The Analogical Foundations of Cooperation*

Philippe Jehiel
Department of Economics
PSE and UCL
Paris, France and London, England

Larry Samuelson
Department of Economics
Yale University
New Haven, CT 06520 USA

August 24, 2021

Abstract. We offer an alternative approach to cooperation in repeated games of private monitoring in which prior beliefs are formed by observing the frequencies of play in a record of past plays of the game. The record is incomplete, reporting past actions but not signals. Players group the histories in the record into a relatively small number of analogy classes to which they attach probabilities. We provide conditions for the existence of equilibria supporting cooperation and supporting high payoffs, and show that more detailed analogy classes (i.e., a "better specified" model) need not lead to better outcomes.

Contents

*This work began when Larry Samuelson was visiting the Paris School of Economics, whom we thank. Jehiel thanks the European Research Council for funding.

The Analogical Foundations of Cooperation

1 Introduction

It is intuitive that repeated interactions, by allowing the participants to link future behavior to current actions, can give rise to different incentives than those of isolated interactions. People readily understand that they should treat differently the purchase of a diamond ring drawn from a display case in Tiffany's Fifth Avenue showroom and the purchase of a diamond ring drawn from the pocket of a trench coat on a nearby 23rd street subway platform. Models of repeated games of perfect monitoring (Fudenberg and Maskin [9]) capture this intuition well, most simply in the Nash reversion equilibria presaged by Friedman [5], in which players cooperate as long as there has been no defection, and then revert to the perpetual play of a Nash equilibrium of the stage game upon the first defection.

One might hope that analogous arguments would continue to hold in the face of the imperfections that inevitably complicate the monitoring of others' actions, as long as the monitoring is informative enough. Unfortunately, there is no counterpart of the Nash reversion equilibrium under private monitoring, no matter how precise the monitoring.[1] A player who receives a signal suggesting that her opponent defected in the first period will cling to the equilibrium hypothesis (that the opponent has cooperated) rather than trigger a punishment, attributing the signal to an unlikely draw from the noisy monitoring technology. This in turn gives players a license to defect in the first period, disrupting the equilibrium.

The standard response provided by game theorists is to construct equilibria in which players mix between cooperation and defection in the first period, so that adverse signals can be interpreted as signals of defection. (See Ely and Valimaki [4], Piccione [16], and Sekiguchi [20] for early contributions.) However, these equilibria typically sacrifice the intuitive link between current and future behavior that motivates much of the interest in repeated games—for example, belief-free equilibria are constructed so that player 1's beliefs about player 2's behavior are *irrelevant* to 1's choice of action—and can depend upon finely-tuned beliefs that are seemingly plucked out of thin air.

In this paper, we suggest an alternative approach to cooperation, centered around four features. First, rather than having their prior beliefs spring to life as part of an equilibrium, players form their prior beliefs by observing the frequencies of the various outcomes in a record of past plays of the game. Second, this frequentist foundation forces prior beliefs to be coarse. No record of previous interactions will allow the estimation of a potentially distinct probability for each of the infinite number of histories in a repeated game. Instead, players must group histories into a relatively small number of analogy classes to which they attach probabilities. Third, the record reports play from a variety of games

[1] As Fudenberg, Levine and Maskin [8] show, *noisy* monitoring need not pose difficulties as long as the monitoring is public.

whose peculiar details shaped the behavior of the players in the game, but are lost in the record. Finally, the record may lack some information. We focus on a record which reports the actions taken in previous plays of the game, but not the private signals received by past players, which we motivate on the grounds that such private signals can rarely be regarded as hard evidence and as such are hard for outside observers to access or interpret.

Formally, we examine analogy-based expectation equilibria (Jehiel [12]) in which the analogy partitions do not distinguish histories according to the private signals, nor according to whether the interaction is likely be short-lived or long-lived. We begin in Sections 2–3 with the prisoners' dilemma, with two analogy classes distinguishing histories according to whether any player has previously defected. We analyze the extent to which cooperation can be achieved when varying the precision of the monitoring and the share of short-lived interactions. We note that the reliance on analogy classes inevitably leads players to a misspecified model of their interaction, and explain how this misspecification is essential in creating the incentives supporting cooperation. Subsequent sections of the paper discuss the applications to which the analysis applies and argue that the intuition gained from the analysis can be extended more generally.

2 Analogical Reasoning and Cooperation

2.1 The Stage Game

We develop the basic ideas in the context of the workhorse model of cooperation, the repeated prisoners' dilemma. We focus in Sections 2 and 3 on the stage game given by

	C	D
C	$1,1$	$-k, 1+k$
D	$1+k, -k$	$0,0$

. (1)

It is a normalization to choose the payoffs of mutual defection and mutual cooperation to be 0 and 1. Our formulation then restricts attention to the commonly examined one-parameter class of games in which the payoff premium to defecting, given by k, is independent of the actions of one's opponent. The larger k, the more tempting is defection, making it more difficult to sustain cooperation.

If this game is infinitely repeated under perfect monitoring and with common discount factor β, then there exists an equilibrium supporting permanent mutual cooperation if and only if the players are relatively patient and the premium on defecting is relatively small, i.e., if and only if

$$\beta \geq \frac{k}{1+k}. \tag{2}$$

Perhaps the best known strategy supporting cooperation is the Nash reversion or "grim" strategy, in which both players cooperate after any history featuring no defections, and defect otherwise.

2.2 The Repeated Game

We now suppose that a pair of players is matched to play the repeated prisoners' dilemma, playing the stage game given in (1) in each of period $0, 1, \ldots$. With probability α, the players are both drawn to have a high discount factor that we denote by $\overline{\beta}$, and with probability $1 - \alpha$ they are drawn to have a low discount factor that we denote by $\underline{\beta}$. We interpret (and hereafter refer to) the discount factor as a continuation probability, governing the random length of the game.

Both players in a game observe the realized continuation probability. Players do not observe other's actions. Instead, if player i plays C in some period t, then player j observes the signal c with probability $1 - \varepsilon$ and observes signal d with probability ε. Similarly, when i plays D in some period t, then player j observes the signal c with probability ε and observes signal d with probability $1 - \varepsilon$. The signals are drawn independently across players and periods. Some of the analysis presented below will assume that ε is small.

We assume that the continuation probability $\underline{\beta}$ is sufficiently low that the only equilibrium in such a game features defection after every history, and so our analysis focuses on games with high continuation probabilities.[2]

2.3 Beliefs

The analogy-based expectations equilibrium (Jehiel [12]) that we examine rests on the two common pillars of equilibrium concepts—best responses and consistent beliefs. The departure from more familiar concepts lies in providing an explicit model of belief formation rather than assuming that these beliefs simply appear as part of the equilibrium concept.

We assume that the players have access to a record of past plays of the game, and adopt the frequencies of behavior observed in the record as their prior belief about their opponent's behavior in their current interaction. This record exhibits two imperfections.

First, we cannot expect the interactions in the record to correspond *exactly* to the current interaction—any pair of interactions will inevitably exhibit some differences, if nothing else reflecting the fact that previous interactions occurred earlier. The players will be unaware of many of these differences, and will ignore some differences in order to include observations they consider sufficiently similar to the current interaction—if the players are too exacting in the interactions they consider relevant, the record may be too sparse to be useful. Hence, whether the players cannot or choose not to distinguish them, the record will include some cases that are analogous to their own, even if not exactly the same. We are especially interested in games which differ in the readiness with which they support cooperation, which we capture in a quite stylized fashion by assuming that the record includes interactions with varying continuation probabilities, i.e., by taking $\alpha \in (0, 1)$.

[2] For our purpose, the universal defection arising in such games is what matters. Any alternative specification (for example, based on varying the monitoring technology) that would give rise to the same behavior would be equivalent.

Second, the record reports the actions played in each game, but not the private signals observed by the players. One interpretation is that these signals are unobservable, perhaps for the same reason that the signals are private. Our preferred interpretation is that the players may not even understand what form these signals might take, or what means the players in previous interactions may have for collecting and using information.[3]

Each previous play of the game contributes an observation to the record listing the actions taken in each period of the game. These observations will be of various lengths (recalling that β is a continuation probability), though all will be finite. A game that ends in period t contributes an observation to the record that contains $t+1$ *cases*, of lengths 1 to $t+1$. A case of length τ specifies the actions both players played in periods 0 through $\tau - 1$ (referred to as the history, taken to be null if $\tau = 1$) as well as the actions both players played in period τ.

A player uses the record to attach a probability that her opponent in her current interaction will cooperate, given that this interaction has reached some period t with history h_t. This probability is taken to be the empirical frequency of cooperation in those cases in the record corresponding to history h_t.

In principle, the player might aspire to attach a different probability to each history, just as the strategies in a standard repeated game can attach different probabilities to different histories. However, the number of such histories is (countably) infinite, putting the estimation of a probability for each beyond the reach of any plausible data set, regardless of recent advances in big data. Hence, the player classifies histories into categories (or analogy classes), and then calculates the empirical frequency of C and D actions for each category.

The equilibrium concept itself provides no guidance as to how many categories a player is likely to use in examining the data, nor how these categories are to be determined. Our intuition is that the number of categories is likely to be small. The categories may be chosen by the player, with their paucity reflecting limitations of the historical record or parsimony in the players' reasoning. Alternatively, the categories may represent aggregate statistics of the corresponding sort provided by third parties to new comers. From that perspective, a question of interest is how the choices of categories made by such third parties affect the possibility of cooperation.

In Sections 2–3, we suppose that players arrange histories into two categories, clean and dirty. A clean history is one in which no player has defected. A dirty history is one in which at least one player has defected. The player uses the record to calculate the probability p that a player cooperates after a clean history, and the probability q that a player cooperates after a dirty history. Section 4 examining how the results can depend on the specification of categories.

[3]One may sense some tension between the assumption that players cannot observe actions in their current interaction, and yet these actions appear in the record. A more realistic model would assume that the actions are reported in the record with some noise, an elaboration we eschew on the grounds that it adds more notation than insight.

For example, an interaction that ends in period 3 may contribute the following observation to the record:

$$\begin{array}{ll} \text{Player 1}: & CCCD \\ \text{Player 2}: & CCDD \end{array} . \tag{3}$$

This interaction contributes the following four cases to the record:

$$\begin{array}{lcccc} & \text{Case 1}: & \text{Case 2}: & \text{Case 3}: & \text{Case 4}: \\ \text{Player 1}: & C & CC & CCC & CCCD \\ \text{Player 2}: & C & CC & CCD & CCDD \end{array} . \tag{4}$$

These involve four histories, given by

$$\begin{array}{lcccc} & \text{Case 1}: & \text{Case 2}: & \text{Case 3}: & \text{Case 4}: \\ \text{Player 1}: & \emptyset & C & CC & CCC \\ \text{Player 2}: & \emptyset & C & CC & CCD \end{array} . \tag{5}$$

The first three of these histories are clean, as no one has defected, while the final history is dirty. Of the six actions (reported in (4)) taken after clean histories, five are cooperate and one is defect. Were this the only observation in the record, the players would estimate that the probability of cooperation after a clean history is $p = 5/6$ and the probability of cooperation after a dirty history is $q = 0$.

In grouping together the various dirty histories, player i does not distinguish whether it is player i who has defected, player j who has defected, or both. Obviously, this may make a difference—player j may be more likely to defect after histories in which player j has already defected than after histories in which only i has defected—and so player i's categorization of the histories potentially obscures some information. Given that i cannot estimate behavior after every one of the infinite number of histories, this is unavoidable.

