Codes of Conduct, Private Information, and Repeated Games[1]

Juan I. Block and David K. Levine[2]

Department of Economics Washington University in St. Louis

First version: 11/14/10

This version: 8/13/12

Abstract: We examine self-referential games in which there is a chance of understanding an opponent's intentions. Our main focus is on the interaction of two sources of information about opponents' play: direct observation of the opponent's code-of-conduct, and indirect observation of the opponent's play in a repeated setting. Using both sources of information we are able to prove a "folk-like" theorem for repeated self-referential games with private information. This theorem holds even when both sources of information are weak.

[1] The idea of self-referential equilibrium was developed in an earlier collaboration with Wolfgang Pesendorfer and the ideas about approximate equilibrium and private information with Drew Fudenberg. I am especially grateful to them, as well as to Phillip Johnson, Balazs Szentes and George Mailath for the many discussions that motivated this work. I would also like to thank National Science Foundation Grant SES-0851315 and the European University Institute for financial support.

[2] Ph.D. Candidate, Washington University in St. Louis and John H. Biggs Distinguished Professor, Department of Economics, Washington University in St. Louis, http://www.dklevine.com, david@dklevine.com.

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.[3] 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, more appealing 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. Unfortunately it is exactly this type of small gain that approximate equilibrium constructions are based on. Similarly, in mechanism design, it is often possible to base mechanisms on having players report on each other. However, since equilibria must be crafted so that players are indifferent between their reports it is hard to understand why these mechanisms would be robust, for example, to situations where players might have social preferences.

It is also the case that the abstract world of repeated games and mechanism design 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 is a world that is would be surprising to 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 other's intentions – whether through facial expressions or skilled interrogation – was introduced in Levine and Pesendorfer [2007]. They were primarily interested in these 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

[3] See for example Sugaya [2011].

stability properties in an evolutionary setting. These questions were impractical to study in repeated games.

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.

The basic setup here is the Levine-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. In other words, codes of conduct both generate signals and respond to those same 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.

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 information is proven.

2. The Model

We consider an N person *base game* with players $i \in I = \{1,\ldots,N\}$. Player i has finitely many strategies $s_i \in S_i$. Note that we do not allow randomizations over S_i. That is not the same as saying that there are no mixed strategies, just that there are only a finite number of them: for example, only a six-sided dice is available. This assumption avoids inconvenient measure theoretic considerations. Since continuous measurement devices are an abstraction not available in practice it is empirically relevant, nor will we need the computational advantages of convexity for our results. 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. Again, this is empirically relevant, and as we will be able to establish finite folk-theorems, there is no reason for the additional mathematical complexity of allowing an infinite strategy space. We denote by $s \in S$ the corresponding profile of strategies. Utility of player i is given by $u_i(s)$ if the strategy profile s is chosen and with the usual notation $u_i(s_j, s_{-j})$.

The *self-referential game* is defined by a finite common abstract space of *codes of conduct* R_0 for each player, by a finite set of signals Y_i for each player, and a fixed exogenously given probability distribution $\pi(y \mid r)$ where $r = (r^1, \ldots, r^N) \in R$ is a profile of codes-of-conduct and $y \in Y$ is a profile of signals.[4] That is, $\pi(y \mid r)$ is a joint distribution over profiles of signals given the profile of codes-of-conduct. A code-of-conduct $r^i \in R_0$ for player i also induces a map $r^i_j : Y_j \to S_j$ for every player j. If every profile of maps $r^i_j : Y_j \to S_j$ is represented in R we say the code-of-conduct is *complete*. We will remark later why we may wish to allow sets of codes that are not complete. 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 receive about each others' intentions through the probability distribution. We illustrate these implications in the examples below.

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

$$U_i(r) = \sum_{y \in Y} u_i(r^1_1(y_1), \ldots, r^N_N(y_N))\pi(y \mid r).$$

A Nash equilibrium of the self-referential game is defined in the usual way. We do not consider refinements such as subgame perfection or sequentiality for several reasons. First, they are not robust (Fudenberg, Kreps and Levine [1988]) nor do they appear to be empirically relevant.[5] Second, we will focus our attention on strict Nash equilibria – which satisfy all the standard tie-breaking refinements including all versions

---

[4] This is only of interest if the signal structure has an element of richness (see below for specific assumptions). If, say, $Y = \{1\}$ then we cannot exploit self-referentiality as the signal structure is not informative.

[5] See for example Binmore et al [2002].

of perfection and sequentiality, and even much stronger ones such as divinity and strategic stability.

Self-referential games were introduced by Levine and Pesendorfer [2007] in the two person symmetric case. The key idea is that in an evolutionary setting it is a great advantage for a strategy to be able to “recognize itself.” That is: an evolutionary advantage will be derived by altruism towards strategies that are the same as your own, and spite towards those that are not. Before examining more closely the motivation for self-referential equilibrium, several examples illustrate the concept.

3. Examples

We will initially consider two player symmetric games with a very simple informational process and a common strategy space $S_1 = S_2$. The space of signals is $Y_i = \{0,1\}$ for each player, 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.” Specifically $\pi(y \mid r) = \pi_0(y_1 \mid r)\pi_0(y_2 \mid r)$ so that the signals are independent, and $\pi_0(1 \mid r) = p$ if $r^1 = r^2$ and $\pi_0(1 \mid r) = q \geq p$ if $r^1 \neq r^2$. In other words, if the two players employ different codes of conduct they are more likely to receive the signal 1 and they may base their play on whether or not this signal is received. The codes of conduct themselves we will take to be all pairs of maps (r^1, r^2) where $r^i : \{0,1\}^2 \rightarrow S_1^2$.

The specific example we will study is a simple prisoner’s dilemma game, possibly repeated. The actions in the stage game are denoted C for cooperate and D for defect, and the payoffs are given in the table below.

