they will not generally have the requisite form $\bar{s}^i = (s_j^i, s_{-j})$ in which just the player j who detects i deviates and indeed it may be very hard for player j to punish i by himself. This problem, however, can be solved by allowing cheap talk after detection and before play: player j simply announces that he thinks that i has violated the code of conduct, and if he is the only one to make such an announcement, then all players play $\bar{s}^i$. To do this, we may use the message structure presented in Section 5 with identical timing. Yet, for such a procedure to work we need to strengthen the notion of "E, D permits detection" slightly along the lines used in the perfect information case. In particular with 3 or more players, we define the notion of E, D strongly permits detection to mean that if j detects i then i does not detect j.

One class of games that has a very rich structure of approximate equilibrium as [Radner(1980)] pointed out, are repeated games between patient players. In the repeated game setting the idea of choosing "equilibria" that punish the punished a lot and the punishers a little is very close to that used in [Fudenberg and Maskin(1986)] to prove the discounted folk theorem. Hence it is plausible that in these games we can find many strict Nash equilibria of the self-referential game even when E is fixed and not necessarily small.

1.7 Repeated Games with Private Monitoring

In this section, our goal is to prove a folk-like theorem for games with private monitoring. Fudenberg and Levine (1991) consider repeated discounted games with private monitoring that are informationally connected in a way described below. They show that socially feasible payoff vectors that Pareto dominate mutual threat points are ε-sequential equilibria where ε goes to zero as the discount factor δ goes to one. Our goal is to show that if the game is self-referential in a way that allows some chance that deviations from codes of conduct are detected (no matter how small is that chance), then this result can be strengthened from ε-sequential equilibrium to strict Nash equilibrium. We follow [Fudenberg and Levine(1991)] in describing the setup.

1.7.1 The Stage Game

The stage game has finite action spaces $a_i \in A_i$ for each player $i \in \{1, \ldots, N\}$ and these are chosen simultaneously. The corresponding action profiles are denoted $a \in A = \times_i A_i$. We denote by $\Delta(A_i)$ the probability distributions over the set of actions A_i with element α_i, and by $\alpha \in \times_i \Delta(A_i)$ mixed profiles. Each player i has a finite private signal space Z_i with signal profiles written as $z \in Z = \times_i Z_i$. Given any action profile $a \in A$, the probability of a signal profile $z \in Z$ is given by $\rho(z|a)$. For each action profile $a \in A$, we write $\rho_i(z_i|r)$ for the marginal distribution of player i over signals $z_i \in Z_i$. This induces also a probability distribution for mixed actions $\rho(z|\alpha)$ as well as marginals over individual signals $\rho_i(z_i|\alpha)$. Utility for individual players in the stage game is denoted by $w_i : Z_i \to \mathbb{R}$ which depends only on private signal received by that player.[2] This gives rise to the expected utility function $g_i(a)$ constituting the normal form of the stage game

$$g_i(a) = \sum_{z_i \in Z_i} \rho(z_i|a) w_i(z_i).$$

We can extend expected payoffs to mix actions profile $\alpha \in \Delta(A)$ in the standard way, thus

$$g_i(\alpha) = \sum_{a \in A} \alpha(a) g_i(a).$$

A *mutual threat point* is a payoff vector $v = (v_i, \ldots, v_N)$ such that there exists a *mutual punishment action* - this is a mixed action profile α such that $g_i(\alpha_i', \alpha_{-i}) \leq v_i$ for all players i, mixed actions α_i'. We say a payoff vector is *mutually punishable* if it weakly-Pareto dominates a mutual threat point. As is standard, a payoff vector v is enforceable if there is a mixed action profile α with $g(\alpha) = v$, and if for some player i and some mixed action

[2]We may include the players own action in his signal if we wish.

α_i', $g_i(\alpha_i', \alpha_{-i}) > g_i(\alpha)$ then for some $j \neq$ we have $\rho_j(\cdot|\alpha_i', \alpha_{-i}) \neq \rho_j(\cdot|\alpha)$. Note that every extremal Pareto efficient payoff is enforceable.

The *enforceable mutually punishable* set V^* is the intersection of the closure of the convex hull of the payoff vectors that weakly Pareto dominate a mutual threat point and the closure of the convex hull of the enforceable payoffs. We will denote by $int(V^*)$ the interior of the set V^*. Notice that this is generally a smaller set than the socially feasible individually rational set both because there may be unenforceable actions, but also because the minmax point may not be mutually punishable.[3]

We now describe the notion of informational connectedness. Roughly this says that it is possible for players to communicate with each other even when one of them tries to prevent the communication from taking place. In a two player game there is no issue, so we give definitions in the case $N > 2$. We say that player i is directly connected to player $j \neq i$ despite player $k \neq j, i$ if there exists a mixed profile α and mixed action $\hat{\alpha}_i \neq \alpha_i$ for player i such that the marginal distribution of player j satisfies

$$\rho_j(\cdot|\hat{\alpha}_i, \alpha_i', \alpha_{-(i,k)}) \neq \rho_j(\cdot|\alpha) \text{ for all } \alpha_k'.$$

In words, this condition requires that given α being played any player i's deviation will be detected by some player j regardless of player k's play. We say that i is *connected* to j if for every $k \neq i, j$ there is a sequence of players $i_1, \ldots, i_n$ with $i_1 = i, i_n = j$ and $i_p \neq k$ for

[3][Fudenberg and Levine(1991)] prove only that the enforceable mutually punishable set contains approximate equilibria leaving open the question of when the larger socially feasible individually rational set might have this property. They construct approximate equilibria using mutual punishment, so in particular there is no effort to punish the specific player who deviates. This is necessary because they do not impose informational restrictions sufficient to guarantee that it is possible to determine who deviated. With additional informational restrictions of the type imposed in [Fudenberg et al.(1994)Fudenberg, Levine, and Maskin] it is likely that their methods would yield a stronger result. As this is a limitation of the original result, we do not pursue the issue here.

any p such that player i_p is directly connected to player i_{p+1} despite player k. Intuitively, we can always find a "network" between players i and j so that the message goes through no matter what other single player tries to do. The game is *informationally connected* if there are only two players, or if every player is connected to every other player.

1.7.2 The Repeated Game

We now consider the T repeated game with discounting, where we allow both T finite and $T = \infty$. A history for player i at time t is a sequence of actions taken and private signals observed until period t denoted by $h_i^t = (a_i^1, z_i^1, \ldots, a_i^t, z_i^t)$ while $h_i^0 = \emptyset$ is the null history. The set of all t-length private histories for player i is denoted by $H_i^t = (A_i \times Z_i)^t$. We denote by H^t the set of all histories of length t. Consequently we can define the set of histories $H = \bigcup_t H^t$ and the set of all private histories for player i given by $H_i = \bigcup_t H_i^t$. A behavior strategy for player i is a sequence of maps σ_i^t taking his private history h_i^{t-1} to a probability distribution over A_i, that is, $\sigma_i : H_i \to \Delta A_i$. We write σ for the profile of behavior strategies. Players have common discount factor $\delta < 1$. For some discount factor δ we let $u_i(\sigma; \delta, T)$ denote expected average present value for the game repeated T periods.

In this repeated game a strategy profile σ is an ε'-Nash equilibrium for $\varepsilon' \geq 0$ if $u_i(\sigma; \delta, T) + \varepsilon' \geq u_i(\sigma_i', \sigma_{-i}; \delta, T)$ for $\sigma_i' \neq \sigma_i$, for each player i. Combining Theorems 3.1 and 4.1 from [Fudenberg and Levine(1991)] we have the following theorem:

Theorem 1.7.1 ([Fudenberg and Levine(1991)])**.** *In an informationally connected game if $v \in V^*$ then there exists a sequence of discount factors $\delta_n \to 1$, non-negative numbers $\varepsilon_n \to 0$ and strategy profiles σ_n such that σ_n is an ε_n-Nash equilibrium*[4] *for δ_n and $u(\sigma_n; \delta_n, \infty) \to v$.*

[4] [Fudenberg and Levine(1991)] prove a stronger result - they show that σ_n is an ε_n-sequential equilibrium which means also that losses from time t deviations measured in time t average present value and not merely

We will require one other result from [Fudenberg and Levine(1991)]. Their Lemma A.2 together with their construction in the proof of Theorem 4.1 implies that it is possible to construct a communication phase with length L.

Lemma 1.7.1 ([Fudenberg and Levine(1991)])**.** *For any $\beta \in (0,1)$ there exist a pair of strategies σ_i, σ_i' and for each player $j \neq i$ a test $\mathcal{Z}_j \subset \{(z_j^1, \ldots, z_j^L)\}$ such that for any player $k \neq i, j$ and strategy σ_k' by player k, under $(\sigma_i, \sigma_k', \sigma_{-(i,k)})$ we have $Pr((z_{-(i,j)}^1, \ldots, z_{-(i,j)}^L) \in \mathcal{Z}_{-(i,j)}) \leq 1 - \beta$, and under $(\sigma_i', \sigma_k', \sigma_{-(i,k)})$ we have $Pr((z_{-(i,j)}^1, \ldots, z_{-(i,j)}^L) \in \mathcal{Z}_{-(i,j)}) \geq \beta$.*

This says that a player can "communicate" by using his actions whether or not someone has deviated. In fact, such communication between players is guaranteed by the assumption of information connectedness.

1.7.3 The Finitely Repeated Self-Referential Game

In the self-referential case it is convenient to work with finite versions of the repeated game. The T-discrete version of the game has finite time horizon T and players have access each period to independent randomization devices that provide a uniform over T different outcomes.

The self-referential T-discrete game consists of signal spaces Y_i for each player i, codes of conduct space R_T, and the signal probabilities are given by $\pi_T(\cdot|r)$.

Next we state the main result of the paper:

Theorem 1.7.2. *If V^* has non-empty interior, if the game is informationally connected, if for some $E \geq 0, D > 0$ the self-referential T discrete versions E, D permits detection, and if*

time t average present value are no bigger than ε_n. As we do not need it, we do not give the extra definitions required to state the stronger result.

$v \in int(V^*)$ then for any ε_0 there exists a sufficiently large discount factor δ, a discretization T and strategy pairs s_i^0, s_i^j for all players $i, j \in I$ such that s^0 is an ε_0-Nash equilibrium for δ, T, $\varepsilon_1 = \varepsilon_p = \varepsilon_0$, $\underline{P} = \sqrt[3]{\varepsilon_0}$ and $(D(\underline{P} - \varepsilon_1))^2 > 4K\varepsilon$.

Proof. First note that by choosing T large enough for given δ it is immediate that ε_0-Nash equilibria of the base game are $2\varepsilon_0$-Nash equilibria of the discretized game, so Theorem 7.1 applies directly to the discretized game. Theorem 7.1 immediately implies that for all sufficiently large δ we can find a sequence of discount factors with $\delta_n \geq \delta$ and corresponding T_n together with strategies $\bar{s}^0, \bar{s}^1, \ldots, \bar{s}^N$such that these are all $\varepsilon/2$-Nash equilibria, that $|u(\bar{s}^0) - v| < \varepsilon_0/2$, $u_i(\bar{s}^0) - u_i(\bar{s}^i) \geq \sqrt[3]{\varepsilon_0}$ and $|u_j(\bar{s}^i) - u_j(\bar{s}^0)| < \varepsilon_0/2$.