2.4 The Equilibrium Concept

The analogy-based expectations equilibrium concept requires that the players' actions are optimal given their beliefs, and that these beliefs match the frequencies contained in a record generated by the equilibrium strategies. We refer to (Jehiel [12]) for a formal presentation. Formally, the requirement is that players' beliefs match those that would be generated by an *infinite* number of draws from the equilibrium strategies. In practice, the record will be finite. Indeed, this was our motivation for restricting attention to a small number of categories. As a result, the players' beliefs will be perturbed by estimation error. Our intention here is to isolate the implications of assuming that players' beliefs are given by the empirical frequencies of analogy classes, without the confounding effects of estimation error. We do so by working with an effectively infinite record, even while restricting players to a small number of analogy classes. In much the same vein, econometricians often assume away estimation error when considering questions of identification.

3 Equilibrium

3.1 Existence of Equilibrium

The candidate equilibrium behavior is that each player initially views the history as clean (and hence the opponent as cooperating with probability p), and cooperates. As long as i continues to cooperate, player i will update the probability i attaches to the event that the history is clean in light of the signals i receives. Once this probability drops below a threshold, player i switches to defecting. Once player i defects, i views the history as being dirty (and hence the opponent as cooperating with probability q), and defects thereafter.

The estimated probabilities p and q will both be positive, but less than one. The probability p will be positive because players with high continuation probabilities initially cooperate, and so the record will include clean histories exhibiting cooperation. This probability will be less than one because players with low continuation probabilities defect after the (clean) null history, and because there may be points in a high-continuation-probability exhibiting a first defection, giving us another clean history with a defection. In general, it may take some time for the other player in such a high-continuation-probability interaction to become sufficiently pessimistic as to defect, giving us some dirty histories with cooperation and hence positive q. But eventually, all players defect on dirty histories, ensuring $q < 1$.

3.1.1 Restless Bandits

We first fix probabilities p and q and examine an individual player's problem. We formulate the player's problem as a restless bandit problem, defined by the values p and q. These values are ultimately determined as equilibrium phenomena, but are viewed by the player as fixed parameters. There are two arms, a C arm (corresponding to cooperating) and a D arm (corresponding to defecting). The arms are characterized by a state that evolves but is identical across arms.

We let z_t, the probability the player attaches in period t to the event that the history is clean, be the state of both arms at time t. We have $z_0 = 1$, since all interactions start with the empty history, which is clean. As long as the C arm is pulled, z_t will evolve in response to the signals the player receives. If the D arm is pulled at time t, then both arms are in state 0 at time $t + 1$.

If the C arm is pulled at time t, then a c signal is observed and both arms move to state[4]

$$\phi(z, c) = \frac{zp(1-\varepsilon)}{zp(1-\varepsilon) + z(1-p)\varepsilon + (1-z)[q(1-\varepsilon) + (1-q)\varepsilon]} \tag{6}$$

with probability $zp(1-\varepsilon) + z(1-p)\varepsilon + (1-z)[q(1-\varepsilon) + (1-q)\varepsilon]$; while a d

[4]This posterior probability is the probability player i attaches the the event that the state is clean, given prior probability c and given that i played C and observed signal c, and is

signal is observed and both arms move to state

$$\phi(z,d) = \frac{zp\varepsilon}{zp\varepsilon + z(1-p)(1-\varepsilon) + (1-z)[q\varepsilon + (1-q)(1-\varepsilon)]} \tag{7}$$

with probability $zp\varepsilon + z(1-p)(1-\varepsilon) + (1-z)[q\varepsilon + (1-q)(1-\varepsilon)]$.

We can use these expressions to calculate that, given current state z, the expected value of the next state is pz. Hence, as long as $p < 1$, the player expects a decline in the probability that the opponent is clean.

This gives us a restless bandit (the states of unpulled arms evolve) rather than a simple bandit (only the state of the pulled arm evolves). Each time the C arm is pulled, it generates a current payoff of

$$zp + (1-z)q + [z(1-p) + (1-z)(1-q)](-k) \;=\; -k + (zp + (1-z)q)(1+k).$$

When the D arm is pulled, it generates a current payoff of

$$(zp + (1-z)q)(1+k).$$

It is clear that once the D arm is pulled, it is then optimal to thereafter pull the D arm. As a result, it is straightforward to calculate the value of the D arm, which is given by

$$\begin{aligned} W(z) &= (1-\overline{\beta})[(zp + (1-z)q)(1+k)] + \overline{\beta}q(1+k) \\ &= (1-\overline{\beta})z(p-q)(1+k) + q(1+k). \end{aligned} \tag{8}$$

We can view the D arm as paying $q(1+k)$ the first time it is pulled as well as every subsequent time, and can view $(1-\overline{\beta})z(p-q)(1+k)$ as an initial bonus the player receives (only) the first time he pulls the D arm. Only the initial bonus depends on the belief z_t.

3.1.2 Equilibrium in the Bandit Problem

The most interesting case is that in which $p > q$, so that players are more likely to cooperate after clean histories than after dirty histories.

Lemma 1 *For fixed $p > q$, there exists an optimal policy in the bandit problem. An optimal policy is characterized by a cutoff belief $\overline{z}$ such that a player cooperates if the belief z exceeds $\overline{z}$ and defects if z is less than $\overline{z}$.*

---

readily constructed from the following accounting of outcomes:

Currentstate	Probability	Signal/Next state	Probability
Clean	z	c/Clean	$zp(1-\varepsilon)$
Clean	z	d/Clean	$zp\varepsilon$
Clean	z	c/Dirty	$z(1-p)\varepsilon$
Clean	z	d/Dirty	$z(1-p)(1-\varepsilon)$
Dirty	$1-z$	c/Dirty	$(1-z)[q(1-\varepsilon) + (1-q)\varepsilon]$
Dirty	$1-z$	d/Dirty	$(1-z)[(1-q)(1-\varepsilon) + q\varepsilon]$

.

Proof. The existence of an optimal policy is standard, having been established by Whittle [23], and follows from dynamic programming arguments.

The statement that the optimal policy takes the form of a cutoff belief $\overline{z}$ is the intuitive result that if there is a belief at which one is willing to cooperate, then learning that the history is more likely to be clean will also make one willing to cooperate. To establish this, let $V(z)$ be the value of cooperating at belief z, and thereafter proceeding optimally (with the existence result ensuring that this is well defined). Suppose it is optimal to cooperate at belief z, or $V(z) \geq W(z)$. Now consider $z' > z$. We know, from (8), that

$$W(z') - W(z) = (1 - \overline{\beta})(p - q)(1 + k)(z' - z).$$

We also know that V is given by the sum of the current payoff $(1 - \overline{\beta})[-k + (zp + (1 - z)q)(1 + k)]$ plus a continuation payoff, allowing us to write

$$V(z') - V(z) = (1 - \overline{\beta})(p - q)(1 + k)(z' - z) + \overline{\beta}[\mathbb{E}V(\phi(z', \cdot)) - \mathbb{E}V(\phi(z, \cdot))].$$

It is immediate that $V(z)$ is increasing in z, and hence that $\mathbb{E}(V(z, \cdot))$ is increasing in z. A comparison then gives

$$V(z') - W(z') \geq V(z) - W(z).$$

Hence, we must have $V(z') \geq W(z')$, and so we have the desired threshold result. ■

Remark 1 We can convert our restless bandit into a stationary bandit to which an index applies. Let $z(p-q)(1+k)$ be denoted by $h(z)$. In period 0, the player makes no decision, and receives payoff $h(z_0)$. In the period 1, the player chooses either C, for payoff $-(k/\overline{\beta}) + \mathbb{E}h(z_1|z_0)$, or chooses D, for a payoff of 0. In the period 2, assuming C was chosen in period one, the player chooses either C , for payoff $-(k/\overline{\beta}) + \mathbb{E}h(z_2|z_1)$, or chooses D, for a payoff of 0. In general, the D arm gives a payoff of 0 and is an absorbing action, while in each period t the C are gives payoff $-k/\overline{\beta} + \mathbb{E}h(z_t|z_{t-1})$. The idea is that no matter what, the player receives the period-1 bonus $h(z_0)$. Then, in the ordinary representation, the player can pay the cost k in period 0 in order to also receive the bonus in period 1, which from period 0's point of view has the value $\mathbb{E}h(z_1|z_0)$. But we can then represent this as the player paying in period 1 the cost $k/\overline{\beta}$ for the reward $\mathbb{E}h(z_1|z_0)$. Continuing in this way, we obtain a stationary bandit. We can then apply the familiar Gittins index characterization (Whittle [23]) of the optimal policy in the stationary bandit to conclude that this policy is a threshold policy that cooperates above some belief $\overline{z}$ and defects below that belief. ■

3.2 Equilibrium in the Repeated Game

An equilibrium in the repeated game requires not only a solution to the bandit problem for fixed values of p and q, but also that this solution generates a record of past plays that is in turn consistent with p and q. Hence, given p and q, the solution to the bandit problem induces values $\hat{p}$ and $\hat{q}$ in the data, where these are the probability of cooperating after a clean history and after a dirty history. We seek a fixed point with $p = \hat{p}$ and $q = \hat{q}$.

3.2.1 Perfect Monitoring

We first suppose that $\varepsilon = 0$, so that monitoring is perfect, and confirm that we recover familiar results. Suppose each player adopts the strategy of cooperating after clean histories and defecting after dirty histories. Then the observations contained in the record will be either perpetual defection, arising in low continuation probability games, or perpetual cooperation, arising in high continuation probability games.

Given this record, each player will estimate an interior value for p, since they observe cooperation after all of the (clean) histories that appear in high continuation probability games, but defection after the null (and hence clean) history in a low continuation probability game. Each player will estimate $q = 0$, observing dirty histories only in low continuation probability games whose players routinely defect.

When will the proposed behavior constitute an equilibrium? We can calculate the probability p:

$$p = \frac{\alpha \sum_{t=0}^{\infty} \overline{\beta}^t}{\alpha \sum_{t=0}^{\infty} \overline{\beta}^t + (1-\alpha)} = \frac{\alpha}{\alpha + (1-\alpha)(1-\overline{\beta})}.$$

The numerator calculates the frequency of clean histories after which a player cooperates.[5] The denominator calculates the frequency of all clean histories.

To confirm that we have an equilibrium, we need only verify the incentive constraint that a player be willing to cooperate at a clean history. The payoff from cooperating is given by

$$V = [p + (1-p)(-k)] + p\overline{\beta}V = \frac{(1+k)p - k}{1 - p\overline{\beta}},$$

while the payoff from defecting is (1+k)p, and hence the incentive constraint $V(z) \geq W(Z)$ is

$$\frac{1+k}{k} p^2 \overline{\beta} \geq 1,$$

or, using our solution for p and rearranging,

$$\overline{\beta} \geq \frac{k}{1+k} \left(\frac{\alpha + (1-\alpha)(1-\overline{\beta})}{\alpha} \right)^2.$$

[5] Cooperation requires that a high continuation probability be drawn, giving us the initial α. Then, with probability 1 we get a period-0 history added to our list, with probability $\overline{\beta}$ we also get a period-1 history added to the list, with probability $\overline{\beta}^2$ we get a period-2 history, and so on.

This inequality holds for sufficiently large $\overline{\beta}$, but is more demanding than the corresponding requirement (2) from the classical perfect monitoring game. The two criteria coincide when $\alpha = 1$ (and hence $p = 1$). As α falls below one, so does the estimated value of p, and hence the value of cooperation, thus making the equilibrium condition more stringent.

3.2.2 Universal Cooperation?

Given the previous subsection's observation that letting $\alpha \to 1$ sharpens the result, why not view the presence of low-continuation-probability interactions as a nuisance and simply set $\alpha = 1$? Doing so allows an equilibrium to creep in that we view as an artifact of a record that is too sterile.

Suppose $\alpha = 1$, so there are no low-continuation-probability interactions, but now let $\varepsilon \in (0, 1)$, so that monitoring is imperfect. Suppose the players adopt the strategy of cooperating after every history. Then the data will give the estimate $p = 1$, since every clean history leads to cooperation, and will provide no means of estimating q, the probability of cooperation after a dirty history. Suppose the players then take $q = 0$. In this case, player i's view of player j is that j will cooperate as long as i does (so that the history is clean), and that j will switch to permanent defection whenever i defects (giving a dirty history). As long as $\overline{\beta} \geq \frac{k}{1+k}$, we then have an equilibrium that somewhat magically supports cooperation, regardless of what signals the players receive and no matter how precise or imperfect the monitoring.