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

Example 1: The Prisoner’s Dilemma

One equilibrium code-of-conduct is simply to ignore the signal and defect – this is a strict Nash equilibrium of the game exactly as in the non-self-referential version of the game. Let us investigate the possibility of sustaining cooperation through self-referentiality. In particular, we consider the code of conduct that chooses C if the signal 0 is received, and chooses D if the signal 1 is received. If both players adhere to the code-of-conduct, they both receive an expected utility of $5-4p$. A player who chooses instead to always defect (whereas his opponent adheres to the code-of-conduct) gets $(1-q)6+q=6-5q$ and does worse by always cooperating. Hence it is strictly optimal to adhere to the code of conduct when $q>(1/5)+(4/5)p$. This says, in effect, that the signal must be informative enough. This has an interesting extension to the repeated case.

Example 2: The Repeated Prisoners Dilemma

We now consider the game repeated twice without discounting, so we simply use the sum of payoffs between the two periods. Consider first the code of conduct that chooses DD if the signal 1 is received and CC if the signal 0 is received. Since play is not conditioned on what the other player does, the optimal deviation against this code is DD, and 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 $q<(1/5)+(4/5)p$. For simplicity we analyze the case $p=0$. We consider the code of conduct that for both players chooses DD if the signal 1 is received, and if the signal 0 is received plays C in the first period and in the second period plays whatever the other player played in the first period. In other words, following the good signal 0 the player plays tit-for-tat, following the bad signal 1 the player defects in both periods. If both players adhere to the code of conduct they both get 10. Call this code-of-conduct $\hat{r}$.

There are two deviations of interest: to defect in both periods, or to cooperate in the first period then defect in the second. (Other strategies such as defecting in the first period and cooperating in the second period achieve lower utility than these two.) We point out that these deviations do not make use of the signal structure.[6] When we consider a deviation from a code-of-conduct, we look at one particular player who deviates from the code while his opponent follows it.

A player who defects in both periods, that is, plays DD, has a $1-q$ chance of getting 6 in the first period, and a q chance of getting 1, while he gets 1 in the second period for sure. Thus, the expected utility of DD is $(1-q)6+q+1=7-5q$. Since this

[6] Since the signals are independent, your signal contains no information about what the other player will do.

is less than 10 for any q, it is never optimal to play DD given that the opponent follows the code. Next, suppose the deviation in which a player who cooperates in the first period and defects in the second, that is, plays CD, gets expected payoff $(1-q)5+6-5q=11-10q$. From these results we can work out that our code-of-conduct $\hat{r}$ would be chosen over the deviation CD when $q>1/10$. 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. Notice how the signal complements the repetition: by using the signal to provide an incentive to retaliate in period 2 – something that with $p=0$ has no cost – a deviator is given incentive to cooperate in the first period, reducing the gain to deviation. In this sense, direct observation of opponent's intentions is a complement for repetition. It is worth noting that in the one-period game the probability of being caught needs to be 20%, while in the two period game we only require this probability to be at least 10%. So, it is fairly clear that we strengthen this complementary as more arbitrary periods are considered. 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 $\hat{r}$ is a strict Nash equilibrium except in the boundary case when $q=1/10$.

It is interesting also to see what happens in the T period repeated game with no discounting. For simplicity of exposition let us consider the time-average payoff. Consider the code of conduct that says that both players should play the grim-strategy on the good signal, and always defect on the bad signal. This gives a payoff of 5. The optimal deviation against this code-of-conduct is to play the grim-strategy until the final period, then defect. This gives a payoff of $1/T[(1-q)(5T+1)+q(T-1)]$. Hence it is optimal to adhere to the code-of-conduct when $q\geq 1/(4T+2)$. The salient fact is that as $T\rightarrow\infty$ only a very tiny probability of "getting caught" is needed to sustain cooperation.

Example 3: Changing Partners

Lots of mechanisms and equilibria are possible if we can get third parties to tell the truth. In a sense this is easy since third parties are generally indifferent. However, it seems a lot to build a society on the idea that indifferent third parties will tell the truth.

Self-referentiality breaks the tie and makes it strictly optimal to tell the truth. A simple example, based on Kandori's [1992] work on social norms is the repeated gift-giving game of Johnson, Levine and Pesendorfer [2001]. However, while Johnson, Levine and Pesendorfer [2001] showed the (exogenous) information system selected to sustain equilibrium, they did not show how this information system arises, which self-referentiality enables us to do.

Overlapping generations of players who live three periods are randomly matched to play a gift-giving game. Young productive players in the first period of life are referred to as "givers", middle-aged unproductive players in the second period of their life are referred to as "receivers" and old unproductive players in the third period of their life are referred to as "witnesses." Every period the witnesses die and an equal number of givers are born, the givers become receivers and the receivers become witnesses. Each giver is endowed with one indivisible unit of a good, a gift. Only receivers get utility from consumption.

Each period receivers are randomly matched with givers. Once a pair is formed it remains intact until one player dies. So when a giver becomes a receiver and is matched with a giver, the person who was his previous receiver becomes his witness. This structure insures that it is possible for a giver to get information about the past play of his receiver from the witness – the key question is whether that information will be provided and how it will be used.