To construct s_i^0, s_i^j we begin the game with a series of communication phases. We go through the players $j = 1, \ldots, N$ in order each phase lasting L periods. In the first j-th phase the player $i \neq j$ who is able to detect deviations by player j has two strategies $\hat{s}_i^j, \hat{s}_i^{j'}$ and players $k \neq i, j$ have a strategy $\hat{s}_k^j$ from Lemma 7.2. In s_i^0 player i plays the L truncation of $\hat{s}_i^j$, alternatively in s_i^j he plays the L truncation of $\hat{s}_i^{j'}$. The remaining players play the L truncation of $\hat{s}_k^j$.

In s^0, at the end of these NL periods of communication each player conducts the test in Lemma 7.2 to see who has sent a signal. The test is used just like cheap talk in the earlier results. If it indicates that exactly one player i has sent a signal he plays his part of the equilibrium $\bar{s}^j$ punishing player j. If the test indicates that exactly two players i, j sent a signal where i reports that j has deviated then he plays his part of the equilibrium $\bar{s}^j$ punishing j. Otherwise he plays $\bar{s}^0$. By Lemma 7.2 by choosing sufficiently large L the probability β under any of the strategies s^0, s^j that all players agree that a single player i sent a signal (since in fact at most one player has actually sent a signal) or that no signal was sent may be as close to as we wish. In particular we may choose β close enough to 1 that play following disagreement or agreement on more than one player sending a signal has no more than an $\varepsilon_0/4$ effect on payoffs.

Observe that this choice of L does not depend at all on δ_n, T_n, so we may choose δ_n and T_n large enough that nothing that happens in the communications phase makes more than a $\varepsilon_0/4$ difference to payoffs. This shows that s_i^0, s_i^j have the desired properties. □

Notice that the structure of Theorem 1.7.2 differs from that of Theorem 1.7.1 in an important way. In Theorem 1.7.1 very precise information is accumulated on how players have played, and a mutual punishment is used, but so infrequently on the equilibrium path it has little cost. By way of contrast, in Theorem 1.7.2 we have fixed E. This means that we must make sure that the cost to the punishers is small relative to the cost to the punished. If not, it would be optimal to accept a small punishment in exchange for not having to dish out a costly one. Hence we cannot rely on mutual punishments, but must target them towards the "guilty."

1.8 Conclusion

The standard world of economic theory is one of perfect liars - a world where Nigerian scammers have no difficulty passing themselves off as English businessmen. In practice social norms are complicated and there is some chance that a player will inadvertently reveal his intention to violate a social norm through mannerisms or other indications of lying. Here we investigate a simple model in which this is the case. Our setting is that of self-referential games, which allows the possibility of observing directly opponents intentions.

We characterized the self-referential nature of this class of games by defining codes of conduct. Adhering to a code of conduct represents agreement between players that even have different roles and allows us to extend the setup studied by [Levine and Pesendorfer(2007)] to games with more than two players.

We have examined when and how codes of conduct matter. In particular, we think of codes of conduct as computer algorithm which goes beyond the interpretation that players submit computer programs to play. In fact, recent anthropological research suggests human genes are programmed as computer codes and behavior recognition between humans ranges from odor to visual cues.

Results obtained in the perfect recognition case have the flavor of folk theorems. This is possible because of perfect revealing signals that point at deviations from code of conduct and hence deviators are punished with certainty. Also, we weaken the assumption about who actually observe these signals. If the set of players receive these signals are endowed with the possibility of communication, the results hold. That is, players that detect deviations use a message structure to communicate with other players needed to implement a punishment.

In practice the probability of detection is not likely to be perfect, so we then focus on the case where the detection probability is small. The key idea is that a little chance of detection can go a long way. Small probabilities of detecting deviation from a code of conduct allow us to sustain approximate equilibria as strict equilibria of the self-referential version of the game. An illustrative, but far more important, application of this result is a discounted strict Nash folk-like theorem for repeated games with private monitoring. We conclude that approximate equilibria can be sustained as "real" equilibria when there is a chance of detecting violations of codes of conduct.

1.9 Appendix: Proof of Theorem 1.6.1

Proof of Theorem 1.6.1. First we bound the possibility that the "punishment" event occurs on more than one randomization device at the same time. Recall that each individual operates N randomizing devices in case they should have to report on more than one person. Hence there are N^2 independent randomization devices in operations. Thus, the event "punishment" does not occur to any player has probability $(1-\varepsilon_R)^{N^2}$. The probability that the event "punishment" occurs exactly once is $N^2\varepsilon_R(1-\varepsilon_R)^{N^2-1}$. From these results we find the probability that the event "punishment" occurs twice or more and an upper bound for this probability

$$\begin{aligned}1-(1-\varepsilon_R)^{N^2}-N^2\varepsilon_R(1-\varepsilon_R)^{N^2-1} &\leq N^2\varepsilon_R - N^2\varepsilon_R\left[1-(N^2-1)\varepsilon_R\right]\\ &= N^2(N^2-1)(\varepsilon_R)^2 \leq N^4(\varepsilon_R)^2\end{aligned}$$

Now we define the code of conduct $\hat{r}^i \in R_0$: for all players $i \in I$, if $\bar{y}_i \in \bar{Y}_i$ and the event "punishment" occurs play $\hat{r}^i_i(y_i) = s^j_i$ and $\hat{r}^i_j(y_j) = s^0_j$ for any $y_j \in Y_j$ and all players $j \neq i$, otherwise play $\hat{r}^i_i(y_i) = s^0_i$ for all $y_i \in Y_i \setminus \bar{Y}_i$ and for all $j \neq i$ choose $\hat{r}^i_j(y_j) = s^0_j$ for any $y_j \in Y_j \setminus \bar{Y}_j$. We will show under what circumstances the profile of codes of conduct $hatr$ is a Nash equilibrium of the self-referential game. The following mutually exclusive events can occur to player i when all players $j \in I \setminus \{i\}$ choose the code of conduct $\hat{r}^j$, but he chooses $\tilde{r}^i$ defined below where $\tilde{r}^i_j = \hat{r}^i_j$ for all $j \in I \setminus \{i\}$:

1. Nobody is punished: if the code of conduct $\hat{r}^i$ is followed by player i, he gets $u_i(s^0)$, if i deviates by choosing the code of conduct $\tilde{r}^i$ that prescribes another strategy $\tilde{s}_i \in S_i, \tilde{s}_i \neq s^0_i$ he gets at most $u_i(s^0) + \varepsilon_0$,

2. Player j is the only player punished: if $\hat{r}^i$ is followed i gets $u_i(s^j)$, if i deviates by choosing $\tilde{r}^i$ that prescribes no punishment to player j, $\tilde{r}^i_i(\bar{y}_i) \neq s^j$, he gets at most $u_i(s^j) + \varepsilon_1$,

3. Two or more players are punished: if $\hat{r}^i$ is followed player i gets at worst $\underline{u}$, if i deviates while choosing $\tilde{r}^i(y_i) = \tilde{s}_i$ with $\tilde{s}_i \in S_i$ and $\tilde{s}_i \neq s^i_j \neq s^0_j$ he gets at most $\bar{u}$.

Hence if all players follow the code of conduct $\hat{r}^i$, player i gets no more than

$$u_i(s^0) + (1 - (1-E)^N) \left[\varepsilon_p + N^4(\varepsilon_R)^2(\bar{u} - \underline{u})\right],$$

and no less than

$$u_i(s^0) - (1 - (1-E)^N) \left[\varepsilon_p + N^4(\varepsilon_R)^2(\bar{u} - \underline{u})\right] - \pi_j(\bar{y}_j|\hat{r})\varepsilon_R P_i.$$

Suppose player i violates the code of conduct profile $\hat{r}$ and chooses an alternative code of conduct $\tilde{r}^i$ like the one we described above. He gets no more than

$$u_i(s^0) + \varepsilon_0 + (1 - (1-E)^N) \left[\varepsilon_1 + N^4(\varepsilon_R)^2(\bar{u} - \underline{u})\right] - \left[(\pi_j(\bar{y}_j|\hat{r}) + D)\varepsilon_R - N^4(\varepsilon_R)^2\right](P_i + \varepsilon_1).$$

Consequently, the gain to violating the code of conduct profile $\hat{r}$ is at most

$$\begin{aligned}
&\varepsilon_0 + (1 - (1-E)^N) \left[\varepsilon_1 + \varepsilon_p + 2N^4(\varepsilon_R)^2(\bar{u} - \underline{u})\right] + \pi_j(\bar{y}_j|\hat{r})\varepsilon_R P_i \\
&\quad - \left[(\pi_j(\bar{y}_j|\hat{r}) + D)\varepsilon_R - N^4(\varepsilon_R)^2\right](P_i - \varepsilon_1) \\
&\quad \leq \varepsilon_0 + (N + \bar{u} - \underline{u})E \left[\varepsilon_1 + \varepsilon_p + 3N^4(\varepsilon_R)^2(1 + \bar{u} - \underline{u})\right] - D\varepsilon_R(\underline{P} - \varepsilon_1) \\
&\quad \leq \varepsilon + K\varepsilon_R^2 - D\varepsilon_R(\underline{P} - \varepsilon_1)
\end{aligned}$$

Hence if $(D(\underline{P} - \varepsilon_1))^2 > 4K\varepsilon$ then there is a strict Nash equilibrium with

$$\begin{aligned}\left|u_i(s^0) - U_i(\hat{r})\right| &\leq NE\varepsilon_p + N\left[N^4(\bar{u} - \underline{u}) + 1\right](\bar{u} - \underline{u})\varepsilon_R \\ &\leq \varepsilon + 2K\varepsilon_R.\end{aligned}$$

We conclude by solving the inequality for ε_R. The roots of the quadratic equation are

$$\varepsilon_R = \frac{D(\underline{P} - \varepsilon_1) \pm \sqrt{(D(\underline{P} - \varepsilon_1))^2 - 4K\varepsilon}}{2K}$$

which gives two real roots since $(D(\underline{P} - \varepsilon_1))^2 > 4K\varepsilon$, implying the existence of an ε_R for which $\hat{r}$ is strict Nash equilibrium of the self-referential game. Plugging the lower root into the inequality for the utility difference $|u_i(s^0) - U_i(\hat{r})|$ gives the remainder of the result. □

Chapter 2

Codes of Conduct and Bad Reputation

2.1 Introduction

A variety of economic long-run relationships—such as mechanic-motorists, doctor-patients and advisor-students—in which reputation concerns may lead to inefficiencies and even to complete market breakdowns were modeled by [Ely and Välimäki(2003), henceforth EV]. More recently, however, [Grosskopf and Sarin(2010)] found experimental evidence that shows that reputation is not as harmful as EV propose it should be. In addition, data suggests that markets that fit the assumptions in their model are not subject to breakdowns, for example, mitral valve surgeries or automobile repair.[5]

[5] In mitral valve surgeries there are two main categories: valve repair and valve replacement. Both surgeries are equally successful when properly conducted by the surgeon. To illustrate the idea, suppose that we have normal and bad surgeons. The bad physician always does valve replacement but the normal one may perform the necessary surgery. Patients only observe the type of surgery conducted. This market is analogous to the mechanic-motorists example in EV. In the US, on the other hand, more than 40,000 mitral valve operations are performed every year.

In this paper, I study a bad reputation game that extends the model of [Ely et al.(2008)Ely, Fudenberg, and
who characterize bad reputation in general. The present model incorporates agents that are capable of recognising an opponent's intentions (i.e., self-referentiality) and the possibility that the long-lived agent with reputation concerns can be replaced. These two features are common to the situations mentioned above, and the results here reconcile the fact that these markets remain functioning.