A basic concern with this construction is that it relies too heavily on the beliefs that $p = 1$ and $q = 0$ (or more generally that q is sufficiently small). Both beliefs are problematic.

The estimate $p = 1$ depends on the players including in the record only games identical to their own, a feat we generally regard as impossible. The record provides no means of estimating q, and so we have arbitrarily set $q = 0$. We could just as well have picked other values of q, not all of which support the candidate equilibrium.[6] More conceptually, the vision behind analogy-based equilibria is that players construct their analogy classes so as to effectively organize the information they can extract from the record. In keeping with this, we do not expect players to work with analogy categories for which the record provides no data. Under this view, the very fact that we must seek insight into q from some mechanism outside the model is an indication that something in the model is amiss. We view the reasoning organized around these two analogy classes *and* $\alpha < 1$ as reasonable.

[6] One route to a firmer foundation for q would be to incorporate trembles into the actions (cf. footnote 3), so that the record would always include some diry histories. In at least one natural way of specifying the trembles, in which players assume that it is intended rather than realized actions that shape behavior, the result would be an estimate of $q = 1$, disrupting the equilibrium.

3.2.3 Universal Defection

It is no surprise, given that we are working with the prisoners' dilemma, that there is an equilibrium in the repeated game featuring relentless defection. If the candidate equilibrium strategies specify defection after every history, then the record will including only strings of mutual defection of varying length. Players observing this record will estimate $p = 0$ (having observed only defection after clean histories, which in this case are only the null histories) and $q = 0$ (having observed only defection after all other histories, all of which are dirty). Defection after every history is then indeed an equilibrium strategy, and would be no matter what the analogy classes.

3.2.4 Cooperation

We now turn to the existence of nontrivial (i.e., exhibiting at least some cooperation) equilibria. We assume monitoring is imperfect ($\varepsilon \in (0, 1/2)$) and players sometimes find themselves in circumstances that cannot support cooperation ($\alpha < 1$).

It is intuitive that we can sustain cooperation only if $p > q$:

Lemma 2 *In any nontrivial equilibrium, $p > q$.*

Proof Rewrite the current payoffs in the bandit problem from pulling the C arm, the first pull of the D arm, and subsequent pulls of the D arm, as

$$\begin{aligned} C: &\quad -k + q(1+k) + z(p-q)(1+k) \\ \text{first } D: &\quad q(1+k) + z(p-q)(1+k) \\ \text{subsequent } D: &\quad q(1+k). \end{aligned}$$

No matter what happens, the player receives $q(1 + k)$ in every period. Given this, we might as well normalize payoffs by subtracting this amount from the payoff in every period, and view our system as one in which the payoffs from the various arms are:

$$\begin{aligned} C: &\quad -k + z(p-q)(1+k) \\ \text{first } D: &\quad z(p-q)(1+k) \\ \text{subsequent } D: &\quad 0. \end{aligned}$$

We can then say that in each period the player has the option of paying a cost k in order to receive a bonus of $z(p - q)(1 + k)$. In the first period that player *fails* to pay the cost, the bonus is paid, but the bonus is never again paid. The players will pay the fee only if the bonus is positive, which requires $p > q$. ■

We now show that we have a nontrivial equilibrium as long as either the monitoring is sufficiently precise, the high continuation probability is sufficiently high, or low-continuation-probability interactions are sufficiently few. In each case, we require that the temptation to defect, captured by k, not be too large.[7]

Proposition 1 *Let $\varepsilon \in (0,1)$ and $\alpha \in (0,1)$.*

[1.1] Suppose

$$k < \frac{\overline{\beta}\alpha^2}{4(1-\overline{\beta}\alpha^2)}.$$

Then there exists $\overline{\varepsilon}$ such that for all $\varepsilon < \overline{\varepsilon}$, a nontrivial equilibrium exists.

[1.2] Suppose

$$k < \frac{\alpha^2}{4-\alpha^2}.$$

Then there exists $\overline{\overline{\beta}}$ such that for all $\overline{\beta} > \overline{\overline{\beta}}$, a nontrivial equilibrium exists.

[1.3] Suppose

$$k < \frac{\overline{\beta}}{1-\overline{\beta}}.$$

Then there exists $\underline{\alpha}$ such that for all $\alpha > \underline{\alpha}$, a nontrivial equilibrium exists.

Appendix 7.1 contains the proof.

The argument proceeds by first showing that a necessary and sufficient condition for the existence of a nontrivial equilibrium is

$$\frac{k}{1+k} \leq \overline{\beta}p(p-q).$$

Notice that if $p = 1$ andf $q = 0$, this is equivalent to the criterion (2) found in the conventional repeated prisoners' dilemma. The results then follow by establishing bounds on the values of p and q under the various conditions. In doing so, we find that a large continuation probability plays two roles. First, as is typical in repeated games, we need the future to be sufficiently important. Second, an increase in $\overline{\beta}$ can decrease the estimate of q extracted from the record, making defecting less attractive.

3.2.5 The Value of Cooperation

Proposition 1 ensures the existence of an equilibrium with some cooperation, but makes no comment as to how much cooperation we can expect and makes no statement about the payoff implications of that cooperation. There remains the possibility that cooperation is a fleeting phenomenon with negligible payoff implications. Our next proposition establishes conditions under which the equilibrium payoff approaches 1, the payoff of the grim equilibrium in a game of perfect monitoring. Appendix 7.2 proves:

[7] No such restriction on k is required in the standard perfect-monitoring formulation, and indeed this restriction becomes moot in the limit as both the monitoring technology becomes arbitrarily precise and low-continuation-probability interactions become arbitrarily rare.

Proposition 2

[2.1] Suppose $\overline{\beta}$ satisfies

$$\overline{\beta} > \frac{k}{1+k}. \tag{9}$$

Then there exists $\overline{\alpha} < 1$ such that for $\alpha \in (\overline{\alpha}, 1)$, there exists a sequence of equilibria such that, in the limit as $\varepsilon \to 0$, the equilibrium payoff approaches 1, the payoff of persistent, mutual cooperation.

[2.2] Suppose $\overline{\beta}$ satisfies

$$\overline{\beta} > \frac{2k}{1+k}. \tag{10}$$

Then there exists a sequence of equilibria such that, in the limit as $\alpha \to 1$, the equilibrium payoff approaches 1, the payoff of persistent, mutual cooperation.

In each case, we require that high continuation probabilities be sufficiently high, relative to the temptation to defect k. This is expected—without a sufficiently likely future, we cannot get cooperation off the ground. The additional conditions ensure that this cooperation is persistent rather than transitory.

The key to persistent cooperation is ensuring that the posterior belief that one's opponent is clean does not decline too rapidly. The first result ensures this by requiring that low-continuation-probability interactions be relatively rare and then examining the limit as the monitoring becomes arbitrarily precise. The paucity of low-continuation-probability interactions ensures that the estimate of p drawn from the record is large, in turn ensuring that players think it unlikely that their opponents have spontaneously switched from clean to dirty. The precise monitoring ensures that erroneous (posterior-depressing) signals are unlikely.

This first result requires the probability ε of mistaken signals to be small relative to $1-\alpha$, the probability of low-continuation-probability interactions. This order of limits brings us back to the reasoning that ensures the grim strategy is *not* an equilibrium in a conventional repeated game of private monitoring. To deter defection, adverse signals must be interpreted as reflecting defection. In the first period of a conventional repeated game of private monitoring, the equilibrium hypothesis of grim strategies precludes this, with the players instead interpreting the adverse signal entirely as a whim of the noisy monitoring technology. Letting $\alpha < 1$ in our context ensures that players will consider the possibility that an adverse signal reflects a defection. However, if ε is relatively large, it will still be considered overwhelmingly likely that the noisy monitoring technology is at fault. To create the requisite incentives, the monitoring technology must be relatively precise, captured by the requirement that ε be small relative to $1-\alpha$.

The second condition places no restriction on the precision of the monitoring, requiring only that low-continuation-probability interactions players be relatively few. This result relies on the observation that if low-continuation-probability interactions are relatively rare, then adverse signals will be interpreted as quirks of the noisy monitoring rather than indications of defection.

This allows the posterior that the opponent is clean to remain high, as needed for long-lasting cooperation. However, this gives rise to precisely the type of inference that scuttles cooperation in standard repeated games of private monitoring. The argument in the current setting relies on the misspecificaiton in the players' models of their interaction to verify the incentives to cooperate are not disrupted in the process. Example 3.3.1 provides a particularly stark illustration of the mechanism.

3.3 Examples

We illustrate the results with two examples. The first example illustrates the forces behind Proposition 2.2 while the second illustrates Proposition 2.1. To keep the notation uncluttered, we set $\underline{\beta} = 0$ and denote $\overline{\beta}$ simply by β.

3.3.1 Example 1: Uninformative Monitoring

One's initial reaction is that imprecise monitoring must be inimical to cooperation, as it makes it more difficult to detect and punish deviations. However, the key to supporting cooperation is that players fear their own deviations will be punished, and this fear in turn is a byproduct of the players' misspecified, two-category model of behavior. Imprecise monitoring may then even be helpful, by making players less likely to suspect opponents have deviated.

To illustrate this, we let $\varepsilon = 1/2$, so that signals carry no information. As before, each player believes that the opponent cooperates with probability p after clean histories and probability q after dirty histories, Player i's strategy is to cooperate as long as the posterior probability z_t of a clean history remains above a threshold $\overline{z}$, and defect when $z_t < \overline{z}$.

In equilibrium, we will see an initial period in which both players cooperate (given a high continuation probability). Each player then continues to cooperates, as the posterior probability that the history is still clean continually falls, until some period $T+1$, at which point z_{T+1} dips below the threshold $\overline{z}$, and the players then defect. As a result, the record will consist entirely of interactions in which the two players initially cooperate, and then simultaneously defect for the first time, and then continue to defect. The first simultaneous defection makes the history dirty, and the subsequent defections ensure that the record never exhibits cooperation after a dirty history. The players' estimate q of the probability of cooperation after a dirty history is thus 0.

We first calculate z_t. Since i's model of player j is that in each period of a clean history, j cooperates with probability p, the probability that the history is still clean (given no defection by i) upon having arrived at period t is

$$z_t = p^t. \tag{11}$$

Notice that z_t falls as t grows—player i believes that in period, j cooperates with probability p, and so as t increases, the probability that j has not yet defected declines.

Next, let us calculate the probability p. In equilibrium, a player who has drawn a high continuation probability will cooperate in periods $0, \dots, T$ for some T, and then defect. We then have

$$p = \frac{\alpha(1+\beta+\cdots+\beta^T)}{\alpha(1+\beta+\cdots+\beta^{T+1})+1-\alpha} = \frac{\alpha(1-\beta^{T+1})}{\alpha(1-\beta^{T+2})+(1-\alpha)(1-\beta)}. \quad (12)$$

The numerator calculates the frequency of clean histories after which an player cooperates.[8] The denominator calculates the frequency of all clean histories.

As T grows from 0 to ∞, the value of p grows from $\alpha/(1+\beta\alpha)$ to $\alpha/(1-\beta+\beta\alpha)$. The latter value approaches 1 as either α (because then there are no low-continuation-probability interactions, which are the only ones exhibiting defection after healthy histories when $T=\infty$) or β (because the low-continuation-probability defections are then swamped) approaches 1.

