The Analogical Foundations of Cooperation

Philippe Jehiel, Larry Samuelson

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

WORKING PAPER N° 2022 – 23

The Analogical Foundations of Cooperation

Philippe Jehiel
Larry Samuelson

JEL Codes: C70, C72, C73
Keywords: Analogical reasoning, cooperation, prisoners' dilemma, repeated game, private monitoring

anr agence nationale de la recherche
AU SERVICE DE LA SCIENCE
INVESTIR L'AVENIR

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

June 11, 2022

Abstract. We offer an approach to cooperation in repeated games of private monitoring in which players construct models of their opponents' behavior by observing the frequencies of play in a record of past plays of the game in which actions but not signals are recorded. Players construct models of their opponent's behavior by grouping the histories in the record into a relatively small number of analogy classes to which they attach probabilities of cooperation. The incomplete record and the limited number of analogy classes lead to misspecified models that provide the incentives to cooperate. We provide conditions for the existence of equilibria supporting cooperation and equilibria supporting high payoffs for some nontrivial analogy partitions.

Keywords: Analogical reasoning, cooperation, prisoners' dilemma, repeated game, private monitoring
JEL codes: C70, C72, C73

Contents

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

The Analogical Foundations of Cooperation

1 Introduction

1.1 Motivation

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. Models of repeated games of perfect monitoring (Fudenberg and Maskin [17]) capture this intuition well, most simply in the Nash reversion equilibria presaged by Friedman [13], 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), attributing the signal to an unlikely draw from the noisy monitoring technology, and hence will cooperate rather than risk triggering a punishment by defecting. This ensures that first-period signals and hence actions have no effect on subsequent actions, giving players a license to defect in the first period, disrupting the putative equilibrium.

In this paper, we suggest an alternative approach to cooperation that emphasizes plausible reasoning and simple belief formation. Players form models of their opponents' behavior by observing the frequencies of the various outcomes in a record of past plays of the game by other players. This empirical foundation of players' models of their opponents leads these models to be coarse—no record of previous interactions will allow the estimation of a potentially distinct behavior 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 and estimate the typical behavior in each class. This introduces a misspecification into players' models of their opponents that resolves the tension between the equilibrium hypothesis of cooperation (and hence the desire to continue cooperating despite adverse signals) and the incentive-producing belief that defecting will prompt an adverse opponent reaction.

Formally, we examine analogy-based expectation equilibria (Jehiel [21]) in which histories with the same action profiles must belong to the same analogy class, reflecting the hypothesis that private signals are inaccessible to outside observers and thus do not appear in the record of past play. We begin in Sections 2–3 with the prisoners' dilemma, examining a simple model with two such analogy classes in which histories are bundled according to whether they exhibit instances of defection. The consequent misspecification gives rise to incentives

[1] As Fudenberg, Levine and Maskin [16] show, *noisy* monitoring need not pose difficulties as long as the monitoring is public.

supporting cooperation. Intuitively, because they hold a constant expectation of their opponent's behavior in each analogy class, players subjectively perceive that a defection would trigger a punishment akin to that of the familiar Nash reversion strategy, leading players to initially cooperate. More precisely, their misspecified representation of their opponent's strategy leads players to reason as if they were facing a stopping problem, determining when they should switch from cooperating to defecting, which they do when their subjective belief that their opponent has defected at least once drops below an endogenously-determined threshold. Section 4 discusses in more detail the forces behind the result, the interpretation of analogy classes, applications and extensions.

1.2 Relationship to the Literature

Matsushima [30] exploits reasoning similar to the intuition offered in the previous subsection to establish a precise and general result. If the players in a repeated game of independent private monitoring adopt pure strategies that are measurable with respect to their beliefs,[2] then the only equilibria of the repeated game play a stage-game Nash equilibrium in every period.