In the bad reputation game, a long-run player, whose actions are imperfectly observable, faces a sequence of myopic agents. Previous interactions are revealed through public signals. There are two types of long-run players: strategic and committed players. The main characteristic of these games is that short-run players participate in the game whenever they expect a *friendly* action, conditioning this decision on public history. However, friendly actions may lead to *bad* signals that can be interpreted as evidence of an *unfriendly* action by myopic agents, thereby deciding to exit the game even if the long-lived player has been playing friendly actions. The long-run player forgoes current payoffs for future reputation by playing unfriendly actions that might give evidence of good behavior, as a result, the contemporaneous and subsequent short-run players step out of the game.

The possibility of renewal is, for instance, surgeons that shift during a valve operation. Indeed, renewal has a dual effect. First, the long-run player might be replaced in the subsequent period, providing less incentive for him to play friendly (e.g., conduct the right surgery). Second, short-run players are unaware whether a replacement has occurred, thereby implying that a history of bad signals (i.e., valve replacement) weighs less in the updating of posteriors. Note that renewal would not overcome the problem of perverse reputation because the long-run player finds it more difficult to separate from unfriendly types with the rich commitment type space which is assumed here.

The idea of self-referentiality relies on agents' rules of behavior that are costly to change, motivated by the possibility that humans may reveal their states of mind, for example, through body gestures (as in [Frank(1988)]). Such rules, that were first introduced by [Levine and Pesendorfer(2007)], are modeled in the self-referential game as codes of conduct that are conditioned on signals that in turn are determined by a probability distribution that depends on everyone's choices of strategies. Intuitively, players have the chance of discerning whether opponents conform to the same rule of behavior. Closely related to this work is [Block and Levine(2012)] who build the self-referential framework used here. (See also [Kalai et al.(2010)Kalai, Kalai, Lehrer, and Samet].)

In the benchmark case, I first consider perfect identification about opponents' codes of conduct where self-referentiality overturns the bad reputation effect, thereby showing that the precommitment action outcome can be induced in equilibrium (Theorem 2.3.1). When there is imperfect information about codes of conduct, on the other hand, self-referentiality is no longer sufficient to overcome this problem because myopic players cannot rely only on intentions to discern types. The main result of the paper (Theorem 2.4.1) entails that self-referentiality complements renewal of the long-run player and that there exists an equilibrium in which the long-run player approximately obtains his precommitment payoff.

The literature includes models with the flavor of bad reputation games, such as [Morris(2001)] who shows that an advisor would not report truthfully if he might be thought to be biased and by lying he avoids being tagged as inaccurate. As a result, with reputation concerns there is no transmission of information in equilibrium. Recently, [Bar-Isaac and Deb(2013)] examine a reputation model with heterogenous audiences. Whenever these audiences cannot observe agent's actions simultaneously a bad reputation effect arises. The reason is that in the attempt to build a reputation towards one audience the agent takes unfriendly actions regarding the other audience. In credit markets, [Ordoñez(2013)] finds that under some conditions in the fundamentals of the economy, in equilibrium borrowers might invest in risky projects that are bad from the perspective of the lender.

This paper proceeds as follows. In Section 3.2.1, I describe the framework. In Section 2.3 and 2.4, I present the main results, and in Section 2.5 I relate them to EV example. Section 2.6 shows the necessity of renewal with imperfect observation. In Section 2.7, I conclude. All proofs are included in the Appendix.

2.2 Setup

A long-run player, player 1, faces a sequence of different short-run players, player 2 in a repeated game with finite horizon T. In the stage-game, player 1 chooses an action $a \in A$ and player 2 chooses an action $b \in B$.

There are finite privately known types for player 1, $\theta \in \Theta$. Let the common prior distribution over types be the probability measure $\mu_0 \in \Delta(\Theta)$ with $\mu_0(\theta) > 0$ for all $\theta \in \Theta$. The normal θ_0 is strategic, and the commitment types $\theta(a)$ always plays some action $a \in A$. The stage-game payoff functions are $u_i : A \times B \to \mathbb{R}$ for $i = 1, 2$. The long-run player discounts future with $\delta \in (0, 1)$. In the repeated game, the normal type maximizes the discounted sum of expected payoffs whereas commitment types plays mechanically the stage action a.

There exists the possibility of renewing the long-run player, that is exponentially distributed with probability $\nu \in (0,1)$. Every period, the long-run player might be replaced by either a normal or a commitment type. The new type θ is drawn according to μ_0.[6]

Here I follow the setup developed by [Ely et al.(2008)Ely, Fudenberg, and Levine] when describing bad reputation games.[7] At the end of the stage game, a public signal y is observed from finite space Y. Given an action profile (a,b), $y \in Y$ is drawn according to the probability distribution $\rho(\cdot|a,b) \in \Delta(Y)$. Short-run players are restricted to observe public signals. Let $H^t = Y^t$ be the set of all t-length public history with element $h^t = (y_1, \dots, y_t)$, $h^0 = \emptyset$, and $H = \bigcup_{t \geq 1} Y^t$. The private history of the long-run player is $h_1^t = (a_1, \dots, a_{t-1})$ which belongs to $H_1^t = A^t$, and let $H_1 = \bigcup_{t \geq 1} H_1^t$.

The behavior strategy of the long-run player is a mapping $\sigma_1 : H \times H_1 \times \Theta \to \Delta A := \mathcal{A}$. A strategy for the short-run player is a sequence of maps $\sigma_2(h^t) \in \Delta B := \mathcal{B}$. Then, a short-run action $\beta \in \mathcal{B}$ is a Nash response to action $\alpha \in \mathcal{A}$ if $u_2(\alpha, \beta) \geq u_2(\alpha, \hat{\beta})$ for all $\hat{\beta} \in \mathcal{B}$. Let $\mathbf{B}$ be the best-response correspondence and $\mathbf{B}(\alpha)$ be the set of short-run Nash responses to α.

Bad reputation games are a subclass of participation games in which short-run players may not to participate in the game. They stay out by choosing an exit action $e \in E$, whereas entry actions belong to $B - E$. Let $Y^E \subseteq Y$ be the set of exit signals. Given an exit action, only exit signals are possible. In addition, when entry actions are chosen none of the exit signals can be observed. A game is a participation game if the exit action set is non-empty $E \neq \emptyset$ and there exists some action α with $\mathbf{B}(\alpha) \cap E \neq \emptyset$.

[6] The literature on impermanent types includes, for instance, [Holmström(1999)], [Mailath and Samuelson(2001)], [Phelan(2006)], [Wiseman(2008), Wiseman(2009)], and [Ekmekci et al.(2012)Ekmekci, Gossner, and Wilson].

[7] To lighten notation, some formal definitions are included in the Appendix.

Throughout, $\beta\{E\}$ denotes the probability assigned to E, by Nash response action β. A non-empty finite set of pure actions $F \in A$ is friendly if there is a number $\gamma > 0$ such that, for all $\alpha \in \mathcal{A}$ if the short-run player strategy is a Nash response $\beta \in \mathbf{B}(\alpha)$ and $\beta\{E\} < 1$ then the long-run player assigns positive probability $\alpha(f) \geq \gamma$ for some $f \in F$. An unfriendly set N corresponding to F is any non-empty subset of $A \setminus F$. In words, it says that Nash response of short-run players puts positive probability to non-exit actions when friendly actions are likely to be played. Any non-friendly action makes short-run players choose an exit action. We say α is enforceable if for every $\tilde{\alpha}, \beta$ with $\beta \in \mathbf{B}(\alpha)$ and $\beta\{E\} < 1$, such that $u_1(\tilde{\alpha}, \beta) > u_1(\alpha, \beta)$ then $\rho(\cdot|\tilde{\alpha}, \beta) \neq \rho(\cdot|\alpha, \beta)$.[8]

Let $\Theta(F)$ and $\Theta(N)$ be the friendly and unfriendly commitment types, respectively. It follows that $\Theta(N) \cap \Theta(F) = \emptyset$. A set of signals $\overline{Y}$ is evidence for N if each action in N gives rise to a higher probability for every signal in $\overline{Y}$ than any action not in N. An action a is vulnerable to temptation relative to $\overline{Y}$ if by playing the temptation action d the long-run player reduces the probability of bad signals by at least $\underline{\rho}$ and increases all signals but in $\overline{Y} \cup Y^E$ by factor $1 + \tilde{\rho}$. We say an action d is a *costly temptation* if it is a temptation and $u_1(a, b) - u_1(d, b) \geq c$ with $c > 0$ for all $b \in B - E$. A participation game has exit minmax if

$$\max_{b \in E \cap \mathbf{B}} \max_{a \in A} u_1(a, b) = \min_{\beta \in \mathbf{B}} \max_{a \in A} u_1(a, \beta).$$

A participation game is a bad reputation game if it has exit minmax, and there is a friendly set F, corresponding non-empty unfriendly set N and a set of signals $\overline{Y}$ that are evidence for N, such that every enforceable friendly action f is vulnerable to temptation relative to the set of signals $\overline{Y}$. Signals in $\overline{Y}$ are called bad signals. A bad reputation game with costly temptation exhibits a costly set of temptations.

[8] The idea of identification of actions was proposed by [Fudenberg et al.(1994)Fudenberg, Levine, and Maskin].

2.2.1 The Self-Referential Game

Consider a generic finite base game Γ. There is a set of players, $\mathcal{I} = \{1, \ldots, N\}$, and for each player i a set of strategies, $s_i \in S_i$, and a payoff function, $u_i : \times_i S_i \to \mathbb{R}$.

The self-referential game is defined on Γ so it inherits the set of players $\mathcal{I}$. There is a finite set of private signals $Z_i \ni z_i$ for each player i such that $|Z_i| \geq 2$. The strategy of player i is defined as code of conduct, r^i, which is an $|\mathcal{I}| \times 1$ vector whose jth element corresponds to the mapping from the set of player j's private signals to player j's strategies in the base game. The common space of such codes of conduct is $R_0 \coloneqq \{r^i \mid r^i_j : Z_j \to S_j \text{ for all } i, j\}$. It requires each player to decide how he will play and how he thinks all the other players will play. Denote the profile of codes of conduct $r \in \mathcal{R}$. Given r, the probability distribution of private signals is $\pi(z|r)$. The profile of codes of conduct $\hat{r}$ is a Nash equilibrium of the self-referential game if for all players i we have that $\hat{r}^i \in \text{argmax}_{\tilde{r}^i} \sum_{z \in Z} u_i(r^1_1(z_1), \ldots, r^N_N(z_N))\pi(z|r)$.

The timing of the self-referential game is as follows: before observing any signal and playing the base game, all players simultaneously choose codes of conduct. Given this choice, a profile of private signals is drawn from the probability distribution $\pi(z|r)$. After observing private signals, players execute codes of conduct $r^i_i(z_i) = s_i$. Let $G(\Gamma)$ denote the self-referential game defined upon the base game Γ that is a bad reputation game throughout this paper.

2.3 Benchmark Case: Perfect Identification

As a benchmark, it is useful to examine the case when players have perfectly revealing signals about codes of conduct that were chosen by opponents. I show that we can sustain "good" equilibria regardless of the long-run player's patience and how frequently he leaves the game.