Our task is to find T such that the induced values of $z_T \geq \overline{z} \geq z_{T+1}$ and p satisfy the incentive constraints. The incentive constraints for these periods will ensure the incentive constraints hold in other periods. The value of cooperation in period T is $z_T[p(1+\beta p(1+k))+(1-p)(-k)]+(1-z_T)(-k)$. The value of defecting in period T is $(1+k)z_T p$. Subtracting the second from the first, player i prefers to cooperate if $z_T[p(-k+\beta p(1+k))+(1-p)(-k)]+(1-z_T)(-k) \geq 0$, or

$$z_T \geq \frac{k}{(1+k)\beta p^2}.$$

The equilibrium condition is then

$$z_T \geq \frac{k}{(1+k)\beta p^2} \geq z_{T+1}. \quad (13)$$

We can use (11) and rearrange to obtain

$$(p(T))^{T+2} \geq \frac{k}{(1+k)\beta} \geq (p(T))^{T+3}.$$

We clearly have $(p(T))^{T+2} > (p(T))^{T+3}$. Both functions initially increase in T and then decline to zero as $T \to \infty$. As long as k is not too large and β not too small, there will exist a value T satisfying (13) and hence an equilibrium in which cooperation persists for the first T periods.

One might wonder whether, as the continuation probability β approaches one, this initial cooperation fades into insignificance, with payoffs approaching zero ($\beta^T \to 1$), or whether are payoffs bounded away from zero ($\beta^T < 1$)? The latter is the case.[9]

---

[8]Cooperation requires that a high continuation probability, giving us the initial α. Then, with probability 1 we get a period-0 history added to our list, with probability β we also get a period-1 history added to the list, with probability β^2 we get a period-2 history, and so on, through probability β^T that we get a period-T history. After that, there is no more cooperation.

[9]To see this, we first note that

$$p^T \approx \frac{k}{1+k} \quad (14)$$

Now consider what happens as $\alpha \to 1$, as in Proposition 2.2. It remains the case that for fixed α, we have $\lim_{T\to\infty}(p(T))^{T+2} = 0$, but also the case that for large T, we have $\lim_{\alpha\to 1}(p(T))^{T+2} = 1$. Hence, as α approaches 1, as long as $k/(1+k)\beta < 1$, the equilibrium value of T will grow arbitrarily large. The equilibrium payoff will thus approach the payoff of permanent mutual cooperation, as in Proposition 2.2.

Two conditions are required in order for cooperation to last indefinitely (relative to the expected length of the interaction). First, player i must persist in the belief that player j is sufficiently likely to cooperate (as long as i does). This is ensured by the assumption that α is close to one, so that the value of p estimated from the record is close to one. Second, player i must believe that a defection will trigger a punishment. Literally speaking, this cannot be the case. Player j observes nothing about i's behavior, and so i's defection cannot affect j's future play. However, the record contains many cases in which defections are followed by punishments, and i's coarse modeling of the record interprets these correlations as causations. In effect, the misspecified model of behavior that i has extracted from the record supplies the requisite fear of punishment.

A situation of absolutely no monitoring is clearly an extreme one, and we do not offer this as a realistic model of cooperation. However, it helps isolate the link between the coarse inferences drawn from the record and equilibrium behavior. These forces persist in settings with more informative modeling, as in Proposition 2.2, where we view them as reasonable.

3.3.2 Example 2: Informative Monitoring

Now consider the informative case in which $\varepsilon \in (0,1)$. We can calculate the posterior probability (denoted by $\phi(z,c)$) that the history is clean, given that i has hitherto always cooperated, that i has a prior probability z that the history

---

as β gets close to 1. To verify this, note that it suffices for (14) that p converges to one. But if p does not converge to one, then (13) ensures that T would remain bounded, and which point (12) ensures that p converges to one, leading to a contradiction. We can thus use (14) to write

$$p \approx 1 - \frac{y}{T} \tag{15}$$

where $e^{-y} = \frac{k}{1+k}$. Rewrite (12) as

$$p = 1 - \frac{\alpha\beta^{T+1} + 1 - \alpha}{\alpha(1-\beta^{T+2}) + (1-\alpha)(1-\beta)}(1-\beta) \tag{16}$$

and postulate that $\beta \approx 1 - \frac{x}{T} + o(\frac{1}{T})$, implying that $\beta^T \to e^{-x}$ as β converges to 1. We then have that

$$\frac{\alpha\beta^{T+1} + 1 - \alpha}{\alpha(1-\beta^{T+2}) + (1-\alpha)(1-\beta)}(1-\beta) \approx \frac{\alpha e^{-x} + 1 - \alpha}{\alpha(1-e^{-x})}\frac{x}{T}\ (+o(\frac{1}{T})).$$

Inserting into (16) and identifying the $1/T$ terms in (15) and (16), we have

$$\frac{\alpha e^{-x} + 1 - \alpha}{\alpha(1-e^{-x})}x = -\ln\frac{k}{1+k}.$$

This gives us a positive value of x, with $e^{-x} < 1$ being the limit of β^T.

is clean, and that i observes a c signal, which is given by (6). Two forces appear in forming this posterior belief. First, a c signal is an indication that it is likely the history is clean, and so tends to push the posterior upward. However, there is always the $1-p$ probability that a player defects at a clean history and hence turns dirty, and this pushes the posterior downward. When the prior z is very large, we expect the second force to dominate, as the good signal carries almost no information, and so the primary forces is that even clean histories have a constant hazard of turning dirty. When z is relatively small, the c signal is more informative, and so we expect the first force to dominate. This suggests that we can find a fixed point z^* as the value of z that solves

$$z = \frac{zp(1-\varepsilon)}{zp(1-\varepsilon) + z(1-p)\varepsilon + (1-z)[q(1-\varepsilon) + (1-q)\varepsilon]}.$$

We can solve for (using the presumption that $p > q$, so that this makes sense)

$$z^* = \frac{p(1-\varepsilon) - [q - 2q\varepsilon + \varepsilon]}{(p-q)(1-2\varepsilon)}.$$

The expression given in (6) is increasing in the prior z. In equilibrium, i's posterior belief that the history is clean starts at 1, and then drifts downward toward z^* as long as i observes a constant stream of c signals. In general, c signals push i posterior either downward toward z^* from above, or upward toward z^* from below.

The posterior probability (denoted by $\phi(z,d)$) that the the history is dirty, given that i has hitherto always cooperated, and that i has a prior probability z that the history is clean and i observes a d signal, is given by (7). One can check that this posterior is always less than z—it is always bad news to observe a d signal. As ε approaches 0, this posterior also approaches 0—when monitoring is arbitrarily close to perfect, a d signal makes it arbitrarily likely that the opponent has defected and hence the history is dirty.

A pure strategy for player i is a function that maps from the collection of finite strings of c and d signals into the set of actions $\{C, D\}$. We can immediately add the restriction that if any string maps to D, then so does every continuation of that string. Once player i defects, i takes it for granted that the history is dirty, hence j's behavior is thereafter impervious to any actions of i, ensuring that i finds it optimal to thereafter defect.

We first argue that we can find values of our parameters such that it is an equilibrium for player i to cooperate after any sequence containing only c signals, and to defect upon observing the first d signal, giving us an equilibrium corresponding to Proposition 2.1. The path of equilibrium play is then that each player begins by assigning probability 1 to the history being clean, and by cooperating. The string of mutual cooperation continues, with the posterior that the history is clean drifting downward toward z^*, until the first d signal appears. The player receiving the d signal than attaches probability at most z^- to the event that the history is clean, and thereafter defects.

The posterior that the history is clean, following a d signal, is higher when the prior probability of being clean is higher (this requires $p > q$, which we will verify), and hence we can give an upper bound on this posterior by looking at the update when the prior is 1:

$$z^{-} \quad \leq \quad \frac{p\varepsilon}{p\varepsilon + (1-p)(1-\varepsilon)}.$$

We have a bound on q, given by

$$q \quad \leq \quad \varepsilon/2.$$

To see this, consider a dirty history in which just one player (say j) has defected. Then in the next period j defects with probability 1 and i cooperates with probability ε (the probability that i has seen a c signal in the most recent period, despite j's defection), giving a probability of cooperation of $\varepsilon/2$. The value of q is less than this, since the record also contains dirty histories in which both players are dirty, and hence both defect in the next period with probability 1.

We can calculate p, obtaining:[10]

$$p = \frac{\alpha + \alpha(1-\varepsilon)\sum_{n=1}^{\infty} \beta^n (1-\varepsilon)^{2(n-1)}}{\alpha + \sum_{n=1}^{\infty} \beta^n (1-\varepsilon)^{2(n-1)}}$$

Calculating that

$$\sum_{n=1}^{\infty} \beta^n (1-\varepsilon)^{2(n-1)} = \frac{\beta}{(1-\varepsilon)^2} \sum_{n=1}^{\infty} \beta^n (1-\varepsilon)^{2n} = \frac{\beta}{(1-\varepsilon)^2} \frac{1}{1-\beta(1-\varepsilon)^2},$$

we can insert this to obtain

$$p = \frac{\alpha + \alpha(1-\varepsilon)\frac{\beta}{(1-\varepsilon)^2}\frac{1}{1-\beta(1-\varepsilon)^2}}{\alpha + \frac{\beta}{(1-\varepsilon)^2}\frac{1}{1-\beta(1-\varepsilon)^2}} = \alpha \frac{(1-\varepsilon)^2(1-\beta(1-\varepsilon)^2) + (1-\varepsilon)\beta}{(1-\varepsilon)^2(1-\beta(1-\varepsilon)^2) + \alpha\beta}.$$

[10] The following table gives the relative frequencies with which clean histories of various lengths appear in the record, and the probability of cooperation after histories of such length:

Length	Frequency of History	Probability of Cooperation
0	1	α
1	$\alpha\beta$	$(1-\varepsilon)$
2	$\alpha\beta^2(1-\varepsilon)^2$	$(1-\varepsilon)$
3	$\alpha\beta^3(1-\varepsilon)^4$	$(1-\varepsilon)$
4	$\alpha\beta^4(1-\varepsilon)^6$	$(1-\varepsilon)$
5	$\alpha\beta^5(1-\varepsilon)^8$	$(1-\varepsilon)$
	$\vdots$	

To see obtain the first term, we note that with probability one, every game contributes a null history to the record, which is clean, and players cooperate after this history if they have a high continuation probability, which occurs with probability α. For the second term, note that with probability $\alpha\beta$, a game also contributes a 1-period history to the record, which is clean. After this history, each player cooperates with probability $1-\varepsilon$, which is the probability they received a c signal in the previous period. Then, with probability $\alpha\beta^2(1-\varepsilon)^2$, a game also contributes a clean 2-period history to the record, with the additional β term reflecting the fact that the game must have reached another period, and the $(1-\varepsilon)^2$ capturing the fact that both players must have cooperated in the previous period in order for the history to be clean.

The key characteristic we will use is that p goes to $\frac{\alpha}{1-\beta+\alpha\beta}$ as ε goes to zero. Hence, as long as $\varepsilon < 1/2$ is sufficiently small, we have $p > q$, as needed. In addition, this gives

$$\lim_{\varepsilon \to 0} z^* = 1.$$

We can also calculate

$$\lim_{\varepsilon \to 0} z^- = 0.$$

This calculation reflects the fact that we have fixed α, which in turn ensures that $p < 1$, while letting ε approach zero. As $\varepsilon \to 0$, a d signal is overwhelmingly likely to have come from a defection. This alone is not enough to ensure that z^- approaches zero, because (if p approaches 1) defections may themselves be yet more overwhelmingly unlikely. However, p is approaching $\frac{\alpha}{1-\beta+\alpha\beta}$, ensuring that a d signal is interpreted as a defection, and hence that z^- is arbitrarily small.

We know from Lemma 1 that there is a cutoff belief $\overline{z}$ such that player i cooperates for higher beliefs and defects for lower beliefs. We then need to show that $z^- < \overline{z} < z^*$.