In each match there are three players, one of each type. The giver must decide whether or not to give a gift to the receiver. If he does not, both receive a payoff of 0; if he does, he receives a payoff of –1, while the gift is worth $\alpha > 1$ to the receiver. Witnesses do not give or receive gifts and receive utility zero, but may choose either not to report, in which case they receive utility 0 or they may make one of two public statements which we refer to as "green" or "red" in which case they receive a utility of $-c$. Players maximize their lifetime utility without discounting. The possibility of reporting is important givers can use reports from witnesses to extract information about receivers' past behavior and exploit the linkage between generations. It is evident that since reporting is costly and witnesses are in the final period of their life, they will never choose to report if there is no self-referentiality in the game. We assume instead as we did before that the signal structure $Y_i = \{0,1\}$ for each player, that these signals are independent, and that the probabilities are given by $\pi_0(y = 1 \mid r) = q$ if $r^1 \neq r^2$ and $\pi_0(y = 1 \mid r) = p$ if $r^1 = r^2$ with $q \geq p$. In the second period of his life, a receiver has a red flag, a green flag or nothing. These states can be used to make the gift-giving decision by the giver.

For comparison note that with no self-referentiality whenever the receiver participates in a match, there is not much left to do. He already made the decision. On the other hand, the giver decides whether to give the gift to his partner or not regardless of whether his partner gave it before. He would like to have some sort of "reward" in the future if he behaves "nice" (and by reward in this game we mean getting the gift in his second period of life). Similarly, the receiver wishes to have a way to tell the giver how he behaved in the previous match in order to justify that he deserves the gift, if so.[7] Now, let us reconsider the game described above. We emphasize that strategies tell how a person will play in his entire life, so not only the circumstances under which he will give the gift, but also what he will say about the person he met when he is a witness. Hence, if for instance part of the code of conduct says "report the truth," either player adheres to it, or there is a chance when he is a giver that the receiver he faces will know that he is deviating from this code and is a liar.

We start our analysis by considering the code of conduct $\hat{r}$ that says:

– if you are a giver and observe $y = 1$, do not give the gift;

– if you are a giver and observe $y = 0$, give to a receiver with a green flag, do not give to a receiver with a red flag;

– report a green flag if you received the gift and had a green flag when old, or if you did not receive the gift and had a red flag when old;

– otherwise report a red flag.

Note that the receiver ignores the signal about whether the young person has adhered to the code of conduct. Following this code of conduct gives a payoff of $(1-p)(\alpha-1)-c$ if he meets a player with a green flag, and $(1-p)\alpha-c$ if he meets a player with a red flag. We assume that c is sufficiently small, that is, $(1-p)(\alpha-1) > c$. One way of violating the code of conduct is by not giving a gift to a green flag. This is not incentive compatible. Intuitively, this is because it will always be

[7] Note that analogously a giver would like to avoid any kind of punishment because of not giving the gift, and consequently, the receiver prefers to have a way of punishing such behavior.

reported. It remains to check that reporting the truth is incentive compatible. To do so, let us study the situation in which a player follows exactly the code of conduct except that when receiver he does not report. Notice that first we are examining another violation of our code of conduct, and second that we are considering "not report" as a lie. Note also that reporting something different from what the code of conduct requires always yields a lower payoff because $q \geq p$. This gives a payoff of $(1-q)(\alpha-1)$ if he meets a player with a green flag and $(1-q)\alpha$ if red flag has been reported in the match. Consequently, the receiver finds optimal to report the truth only if $q > c / (\alpha - 1) + p$. In other words, the signal should be again informative enough.

Next, we examine different alternatives to the code of conduct. We begin with *tit-for-tat* code of conduct whose only variation from the one presented above is: give the gift to a player if and only if he has a green flag provided he observed the good signal, and report a green flag if and only if player received gift and had a green flag. This code of conduct tells giver to give the gift if he observes a green flag and witness to report a green flag only if the gift was given. That is, if the giver suspects the receiver on account of the bad signal and consequently decides not to give the gift, in the eyes of the tit-for-tat code of conduct this is considered an action to be punished. It is important to note that players expect to encounter players with a red flag since there is a positive probability of not giving the gift, namely, if the bad signal is realized. This says that presumed evidence of deviation from the code of conduct will be punished. Moreover, adhering to this code of conduct is not a best response to itself because of this positive probability. The idea is that givers will get a red flag whenever the outcome is the bad signal because they are not giving the gift and hence are punished. In fact, other players following this code of conduct would punish this behavior and therefore it is not an optimal response.

Alternatively, consider *weak* code of conduct that is a mild version of our code of conduct. Specifically, witness reports a red flag only if he did not receive gift and had a green flag, otherwise report a green flag, in addition giver gives to the receiver with a green flag only if the good signal is observed. Indeed, even if the giver gives the gift to someone with a red flag there is no punishment – the idea here is that you are nice with your match whenever you see a good signal. The simplest way to see this code of conduct does worse is to suppose that some proportion of the population is adhering to tit-for-tat code of conduct. In the case that a giver does not give the gift to the receiver with a red flag he would get a red flag if he encounters someone that adheres the tit-for-tat code, then he would be punished since he would be reported with a red flag. The key point is that the weak code of conduct would have reported this behavior with a green flag and consequently it would not have been punished. This is the reason the weak code of conduct will not be adhered by players if they are aware of the existence of some people following the tit-for-tat code. Notice the subtle difference between the proposed code of conduct $\hat{r}$ and the weak code. Under the code-of-conduct $\hat{r}$ players have a strict incentive to report people who did give the gift to players with a red flag. On the other hand, the weak code of conduct does not provide players with strict incentives to do so. This means that the code of conduct $\hat{r}$ will still be adhered by players even in the presence of players following the tit-for-tat code.

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.

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. Fix a signal space Y and break the program into two parts, one of which generate y based on analyzing the programs, the other of which maps Y to S. The programs are "self-referential" in the sense that they also receive as input the program of the other player. That is, each program takes as input itself and the program submitted by the other player. 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 $p^i : P \times P \to \{1,2,\ldots,\infty\} \times S$.[8] The interpretation is that

[8] This analysis can be easily extended to the case that there are more than two players.