Formally, we say a self-referential game permits detection if for every player i, all players $j \neq i$ there exists a set $\overline{Z}_j^i \subset Z_j$ such that for every $r \in \mathcal{R}$ and any code of conduct $\tilde{r}^i \neq r^i$, we have $\pi_j(\overline{Z}_j^i|\tilde{r}^i, r^{-i}) = 1$ and $\pi_j(\overline{Z}_j^i|r) = 0$. This definition says that any deviation is detectable by all players in the game. Let $f^* \in F$ be the precommitment friendly action of the long-lived player.

Theorem 2.3.1. *Consider a bad reputation game* Γ *with any* ν *and* δ*. Suppose that the self-referential game* $G(\Gamma)$ *permits detection. Then, there exists a Nash equilibrium of* $G(\Gamma)$ *where the payoff of the normal long-run player is* $u_1(f^*, \beta)$*, and* $\beta \in \mathbf{B}(f^*)$*.*

See the Appendix for the proof.

In the canonical bad reputation game Γ, short-run players try to anticipate whether player 1 would choose a friendly action $f \in F$, had he entered the game $b \in B - E$. With perfectly revealing signals, on the other hand, public history h^t is irrelevant to myopic players because they can condition actions only on what the long-run player will be doing through signals $z_2 \in Z_2$. Thus, player 1 would not need to regain the short-run players' faith after a history of bad signals (i.e., $\overline{y}_s \in h^t$ for many $s \leq t$) but he might find it optimal to choose an unfriendly action $a \in N$. However, he will never choose $a \in N$ in equilibrium, since doing so would ensure him a punishment immediately and thereafter, getting his minmax payoff 0. By focusing on this benchmark case, we may consider restrictive situations, but most importantly it shows that there is no room for reputation building.

2.4 Identification with Noise

In this section, I consider the case where private signals about codes of conduct are noisy. A self-referential game is said to λ, η permits detection with constants $\lambda, \eta \in [0,1]$ if for every player i there exists some player $j \neq i$ and a set $\overline{Z}_j \subset Z_j$, such that for any $r \in \mathcal{R}$, any signal $\overline{z}_j \in \overline{Z}_j$ and any code of conduct $\tilde{r}^i \in R_0$ where $\tilde{r}^i \neq r^i$ we have $\pi_j(\overline{z}_j|\tilde{r}^i, r^{-i}) - \pi_j(\overline{z}_j|r) \geq \eta$ and $\pi_j(\overline{z}_j|r) \leq \lambda$. In words, η measures the probability of detection if any player i deviates, and λ can be interpreted as the probability of accusing someone who is being honest. It is convenient to assume that all short-run players obtain the same private signal about the long-lived player's intentions in turn.[9]

It is necessary to impose uniformity to avoid the possibility of a relatively too high prior probability of unfriendly types. We say a bad reputation game with friendly set F and unfriendly set N has *uniformly friendly commitment size* ψ, χ with $\chi > 0$ if the prior probability of friendly and unfriendly types $\mu_0[\Theta(N)], \mu_0[\Theta(F)]$ satisfy

$$\mu_0[\Theta(N)] \leq \psi \left(1 - \mu_0[\Theta(F)] + \mu_0[\Theta(F)]^{\frac{1+\eta}{\eta}}\right) - \chi.$$

The constant ψ reflects the uniformity of friendly types $\mu_0[\Theta(F)]$ relative to unfriendly types $\mu_0[\Theta(N)]$. Players use their private signals to update their prior probabilities about commitment types to a limited extent, therefore a relatively high likelihood of friendly commitment types is necessary to guarantee participation.

Next, the main result of the paper:

[9] This assumption ensures that myopic players do not infer anything about others' draw.

Theorem 2.4.1. *Assume a bad reputation game* Γ *with uniformly friendly commitment size* $1-\gamma, \chi$ *for some* $\chi > 0$ *and with costly temptation. Suppose that* $G(\Gamma)$ *the self-referential game* η, λ *permits detection. For* $f^* \in F$ *and any* δ*, there exist* $\tilde{\eta} > 0$*,* $\tilde{\lambda}$ *and* $\tilde{\nu}$ *such that for* $\eta \geq \tilde{\eta}$*,* $\lambda \leq \tilde{\lambda}$*, and* $\nu \geq \tilde{\nu}$ *there is Nash equilibrium of* $G(\Gamma)$ *where the payoff of the normal long-run player is approximately* $u_1(f^*, \beta)$ *with* $\beta \in \mathbf{B}(f^*)$*.*

The outline of the proof is as follows. By the η-λ detection technology, agents may adhere to a code of conduct that prescribes a profile of friendly actions and participation unless there is evidence of deviations. Unlike the case with perfectly revealing signals, short-run players cannot rely only on signals z_2 to predict whether friendly actions are likely to be chosen. That is, they consider both public history, h^t, and private signals, z_2, when updating beliefs on player 1's type, i.e. $\mu(\theta)(h^t, z_2)$ for $\theta \in \Theta$.

Then, for any T and any $z_2 \notin \overline{Z}_2$, the probability of renewal $\tilde{\nu}$ (under any history of just bad signals $\overline{y}_t$) provides an upper bound on posteriors such that any myopic player participates in the game, that depends on parameters η and λ. As a result, it guarantees that players will not exit whenever informative signals point out adherence, notwithstanding the realisation of bad public signals. As is standard, while constructing equilibrium we focus on patient enough long-run players, implying that ν affects such player's behavior. In fact, the critical value $\tilde{\nu}$ was calculated so that myopic players do not exit with bad histories. It suffices then to restrict attention to stage-game incentives, therefore finding a parameterization of $\tilde{\eta}$ and $\tilde{\lambda}$ such that the long-run player chooses friendly actions. Because we separate the effects of renewal from the chances of being detected whenever player 1 deviates, the proposed code of conduct profile forms a Nash equilibrium.

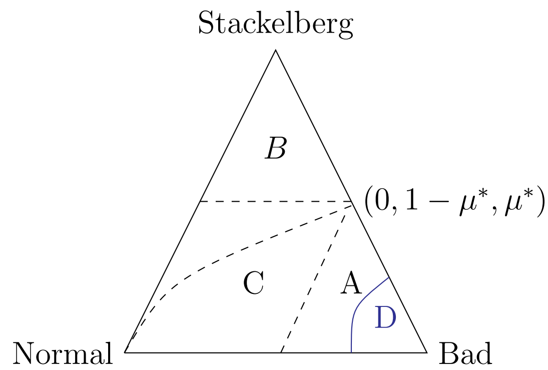

Figure 2.1: Space of prior distributions.

I end this section by illustrating the main result in [Ely et al.(2008)Ely, Fudenberg, and Levine], and relating them to those presented here. Figure 2.1 represents the set of possible prior probabilities on normal, bad and Stackelberg types. The upper triangle region (region B) depicts the cases where short-run players enter irrespective of the behavior of the normal type long-run player. In this case reputation is always good. On the other hand, region A represents the cases when the probability of the bad type is so high that none of short-run players participate. [Ely et al.(2008)Ely, Fudenberg, and Levine] show that in the region below the dashed curve (region C) the bad reputation effect also arises. Moreover, the curve asymptotically reaches the lower left vertex. This implies that for any arbitrary perturbation of the complete information case, the bad reputation effect occurs.

In contrast, with self-referentiality and renewal a reputation may not be bad for any epsilon neighborhood of the complete information case. From their result we cannot delimit where reputation is bad or good (between regions B and C), whereas the characterisation proposed here holds for the entire space of priors. More specifically, I have shown that the bad reputation effect occurs with self-referentiality (Region D), which is a subset of their region A. Finally, the higher the precision of the probability of detecting deviation from the code of conduct, the easier it is to overcome the bad reputation effect in the self-referential game.

2.5 Ely-Välimäki Example

In this section, I explore the implications of the previous results in the context of EV's example. The bad reputation game in EV consists of a long-run player, a mechanic, and a sequence of short-run players, the customers. There are two equally likely i.i.d. states of the world $\omega \in \{\mathcal{E}, \mathcal{T}\}$ that are not directly observable by customers and these states represent what type of repair the car needs: engine replacement $\mathcal{E}$, or tune-up $\mathcal{T}$. The action space of the mechanic is $A = \{ee, et, te, tt\}$ where te reads for tune-up in state $\mathcal{E}$ and engine replacement in state $\mathcal{T}$. The customer's action choices are to hire the mechanic In or to not hire the mechanic Out. The public outcomes are $Y = \{e, t, Out\}$ with distribution ρ described by $\rho(Out|(\cdot, Out)) = 1$, and the corresponding announcements $\rho(e|(et, In)) = \rho(e|(te, In)) = \frac{1}{2}$, $\rho(e|(ee, In)) = 1$ and $\rho(e|(tt, In)) = 0$.

If the mechanic performs the correct repair in each state and the costumer plays In, both receive a payoff of u. Otherwise, given participation, both get $-w$ with $w > u > 0$. Alternatively, if the customer plays Out both players get utility 0. I consider the finite horizon version of this game. The set of friendly and unfriendly actions are $F = \{et\}$ and $N = \{ee\}$, respectively. Where e is evidence for the set $\{ee\}$. The enforceable friendly action et is vulnerable relative to e and the temptation action is tt.

Suppose that there are two types, a normal type with payoffs specified above and a commitment bad type that always plays ee. EV show that the normal long-run player's payoffs vanish to zero as $\delta \to 1$. Roughly, the myopic players' posterior probability that the mechanic is of the bad type is increasing in realizations of the signal e. Then, there exists some threshold on the number of bad signals e such that the myopic player decides to exit the game, whereby the mechanic would choose the action tt if he were to delay this critical period. In contrast to EV, in this paper the mechanic's type may change over time and all commitment types are possible, i.e. $\Theta = \{\theta_0, \theta_{ee}, \theta_{et}, \theta_{te}, \theta_{tt}\}$. Observe that renewal would not mitigate the bad reputation effect because if the long-lived player mimics the Stackelberg type, θ_{et}, the resulting empirical distribution is the same as if he were a commitment type θ_{te}.

Consider the self-referential game that consists of the mechanic's set of signals $Z_1 = \{c, nc\}$ and customer's $Z_2 = \{m, g\}$. Denote $\mathbf{m}$ the $T \times 1$ vector with all entries filled by m. Starting with perfect revealing signals, let the distribution π be characterized as follows: $\pi((nc, \mathbf{m})|\tilde{r}^i, r^{-i}) = 1$ and $\pi((nc, \mathbf{m})|r) = 0$ for all players i. The next result states that the good equilibrium *can* be recovered in EV example and is an immediate consequence of Theorem 2.3.1.

Proposition 2.5.1. *Suppose there are normal and unfriendly commitment types for the mechanic, i.e.* $\Theta = \{\theta_0, \theta_{ee}\}$*, then the mechanic gets a normalized discounted payoff of* u *in the Nash equilibrium of the self-referential game.*

As mentioned above, when signals are perfectly informative in the equilibrium of the self-referential game the long-run player does not build reputation. Therefore, the normal mechanic need not choose a temptation action while pursuing to separate from the bad type because customers can distinguish his type without error. This proposition suggests that a normal mechanic would prefer to adhere to a code of conduct under which he chooses the friendly action *et* if the institutional environment in which the interaction with customers happens is sufficiently transparent.

In the imperfect detection case, to clarify the role of renewal I confine attention to the stage-game for two reasons. First, it simplifies the details of the proof without computing the required probability of renewal as in Theorem 2.4.1. Second, the constructed code of conduct profile would highlight the connection between noisy signals and the replacement of the mechanic (and more general of the long-run player, see Section 2.6).