Let $V(z)$ be the value for a player who has hitherto not defected and observed no d signals, believes the history to be clean with probability z $(\geq z^*)$, and who cooperates in the current period. Then we have

$$\begin{aligned} V(z) \;=\; & (1-\beta)[(zp+(1-z)q)+(1-(zp+(1-z)q))(-k)] \\ & + \beta[zp(1-\varepsilon)+z(1-p)\varepsilon+(1-z)q(1-\varepsilon)+(1-z)(1-q)\varepsilon]V(\phi(z,c)) \\ & + \beta[1-(zp(1-\varepsilon)+z(1-p)\varepsilon+(1-z)q(1-\varepsilon)+(1-z)(1-q)\varepsilon)]W(\phi(z,d)), \end{aligned}$$

recalling $\phi(z,c)$ is the posterior that the history is clean following prior z and signal c. The first line is the current-period payoff, the second line is the discounted value of the probability of a c signal times the continuation payoff $V(\phi(z,c))$ in the event of such a signal, and the third line is the discounted probability of a d signal times the continuation payoff $q(1+k)$ in the event of such a signal. This value is decreasing in z, and obtains its infimum in the limiting case of $z = z^* = \phi(z^*, c)$. Letting V^* denote this value, it is the solution to

$$\begin{aligned} V^* \;=\; & (1-\beta)[(z^*p+(1-z^*)q)+(1-(z^*p+(1-z^*)q))(-k)] \\ & + \beta[z^*p(1-\varepsilon)+z^*(1-p)\varepsilon+(1-z^*)q(1-\varepsilon)+(1-z^*)(1-q)\varepsilon]V^* \\ & + \beta[1-(z^*p(1-\varepsilon)+z^*(1-p)\varepsilon+(1-z^*)q(1-\varepsilon)+(1-z^*)(1-q)\varepsilon)]W(\phi(z^*,d)). \end{aligned}$$

From (8), we have

$$W(z) = (1-\beta)z(p-q)(1+k) + q(1+k).$$

The incentive constraints for equilibrium are

$$\begin{aligned} V(z^*) &\geq W(z^*) \\ V(z^-) &\leq W(z^-). \end{aligned}$$

To check these conditions, we first note that as ε gets small, we have $z^* \to 1$, $z^- \to 0$, and $q \to 0$, and hence we have limiting values for $W(\phi(z,d))$ and $W(z^-)$ of 0. This in turn ensures that $V(z^-) = (1-\beta)(-k)$, giving the second incentive constraint—players will prefer to defect when the strategies call for them to do so. We can also solve for

$$V(z^*) = p + (1-p)(-k).$$

The first incentive constraint, given by $p + (1-p)(-k) \geq (1-\beta)p(1+k)$, then becomes

$$\beta p \geq \frac{k}{1+k},$$

which, using our limiting expression for p, becomes

$$\frac{\alpha\beta}{1-\beta+\alpha\beta} \geq \frac{k}{1+k}.$$

If we were to now let α approach one, then we would recover the limit $\beta \geq \frac{k}{1+k}$ from the traditional repeated game of perfect monitoring. This gives us an equilibrium of the type described in Proposition 2.1, with a value approaching one, the value of permanent cooperation.

Much more briefly, we note that we could alternatively keep ε fixed, so that monitoring is inherently noisy. We would then have equilibrium strategies exhibiting cooperation as long as the probability the history is clean remains above a cutoff $\overline{z}$, with the first d signal no longer necessarily prompting defection. If any of the conditions of Proposition 1 are met, we will have $\overline{z} < 1$ and hence the equilibrium will exhibit at least some cooperation. If α approaches 1 (now with ε fixed), the value of this cooperation will again approach one, as in Proposition 2.2.

4 What Difference Does an Analogy Make?

The players' analogy classes reflect their model of their strategic interaction and shapes their behavior. This section explores the implications of different analogy classes. To keep things simple, we continue to let $\underline{\beta} = 0$ and to denote $\overline{\beta}$ simply as β.

Equilibrium cooperation rests on three pillars. First, player i must believe that at least under the right circumstances, player j will cooperate. Second, player i must believe that if i defects, then j will be more likely to defect. Third, the difference in j's behavior must be large enough to make it worthwhile for i to forsake the immediate payoff gains from defection.

These conditions clarify why the grim strategy is not an equilibrium in a conventional repeated game of private monitoring—condition two fails in the first period. The basic difficulty is that it is difficult to ensure both conditions one and two. Suppose player j receives a d signal in the first period. Under the equilibrium hypothesis, j's interpretation is that i cooperated and the signal is erroneous. Given this, j will continue to cooperate, recognizing (in keeping with condition one) that doing so makes it more likely that i receives good signals and continues to cooperate. However, condition two then fails for i, as i now does *not* fear that a first-period defection will make it more likely that her opponent defects.

In our setting, player i's model of the interaction, captured in her analogy classes, is that all interactions start clean, giving rise to the prospect of cooperation and hence the first condition. In addition, i believes that a defection renders the history dirty, and hence defection more likely, giving the second condition. Finally, the proofs of Propositions 1 and 2 involve showing that the difference $p - q$ in the probability of defection after clean and dirty histories is sufficiently large, giving the third condition.

Player i's assessment of the adverse consequences of a defection reflect the misspecification inherent in i's analogy classes. Player j cannot observe i"s action, and indeed the monitoring may be sufficiently noisy (as allowed in Proposition 2.2) that j receives virtually no information about i. If i formed a separate estimate of j's behavior for every possible history, i would find some histories where a defection would be relatively innocuous, as in the first period of a conventional repeated game of private monitoring. Instead, i averages many histories together into a single estimate of the effects of a defection, obtaining an average effect large enough to support cooperation.

In the next two subsections, we have in mind that the record of past play does not contain (or the players do not exploit) information about the exact timing at which the previous actions were chosen, as is the case with the clean-and-dirty two-category case we have been examining. This makes it impossible to condition the categorization of histories on how the actions of the two players compared in the same period. In the third subsection, we briefly suggest how a simple categorization based on the profile of actions could help coordination in some cases.

4.1 Proportions of Defection

Suppose players classify histories into two categories, with the dirty category being that more than θ proportion of the actions in the history have been defections, and the clean category being those with fewer than θ proportion of defections. This might be appropriate for a case in which the record does not contain detailed information on individual actions, perhaps instead simply giving an indication of whether an industry or group of countries is currently in a cooperative or contentious phase of their relationship. We can view our first analysis as corresponding to the case in which $\theta = 0$.

We ask here whether, for a fixed continuation probability β, in the limits as $\alpha \to 1$ and $\varepsilon \to 0$, we can support an equilibrium in which players cooperate until receiving their first d signal. If so, we have an equilibrium of the type described in Proposition 2.1 and Section 3.3.2.

Let us accordingly fix the (high) continuation probability β and assume that players in high-continuation-probability interactions cooperate until received their first d signal, and defect thereafter. As ε becomes small, the typical experience in a high-continuation-probability interaction is that both players cooperate until one player (say) i receives an errant d signal, with i then defecting in the next period, causing j to receive a d signal, after which i and j both continually defect. The first dirty history that appears in this sequence arises after j's defection, and both players thereafter defect. As a result, the estimated probability q of cooperation at a dirty history will converge to zero.

A similar argument shows that p will converge to one: As α converges to one, low-continuation-probability interactions, in which the (clean) null history is followed by defection, become arbitrarily rare. As ε converges to zero, d signals and hence defections after clean histories in high-continuation-probability interactions also become rare, ensuring that p approaches one.

It is then straightforward to see that the proposed behavior will *not* constitute an equilibrium. Fix a candidate value θ and let α be sufficiently large and ε sufficiently small that $1 - p < \theta$. Then for sufficiently large t, a player who has invariably played C and has seen all c signals (or has not seen too many d signals) can invoke a law-of-large-numbers argument to conclude that with very high probability, the proportion of defections in the history falls enough short of θ that another defection engenders virtually no risk of rendering the history dirty. This destroys the incentive to play C, disrupting the equilibrium. We believe that the same argument precludes the existence of any equilibrium with almost permanent cooperation, even when α gets close to 1 and ε gets close to 0.

By contrast, our two-category equilibrium avoids this difficulty by working in the extreme case of $\theta = 0$, ensuring that player i always fears that a single defection will render the history dirty. Again, if analogical thinking is to support cooperation, it is important that it fulfill the second condition that a defection is sufficiently likely to cause the history to switch categories. As the next setting illustrates, this switch must then have a sufficient effect on estimated behavior.

4.2 Three Analogy Classes

In our examination of two analogy classes, player i assumes that a defection renders the history dirty and instantly induces a jump in the probability that j defects, even though this cannot literally be the case. We examine here a three-category alternative that mutes this tension. One might reasonably argue that the movement to three categories gives players a more faithful representation of their interaction. However, we find that it can make cooperation more difficult to sustain, with the primary affect appearing in the estimate of the counterpart of q, the probability of cooperating after a dirty history.

A history for player i is deemed *healthy* if there has been no past defection by either player. A history for player i is *infected* if the player has defected at least once in the history. A history for player i is *exposed* if player i has cooperated throughout, but player j has defected at least once.

To illustrate, suppose the record contains the four-period observation introduced in Section 2.3. As before we have four cases in the record, given by

(4), with the attendant histories given by (5). The three cases in the three leftmost columns give rise to six histories, one per player per case, all of which are healthy. Five of these histories are followed by cooperation, and so the probability of cooperation after a healthy history is thus taken to be 5/6. The final case gives rise to two histories. The attendant player-1 history is exposed and exhibits cooperation, leading to the probability of cooperation of 1. The player-2 history is infected and exhibits defection, leading to a the probability of cooperation 0.

Player i's model of player j is that j is initially healthy, and then is either healthy (in which case j cooperates with probability p), exposed (in which case j cooperates with probability q), or infected (in which case j cooperates with probability r). As long as i continues to cooperate, i will view j as being either healthy or infected, and will update the probability i attached to the event that j is healthy in light of the signals i receives. Once i defects, player i now views player j as being either exposed or infected, and will update the probability i attaches to the even that j is exposed in light of the signals i receives.

The candidate equilibrium strategies are that each player begins by cooperating, and continues to do so as long as they think it sufficiently likely that their opponents are cooperating. However, the probability that player i attaches to j being healthy will fall over time, reducing the probability that j is cooperating, until i switches to thereafter defecting. At this point, i is infected. Player j is either exposed or infected, and will at some point switch to being infected.

A helpful first observation is that being infected is an absorbing state—once player i defects, then player i will thereafter defect, no matter what signals i receives. Player i's model of player j is such that once i is infected, i's actions have no effect on j's transition from exposed to infected. Instead, player i models j as making the transition to infected the first time j's draw of an action comes up with the probability $1-q$ action D. Hence, once i defects, no subsequent signals or beliefs will cause i to cooperate. This in turn allows us to conclude that $r=0$.

Once again we can formulate player i's maximization problem as a restless bandit problem. with details in Appeindix 7.3. We can then show that we have a nontrivial equilibrium as long as players are sufficiently patient and the monitoring is sufficiently precise. Appendix 7.3 proves:

Proposition 3 *Suppose*

$$\beta > \frac{4k}{\alpha^2(1+k)}. \tag{17}$$

The there exists $\bar{\varepsilon}$ such that for all $\varepsilon < \bar{\varepsilon}$, a nontrivial equilibrium exists.

The argument behind this result first shows that a necessary and sufficient condition for the existence of a nontrivial equilibrium is

$$\frac{p\beta(p-q)}{1-q\beta} \geq \frac{k}{1+k}. \tag{18}$$

The values of p and q are endogenous, and so the next step is to place some bounds on their values. It is not surprising that we need the monitoring to be sufficiently informative, with the bound $\bar{\varepsilon}$ becoming less stringent the higher is the continuation probability. The important relationship here is that q approaches zero as does ε. We then argue that $\alpha/2$ is a lower bound on p. Inserting this bound in (18) and letting ε and hence q approach zero, we obtain (17).