Subsequent approaches to repeated games of private monitoring accordingly exploit relaxations of Matsushima's three conditions. Mailath and Morris [27, 28] relax the independent-signals condition, providing a folk theorem for repeated games in which the players' signals are almost public. Adverse signals are now likely to be (highly) correlated, allowing players to create incentives by attaching punishments to such signals. Sekiguchi [38] and Bhaskar and Obara [4] relax the restriction to pure strategies, providing folk theorems for the repeated prisoners' dilemma based on "belief-based" equilibria in mixed strategies. For example, if players mix between a Nash reversion strategy and the strategy of always defecting, then adverse signals are an indication that the latter is likely to have been realized, prompting changes in future behavior that create incentives. The belief-free approach pioneered by Ely and Välimäki [9] and Piccione [35] similarly allows mixed strategies (though first-period actions are pure, unlike the belief-based approach), providing folk-theorem results supported by equilibria in which players mix between actions so as to ensure that their opponents are always indifferent between cooperating and defecting. As the name suggests, beliefs are irrelevant in belief-free equilibria, avoiding the potentially intricate updating of beliefs in the belief-based approach of Sekiguchi [38] and Bhaskar and Obara [4], while raising potential difficulties in purifying the equilibrium mixtures (cf. Bhaskar, Mailath and Morris [3]).

The belief-free folk theorems developed by Ely and Välimäki [9] and Piccione [35] require that monitoring be nearly perfect. Matsushima [31] and Yamamoto [45] relax this requirement. Intuitively, they replace a single period in the belief-free equilibria of Ely and Valimaki [9] and Piccione [35] with a multi-period review phase, allowing precise information to be extracted from a sequence of

[2]The requirement is that if two histories for player i induce identical beliefs over the opponents' histories, then i must take identical actions at these histories.

individually less informative signals. For this to work, however, it is important that signals be independent, so that the information received by a player in the midst of the review phase does not provide clues as to how likely she is to pass the review (and hence how important is adherence to the prescribed equilibrium). Sugaya [41] retains the convention of a review phase, while extending the folk theorem to the case of correlated signals.

We focus on the ability to maintain cooperation (rather than a full folk theorem) in the repeated prisoners' dilemma with independent private monitoring. The most relevant comparisons are then the belief-based equilibrium of Sekiguchi and the belief-free equilibria of Ely and Välimäki. Some of our results do not require that monitoring be nearly perfect. These results depend on a misspecification in the players' model of their interaction, viewing opponents' behavior as constant across the analogy classes used to partition the historical record, rather than the review phases of Matsushima [31] and Yamamoto [45]. Compte and Postlewaite [7] similarly work with a model in which histories are grouped together into categories, though with a different motivation for the grouping and a different equilibrium construction. Section 4.3 explains the features our approach shares with that of Sekiguchi [38], Ely and Välimäki [9], and Compte and Postlewaite [7].

2 Analogical Reasoning and Cooperation

2.1 The Stage Game

We examine the workhorse model of cooperation, the repeated prisoners' dilemma, with the stage game given by

$$\begin{array}{c|c|c} & C & D \\ \hline C & 1,1 & -k,1+k \\ \hline D & 1+k,-k & 0,0 \end{array} \quad . \tag{1}$$

It is a normalization to choose the payoffs of mutual defection and mutual cooperation to be 0 and 1. We simplify the analysis by restricting 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 is 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 δ, then there exists an equilibrium supporting permanent mutual cooperation if and only if the players are sufficiently patient and the premium on defecting is sufficiently small, i.e., if and only if

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

Perhaps the best known strategy supporting cooperation is the Nash reversion 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 period $0, 1, \ldots$. The players have a common discount factor δ, which we interpret (and hereafter refer to) as a continuation probability, governing the random length of the game. Hence, after each period, an independent draw is taken determining whether the game continues (probability δ) or terminates. The continuation probability may be either high (denoted by $\overline{\delta}$) or low (denoted by $\underline{\delta}$). At the beginning of the game, the continuation probability is randomly drawn, with probability α of continuation probability $\overline{\delta}$, and is known by both players.

Our interpretation is that some relationships are likely to last longer than others, depending on contextual features of the relationship that are known by both players. Members of a special commission assembled by the US Congress to report on a specific project will have a low continuation probability. Staff members of US Senators have high continuation probabilities. The summer interns at a firm have a low continuation probability, while the firm's partners have a higher continuation probability. We capture this diversity as simply as possible, in the form of two continuation probabilities.

We assume that the continuation probability $\underline{\delta}$ 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.[3]

If player i plays C in some period t, then player j privately observes the signal c with probability $1-\varepsilon$ and observes signal d with probability $\varepsilon \in [0, 1/2]$. Similarly, when i plays D in some period t, then player j privately observes the signal c with probability ε and observes signal d with probability $1-\varepsilon$. The signals are drawn independently across players and periods. Players do not observe other's signals.