$p^i(p^1,p^2) = (\nu^i, s^i)$ 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 is given itself and the program submitted by the opposing player 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 $\nu^i \le \bar{\nu}$, 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 $Y_i = S$. Then $\pi(y \mid r) = 1$ if $y_i = r^i(p^i, p^{-i})$ for $i = 1,2$, and 0 otherwise.

Kind of example that does not exist: read the other guys program and make a best response. Kind of example that does exist: make one response if same and an alternative response if different. Must be based on actual code, not on function of program (see Levine and Szentes [2006]).

Definition 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 $s = (d,p) \in S$ 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 action space $A = \{0,1\}$. The default action is 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. Although in this listing all the actual work is done by the “comp” command it is easy enough to write a program that compares two files, and takes a number of steps proportional to the length of the shorter file. In other words, the program works in finite, and relatively short time.

Note that this mechanical procedure is feasible. The idea here is that the computer program takes as input the opponent program, reads it line by line and then decides what to do. Basically the opponent code is just a list of characters, the computer program compares it to itself and checks if they are the same. What do we mean to be the “same?” In our setting be the same means using the same language written in exactly the same way.

It is worth mentioning that if both players know in advance they have similar abilities then there is no reason to believe that some of them would try to “obscure” their program. That is, both players have the same coding technology to check whether codes are equal or not it is difficult to commit fraud. However, if one of the players knows that the other player has a relatively bounded memory size and is not a skilled programmer who will not be able to entirely compare the programs, then it might be beneficial for him to exploit this. Consider for example the case of fraud. It demands a lot of skills to pretend to be a rich successful businessman who offers you a great investment opportunity when in reality nothing could be further from being true. By the same token, writing a program that seems to be the same but actually is different requires large memory and clever computational skills – it is not costless. In addition, the use of these more complex 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.

In the face of limitations, we might imagine that rather than writing a program that compares the opponent program to itself line by line, it takes sample lines at random and compares them to itself. Consequently answers are noisy. Note that there is always a chance of detecting a different code, perhaps very small but not negligible.

People playing games do not often do so by submitting computer programs. From an evolutionary perspective, however, 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.

From Two Players to Many

In a two player symmetric game, like the prisoner's dilemma game above, there is a notion of similarity between players. Because of symmetry a player can simply 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. We emphasize that this definition is convenient when we lose symmetry: 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. The key element of adherence to code of conduct is that players do not only agree about how they would behave, but also about how third parties would behave. One example that highlights the importance of this extension is the case of interaction between individuals and the Government. Clearly, ordinary citizens and politicians play different roles and have different set of strategies. However under a state of law, they both agree that the ones elected by the majority should make the decisions that affect the course of everyone's actions and that everyone should obey the law. This complex environment can be easily captured by our notion of adhering to code-of-conduct.

Do People Use Self-Referential Strategies?

Here is one possible motivatation for this 'recognition technology.' Strategies govern the behavior of agents over many 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 example, 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].)

As an example, suppose that $Y = \{0,1\}$ for both players. Further assume that $\pi(y = 0 \mid s',s) = 1$ if $s' = s$ and $\pi(y = 1 \mid s',s) = 1$ if $s' \neq s$. Thus, if two players meet who use the same strategy then both receive the signal 0 whereas when two players meet who use different strategies then both receive the signal 1. In other words, players recognize if their opponents use the same or a different strategy prior to play. This example is important, because it turns out that strategies that recognize themselves are likely to emerge in the long-run equilibrium.

It bears emphasis that the space of signals is necessarily smaller than the set of strategies: the cardinality of the space of strategies is at least $|A|^{|Y|}$, which is greater than that of Y provided that there are at least two actions. This relative coarseness of the signal space means that it is not possible that a signal could reveal the precise strategy of an opponent for every possible strategy profile.

5. Perfect Information

We now focus on the case of perfect information. Throughout this section we assume that signals are perfectly revealing. In this setting, players directly observe the code-of-conduct chosen by their opponents by receiving these completely reliable signals. Notice that we relax the assumption about the relative coarseness of the space of signals.

Let us begin with a static two player game with finite spaces of strategies S_i. They will simultaneously choose strategies $s_i \in S_i$ and get payoff $u_i(s)$. The self-referential game consists of a fixed set of signal spaces Y_i for $i = 1,2$, a complete code-of-conduct space R and probabilities over profiles of signals given by $\pi(y \mid r)$. To analyze this case we assume that there exists a profile of signals with $y_i^c \in Y_i$ such that $\pi_i(y_i^c \mid r) = 1$ for $r^1 = r^2$, and $\pi_i(y_i^c \mid r) = 0$ for $r^1 \neq r^2$. This assumption says that the signal y_i^c is realized if and only if both players adhere to the same code of conduct. Putting this differently, players are able to perfectly identify those opponents who share the same expectations of others' behavior and they both agree how the two of them behave. We define the (possibly mixed) minmax strategy of the base game against player i by $\underset{\sim}{s}_{-i}^i$ as the argument of $\min_{s_{-i}} \max_{s_i} u_i(s_i, s_{-i})$. Let $\underset{\sim}{u}_i = u_i(\underset{\sim}{s}_i^i, \underset{\sim}{s}_{-i}^i)$ be the smallest payoff that his opponent can keep player i below and $\underset{\sim}{s}_i^i$ denote i's best response to $\underset{\sim}{s}_{-i}^i$.

Our first result in the perfect information case is very similar to the Levine and Pesendorfer [2007]:

Theorem 5.1: *For any $v_i = u_i(s_1, s_2) \geq \underset{\sim}{u}_i$ for all $i = 1,2$ and $(s_1, s_2) \in S$, there exists a profile of codes of conduct r such that (v_1, v_2) is a Nash self-referential equilibrium payoff.*

Proof: Let $(s_1, s_2) \in S$ with $u_i(s) \geq \underset{\sim}{u}_i$ for $i = 1,2$. Suppose the code of conduct r says player i should play s_i if $y_i = y_i^c$, and play s_i^{-i} otherwise. If both players follow this code of conduct player i gets $U_i(r) = u_i(s_1, s_2)$ for $i = 1,2$. Since any deviation from this code-of-conduct will be detected with probability one and punished with the minmax strategy, r is a Nash equilibrium of the self-referential game with payoff v_i for all i.