Proposition 3 ensures the existence of an equilibrium with some cooperation, but makes no comment as to how much cooperation we can expect and makes no statement about the payoff implications of that cooperation. Now we fix the continuation probability β and show that, as the monitoring structure becomes increasingly precise and the proportion of impatient players shrinks, then the there exists an equilibrium with payoff approaching 1, the payoff of the grim equilibrium in a game of perfect monitoring.

Proposition 4 *Fix $\beta \geq \frac{k}{1+k}$. Then there exists a sequence of equilibria such that, in the limit as first $\varepsilon \to 0$ and then $\alpha \to 1$, the equilibrium payoff approaches 1, the payoff of persistent mutual cooperation.*

The bound $\beta \geq \frac{k}{1+k}$ on the continuation probaibility is precisely the bound for the grim strategy to be an equilibrium in the standard repeated game of perfect monitoring. The argument, which follows that of Proposition 2 and is hence omitted, proceeds by stablishing the sufficient condition that, for small ε and large α, we have

$$\frac{p(1-\varepsilon)-\varepsilon}{p(1-\varepsilon)+(1-p)\varepsilon-\varepsilon} \geq \frac{k}{1+k}\frac{1-q\beta}{\beta p(p-q)}$$

We then show that as ε goes to zero so does q, and then as α goes to 1 so does p, reducing this condition to $\beta \geq \frac{k}{1+k}$. It is expected that we need ε to be small and α to be large. If ε is small, then the noisy monitoring technology will sometimes provide seemingly convincing evidence that an opponent is infected, even if this is not the case. If incentives are to be preserved, these signals must draw punishment, and the fixed continuation probability ensures that these punishments are costly. If α is not close to 1, then each player believes the opponent is subject to the constant hazard of infection, again precluding universal cooperation.

We can examine the forces behind the differences in the two-category and three-category case. First, as is familiar from thinking about repeated games, one step toward constructing an equilibrium is to make the punishment payoff small. In the two-category case, the relevant punishment payoff is

$$q(1+k),$$

while in the three-category case it is

$$q(1+k)\frac{1-\beta}{1-q\beta}.$$

It is immediate that $(1-\beta)/(1-q\beta) < 1$, and so the punishment payoff seems less attractive under three categories. This is intuitive. Under two categories, the punishment consists of receiving $q(1+k)$ forever, while under three categories, $q(1+k)$ is received only temporarily, until the opponent switches from exposed to infected. In light of this, it is initially counterintuitive that, as the results indicate, cooperation is more difficult to achieve with three categories, where Proposition 3 requires ε small while Proposition 1 does not, and Proposition 4 requires an order of limits while Proposition 2 does not.

The resolution is that the comparison above holds q fixed, while in fact the two-category system gives a smaller estimate of q than does three categories. The two-category case allows us to identify cases in which $q = 0$ while three categories typically give $q > 0$, possibly much larger. The difference arises because with three categories, the estimate of the probability q is taken from the set of exposed histories, which quite often lead to cooperation, with the first defection converting the history to infected. With two categories, in contrast, the set of histories from which q is estimated includes (among others) every history in which both players have defected, contributing many instances of defection to the frequency calculations. It is then no surprise that the estimated value of q is smaller under two categories.

4.3 Simultaneous Defection

Suppose the players adopt two analogy classes, one involving histories in which there has never been a simultaneous defection, and one involving histories in which there has been a simultaneous defection. The former are clean and the latter dirty, with p and q being the respective probabilities of cooperation. Let the players employ the strategy of cooperating as long as the probability z of the history being clean is sufficiently high. To isolate the key considerations, we examine the extreme case in which signals are uninformative ($\varepsilon = 1/2$).

We identify conditions under which equilibrium play in high-continuation-probability interactions exhibits permanent cooperation. If this is the case, then we will have $q = 0$, since dirty histories then occur only in low-continuation-probability interactions, which exhibit only defection. Because players in low-continuation-probability interactions defect after the (clean) null history, we will have $p < 1$.

Let z be the probability that the history is clean. As long as player i cooperates, there is no chance for the history to change to dirty, and so if cooperation is optimal at belief z, is is optimal thereafter. We can then calculate that the value of doing so is

$$z(p + (1-p)(-k)) + (1-z)k,$$

obtained by noting that the opponent cooperates (for payoff 1) with probability zp and otherwise defects (for payoff $-k$). If player i defects, then the history is clean in the next period with probability zp (the probability that it is clean now and j cooperates). If is optimal to defect at belief z, it will be optimal to defect thereafter, as the probability of a clean history continually falls, allowing us to calculate a payoff of

$$(1-\beta)(zp(1+k)+\beta zp^2(1+k)+\beta^2 zp^3(1+k)+\ldots) \;=\; \frac{(1-\beta)zp(1+k)}{1-\beta p}.$$

The condition for cooperation is then

$$z(p+(1-p)(-k))+(1-z)k \geq \frac{(1-\beta)zp(1+k)}{1-\beta p}.$$

As expected, this condition will fail for an interval of values of z of the form $[0,\overline{z}]$, confirming that players will not cooperate if the history is not sufficiently likely to be clean. The inequality will hold for large z, which suffices for an equilibrium exhibiting perpetual cooperation, if k is sufficiently small. However, the upper bound on k shrinks to zero as α and hence p approaches 1.

In this setting, it is impossible for the history to switch from clean to dirty if player i cooperates, with the incentive to cooperate provided by the possibility that the history will switch to dirty if i defects. In our previous analysis, a defection on i's part *ensured* that the history would switch to dirty, but here this switch occurs only if j also defects. As the proportion of low-continuation-probability interactions shrinks and hence p approaches one, the probability of such a double defection shrinks to zero, causing the incentive to cooperate to shrink disappear along with it. As before, the incentive to cooperate arises out of a misperception, since the putative equilibrium calls for players in high-continuation-probability interactions to invariably cooperate, while player i estimates that such players cooperate only with probability $p<1$. Unlike the previous case, the misperception here depends upon there being many low-continuation-probability interactions, and hence dissipates as such interactions become rare. However, when cooperation is possible, we note that permanent cooperation can be achieved even as the monitoring technology is not very precise (as ε away from 0 is allowed).

5 Discussion

5.1 Which Analogies?

We have identified conditions under which the analogues of familiar Nash reversion strategies support cooperation in repeated games of private monitoring. Our model of players' behavior is conventional, except that we ask that equilibrium beliefs be grounded in the empirical frequencies in a record of past plays of the game, rather than being drawn out of thin air. Much the same reasoning motivates the notion of self-confirming equilibrium (Fudenberg and Levine [6, 7]). Applications of self-confirming equilibria are often concerned with the fact that the record of play may place no discipline on beliefs at out-of-equilibrium histories. In contrast, we emphasize the idea that players will organize histories into analogy classes in such a way that they have data to estimate all of the probabilities they need, while recognizing that this is possible only with a coarse categorization that lumps some histories together and inevitably leads to some misspecification.

This misspecification plays a central role in supporting cooperation in the simple settings we have examined. A player observes in the record that once a defection appears, the probability of cooperation falls, and captures this in the estimate that dirty histories give rise to lower probabilities of cooperation than do clean histories. In doing so, player i misses some details of how subsequent play depends on *which* player first defecting, leading i to interpret the correlation between dirty histories and attenuated cooperation as in indication that should i defect, the probability that j will cooperate deteriorates, regardless of the monitoring structure. Our setting is sufficiently stark that one might expect players to correct such misspecification. However, misspecification will be less obvious amid the complications of an actual interaction, and is inevitable as long as one deals with a finite record.

These observations naturally direct attention to the question of how analogy classes are determined. One approach would be to assume that the players construct the analogy classes themselves, perhaps to the point that this construction becomes part of the strategic interaction. However, we find it somewhat counterintuitive to think of players as optimally choosing a model and the proceeding as if unaware of the misspecification inherent in that model. We are accordingly more inclined to think that the classes are constructed outside the relationship, perhaps unknowingly, often in the form of conventions governing the sorts of information that are readily available in the record.

When taking this view, one might naturally think that a more detailed record is better, and it is better for the players to work with more categories. However, we have already seen that refining the categories need not facilitate cooperation. The intuition that more precise monitoring facilitates cooperation arises out of the fact that more precise monitoring makes it easier to detect opponent deviations. In contrast, coarse analogy classes may facilitate cooperation by obscuring situations in which deviations will *not* prompt responses form others. In this sense, at least partial ignorance may indeed be bliss.

5.2 Beyond the Prisoners' Dilemma

We have exploited the particular structure of the prisoners' dilemma, but we believe the analysis applies much more generally. The key is that just as the players are likely to assess the record by organizing histories into analogy classes, so are they likely to think of the game itself in terms of analogy classes.

Let us fix a stage game, perhaps capturing the interactions of a pair of Cournot duopolists. Let D denote an action corresponding to a Nash equilibrium of the stage game, and let C denote the collusive action. Let c be a signal interpreted as reflecting the play of C, and let d be a signal interpreted as "not cooperating." Then we can perform much the same analysis as before.

Unlike the prisoners' dilemma, D is no longer a stage-game best response to C, and so deviations from C need not necessarily induce the same signal distribution as does playing D. However, we need only that profitable deviations from C push the signal distribution in the same direction as does the action D. In such a case, we view it as plausible that players might organize the actions of others into the categories of cooperating and not cooperating. Indeed, players may have only an imperfect understanding of the actions available to their opponents and the implications of these actions, and may find it impossible to form a more precise understanding of their opponent. As a result, players may form a fairly precise view of what it means to cooperate, and keep track of that rather finely, but may be less discriminating once the opponent has fallen into the not-cooperating category.

5.3 Simplicity in Context

Even in repeated ganes of perfect monitoring, there is an intuition that some equilibrium strategy profiles are more plausible than others. A commonly voiced sentiment is that simple strategies are more plausible than more complicated ones, and especially that simple punishments are more plausible than more complex punishments. One sees this view reflected in applied work, where Nash revision is the most commonly-invoked punishment. In some cases, the appeal to Nash reversion may reflect only a desire to keep the theoretical and computation analysis manageable. However, Nash reversion has an appeal beyond tractability, often supported by arguments that "it's unreasonable to think that firms cooperate in punishing for not cooperating."

The early literature, inspired by Rubinstein [18], tried to make this intuition precise by introducing measures of strategic complexity and objectives that balanced complexity with payoffs in the repeated game. The complexity measures were often expressed in terms of the ability to express a strategy in terms of an automaton or similar device, and the analysis showed that even lexicographically inferior concerns for complexity could have important implications. This literature foundered on two difficulties. There quickly appearedd a host of simplicity definitions, with different implications. In addition, one readily encountered the seeming paradox that cooperation is sustained by a punishment, but if this is successful, then the punishment is never used, at which point the punishment capability will be discarded on simplicity grounds, at which point the cooperation cannot be sustained.

Our approach leads to an alternative notion of simplicity, based on the analogy classes used by the players in interpreting the record. We have simplified our presentation by giving players an arbitrarily rich record, but in practice players are likely to be forced into simple strategies by a limited record, ensuring that they can obtain useful estimates of only a small number of probabilities. Moreover, the need to sustain punishment capabilities arises naturally in our setting.

The initial approaches to strategic complexity examined games in isolation, with the complexity of a strategy assessed in the context of the single game in which it was applied. In contrast, we suspect that people maintain a suite of strategies that are applied to various games as needed, as in Samuelson [19]. This is reflected in our analysis in the fact that games are drawn to have either high or low continuation probabilities. It is then important for strategies to be capable of "Nash reversion," as this is the only equilibrium in some interactions, and Nash reversion then plays a natural role as a punishment in other interactions.

6 The Foundations of Cooperation?

The initial results in the study of repeated games were influential not only for showing that cooperative payoffs can be supported, but for exposing the intuitive mechanism behind this cooperation, namely that deviations from cooperation in a current period lead to payoff-lowering punishments in future periods. This link between the structure of the equilibrium and intuition about how cooperation is supported breaks down in games of private monitoring, where the various versions of belief-free equilibria are wonderfully elegant but devoid of any notion that a person refrains from defecting today in order to avoid triggering a shift to lower payoffs tomorrow. The approach introduced in this paper is an attempt to recover this link. Does this approach provide insight into cooperation in repeated interactions?