Next, assume that $\pi((nc, \mathbf{m})|r) = p$ and $\pi((nc, \mathbf{m})|\tilde{r}^i, r^{-i}) = q$ with $q \geq p$ for all players i. It says that the profile of signals $(c, \mathbf{g})$ is more likely to be observed if all players adhere to the same code of conduct. Suppose that there is also a Stackelberg type, θ_{et}. Let μ^* be the prior probability of unfriendly type that induces entry of short-run players whenever there are only Stackelberg and bad types in the version of the game without self-referentiality.

Proposition 2.5.2. *Assume that $\Theta = \{\theta_0, \theta_{ee}, \theta_{et}\}$ in the one-shot version of the game. Then, there exists a self-referential equilibrium code of conduct profile r such that $\mu^* \leq \mu^*(g)$ and the mechanic is hired.*

The reason why the requirement on the prior of the bad type is relaxed is as follows. In equilibrium, signal g is associated with the possibility of friendly action et, as long as customers hire the mechanic. Therefore, the realization of g may identify adherence both by the normal and by the Stackelberg type. At the same time, the bad type θ_{ee} would not adhere to the proposed code of conduct. As a result, the fact that the normal type is pooling with the Stackelberg type makes the upper bound on the customers' beliefs less stringent. This code of conduct vector, on the other hand, cannot be a Nash equilibrium in the self-referential repeated game (see Section 2.6 for a further discussion).

2.6 Bad Reputation Games without Renewal

Now, I turn to the discussion about the assumption that the long-run player can be renewed given by the probability ν. Suppose that detecting deviations from the code of conduct profile is imperfect, η-λ permit detection, and that there is no renewal, $\nu = 0$, I will show that the bad reputation effect persists for an arbitrary long horizon.

As the number of signals $\overline{y}$ grows, myopic players update their beliefs regarding the long-run player's type, becoming virtually convinced that they are facing a bad type. With informative signals from the self-referential framework, such histories would be undermined only by a minuscule portion, eventually, these signals shake beliefs just not enough to make short-run players participate in the game. Because I am interested in arbitrary long horizon T, with probability one in any equilibrium of the self-referential game the short-run players would exit, thereby unraveling the possibility that the long-run player gets an expected payoff greater than his minmax payoff.

More specifically, at the beginning of each period t, the short-run player combines information contained in the public history, h^t, and information about player 1's intentions of play, $z_2 \in Z_2$, to infer the likelihood of friendly actions $f \in F$. For an arbitrary long horizon T, a history full of bad signals $\overline{h}^t = (\overline{y}_1, \dots, \overline{y}_t)$ would make short-run players exit the game because self-referential signals modify posterior beliefs by a fixed boost while not overturning the incentives for the long-run player to separate from the bad types. In addition, the assumption of replacement is crucial for short-run players updating beliefs so as the following proposition focuses on short-run players' behavior.

Proposition 2.6.1. *Consider the code of conduct profile constructed in Theorem 2.4.1. Then,* *(a) For $z_2 \notin \overline{Z}_2, h^t \in H$, we have $\mu(h^t, z_2)[\Theta(N)] < \mu(h^t)[\Theta(N)]$;*

(b) For $\overline{z}_2 \in \overline{Z}_2, h^t \in H$, $\mu(h^t, \overline{z}_2)[\Theta(N)] > \mu(h^t)[\Theta(N)]$;

(c) Assume an arbitrary long horizon T, there exists a number k^* of bad signals $\overline{y}$ such that short-run players do not participate in the game.

When players observe noisy signals, the posterior probability $\mu(h^t, z_2)$ weighs a bad signal $\overline{y}_t$ as evidence of an unfriendly commitment type, $\theta \in \Theta(N)$. The code of conduct vector reduces the weight on such types whenever that realization is complemented with a signal $z_2 \notin \overline{Z}_2$, thereby requiring a higher number of bad signals to convince short-run players to exit. Conversely, private signals in $z_2 \in \overline{Z}_2$ jointly with a history $\overline{h}^t$ imply a greater posterior probability. Finally, part (c) points out a potential problem with pursing to sustain friendly actions by using the code of conduct profile used in Proposition 2.5.2. Since short-run players increase the posterior probability of unfriendly commitment types with each observation of a bad signal, eventually $\mu(h^t, z_2)[\Theta(N)] > 1 - \gamma$ for any $z_2 \in Z_2$ and any h^t with k^* signals in $\overline{Y}$.

2.7 Conclusion

I have presented a model of bad reputation games with impermanent types and agents' ability to detect intentions. The results suggest that the perverse effect of reputation is attenuated in this class of games. In addition, I have identified conditions on the likelihood of commitment types for such results to hold. In that sense this characterisation is more complete than previous research. As mentioned, there is experimental evidence showing that reputation is not as bad as the original model suggests it should be. Here, I have proposed the idea that complementarity between the possibility of indirect observation of opponents' intentions and frequency of renewal are forces behind the apparently troubling general result. Both ideas are used to reconcile predictions from bad reputation models and the existence of markets with those features. It is also important that the results hold for arbitrarily long finite horizon games.

2.8 Appendix: Proofs

Definitions. For each exit action $e \in E$, the probability distribution over public signals satisfy these two conditions: $\rho(y|a, e) = \rho(y|e)$ for all $a \in A, y \in Y$, and $\rho(Y^E|e) = 1$. If short-run player chooses any entry action $b \notin E$, then $\rho(Y^E|a, b) = 0$ for all $a \in A$. A set of signals $\overline{Y}$ is evidence for a set of actions N if N is non-empty and $\rho(\overline{y}|n, b) > \rho(\overline{y}|a, b)$ for all $b \notin E, \overline{y} \in \overline{Y}, n \in N$ and $a \notin N$. An action a is vulnerable to temptation relative to a set of signals $\overline{Y}$ if there exist numbers $\underline{\rho}, \tilde{\rho} > 0$ and an action d such that (i) If $b \notin E, \overline{y} \in \overline{Y}$, then $\rho(\overline{y}|d, b) \geq \rho(\overline{y}|a, b) - \underline{\rho}$; (ii) If $b \notin E$ and $y \notin \overline{Y} \cup Y^E$ then $\rho(y|d, b) \geq (1 + \tilde{\rho})\rho(y|a, b)$; and (iii) For all $b \in E, u_1(d, b) \geq u_1(a, b)$. Define the effective discount factor $\tilde{\delta} = (1 - \nu)\delta$. Let the constant κ be interpreted as a measure of how revealing the evidence is, and defined by $\kappa = \min_{n \in N, a \notin N, \beta\{E\} < 1, \overline{y} \in \overline{Y}} \frac{\rho(\overline{y}|n,\beta)}{\rho(\overline{y}|a,\beta)}$. Note that κ is finite with $\kappa > 1$. Let $\varphi > 0$ be the minimum of the temptation bounds $\underline{\rho}$ and finally the signal lag given by $\eta = -\log(\gamma\varphi) \backslash \log \kappa$.

Proof of Theorem 2.3.1. Let us define the minmax payoff $\underline{u}_2$ for short-run player 2 given by $\underline{u}_2 = \min_{\alpha \in \mathcal{A}} \max_{b \in B} u_2(\alpha, b)$, and let α^2 be the long-lived player's strategy that minimizes the short-run player in the stage-game. Pick the profile of code of conduct $r \in \mathcal{R}$ such that for all players $i = 1, 2$, $r^i \in R_0$ prescribes:

$$r_1^i(z_1) := \begin{cases} f^* & \text{for all } z_1 \in Z_1 \setminus \overline{Z}_1, \\ \alpha^2 & \text{otherwise,} \end{cases} \quad \text{and} \quad r_2^i(z_2) := \begin{cases} \beta \in \mathbf{B}(f^*) & \text{for all } z_2 \in Z_2 \setminus \overline{Z}_2, \\ e \in E & \text{otherwise.} \end{cases}$$

With some abuse of notation, we denote by f^* the strategy for long-run player which prescribes the play of friendly action f^* every period, similarly, for strategy α^2. It remains to show that this profile of codes of conduct $r \in \mathcal{R}$ constitutes a Nash equilibrium in the self-referential game. Note that the long-run player gets an expected payoff of $u_1(f^*, \beta)$ which is the most he can get by playing his Stackelberg friendly action, any other action will cause the short-run player to exit game so he does not have incentives to deviate from this code. Similarly for short-run player, by adhering to this code he expects a friendly action to be played and avoids being minmaxed by the long-lived player. □

Proof of Theorem 2.4.1. For any positive probability public history $h^t \in H^t$, let $\mu(h^t)[\Theta_0]$ be the posterior beliefs over types $\Theta_0 \subseteq \Theta$. For player 2 and any private signal $z_2 \in Z_2$, let $\mu(h^t, z_2)[\Theta(N)]$ be the posterior beliefs on unfriendly types after incorporating the information of the self-referential game using Bayes' rule, and let $\mu_0(z_2)[\Theta_0]$ be the posterior beliefs for the null history h^0. We write $\tilde{\mu}(h^t, z_2)[\Theta(N)]$ for the posterior beliefs at the beginning of period $t+1$ taking into account private signals and the probability of renewal of the long-lived player, and its formal expression can be found below. We next construct the profile of code of conduct $r \in \mathcal{R}$ with $r^i \in R_0$ for all players i such that for the long-run player

$$r_1^i(z_1) := \begin{cases} f \in F & \text{for all } z_1 \in Z_1 \setminus \overline{Z}_1, \\ a \notin F & \text{otherwise.} \end{cases}$$

Same disclaimer about notation as in the previous proof applies here. For short-run players we have

$$r_2^i(z_2) \coloneqq \begin{cases} \beta \in \mathbf{B}(f^*) & \text{if } \tilde{\mu}(h^t, z_2)[\Theta(N)] < 1-\gamma \text{ for all } z_2 \in Z_2 \setminus \overline{Z}_2, \\ e \in E & \text{otherwise.} \end{cases}$$

Suppose that all players i adhere to the proposed code of conduct $r^i \in R_0$. Consider any positive probability history $h^t \in H^t$, and assume $z_2 \in Z_2 \setminus \overline{Z}_2$. Then the posterior beliefs of short-run player on the unfriendly commitment types at the beginning of period t can be written as

$$\tilde{\mu}(h^t, z_2)[\Theta(N)] = \nu\mu_0(z_2)[\Theta(N)] + (1-\nu)\mu(h^t, z_2)[\Theta(N)]$$

The posterior beliefs is a linear combination of two components. The first component takes into account the possibility of renewal of the long-run player. The second component combines the information in the public history up to period t and the information in the private signal. We define the constant $\Lambda = (1-\lambda)/(1-(\lambda+\eta))$ for natational convenience, and note that $\Lambda > 1$ as $\eta > 0$. By Bayes' rule we obtain

$$\begin{aligned} \tilde{\mu}(h^t, z_2)[\Theta(N)] = & \frac{\nu(1-\pi_2(\overline{z}_2|\tilde{r}^1, r^{-1}))\mu_0[\Theta(N)]}{(1-\pi_2(\overline{z}_2|\tilde{r}^1, r^{-1}))\mu_0[\Theta(N)] + (1-\pi_2(\overline{z}_2|r))(1-\mu_0[\Theta(N)])} \\ & + \frac{(1-\nu)(1-\pi_2(\overline{z}_2|\tilde{r}^1, r^{-1}))\mu(h^t)[\Theta(N)]}{(1-\pi_2(\overline{z}_2|\tilde{r}^1, r^{-1}))\mu(h^t)[\Theta(N)] + (1-\pi_2(\overline{z}_2|r))(1-\mu(h^t)[\Theta(N)])} \end{aligned}$$