☑

We can extend the previous result to the case in which we have more than two players. The proof of the next result, which is similar to that of Theorem 5.1, is omitted.

The intuition of the proof is again to use the idea of a code of conduct that punishes with the minmax strategy in case of detecting deviation by one of the players and play the prescribed strategy if the signal associated to this strategy is received.

Theorem 5.2: *If* $v_i = u_i(s) \geq \underset{\sim}{u}_i$ *for all players* i *and strategy profile* $s \in S$ *, then* $(v_1, \ldots, v_N)$ *is a Nash equilibrium payoff of the self-referential version of the 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. Because we relax the assumption that everyone sees everything, we need to establish the sense in which a player who deviates is detected.

We say that the self-referential game *permits detection* if for every player i there exists some player j and a set $\bar{Y}_j \subset Y_j$ such that for any profile of codes of conduct $r \in R$, any signal $\bar{y}_j \in \bar{Y}_j$ and any $\tilde{r}^i \neq r^i$ we have $\pi_j(\bar{y}_j \mid \tilde{r}^i, r^{-i}) = 1$. Intuitively, this says that if player i deviates we can always find another player j who detects this deviation with certainty.

The fact that we have a selected pool of people who receives perfectly revealing signals raises the issue of what happens if a player from that pool needs some other player to punish the deviator. Of course if that player can unilaterally punish the deviator there is no problem. Hence, we first define the notion of this group by saying that a self-referential game is *locally perfectly informative* if there exists a proper nonempty subset $I_k \subset \{1, \ldots, N\}$ such that all player k's contained in that set receive perfect signals.

In this specification of the self-referential game, there is a possible scenario in which one of players who receive the perfectly revealing signal needs some other player to punish the deviator. Thus, in order to have the deviator punished this player must have some sort of communication device to inform the presence of deviation to the player he needs to implement the punishment. To provide this communication we assume cheap talk after receiving signals and before play.

The game has the following timing. First, players select their code-of-conduct. Second, after signals are received, players make announcements on violation of the code-of-conduct. Finally, players choose their actions and play the base game.

Formally the model is as follows. Each player adheres to a code of conduct, $r^i \in R_0$ and that induces a probability distribution over profiles of signals given a profile of code-of-conduct $\pi(y \mid r)$. After receiving signals, players send cheap talk signals defined as *announcement* taken from a finite set $\tilde{y}_i \in \tilde{Y}_0$, and a profile of announcements is defined as $\tilde{y} \in \tilde{Y}$. A *message* from player i is a map $m_i : Y_i \to \tilde{Y}_0$ with message profile denoted by m. Players choose actions simultaneously from a finite space, $a_i \in A_i$ with action profile $a \in A$. The payoff associated to action profile a is $u_i(a)$. A strategy for player i is the decision about an action a_i to take and a message m_i to send, that is, $s_i = (a_i, m_i)$.

We remark that players respond to such announcements, namely, if player j announces player i has violated the code-of-conduct and this was the only announcement, all players play the prescribed action required to implement punishment to i. Moreover player i may try to take advantage of this information structure by announcing somebody else has violated the code-of-conduct. However, this does not look like a good strategy since he will certainly be detected and consequently punished. On the other hand, there exists the possibility that two players are pointing to each other and it is not possible to tell who actually deviated. Hence, 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. In words, player j detects player i, but not vice versa.

Having stated the model, we prove the next theorem that is in same spirit as the previous results.

Theorem 5.3: *If $v_i = u_i(a) \geq \underset{\sim}{u}_i$ for all i and actions profile a, if the self-referential version strongly permits detection and is locally perfectly informative, then $(v_1, \ldots, v_N)$ is a Nash equilibrium of the self-referential version of the game.*

Proof: Let $\underset{\sim}{a}^i_{-i}$ be the minmax action for player i. Fix an arbitrary profile of actions $a \in A$ such that $u_i(a) \geq \underset{\sim}{u}_i$ for all i. Given the profile of actions, we begin with constructing the code of conduct, $\hat{r}$, that implements $u_i(a)$ for all i.

The code of conduct $\hat{r}$ says for any player j: if he receives the signal $\bar{y}_j$ and is able to unilaterally punish player i, play $\underset{\sim}{a}^i_j$ and send the message $m_j = \varnothing$ (he does not send any message). Alternatively, if player j observes the signal $\bar{y}_j$ and needs player k's to punish i, he sends message $m_j = \tilde{y}_0$ and plays $\underset{\sim}{a}^i_j$. If he does not receive any signal from the set $\bar{Y}_j$ but the message to punish player i, play $\underset{\sim}{a}_j^i$. In any other case, play a_j.

We now proceed to show that this code of conduct implements the payoff proposed. With this code of conduct, if players adhere to $\hat{r}$ their payoffs are $U_i(r) = u_i(a)$ for all i where the profile of messages is the empty set.

It is seen immediately that not sending the message in case of detecting deviation is never chosen since messages are costless and failure to send messages will be punished as a violation of the code of conduct. Nevertheless, a possible deviation from this code-of-conduct is not to play the prescribed action and to announce somebody else has deviated. There are two possible cases to consider here. First, player j announces that a player k who cannot observe $j's$ play has deviated. Since there is some other player l points out player j's deviation and j is correspondingly punished j only loses by violating the code of conduct. More interestingly, suppose player j accuses the player k who does observe j's play and who accuses j. In this case j and k point the finger at each other: one is guilty – but who? Strong detection enables us to sort this out. Detection is unidirectional and everybody knows which way – so everyone knows that while k observes j's play, j does not observe k's play, hence it must be j who is lying and who should be punished and obtains $\underset{\sim}{u}_j$. Because of strong detection the deviator will suffer the punishment irrespective of his announcement. 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 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. Hence if you deviate and are caught you will be punished regardless of accusations you might make about others. 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.