The most striking feature of our equilibria is that they in general feature only temporary cooperation, until the posterior belief about the opponent dips below the threshold $\overline{z}$.[11] This contrasts with an emphasis in the literature on models of permanent cooperation. Should a "foundation of cooperation" for games of private monitoring command our interest that supports only temporary cooperation?

Much of the motivation for studying repeated games comes from thinking about collusion in oligopolies. Marshall and Marx and [14] provide a comprehensive discussion of collusion, drawing on theoretical sources and case studies. Their work suggests that neither a standard model of a repeated game nor the approach outlined here is relevant.

Marshall and Marx emphasize that two features of collusive arrangements are endemic. First, monitoring is noisy and private. Realized quantities never precisely match their target levels, firms have only imperfect information about their rivals' quantities, and firms have no way of ascertaining how well their own actions are tracked by others, or whether the seemingly anomalies quantities of others reflect deliberate actions or the capriciousness of the market. As a result, Marx and Matthews stress that collusion can be sustained only if the firms introduce some explicit means (in the form of an industry trade association, common accounting firm, or some similar arrangement) of collecting and disseminating information, allowing communication, and coordinating redress for anomalous outcomes. In effect, their message is that oligopolies support persistent cooperation only if they convert the game of private monitoring into a game of public monitoring.

[11] Based on Section 4.3, we suspect that this will necessarily be the case whenever the precise timing of past actions is not accessible in the record (thereby not allowing the categorization to depend on the precise details of the profile of actions).

Porter [17] and Ulen [21, 22] provide a detailed case study of the Joint Executive Committee, the governing body of a railroad cartel operating in the 1880s. The cartel operated without hindrance from the Interstate Commerce Commission (founded in 1887) and antitrust legislation (originating in the Sherman Antritrust Act of 1890). The Joint Executive Committee published weekly shipping statistics, verified by station agents and employees of the Chicago Board of Trade, hired a prominently-staffed Board of Arbitrators to settle disputes, and assigned punishments in response to cheating on agreements. Once again, successful collusion hinged on public monitoring. Levenstein and Suslow [13] and Harrington and Skrzypacz [10, 11] stress the importance cartels place on disseminating public sales information to their members, perhaps through the creation of joint sales agencies or trade associations. In contrast to the case of the Joint Executive Committee, this coordination can be hampered by the specter of illegality.

A second approach to cooperation is exemplified by the work of Elinor Ostrum (e.g., [15]; see also Ellickson [3]) on managing common resources. There are no legal obstacles to cooperation in these cases, though there are typically many more people involved than in an oligopoly. Ostrum again emphasizes the importance of continual communication and the creation of informal or informal mechanisms to monitor behavior and impose sanctions when needed. The arrangements often involve explicit agreements as to how the participants are to monitor and verify other's behavior. Blomquist, Schlager, Tang and Ostrum [2] reinforce the importance of monitoring. They search for the features that are common to a large number of cases in which cooperation has been sustained in the use of common-pool resources, such as fisheries, irrigation systems, groundwater systems, and forests. These common features include organized procedures for monitoring actions, making deviations known, and assessing and enforcing sanctions. These monitoring arrangements were often backed by formal institutions. In essence, actions become public.

Our interpretation of this literature is that one cannot typically expect to sustain cooperation permanently without converting private monitoring into public monitoring.[12] We view our model as applying to cases in which monitoring is inherently private, communication is unreliable or communication alone is ineffective in supporting cooperation. Here, we are not surprised that punishments may eventually get triggered and then will be permanent. Relations between countries, where institutions to provide monitoring are sparse, are one

[12] Aoyagi, Bhaskar and Frechette [1] report experimental results for the repeated prisoners' dilemma with private monitoring. Two of the three most popular strategies (consisting of always-defect and a lenient version of Nash reversion, and together comprising 56% of all strategies) necessarily eventually always defect, and this is one possibility for the second most popular strategy (which they refer to as Sum2). This is qualitatively consistent with our equilibria, in which cooperation eventually dissipates. The also find that overall cooperation in private monitoring games reaches approximately the levels found in perfect-monitoring games, partly because the lenient Nash reversion strategy requires up to three successive d signals before switching to always defect, and partly because the Sum2 strategy can also settle into persistent cooperation (regardless of subsequent history). This is consistent with our finding that seeming temporary cooperation can be quite valuable.

obvious area of application, as are relations between firms when the specter of antitrust enforcement is sufficient to deter effective communication.

We find that cooperation can still be immensely valuable. The time scale on which cooperation breaks down may be so long as to make the payoff effects of cooperation effectively permanent; this is the implication of Proposition 2. We thus have a situation in which institutions may last a very long time, and cooperation may also last a very long time, but eventually either the institution or the cooperation degenerates or disappears. One could argue that in the Roman empire, cooperation broke down (which then lead to the demise of the empire) while in the British empire, cooperation persevered but the empire withered away.

7 Appendix: Proofs

7.1 Proof of Proposition 1

[Step 1] We first establish conditions under which a player will optimally pull the C arm in period 1 of the modified bandit of Remark 1. A sufficient condition for this to be the case is that the period-1 reward from the C arm exceed that of the D arm, or

$$\frac{k}{\overline{\beta}} \leq \mathbb{E}h(z_1|z_0) = \mathbb{E}\{z_1|z_0\}(p-q)(1+k) = pz_0(p-q)(1+k) = p(p-q)(1+k),$$

where the second equality uses the fact that $\mathbb{E}\{z_1|z_0\} = pz_0$ and the next uses the fact that $z_0 = 1$. We can rearrange this as

$$\frac{k}{1+k} \leq \overline{\beta}p(p-q). \tag{19}$$

Remark 2 An alternative derivation of (19) helps illuminate the underlying forces. If cooperation is ever to be optimal, it must be better to cooperate in the first period and defect thereafter than to defect immediately (and permanently). The payoffs from these two strategies, arranged by period, are:

$$\begin{array}{llllllll} CDD\ldots: & p+(1-p)(-k) & + & \overline{\beta}[p^2(1+k)+(1-p)q(1+k)] & + & \overline{\beta}^2 q(1+k) & + & \overline{\beta}^3 q(1+k)+\ldots \\ DDD\ldots: & p(1+k) & + & \overline{\beta}q(1+k) & + & \overline{\beta}^2 q(1+k) & + & \overline{\beta}^3 q(1+k)+\ldots \end{array}$$

All of the payoff differences occur in the first two periods. The first strategy sacrifices some payoff in the first period, in order to obtain a larger payoff in the second period. The condition that the first strategy give a higher payoff is

$$-\frac{k}{1+k} + \overline{\beta}p(p-q) \geq 0,$$

which is (19). The first term captures the payoff reduction in the first period from cooperating, while the second captures the payoff gain in the second period. ■

[Step 2] We now derive some bounds on p and q The probability of p of cooperation after clean histories is bounded below by $\alpha/2$. So confirm this, we note that in a patient interaction, both players cooperate in the first period, and there can be at most one (if the first defection is unilateral) or two (if the first defection is mutual) clean histories after which players defect. Hence, the probability of cooperation after clean histories in patient interactions is at least 1/2, ensuring that p is at least $\alpha/2$. A sufficient condition is then

$$\frac{k}{1+k} \leq \overline{\beta}\frac{\alpha}{2}\left(\frac{\alpha}{2}-q\right).$$

Now fix α and $\overline{\beta}$. As ε approaches zero, so does q (as we confirm in the next step). The sufficient condition then becomes

$$\overline{\beta} > \frac{4k}{\alpha^2(1+k)}$$

which rearranges to give the first result.

Alternatively, fix α and ε. As $\overline{\beta}$ approaches one, either p approaches 1 or q approaches 0 (as verified in the next step). The latter case gives the more demanding condition, which is (substituting $\overline{\beta}=1$, $p=\alpha/2$ and $q=0$)

$$\frac{k}{1+k} < \frac{\alpha^2}{4},$$

which is equivalent to $k \leq \frac{\alpha^2}{4-\alpha^2}$.

Statement [1.3] is a special case of Proposition 2.2, and so we defer its proof.

[Step 3] The first two steps identify conditions under which players will initially cooperate. To establish the existence of such an equilibrium with this property, we now construct a correspondence Φ that maps values of $(p,q,\overline{z}) \in [\frac{\alpha}{2},1]\times[0,\frac{\alpha}{2}]\times[\underline{z},1]$ into new values of $(\hat{p},\hat{q},\hat{\overline{z}}) \in [\frac{\alpha}{2},1]\times[0,\frac{\alpha}{2}]\times[\underline{z},1]$, for some $\underline{z}>0$. The mapping is defined as follows. First, given (p,q), a player solves for the the optimal value $\hat{\overline{z}}$ in the modified bandit. Then, given this value of $\hat{\overline{z}}$ and working with the updating rules defined by (p,q), we construct the distribution over histories, and from this infer new values $(\hat{p},\hat{q})$. When doing the latter step, we consider all possible values of $(\hat{p},\hat{q})$ by allowing a player to mix between cooperating and defecting when indifferent.

The argument now involves verifying that for sufficiently large $\overline{\beta}$, this correspondence indeed maps into $[\frac{\alpha}{2},1]\times[0,\frac{\alpha}{2}]\times[\underline{z},1]$, and that it has a fixed point.

First, we argue that the map takes values from $(p,q,\overline{z}) \in [\frac{\alpha}{2},1]\times[0,\frac{\alpha}{2}]\times[\underline{z},1]$ into $(p,q,\overline{z}) \in [\frac{\alpha}{2},1]\times[0,\frac{\alpha}{2}]\times[\underline{z},1]$. Because they are probabilities, p cannot exceed 1, q cannot fall short of 0, and $\overline{z}$ cannot exceed 1, giving three of the required bounds. As we argued in the previous step, the probability p is bounded below by $\frac{\alpha}{2}$ (because with probability α, we have patient players who both cooperate after the (clean) null history). Somewhat similarly, q is at most $\frac{\alpha}{2}$, because only patient players ever cooperate after a dirty history, after which at most one can cooperate. Finally, we need the lower bound $\underline{z}$. Notice that we cannot simply take this to be zero. The argument in step 1 requires that q goes to zero as $\overline{\beta}$ gets large. We can be sure of this only if signals are informative and $\overline{z}$ is bounded away from zero. An upper bound on the continuation payoff from cooperating and an exact calculation of the payoff from defecting are given by:

$$\begin{array}{ll} C: & [zp + (1-z)q] + [(1-(zp+(1-z)q))](-k) \\ D: & (1-\overline{\beta})[zp+(1-z)q](1+k) + \overline{\beta}q(1+k). \end{array}$$

Given these payoffs, the condition that cooperation have a higher payoff is

$$z \geq \frac{k}{(1+k)\overline{\beta}(p-q)}.$$

A lower bound on the value of z that solves this equation with equality, and hence (given that we have overestimated the payoff of cooperation) a lower bound on $\underline{z}$ (and hence $\overline{z}$), is given by (setting $\overline{\beta} = p = 1$ and $q = 0$)

$$\frac{k}{1+k}.$$

Given that we allow the player to mix when indifferent between cooperating and defecting, our map is a convex-valued, upper hemicontinuous correspondence, ensuring that it has a fixed point.

Finally, we note that as ε approaches zero, so does $\phi(z,d)$ for all $z \in [0,1]$. Intuitively, as monitoring becomes arbitrarily precise, a bad signal is taken as convincing evidence of defection. Combining this with the lower bound on $\overline{z}$, small values of ε ensure that (for fixed $\overline{\beta}$) the first d signal in an interaction is with arbitrarily high probability produced by a D action, and prompts a D action from the recipient in the next period. This in turn ensures that with arbitrarily high probability we observe only defection after dirty histories, causing q to approach 0.