Observe that all unfriendly commitment types would be violating the code of conduct r. Pick the history $\hat{h}^T \in H$ where all signals $y_t \in \hat{h}^T$ belong to the set of bad signals $\overline{Y}$. Since the self-referential game has uniformly friendly commitment size $1-\gamma, \chi$ for some constant

$\chi > 0$, and probability of replacement is ν the posterior beliefs then can be bounded by

$$\tilde{\mu}(\hat{h}^T, z_2)[\Theta(N)] \leq \frac{\nu[1-(\gamma+\chi)]}{1-(\gamma+\chi)(1-\Lambda)} + \frac{(1-\nu)[1-\nu(\gamma+\chi)]}{1-\nu(\gamma+\chi)(1-\Lambda)}$$

By adhering to the code r, short-run players do not exit the game as long as the posterior beliefs satisfy $\tilde{\mu}(\hat{h}^T, z_2)[\Theta(N)] < 1-\gamma$. Using the last expression, this is equivalent to

$$-\nu^2(\gamma+\chi)\Lambda + \nu(\gamma+\chi)b - \gamma(1-(\gamma+\chi)(1-\Lambda)) > 0$$

where the constant b is given by $b \equiv (\gamma+\chi)[\Lambda-(\gamma+\chi)(1-\Lambda)(\Lambda(1-\gamma)+\gamma)+\gamma+\Lambda(1-\gamma)]$. Thus, from the second-order polynomial we obtain

$$\tilde{\nu} = \frac{-b-\sqrt{b^2-4\Lambda\gamma(1-(\gamma+\chi)(1-\Lambda))}}{-2(\gamma+\chi)\Lambda}$$

For all probabilities of renewal ν with $\nu \geq \tilde{\nu}$ we have $\tilde{\mu}(h^t, z_2)[\Theta(N)] < 1-\gamma$ for any history $h^t \in H$ that guarantees short-run player would participate in the game. Observe that since all short-run players draw the same signal about long-lived player's intentions, his incentives to follow the code are driven at a stage-game level. Suppose that all short-run player adhere to the proposed code of conduct. Let $M = \max_{i\in\mathcal{I}} u_i$ and $m = \min_{i\in\mathcal{I}} u_i$ be the highest and lowest payoffs in the stage-game. If the normal long-run player adheres to the code of conduct r he obtains at least

$$u_1(f^*,\beta) + (1-(1-\lambda)^2)(u_1(a,\beta)-u_1(f^*,\beta)-(M-m)) + \pi_2(\overline{z}_2|r)(u_1(f^*,e)-u_1(f^*,\beta))$$

On the other hand, if he optimally deviates and chooses the code of conduct $\tilde{r}^1$ in which he plays some action $\tilde{a} \in A$ he gets at most

$$u_1(\tilde{a}, \beta) + (1-(1-\lambda)^2)(u_1(\tilde{a}, \beta) - u_1(a, \beta) + M - m)$$
$$+ (\pi_2(\overline{z}_2|r) + \eta)(u_1(f^*, e) - u_1(f^*, \beta) + u_1(a, \beta) - u_1(\tilde{a}, \beta))$$

Then the gain to deviation can be bounded above by

$$\begin{aligned} &u_1(\tilde{a}, \beta) - u_1(f^*, \beta) + (1-(1-\lambda)^2)(u_1(\tilde{a}, \beta) - 2u_1(a, \beta) + u_1(f^*, \beta) + 2(M-m)) \\ &\quad + (\pi_2(\overline{z}_2|r) + \eta)(u_1(f^*, e) - u_1(f^*, \beta) + u_1(a, \beta) - u_1(\tilde{a}, \beta)) + \pi_2(\overline{z}_2|r)(u_1(f^*, \beta) - u_1(f^*, e)) \\ &\quad \leq u_1(\tilde{a}, \beta) - u_1(f^*, \beta) + 2\lambda(u_1(\tilde{a}, \beta) - 2u_1(a, \beta) + u_1(f^*, \beta) + 2(M-m)) + \eta\mathcal{C}_1 \end{aligned}$$

where $\mathcal{C}_1 = u_1(f^*, e) - u_1(f^*, \beta) + u_1(a, \beta) - u_1(\tilde{a}, \beta)$. Adherence to the code of conduct r by the normal long-run player requires

$$\eta > \frac{u_1(\tilde{a}, \beta) - u_1(f^*, \beta) + 2\lambda\mathcal{C}_2}{\mathcal{C}_1}$$

we have defined constant $\mathcal{C}_2 = u_1(\tilde{a}, \beta) - 2u_1(a, \beta) + u_1(f^*, \beta) + 2(M-m)$. This shows that the code of conduct r is a self-referential Nash equilibrium. □

Proof of Proposition 2.5.2. If $\Theta = \{\theta_{ee}, \theta_{et}\}$, in Ely-Välimäki example without self-referentiality the prior probability of bad type is bounded above by $\mu^* \leq 2u/w + u$. Suppose now we look at the self-referential game. For the case with $\Theta = \{\theta_0, \theta_{ee}, \theta_{et}\}$, the code of conduct profile r states that the long-run player chooses et for c, unless he observes nc in which case he chooses ee. In addition, player 2 may choose In if he receives the signal g and stays out Out if b is realized. The prior probability of bad type required for having short-run player participating is given by $\mu^*(g) \leq (1-p)2u/(w+u-q(w-u)-p2u)$. It follows that $\mu^*(g) \geq \mu^*$ as $q \geq p$. We need to find the parameterization for p and q so that this code of conduct constitutes a Nash equilibrium. □

Proof of Proposition 2.6.1. We begin with part (a) of the Proposition. Suppose that the short-run players observe only signals $z_2 \in Z_2 \setminus \overline{Z}_2$. Recall that

$$\begin{aligned}\mu(h^t, z_2)[\Theta(N)] &= \frac{(1-\pi_2(\overline{z}_2|\tilde{r}^1, r^{-1}))\mu(h^t)[\Theta(N)]}{(1-\pi_2(\overline{z}_2|\tilde{r}^1, r^{-1}))\mu(h^t)[\Theta(N)] + (1-\pi_2(\overline{z}_2|r))(1-\mu(h^t)[\Theta(N)])} \\ &\geq \frac{1}{\mu(h^t)[\Theta(N)] + \Lambda(1-\mu(h^t)[\Theta(N)])}\mu(h^t)[\Theta(N)]\end{aligned}$$

Observe that $\mu(h^t, z_2)[\Theta(N)] < \mu(h^t)[\Theta(N)]$ as $\Lambda > 1$. Suppose that for some $\epsilon > 0$, $\mu(h^t)[\Theta(N)] = 1-\gamma+\epsilon$ and $\mu(h^{t-1})[\Theta(N)] < 1-\gamma$. Thus

$$\mu(h^t, z_2)[\Theta(N)] \geq \frac{1}{1-\gamma+\Lambda\gamma-\epsilon(\Lambda-1)}\mu(h^t)[\Theta(N)]$$

Next, we borrow from Lemma 2 in [Ely et al.(2008)Ely, Fudenberg, and Levine] the following lower bound on the posterior probability of unfriendly commitment types[10]

$$\mu(h^t)[\Theta(N)] \geq \left(\frac{1}{1-\gamma+\frac{\gamma}{\kappa}}\right)^k \mu_0[\Theta(N)]$$

This shows that the lower bound for $\mu(h^t, z_2)[\Theta(N)]$ is below the lower bound found in the last expression (without self-referentiality) which implies that signals in $Z_2 \setminus \overline{Z}_2$ allow for a greater number of bad signals $\overline{Y}$ given the same history.

[10] To find the number k they use the argument in [Fudenberg and Levine(1989), Fudenberg and Levine(1992)].

Suppose short-lived players' private signals $z_2 \in \overline{Z}_2$. By similar arguments, it must be the case that $\mu(h^t, \overline{z}_2)[\Theta(N)] > \mu(h^t)[\Theta(N)]$. That is, for a given number of bad signals, observation of private signals in $\overline{Z}_2$ pushes the posterior probability upward relatively more to the case without code of conduct. Finally, note that the updating beliefs formula with self-referential signals is characterized by the factor $\Upsilon \equiv 1/\mu[\Theta(N)] + \Lambda(1 - \mu([\Theta(N)])$, from which we stress $\Upsilon \to 1$ as $\mu \to 1$. $\square$

Chapter 3

Timing and Codes of Conduct

3.1 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(1987)]'s 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.

3.1.1 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 et al.(2010)Kalai, Kalai, Le 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)Kalai, Kalai, Lehrer, and Samet]'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.

[11] [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)Kalai, Kalai, Lehrer, and Samet] characterize the conditional commitment devices space. In self-referential games, this space is assumed to be common and very general.

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 et al.(2013)Calcagno, Kamada, Lovo, and Sugaya] show that revision games narrow 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 3.2, I present the framework of the base and the self-referential game, and the information technology. Section 3.3 studies self-referential games in which recognition occurs only in advance. In section 3.4, I analyze finitely repeated games and characterize the equilibrium set in terms of the timing of signals. In Section 3.5, I examine exit games showing the implications of the timing on equilibrium behavior. Section 3.6 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 3.7. The Appendix collects the proofs.

3.2 The Model

I next outline the general framework. In section 3.2.1 I describe multistage games with observed actions. Section 3.2.2 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.

3.2.1 Setup and Notation

I concentrate on multistage games with observed actions as defined in [Fudenberg and Tirole(1991)].[12] There are a set of players $\mathcal{I}$ with cardinality $|\mathcal{I}| = N$ and $T+1$ stages.[13]

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

[12] 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 et al.(1996)Greenberg, Monderer, and Shitovitz]. 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.

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

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 $H \coloneqq \bigcup_{t=0}^{\infty} H^t$ 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 $\mathcal{S}_i^t \coloneqq \Delta A_i(h^t)$ for notational convenience. Let $\Delta A(h^t)$ 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.

3.2.2 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 $|Y_i| \geq 2$.[14] 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, $s_i : H \times H_i \rightarrow \mathcal{S}_i^t$. Let $s \in S$ be a profile of extended strategies.[15]

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, $r_j^i : H \times H_j \rightarrow \mathcal{S}_j^t$. 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

[14] 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.

[15] 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.[16] 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

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

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.[17] Note also that it gives the identity of the deviator but not the magnitude of the deviation.

3.3 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.

[17] 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(2005)]'s 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.

Let Γ^∞ be an infinite horizon multistage game with observed actions described in Section 3.2. 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.[18] 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 3.3.1. *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 $\tau \to \infty$.*

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

See Appendix 3.8.1 for the proof.

The literature on the connection between infinite and finite horizon games has focused on perfect ε equilibria of the finite truncation Γ^{τ}.[19] In contrast, Theorem 3.3.1 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 3.3.1 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.

[19] See, for instance, [Radner(1981)], [Fudenberg and Levine(1986)], [Harris(1985a), Harris(1985b)] and [Börgers(1989), Börgers(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 et al.(2005)Mailath, Postlewaite, and Samuelson]. 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.

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.[20]

3.4 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 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.

3.4.1 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

[20] 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 $\lambda = 0$.