The code of conduct constructed in the proof of the last theorem incorporates the idea about communicating what other players cannot see using cheap talk signals. Even though all players do not receive perfect signals, the existence of messages allows the detector to communicate and point out a deviation to the players required to materialize the punishment. Recall that in the changing partners' example, the young generation that decides to give or not the gift would be supported (if they follow the code-of-conduct) or exposed (if they did not) by their witnesses. Furthermore, witnesses are able to send an almost cheap talk signal – the report – to the incoming generation of givers about the interaction that they previously perfectly observed.

We emphasize that when a player adheres to the code of conduct $\hat{r}$, if he receives the message that calls for a punishment he is expecting all players involve in it will play accordingly. Moreover, the chance of observing a deviation from this code is captured by the signals observed after selecting codes of conduct. Clearly this illustrates the notion of agreeing about how others should behave that is present in the code of conduct.

6. Approximate Equilibria

We now 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 assume that in the base game all players have access to N individual randomizing devices each of which has an independent probability $\varepsilon_R > 0$ of an outcome we call *punishment*. Since the base game is finite, we can denote by $\underline{u}, \overline{u}$ be the highest and lowest payoffs to any player in the game.

We assume fixed signal spaces Y_i, that R is a complete code-of-conduct, and that the signal probabilities are $\pi(y \mid r)$. The self-referential game is said to E,D *permit detection* where $1 \geq E,D \geq 0, 1 \geq E + D$ if for every player i there exists a player j and a set $\overline{Y}_j \subset Y_j$ such that for any profile code of conduct $r \in R$, any signal $\overline{y}_j \in \overline{Y}_j$, and any $\tilde{r}^i \neq r^i$ we have $\pi_j(\overline{y}_j \mid \tilde{r}^i, r^{-i}) - \pi_j(\overline{y}_j \mid r) \geq D$ and $\pi_j(\overline{y}_j \mid r) \leq E$. 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. In addition, E represents an upper bound for the probability of false positive. You can think of E as the probability of someone who is being falsely accused of cheating when he behaves honestly.

We start by supposing that the strategy profile s^0 is a ε_0-Nash equilibrium giving utility profile u^0 in the base game. Suppose for some strategy profile s and strategies s_j^i played by player j for any pair of players $i,j = 1,\dots,N$ that $s_{(j)}^i = (s_j^i, s_{-j})$ are ε_1-Nash equilibria satisfying for each player i that $P_i = u_i(s^0) - u_i(s_{(j)}^i) \geq \underline{P} \geq 0$ and for some $\varepsilon_P \geq 0$ that $\left| u_j(s_{(j)}^i) - u_j(s^0) \right| \leq \varepsilon_P$. 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 $s_{(j)}^i$ to s^0, that is, a measure of how far the punishment equilibria are from the original equilibrium. Define two parameters ε and K

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

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

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 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* r *with*

$$\left| u_i(s^0) - u_i(r) \right| \leq \varepsilon + D(\underline{P} - \varepsilon_1) - \sqrt{(D(\underline{P} - \varepsilon_1))^2 - 4K\varepsilon}, \text{ for all } i.$$

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. If E is small the cost of punishment and lack of credibility do not matter, because punishments must only be carried out infrequently on the equilibrium path, so there is no sense in risking getting caught violating the code-of-conduct to attain what is only a small gain. At the extreme case, when $E = 0$, the parameter ε turns out to be ε_0. This implies that punishments are not carried out on the equilibrium path if nobody deviates. If that the case, the closeness of the strict equilibrium of the code-of-conduct depends solely on the 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 $\bar{s}_{(j)}^i$ which punish player i, but they will not generally have the requisite form $s_{(j)}^i = (s_j^i, s_{-j})$ 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 $\bar{s}_{(j)}^i$. 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. We remark that the definition of E, D strongly permits detection precludes the puzzling scenario in which two players are pointing to each other, but it captures the same intuition discussed after Theorem 5.3 (see further discussion in section 5 on strongly permits detection). The main difference is that now players who deviate are not detected with certainty.

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.

7. Repeated Games with Private Information

Our goal is to prove a folk-like theorem for games with private information. Fudenberg and Levine [1991] consider repeated discounted games with private information 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.

The Stage Game

The stage game has finite action spaces $a_i \in A_i$ for each player $i, i = 1,...,N$ and these are chosen simultaneously. The corresponding action profiles (vector of actions) are denoted $a \in A$, while α_i, α denote mixed actions and profiles. Each player has a finite private signal space $z_i \in Z_i$ with signal profiles written as $z \in Z$. Given an action profile the probability of a signal profile is given by $\rho(z \mid a)$. This induces also a probability distribution for mixed actions $\rho(z \mid \alpha)$ as well as marginals over individual signals $\rho_i(z_i \mid \alpha)$. Utility for individual players $w_i(z_i)$ depends only on private signal received by that player.[9] This gives rise to the expected utility function $g_i(\alpha)$ constituting the normal form of the stage game.