Similarly, fix α and ε in $(0,1)$. Suppose that p is bounded below 1 as $\overline{\beta} \to 1$. Because $\overline{z}$ is bounded below, for any $\eta > 0$ there is a number τ such that, once player i defects, player j will defect within τ periods with probability at least $1-\eta$. This places a bound on the number of dirty histories after which players cooperate. However, as $\overline{\beta} \to 1$, each defection gives rise to an arbitrarily large number of dirty histories, ensuring that q converges to zero. ■

7.2 Proof of Proposition 2

If the opponent is currently clean with probability z, a player who cooperates and receives a c signal forms the posterior $\phi(z,c)$ that the opponent is clean given by (6). A consistent string of c signals will lead to the posterior z^* solving $z^* = \phi(z^*, c)$, given by

$$z^* = \frac{p(1-\varepsilon) - [q - 2q\varepsilon + \varepsilon]}{(p-q)(1-2\varepsilon)}.$$

The generalization of (19), giving a sufficient condition for an player to cooperate, holding posterior z (equal to one in (19)) that the opponent is clean, is

$$\overline{\beta} z p(p-q) \geq \frac{k}{1+k},$$

or, equivalently

$$z \geq \frac{k}{1+k} \frac{1}{\overline{\beta} p(p-q)} \equiv \overline{z}. \tag{20}$$

The condition that $z^* \geq \overline{z}$ is then

$$\frac{p(1-\varepsilon) - [q - 2q\varepsilon + \varepsilon]}{(p-q)(1-2\varepsilon)} \quad \geq \quad \frac{k}{1+k} \frac{1}{\overline{\beta} p(p-q)} \tag{21}$$

First, let $\varepsilon \to 0$. Because (i) $\overline{\beta}$ is fixed, (ii) the candidate equilibrium strategies are that players cooperate as long as their posterior exceeds $\overline{z}$, and (iii) erroneous d signals become arbitrarily rare as ε falls, we can conclude that interactions between patient players will contribute to the record primarily cases in which mutual cooperation persists throughout the interaction. This ensures that p will approach 1 as does α. This also ensures that (given fixed α) virtually all dirty histories will occur among impatient players, whose defection then causes q to approach zero. Hence, (21) becomes (9). If this condition holds, then for values of α larger than some $\overline{\alpha} < 1$, we have a sequence of equilibria in which the probability of cooperation throughout the life of the interaction becomes arbitrarily large as ε approaches one.

Next, let us fix ε. Let us hypothesize that as $\alpha \to 1$, we have $p \to 1$, while remembering the bound $q \leq 1/2$. This will be the case if the probability that z_t dips below $\overline{z}$ before the interaction ends becomes vanishingly small. For this to be the case, we require two conditions. First, we need $\overline{z} < 1$, which (from (20)) will be the case (using $p \to 1$ and $q \leq 1/2$) if (10) holds. Second, we need

$$\frac{z\varepsilon^n}{z\varepsilon^n + (1-z)(1-\varepsilon)^n}$$

to converge to 1 as does z, for all n. This is the posterior probability that the opponent is clean, given a prior of z and given that n consecutive d signals have been received, calculated in the limit as p takes on the value 1 and calculated in the worst-case scenario in which q is set to 0. This condition is obviously met. This in turn ensures that very large values of p, even the worst case of a relentless string of bad signals does not drive the posterior probability z below the defection threshold $\overline{z}$ before the interaction ends. But then, given that α is arbitrarily close to one, the record will indeed produce an estimate of p arbitrarily close to one. Coupling this with $q \leq 1/2$, (21) gives (10). The result is an equilibrium in which cooperation persists throughout virtually all interactions, as desired. ■

7.3 Details for Section 4.2

We first formulate the bandit problem. Let z_t be the probability that i attaches to the event that j is not infected in period t. This will be either the probability that j is healthy (if i has not yet defected) or exposed (if i has defected).

The bandit is defined by two parameters, p and q. There are two arms, a C arm (corresponding to cooperating) and a D arm (corresponding to defecting). We let z_t be the state of both arms at time t. If the D arm is pulled at time t, then both arms are in state 0 at time $t+1$. If the C arm is pulled at time t, then both arms move to state

$$\phi(z,c) = \frac{zp(1-\varepsilon)}{zp(1-\varepsilon)+z(1-p)\varepsilon+(1-z)\varepsilon}$$

with probability $zp(1-\varepsilon)+z(1-p)\varepsilon+(1-z)\varepsilon$ and to state

$$\phi(z,d) = \frac{zp\varepsilon}{zp\varepsilon+z(1-p)(1-\varepsilon)+(1-z)(1-\varepsilon)}$$

with probability $zp\varepsilon+z(1-p)(1-\varepsilon)+(1-z)(1-\varepsilon)$.

Each time the C arm is pulled, it generates a current payoff of

$$zp+(1-zp)(-k) \;=\; -k+zp(1+k).$$

When the D arm is first pulled, it generates a current payoff of

$$zp(1+k).$$

We have noted that once the D arm is pulled, it is then optimal to thereafter pull the D arm. This allows us to calculate the expected value of a path of play that begins with player i's first defection. Suppose i defects for the first time in period $t-1$, and as a result, period t begins with i attaching probability z_t to the event that j is exposed, and probability $1-z_t$ to the event that j is infected. Then i's continuation payoff is[13]

$$(1-\beta)z_t(1+k)[q+q^2\beta+q^3\beta^2+q^4\beta^3+\ldots].$$

We can solve for the value of

$$z_t q(1+k)\frac{1-\beta}{1-\beta}.$$

As a result, it is straightforward to calculate the value of the D arm, which is given by

$$W(z) = (1-\beta)zp(1+k)+\beta pzq(1+k)\frac{1-\beta}{1-q\beta} = pz(1+k)\frac{1-\beta}{1-\beta q}. \tag{22}$$

[13] To see this, we note that with probability $1-z_t$, player j is infected and defects thereafter, giving i a 0 payoff (since i is also defecting). With probability z_t player i receives a payoff $1+k$ each time j cooperates (and 0 otherwise). With probability q, j cooperates in period t. With probability $q^2\beta$, the game lasts another period and j again cooperates. With probability $q^3\beta^2$, the game lasts yet another period, and j again cooperates, and so on.

Proof of Proposition 3 We establish conditions under which a player will optimally pull the C arm in period 1, with the remainder of the argument mimicking that of Proposition 1. A sufficient condition for this to be the case is that pulling the C arm in the first period and thereafter defecting is better than defecting immediately. This comparison is (using the facts that $z_0 = 1$, the expected value of z_1 is p, and the value of defecting is linear in z):

$$(1-\beta)(-k+p(1+k))+p\beta W(1) \geq W(1)$$

where the left side sums the current payoff from playing C plus the discounted expected value of defecting next period ($\beta W(p) = p\beta W(1)$) and the right side is the value of immediate defection. We can rewrite this successively as

$$\begin{aligned} (1-\beta)[-k+p(1+k)]+\beta p\left[p(1+k)\frac{1-\beta}{1-\beta q}\right] &\geq \left[p(1+k)\frac{1-\beta}{1-\beta q}\right] \\ (1-\beta)(-k+p(1+k)) &\geq (1-p\beta)p(1+k)\frac{1-\beta}{1-q\beta} \\ p\left(1-\frac{1-p\beta}{1-q\beta}\right) &\geq \frac{k}{1+k} \\ \frac{p\beta(p-q)}{1-q\beta} &\geq \frac{k}{1+k}. \end{aligned} \tag{23}$$

Now, for a fixed β, let ε approach 0. This will ensure q approaches 0. (The key to this conclusion is that $\alpha < 1$, and so p remains bounded below 1. As a result, as ε gets arbitrarily small, a d signal is arbitrarily more likely to have come from an action of D (and hence an infected opponent) than from an action of C.) When player i defects, it becomes arbitrarily likely that j's posterior that i is infected gets arbitrarily close to 1, ensuring that j will defect, and hence q will be arbitrarily close to 0. Then, letting β approach one completes the argument, as before. In the limit as β approaches 1, the sufficient condition is then (substituting $\beta = 1$, $p = \alpha/2$ and $q = 0$)

$$\frac{\alpha^2}{4} \geq \frac{k}{1+k},$$

which is equivalent to $k \leq \frac{\alpha^2}{4-\alpha^2}$.

■

References

[1] Masaki Aoyagi, V. Bhaskar, and Guillaume R. Frechette. The impact of monitoring in infinitely repeated games: Perfect, public, and private. *American Economic Journal: Microeconomics*, 11(1):1–43, 2019.

[2] Willaim Blomquist, Edella Schlager, Shui Yan Tang, and Elinor Ostrum. Regularities from the field and possible explanations. In Elinor Ostrum,

Roy Gardner, and James Walker, editors, *Rules, Games and Common-Pool Resources*, pages 301–316. University of Michigan Press, Ann Arbor, Michigan, 1994.

[3] Robert C. Ellickson. *Order without Law: How Neighbors Settle Disputes.* Harvard University Press, Cambridge, Massachusetts, 1991.

[4] Jeffrey C. Ely and Juuso Välimäki. A robust folk theorem for the prisoner's dilemma. *Journal of Economic Theory*, 102(1):84–105, 2002.

[5] James W. Friedman. A noncooperative equilibrium for supergames. *Review of Economic Studies*, 38(1):1–12, 1971.

[6] Drew Fudenberg and David K. Levine. Self-confirming equilibrium. *Econometrica*, 61:523–546, 1993.

[7] Drew Fudenberg and David K. Levine. Steady state learning and Nash equilibrium. *Econometrica*, 61:547–574, 1993.

[8] Drew Fudenberg, David K. Levine, and Eric Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1031, 1994.

[9] Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

[10] Joseph E. Harrington, Jr. and Andrzej Skrzypacz. Private monitoring and communication in cartels: Explaining recent collusive practices. *American Economics Review*, 101(6):2425—-2449, 20011.

[11] Joseph E. Harrington, Jr. and Andrzej Skrzypacz. Collusion and monitoring of sales. *The RAND Journal of Economics*, 38(2):314–331, 2007.

[12] Philippe Jehiel. Analogy-based expectation equilibrium. *Journal of Economic Theory*, 123(2):81–104, 2005.

[13] M. Levenstein and V. Suslow. What determines cartel success? *Journal of Eonomics Literature*, 44(1):43–95, 2006.

[14] Robert C. Marshall and Leslie M. Marx. *The Economics of Collusion.* MIT Press, Cambridge, Massachusetts, 2012.

[15] Elinor Ostrum. *Governing the Commons.* Cambridge University Press, Cambridge, 1990.

[16] Michele Piccione. The repeated prisoner's dilemma with imperfect private monitoring. *Journal of Economic Theory*, 102:70–83, 2002.

[17] Robert H. Porter. A study of cartel stability: The joint executive committee, 1880–1886. *Bell Journal of Economics*, 14(2):301–314, 1983.

[18] Ariel Rubinstein. Finite automata play the repeated prisoners' dilemma. *Journal of Economic Theory*, 39(1):83–96, 1986.

[19] Larry Samuelson. Analogies, adaptation, and anomalies. *Journal of Economic Theory*, 97:320–366, 2001.

[20] Tadashi Sekiguchi. Efficiency in repeated prisoner's dilemma with private monitoring. *Journal of Economic Theory*, 76(2):345–361, 1997.

[21] Thomas S. Ulen. *Cartels and Regulation.* Stanford University, Unpublished Ph. D. Disseration, 1978.

[22] Thomas S. Ulen. Cartels and regulation: late nineteenth-century railroad collusion and the creation of the interstate commerce commission. *Journal of Economics History*, 40(1):179–181, 1980.

[23] Peter Whittle. *Optimization Over Time: Dynamic Programming and Stochastic Control, Vol II.* John Wiley and Sons, New York, 1983.