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

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

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

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

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

$$U := \{(v_1, \ldots, v_N) : \exists a \in A, \forall i, g_i(a) = v_i\},$$

$$V := co(U),$$

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

V is the set of feasible payoff vectors and V^* is the set of feasible and strictly individually rational payoff vectors.[21] 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.

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

Theorem 3.4.1. *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 $v \in V^*$ 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 3.8.2 along with the rest of the proofs corresponding to the results in this section.[22]

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)Kalai, Kalai, Lehrer, and Sar
In their setting, players observe the choice of conditional commitment devices and use jointly

[22] 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(2007)]'s result.

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)Kalai, Kalai, Lehrer, and Samet], Theorem 3.4.1 is approximate since recognition is correct only probabilistically on the equilibrium path.

3.4.2 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 $A^t := A^0 \times \ldots \times A^{t-1}$ 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.[23] Players discount future with common discount factor $\delta \in (0,1)$. 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

[23] 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.

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 3.4.2. *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)].[24] 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 3.4.2 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 3.4.2. 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

[24] 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)].

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 ($\eta \nrightarrow 0$) in the asymptotic limit of the time horizon $T \to \infty$.

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 3.3 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 3.3.1, let us assume that $\lambda = 0$.[25] Indeed, I will show that any $v \in V^*$ is attainable in the 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 3.4.2.

Proposition 3.4.1. *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 $\eta_{T-k} \leq \eta_T$.*

[25] 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.

The detection probability threshold η_{T-k} is relaxed relative to the threshold η_T found in Theorem 3.4.2 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 $\breve{r}^T$ 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 $\breve{r}^T$ be the self-referential equilibrium profile from Proposition 3.4.1. The approximate expected payoff of each player i is given by*

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

This is a corollary to Proposition 3.4.1. 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 3.4.2.

The findings of Theorem 3.3.1 and Proposition 3.4.1 can be combined to compute the speed of convergence of η_T. Recall that Theorem 3.3.1 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 3.4.1 and the corollary below follows this proposition.[26]

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* $O(\delta^T)$*.*

Proof. By Proposition 3.4.1, 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.

3.5 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*.[27] 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.

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

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

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 $\mathcal{F}_i(h^t)$ 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.

3.5.1 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 3.5.1. *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 3.8.3.[28]

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 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

[28]The reader may also find all the proofs of the results in this section in Appendix 3.8.3.

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 3.5.2. *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 3.5.1, 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 3.5.1. *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.

3.5.2 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

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

where $\mathbb{1}_{\{.\}}$ denotes the indicator function.[29] 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

[29] I write $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$.

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'.[30]

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.

[30]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.

Theorem 3.5.3. *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 3.5.1 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 3.5.4. *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 3.5.4. 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 3.5.4 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 $h_j^t \notin \overline{H}_i$. □

3.6 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 3.2.2. 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.

3.6.1 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 3.6.1. *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 3.8.4.[31]

[31] The rest of the proofs related to this section are also in Appendix 3.8.4.

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 3.4.2.

Recall that Theorem 3.4.2 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.

3.6.2 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.

Proposition 3.6.2. *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 3.5.2, especially, Theorem 3.5.4 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.

3.6.3 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 3.5.3 obtained for synchronous monitoring structure.

Proposition 3.6.3. *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 3.6.4. *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 3.6.4 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 $r^{\iota}_{\iota}(h^t, h^t_{\iota}) = e^t_{\iota}$ for some $e^k_{\iota} \in \mathcal{E}^k_{\iota}$ and all $h^k_{\iota} \notin \overline{H}_{\iota}$.*

3.7 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 pre-game 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.

3.8 Proofs

3.8.1 Proof of Theorem 3.3.1

Before proving Theorem 3.3.1, 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 $\lim_{\tau \to \infty} \zeta^{\tau} = 0$.[32] 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 $BR_i(\sigma_{-i}) = \text{argmax}_{\sigma_i \in \mathcal{E}_i} u_i(\sigma_i, \sigma_{-i})$ 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

[32] 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 $i \in \mathcal{I}$.

$\overline{\sigma}$ 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.[33]

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 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

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

[33] 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 $h \in H$.

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

$$\begin{aligned} u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\hat{\sigma}_{\tau}) &\leq u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\hat{\sigma}_{\tau}) + u_i(\hat{\sigma}) - u_i(\sigma_i, \hat{\sigma}_{-i}) \\ &\leq 2\zeta^{\tau} \end{aligned} \tag{3.1}$$

and working similarly on the denominator we find the bound

$$\begin{aligned} u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\sigma_{\tau}, \underline{\sigma}_{-i,\tau}) &\leq u_i(\sigma_{i,\tau}, \hat{\sigma}_{-i,\tau}) - u_i(\sigma_{\tau}, \underline{\sigma}_{-i,\tau}) + u_i(\hat{\sigma}) - u_i(\sigma_i, \hat{\sigma}_{-i}) \\ &\leq 2\zeta^{\tau} + u_i(\hat{\sigma}_{\tau}) - u_i(\sigma_{\tau}, \underline{\sigma}_{-i,\tau}) \end{aligned} \tag{3.2}$$

Hence, combining expressions (3.1) and (3.2) 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 $\eta_{\tau} < \varepsilon$. □

3.8.2 Proofs for Section 3.4

Proof of Theorem 3.4.1. 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 $i \in \mathcal{I}$.[34] 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

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

$a_i \in \text{argmax}_{\tilde{a}_i \in A_i} g_i(\tilde{a}_i, a_j^*) \geq g_i(a^*)$, and for the rest of the players $j \neq i$ it states $\tilde{r}_j^i = \hat{r}_j^i$. Given that $j \in \mathcal{I} \setminus \{i\}$ adhere to the code of conduct $\hat{r}$, then the highest expected payoff for player i, denoted by $\overline{W}_i(\tilde{r}^i)$ is

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

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. $\square$

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 3.4.2. From Theorem 3.3.1, 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 $a^* \in A$.[35] 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 $i, j \in \mathcal{I}$, $\hat{r}_j^i(h^t, h_j^t) = a_j^*$ for all histories $h^t \in H, h_j^t \in H_j^t$. 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

[35]The same argument applies to mixed strategies.

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 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 3.4.1. 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 3.4.2 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 3.4.2. (I omit the calculation of the probability of detection η_T in the analogous result to Theorem 3.4.2 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 $U_i(\breve{r}) = v_i$. □

3.8.3 Proofs for Section 3.5

Proof of Theorem 3.5.1. 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 3.4.1, 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 3.5.2. From Theorem 3.5.1, for any r such that $r_i^i(\dot{}) = e_i^t$ 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 3.5.1, 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 3.5.3. 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 $e_{\iota(h^t)}^t$ and $f_{\iota(h^t)}^k$ 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 3.5.4. 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 $\eta_k := \max_i \eta_{i,k}$ and $\lambda_k := \max_i \lambda_{i,k}$, the profile $\hat{r}$ forms a self-referential equilibrium. □

3.8.4 Proofs for Section 3.6

Proof of Proposition 3.6.1. 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 3.4.2, it follows r is a self-referential equilibrium here as well. □

Proof of Proposition 3.6.2. Suppose that $\tau_1, \tau_2 \leq T$, $\tau_1 \neq \tau_2$ and say $\tau_1 < \tau_2$ provided by (Y, π).[36] 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 ends in period k, player 2 receives her signal too late to materialise any punishment. Let r^i be the code of conduct $\forall i, j$

$$r_j^i(h^t, h_j^t) := \begin{cases} s_j^t(h^t, h_j^t) & \text{if } h_j^t \notin \overline{H}_j^t, \text{ for all } t \geq 0, \\ \underline{s}_j^t(h^t, h_j^t) & \text{if } h_j^t \in \overline{H}_j^t, \text{ for all } t \geq 0. \end{cases}$$

[36]The argument is invariant to permutation.

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 $t \leq k$, $s_i(h^t, h_i^t) = f_i^t$ 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 $t \geq k$ we have $s_i(h^t, h_i^t) = e_i^t$ for any $h^t \in H$, $h_i^t \notin \overline{H}_i$ and $e_i^t \in \mathcal{E}_i^t$. For $\underline{s}_i \in S_i$, for all $t < \tau_2$, $\underline{s}_i(h^t, h_i^t) = f_i^t$ for $h^t \in H$, $h_i^t \in \overline{H}_i$ and $f_i^t \in \mathcal{F}_i^t$, and for all $t \geq \tau_2$ it says $\underline{s}_i(h^t, h_i^t) = e_i^t$ for $h^t \in H$, $h_i^t \in \overline{H}_i$ and $e_i^t \in \mathcal{E}_i^t$. Then, the lower bound in expected payoffs is given by

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

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 $s_1(h^t, h_1^t) = e_1^t$ for all $h^t \in H$, $h_1^t \in H_1$, $t \geq \tau_1$ and $e_1^t \in \mathcal{E}_1^t$, and for his opponent $\tilde{r}_2^1 = r_2^1$ 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 $s_2(h^t, h_2^t) = e_2^t$ for all $h^t \in H$, $h_2^t \in H_2$, $t \geq \tau_1 - 1$ and $e_2^t \in \mathcal{E}_2^t$, and for $t < \tau_1 - 1$ simply $s_2(h^t, h_2^t) = f_2^t$, for all $h^t \in H$, $h_2^t \in H_2$ and $f_2^t \in \mathcal{F}_2^t$. Finally for player 1 it requires $\hat{r}_1^2 = r_1^2$. 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 $r_j^i = s_j$ with $s_j(h^{\cdot} h_j^t) = e_j^t$ for all t, $h^t \in H$, $h_j^t \in H_j^t$ and any $e_j^t \in \mathcal{E}_j^t$. It follows that the only equilibrium outcome in the self-referential game exhibits all players exiting in period $t = 0$. □

Proof of Proposition 3.6.3. 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 $t \geq k$, $s_2(h^t, h^t_2) = f^t_2$ for all $h^t \in H$, $h^t_2 \in H_2$, and $f^t_2 \in \mathcal{F}^t_2$, while for $t \geq k$ it requires $s_2(h^t, h^t_2) = e^t_2$ for all $h^t \in H$, $h^t_2 \in H_2$, and $e^t_2 \in \mathcal{E}^t_2$. In addition, $\underline{s}_2 = s_2$. On the other hand, for player 1 it says for all $t \geq 0$, $s_1(h^t, h^t_1) = f^t_1$ for all $h^t \in H$, $h^t_1 \notin \overline{H}_1$, and $f^t_1 \in \mathcal{F}^t_1$, whereas for any $t \leq k_{-1}$, $\underline{s}_1(h^t, h^t_1) = f^t_1$, for all $h^t \in H$, $h^t_1 \in \overline{H}_1$, and $f^t_1 \in \mathcal{F}^t_1$; and for $t \geq k_{-1}$, $\underline{s}_1(h^t, h^t_1) = e^t_1$, for all $h^t \in H$, $h^t_1 \in \overline{H}_1$, and $e^t_1 \in \mathcal{E}^t_1$.