A *mutual threat point* is a payoff vector v such that there exists a *mutual punishment action* – this is a mixed action profile α such that $g_i(\alpha_i', \alpha_{-i}) \le v_i$ for all players i and mixed actions α_i'. As is standard, a payoff vector v is *enforceable* if there is an α with $g(\alpha) = v$, and if for some mixed action α_i', $g_i(\alpha_i', \alpha_{-i}) > g_i(\alpha)$ then for some $j \ne i$ we have $\rho_j(\cdot \mid \alpha_i', \alpha_{-i}) \ne \rho_j(\cdot \mid \alpha)$. 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 punishment point and the closure of the convex hull of the enforceable payoffs. Notice that this is generally a smaller set than the socially feasible individually rational set both because

---

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

there may be unenforceable actions, but also because the minmax point may not be mutually punishable. 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, Levine and Maskin [1994] 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.

We now describe the notion of informational connectedness. Roughly this says that it is possible for player 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 $k \neq i, j$ if there exists a mixed profile α and mixed action $\hat{\alpha}_i$ such that

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

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 any p such that player i_p is directly connected to player i_{p+1} despite player k. The game is *informationally connected* if there are only two players, or if every player is connected to every other player.

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 $h_i(t) = (a_i(1), z_i(1), \ldots, a_i(t), z_i(t))$ while $h_i(0)$ is the null history. A behavior strategy for player i is a sequence of maps $\sigma_i(t)$ taking his private history $h_i(t-1)$ to a probability distribution over A_i. For $0 \leq \delta < 1$ 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 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*[10] *for δ_n and $u_i(\sigma_n;\delta_n,\infty) \to v_i$.*

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 *communications phase* with length L such that

Lemma 7.2 (Fudenberg and Levine [1991]): *For any $0 < \beta < 1$ there exist a pair of strategies σ_i, σ_i' and for each player $j \neq i$ a test $\boldsymbol{Z_j} \subseteq \{(z_j(1),\ldots,z_j(L))\}$ 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 $\text{Prob}[(z_{-i-j}(1),\ldots,z_{-i-j}(L)) \in \boldsymbol{Z_{-i-j}}] \leq 1-\beta$, and under $(\sigma_i', \sigma_k', \sigma_{-i-k})$ we have $\text{Prob}[(z_{-i-j}(1),\ldots,z_{-i-j}(L)) \in \boldsymbol{Z_{-i-j}}] \geq \beta$.*

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.

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.

Thus, the self-referential T-discrete game consists of signal spaces Y_i and complete codes of conduct space R_T, moreover, the signal probabilities are $\pi_T(y \mid r)$.

---

[10] 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 time 0 average present value are no bigger than ε_n. As we do not need the stronger result, we do not give the extra definitions required to state the stronger result.

Theorem 7.3*: 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 strongly permits detection, and if $v \in V^*$ then there exists a sequence of discount factors $\delta_n \to 1$, discretizations T_n and codes of conduct r_n such that r_n is a strict Nash equilibrium for δ_n, T_n and $U_i(r_n; \delta_n, T_n) \to v_i$.*

Notice that we do not allow ε to go to zero with T. In other words even as the game becomes more complex and codes of conduct potentially more elaborate, there is still an ε chance of detecting a deviation.

By Theorem 6.1 it suffices to prove the following result:

Theorem 7.4*: 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 $v \in \text{int}(V^*)$ then for any $\varepsilon_0 > 0$ there exists a discount factor δ, a discretization T and strategy pairs s_i^0, s_i^j such that s^0 is an ε_0-Nash equilibrium for δ, T, $\varepsilon_1 = \varepsilon_P = \varepsilon_0$ and $\underline{P} = \sqrt[3]{\varepsilon_0}$.*

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 8.1 applies directly to the discretized game. Theorem 8.1 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 $\varepsilon_0 / 2$-Nash equilibria, that $\left| u(\bar{s}^0) - v \right| < \varepsilon_0 / 2$, $u_i(\bar{s}^0) - \bar{u}_i(s_{(i)}^j) \geq 2\sqrt[3]{\varepsilon_0}$ and $\left| u_j(\bar{s}^i) - u_j(\bar{s}^0) \right| \leq \varepsilon_0 / 2$.

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 $\hat{s}_i^j, \hat{s}_i^{j\prime}$ 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 $\hat{s}_i^{j\prime}$. 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 1 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 7.4 differs from that of Theorem 7.1 in an important way. In Theorem 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 7.4 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."

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 – which determine signals conveying information about players' intentions to play, and specify how all players should behave according to those signals. This is important because 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. Several examples explored in our analysis illustrate the applicability of self-referential games.

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. When describing this motivation, we pointed out that playing through computer programs is not exclusively relevant per se. 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 (cheap talk signals) to communicate with other players needed to implement a punishment. This networking idea was introduced by Fudenberg and Levine [1991] and is used later on while proving our folk theorem result. Yet we encounter the issue of mutually accusation. As we pointed out only a slightly stronger version of permits detection is needed to overcome this issue.

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. Our requirement of approximate punishment strategies having only the detector punishing the deviator may look restrictive, however, those strategies are ensured simply by allowing cheap talk before play in the same way we did in the perfect information case. An illustrative, but far more important, application of this result is a discounted strict Nash folk-like theorem. Here playing the game repeatedly and likelihood of detecting deviation from a code-of-conduct as information resources reinforce each other. Hence we conclude that approximate equilibria can be sustained as "real" equilibria when there is a chance of detecting violations of codes-of-conduct.

References

Binmore, K., J. McCarthy, G. Ponti, L. Samuelson and A. Shaked [2002]: "A Backward Induction Experiment," *Journal of Economic Theory*, 104: 48-88.

Binmore, K. G. and L. Samuelson [1992]: “Evolutionary Stability in Repeated Games Played by Finite Automata,” *Journal of Economic Theory*, 57: 278-305.

Frank, R. H. [1987]: “If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?,” *American Economic Review*, 77: 593-604.

Fudenberg, D., D. Kreps and D. K. Levine [1988]: "On the Robustness of Equilibrium Refinements," *Journal of Economic Theory*, 44: 354-380

Fudenberg, D. and D. K. Levine [1991]: "An Approximate Folk Theorem with Imperfect Private Information," *Journal of Economic Theory*, 54: 26-47.

Fudenberg, D., D. K. Levine and E. Maskin [1994]: "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62: 997-1039.

Fudenberg, D. and E. Maskin [1986]: “The Folk Theorem in Repeated Games with Discounting or with Incomplete Information,” *Econometrica*, 54: 533-554.

Kandori, M. [1992]: “Social Norms and Community Enforcement,” *Review of Economic Studies*, 59: 61-80.

Levine, D. K. and W. Pesendorfer [2007]: "Evolution of Cooperation through Imitation," *Games and Economic Behavior*, 58: 293-315.

Levine, D. K. and B. Szentes [2006]: "Can A Turing Player Identify Itself?," *Economics Bulletin*, 1: 1-6.

Radner, R. [1980]: “Collusive Behavior in Noncooperative Epsilon-equilibria of Oligopolies with Long but Finite Lives,” *Journal of Economic Theory*, 22: 136-154.

Robson, A. J. [1990]: “Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake,” *Journal of Theoretical Biology*, 144: 379-396.

Rubinstein, A. [1986]: “Finite Automata Play the Repeated Prisoner's Dilemma,” *Journal of Economic Theory*, 39: 83-96.

Sugaya, T. [2011]: “Folk Theorem in Repeated Games with Private Monitoring,” mimeo, Princeton.

Tooby, J. and L. Cosmides [1996]: “The computationalist theory of communication,” Paper presented at the meeting of the Human Behavior and Evolution Society Meetings, University of Arizona, Tucson.

Appendix: Proof of Theorem 6.1

To prove Theorem 6.1 we rewrite the necessary information here. Recall that we assume fixed signal spaces Y_i, that R is a complete code-of-conduct, and that the signal probabilities are $\pi(y \mid r)$. The self-referential game is said to E,D *permit detection* where $1 \geq E,D \geq 0, 1 \geq E+D$ if for every player i there exists a player j and a set $\bar{Y}_j \subset Y_j$ such that for any code of conduct $r \in R$, any signal $\bar{y}_j \in \bar{Y}_j$, and any $\tilde{r}^i \neq r^i$ we have $\pi_j\left(\bar{y}_j \mid \tilde{r}^i, r^{-i}\right) - \pi_j\left(\bar{y}_j \mid r\right) \geq D$ and $\pi_j\left(\bar{y}_j \mid r\right) \leq E$.

Let the strategy profile s^0 be a ε_0- Nash equilibrium giving utility u^0 in the base game. Suppose for some profile s and strategies s_j^i for $i,j = 1,\ldots,N$ that $s_{(j)}^i = (s_j^i, s_{-j})$ are ε_1-Nash equilibria satisfying for each player i $P_i = u_i(s^0) - u_i(s_{(j)}^i) \geq \underline{P} \geq 0$ and for some $\varepsilon_P \geq 0$ that $\left|u_j(s_{(j)}^i) - u_j(s^0)\right| \leq \varepsilon_P$. Define two parameters ε and K

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

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

Theorem 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* r *with*

$$\left|u_i(s^0) - u_i(r)\right| \leq \varepsilon + D(\underline{P} - \varepsilon_1) - \sqrt{(D(\underline{P} - \varepsilon_1))^2 - 4K\varepsilon}, \text{ for all } i.$$

Proof: 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

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

Now we define the code-of-conduct r: for all players i, if $y_i \in \overline{Y}_i$ and the event "punishment" occurs play s_i^j, otherwise play s_i^0.

The following mutually exclusive events can occur to player i:

- Nobody is punished: if r is followed i gets $u_i(s^0)$, if i deviates he gets at most $u_i(s^0) + \varepsilon_0$
- Player j is the only player punished: if r is followed i gets $u_i(s^j_{(i)})$, if i deviates he gets at most $u_i(s^j_{(i)}) + \varepsilon_1$
- Two or more players are punished: if r is followed i gets at worst $\underline{u}$, if i deviates he gets at most $\overline{u}$

Hence if all players follow the code player i gets no more than

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

and no less than

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

If i violates the code, and everybody else follows the code, he gets no more than

$$\begin{aligned} &u_i(s^0) + \varepsilon_0 + (1-(1-E)^N)\left[\varepsilon_1 + N^4(\varepsilon_R)^2(\overline{u}-\underline{u})\right] \\ &-\left[(\pi_j(\overline{y}_j \mid r) + D)\varepsilon_R - N^4(\varepsilon_R)^2\right](P_i + \varepsilon_1) \end{aligned}$$

Consequently the gain to violating the code is at most

$$\begin{aligned} &\varepsilon_0 + (1-(1-E)^N)\left[\varepsilon_1 + \varepsilon_P + 2N^4(\varepsilon_R)^2(\overline{u}-\underline{u})\right] \\ &+\pi_j(\overline{y}_j \mid r)\varepsilon_R P_i - \left[(\pi_j(\overline{y}_j \mid r) + D)\varepsilon_R - N^4(\varepsilon_R)^2\right](P_i - \varepsilon_1) \\ &\le \varepsilon_0 + (N + \overline{u} - \underline{u})E\left[\varepsilon_1 + \varepsilon_P + 3N^4(\varepsilon_R)^2(1 + \overline{u} - \underline{u})\right] - D\varepsilon_R(\underline{P} - \varepsilon_1) \\ &\le \varepsilon + K{\varepsilon_R}^2 - D\varepsilon_R(\underline{P} - \varepsilon_1) \end{aligned}$$

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

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

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 r is strict. Plugging the lower root into the inequality for the utility difference $\left|u_i(s^0) - u_i(r)\right|$ gives the remainder of the result.

☑