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 $\tilde{r}^1_2 = r^1_2$. For player 1, $\tilde{r}^1_1 = \tilde{s}_1$ so that for all $t \leq \phi(\overline{n}) - 1$, this strategy is $\tilde{s}_1(h^t, h^t_1) = f^t_1$, for all $h^t \in H$, $h^t_1 \in \overline{H}_1$, and $f^t_1 \in \mathcal{F}^t_1$. For all $t \geq \phi(\overline{n}) - 1$, $\tilde{s}_1(h^t, h^t_1) = e^t_1$, for all $h^t \in H$, $h^t_1 \in H_1$, and $e^t_1 \in \mathcal{E}^t_1$. This gives a lower bound of at least $g_1(e^{\phi(\overline{n})-1}_1, \overline{a}) > g_1(\overline{a}, e^k_2)$ 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 3.6.4. 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

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

Therefore, the strategy for player ι is for all $t < k$, $s^t_\iota(h^t, h^t_\iota) = f^t_\iota$ for all $h^t \in H$, $h^t_\iota \notin \overline{H}_\iota$, and $f^t_\iota \in \mathcal{F}^t_\iota$; and for all $t \geq k$, $s^t_\iota(h^t, h^t_\iota) = e^t_\iota$ for all $h^t \in H$, $h^t_\iota \notin \overline{H}_\iota$, and $e^t_\iota \in \mathcal{E}^t_\iota$. Further, the punishment strategy is such that for all $t < k_{-2}$, $\underline{s}^t_\iota(h^t, h^t_\iota) = f^t_\iota$ for all $h^t \in H$, $h^t_\iota \in \overline{H}_\iota$, and $f^t_\iota \in \mathcal{F}^t_\iota$, moreover, for all $t \geq k_{-2}$, $\underline{s}^t_\iota(h^t, h^t_\iota) = e^t_\iota$ for all $h^t \in H$, $h^t_\iota \in \overline{H}_\iota$, and $e^t_\iota \in \mathcal{E}^t_\iota$. It remains to state the strategies for player i with $i \neq \iota$. For player i, for each $t < k$, $s^t_i(h^t, h^t_i) = f^t_i$ 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 $t \geq k$, $s^t_i(h^t, h^t_i) = e^t_i$ for all $h^t \in H$, $h^t_i \notin \overline{H}_i$, and $e^t_i \in \mathcal{E}^t_i$. Similar to player ι but with different timing the punishment strategy is characterized by: for all $t < k_{-1}$, $\underline{s}^t_i(h^t, h^t_i) = f^t_i$ for all $h^t \in H$, $h^t_i \in \overline{H}_i$, and $f^t_i \in \mathcal{F}^t_i$, moreover, for all $t \geq k_{-1}$, $\underline{s}^t_i(h^t, h^t_i) = e^t_i$ for all $h^t \in H$, $h^t_i \in \overline{H}_i$, and $e^t_i \in \mathcal{E}^t_i$. By parameterizing η_k and λ_k as in Theorem 3.5.4, the proposed code of conduct profile forms a self-referential equilibrium. $\square$

References

[Anderlini and Felli(2001)] Luca Anderlini and Leonardo Felli. Costly bargaining and renegotiation. *Econometrica*, 69(2):377–411, 2001. ISSN 1468-0262.

[Aumann and Maschler(1995)] R.J. Aumann and M.B. Maschler. *Repeated Games With Incomplete Information.* The MIT Press, Cambridge, MA, 1995. ISBN 9780262011471.

[Bar-Isaac and Deb(2013)] Heski Bar-Isaac and Joyee Deb. (good and bad) reputation for a servant of two masters. Working paper, New York University, 2013.

[Benoît and Krishna(1985)] Jean-Pierre Benoît and Vijay Krishna. Finitely repeated games. *Econometrica*, 53(4):905–22, 1985.

[Block and Levine(2012)] Juan I Block and David K Levine. Codes of conduct, private information and repeated games. Working paper, Washington University in St. Louis, 2012.

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

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

[Calcagno et al.(2013)Calcagno, Kamada, Lovo, and Sugaya] Riccardo Calcagno, Yuichiro Kamada, Stefano Lovo, and Takuo Sugaya. Asynchronicity and coordination in common and opposing interest games. *Theoretical Economics*, 2013. forthcoming.

[Caruana and Einav(2008)] Guillermo Caruana and Liran Einav. A theory of endogenous commitment. *Review of Economic Studies*, 75(1):99–116, 2008.

[Chassang(2010)] Sylvain Chassang. Fear of miscoordination and the robustness of cooperation in dynamic global games with exit. *Econometrica*, 78(3):973–1006, 2010. ISSN 00129682.

[Crawford(2003)] Vincent P. Crawford. Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions. *American Economic Review*, 93(1):133–149, 2003. ISSN 00028282.

[Ekmekci et al.(2012)Ekmekci, Gossner, and Wilson] Mehmet Ekmekci, Olivier Gossner, and Andrea Wilson. Impermanent types and permanent reputations. *Journal of Economic Theory*, 147(1):162–178, 2012.

[Ely and Välimäki(2003)] Jeffrey C. Ely and Juuso Välimäki. Bad reputation. *The Quarterly Journal of Economics*, 118(3):785–814, 2003.

[Ely et al.(2008)Ely, Fudenberg, and Levine] Jeffrey C. Ely, Drew Fudenberg, and David K. Levine. When is reputation bad? *Games and Economic Behavior*, 63(2):498–526, 2008.

[Forges(2013)] Franoise Forges. A folk theorem for bayesian games with commitment. *Games and Economic Behavior*, 78(0):64–71, 2013. ISSN 0899-8256.

[Frank(1987)] Robert H. Frank. If homo economicus could choose his own utility function, would he want one with a conscience? *American Economic Review*, 77(4):593–604, 1987. ISSN 00028282.

[Frank(1988)] Robert H. Frank. *Passions Within Reason: The Strategic Role of the Emotions*. W. W. Norton and Company, 1988.

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

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

[Fudenberg and Levine(1989)] Drew Fudenberg and David K Levine. Reputation and equilibrium selection in games with a patient player. *Econometrica*, 57(4):759–78, 1989.

[Fudenberg and Levine(1991)] Drew Fudenberg and David K. Levine. An approximate folk theorem with imperfect private information. *Journal of Economic Theory*, 54(1):26–47, 1991.

[Fudenberg and Levine(1992)] Drew Fudenberg and David K Levine. Maintaining a reputation when strategies are imperfectly observed. *Review of Economic Studies*, 59(3): 561–79, 1992.

[Fudenberg and Levine(1986)] Drew Fudenberg and David Knudsen Levine. Limit games and limit equilibria. *Journal of Economic Theory*, 38(2):261–279, 1986.

[Fudenberg and Maskin(1986)] Drew Fudenberg and Eric Maskin. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54 (3):533–54, 1986.

[Fudenberg and Tirole(1991)] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press Books. The MIT Press, Cambridge, MA, 1991.

[Fudenberg et al.(1994)Fudenberg, Levine, and Maskin] Drew Fudenberg, David K Levine, and Eric Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1039, 1994.

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

[Greenberg et al.(1996)Greenberg, Monderer, and Shitovitz] Joseph Greenberg, Dov Monderer, and Benyamin Shitovitz. Multistage situations. *Econometrica*, 64(6):1415–1437, 1996. ISSN 00129682.

[Grosskopf and Sarin(2010)] Brit Grosskopf and Rajiv Sarin. Is reputation good or bad? an experiment. *The American Economic Review*, 100(5):2187–2204, 2010.

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

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

[Holmström(1999)] Bengt Holmström. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–82, 1999.

[Howard(1988)] John V Howard. Cooperation in the prisoner's dilemma. *Theory and Decision*, 24(3):203–213, 1988.

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

[Kalai et al.(2010)Kalai, Kalai, Lehrer, and Samet] Adam Tauman Kalai, Ehud Kalai, Ehud Lehrer, and Dov Samet. A commitment folk theorem. *Games and Economic Behavior*, 69(1):127–137, 2010.

[Kamada and Kandori(2011)] Yuichiro Kamada and Michihiro Kandori. Revision games. Mimeo, 2011.

[Levine and Pesendorfer(2007)] David K. Levine and Wolfgang Pesendorfer. The evolution of cooperation through imitation. *Games and Economic Behavior*, 58(2):293–315, 2007.

[Levine and Szentes(2006)] David K Levine and Balzs Szentes. Can A Turing Player Identify Itself? *Economics Bulletin*, 1(1):1–6, 2006.

[Mailath and Samuelson(2001)] George J Mailath and Larry Samuelson. Who wants a good reputation? *Review of Economic Studies*, 68(2):415–41, 2001.

[Mailath et al.(2005)Mailath, Postlewaite, and Samuelson] George J. Mailath, Andrew Postlewaite, and Larry Samuelson. Contemporaneous perfect epsilon-equilibria. *Games and Economic Behavior*, 53(1):126–140, 2005.

[Matsui(1989)] Akihiko Matsui. Information leakage forces cooperation. *Games and Economic Behavior*, 1(1):94–115, 1989. ISSN 0899-8256.

[Miyahara and Sekiguchi(2013)] Yasuyuki Miyahara and Tadashi Sekiguchi. Finitely repeated games with monitoring options. *Journal of Economic Theory*, 148(5):1929–1952, 2013. ISSN 0022-0531.

[Morris(2001)] Stephen Morris. Political correctness. *Journal of Political Economy*, 109(2): 231–265, 2001.

[Neyman(1999)] Abraham Neyman. Cooperation in repeated games when the number of stages is not commonly known. *Econometrica*, 67(1):45–64, 1999. ISSN 1468-0262.

[Ordoñez(2013)] Guillermo L. Ordoñez. Fragility of reputation and clustering of risk-taking. *Theoretical Economics*, 8(3):653–700, 2013. ISSN 1555-7561.

[Peters(2013)] Michael Peters. Reciprocal contracting. Mimeo, University of British Columbia, 2013.

[Peters and Szentes(2012)] Michael Peters and Balzs Szentes. Definable and contractible contracts. *Econometrica*, 80(1):363–411, 2012. ISSN 1468-0262.

[Peters and Troncoso-Valverde(2013)] Michael Peters and Cristin Troncoso-Valverde. A folk theorem for competing mechanisms. *Journal of Economic Theory*, 148(3):953–973, 2013. ISSN 0022-0531.

[Phelan(2006)] Christopher Phelan. Public trust and government betrayal. *Journal of Economic Theory*, 130(1):27–43, 2006.

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

[Radner(1981)] Roy Radner. Monitoring cooperative agreements in a repeated principal-agent relationship. *Econometrica*, 49(5):1127–1148, 1981. ISSN 00129682.

[Smith(1995)] Lones Smith. Necessary and sufficient conditions for the perfect finite horizon folk theorem. *Econometrica*, 63(2):425–430, 1995. ISSN 00129682.

[Solan and Yariv(2004)] Eilon Solan and Leeat Yariv. Games with espionage. *Games and Economic Behavior*, 47(1):172–199, 2004. ISSN 0899-8256.

[Sugaya(2011)] Takuo Sugaya. Folk theorem in repeated games with private monitoring. *Economic Theory Center Working Paper*, (011-2011), 2011.

[Tennenholtz(2004)] Moshe Tennenholtz. Program equilibrium. *Games and Economic Behavior*, 49(2):363–373, 2004.

[Tooby and Cosmides(1996)] J Tooby and L Cosmides. The computationalist theory of communication. In *meeting of the Human Behavior and Evolution Society Meetings, University of Arizona, Tucson*, 1996.

[Wiseman(2008)] Thomas Wiseman. Reputation and impermanent types. *Games and Economic Behavior*, 62(1):190–210, 2008.

[Wiseman(2009)] Thomas Wiseman. Reputation and exogenous private learning. *Journal of Economic Theory*, 144(3):1352–1357, 2009. ISSN 0022-0531.