A history at time t is denoted by h^t and includes a choice of discount factor δ as well as action profiles $a^k \in \{C, D\}^2$ and signal profiles $s^k \in \{c, d\}^2$ for all periods $k < t$. The corresponding private history of player i is denoted by h_i^t and consists of δ and the own-action-and-signal pair (a_i^k, s_i^k) of player i for $k < t$. The set of histories is denoted by H, with typical element h. The set of private histories for player i is denoted by H_i, with typical element h_i.

A strategy for player i is denoted by $\sigma_i : H_i \to \Delta\{C, D\}$, describing the (possibly mixed) action taken after every possible private history h_i. The strategy profile is denoted by σ.

2.3 Modeling Opponents

The analogy-based expectations equilibrium (Jehiel [21]) that we examine rests on the three common pillars of equilibrium concepts for sequential games—

[3]For our purpose, the universal defection arising in such games is what matters. Any alternative specification (for example, based on varying the monitoring technology or the stage-game payoffs) that would give rise to the same behavior would be equivalent. For example, these games may be played by sequences of short-lived players.

a consistent model of the opponent's behavior, Bayesian updating of beliefs within this model in response to experience, and best responses to the updated beliefs. The departure from more familiar concepts lies in providing an empirical foundation for the model of opponents' behavior, rather than assuming that these models simply appear as part of the equilibrium concept.

We assume that currently-matched players have access to a record of previous plays of the game by other agents. The current players adopt the frequencies of behavior observed in the record as their model of their opponent's behavior in their current interaction.[4]

This record exhibits two imperfections. First, the record includes both high continuation probability interactions (in proportion α) and low continuation probability interactions (in proportion $1-\alpha$). The current player cannot observe the continuation probability attached to each observation of a previous game in the record. Second, the record reports the actions taken by each player in each period of each observation, but not the signals observed by the players.

There are three potential sources of tension in this formulation of the record. First, players can observe the continuation probability in their own interaction, but not in the interactions in the record. Our view is that we can never expect an interaction 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. Moreover, the players will deliberately ignore other 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 will be too sparse to be useful. Hence, the record will include some games that are analogous to, but not precisely the same as, the current interaction. We are especially interested in differences which affect the ability to support cooperation, which we capture by assuming that the record includes interactions with varying, unobserved continuation probabilities.

Second, the record does not report the private signals observed by the players, though each player obviously observes their signals in their own interaction. Our interpretation is that these past signals are unobservable for much the same reason that current signals are private. Indeed, subsequent observers may not even understand what form these signals might take, or what means the players in previous interactions may have had for collecting and using information. A firm's sales force may hear comments from customers that provide clues as to the behavior of their rivals, while being unable to observe corresponding information flows in past plays of the game.

[4] Young's [46] analysis of conventions is similarly based on the assumption that current players play best responses to the frequencies of play in a record of past play by other agents. Young emphasizes the dynamics that arise because current players take a sample of a record that is perturbed by mutations. Fudenberg and Levine's [14, 15] self-confirming equilibrium similarly assumes that players play best responses to beliefs consistent with frequencies of play in a record, with the leeway in forming out-of-equilibrium beliefs playing an important role.

Third we assume that actions are reported in the record, even though they are unobservable in the current interaction. Here, our presumption is that hindsight can reveal much that is hidden at the time. Just as one can view research in history as a process of making previous actions observable, we expect the record to include observations of actions not available at the time at the time the actions are taken. A firm may be unable to detect secret discounts offered by rivals to their customers, but retrospective analysis may be more informative.

Of course in each of these cases we should expect the reality to be less sharp than our model. Current players will have some inkling as to the relevant continuation probability in a previous interaction, as well as some hints of what signals were received, while some of the actions will be obscure. We view our formulation as a conveniently stylized way to capture the differences in observations between current and historical interactions.

Each previous play of the game thus contributes an observation to the record listing the actions taken in each period of a previous game. These observations will be of various lengths (recalling that δ is a continuation probability), though all will be finite. A previous game that terminated in its tth period specifies the pair of actions taken in that game in each of its periods 0 through t.

A player uses the record to estimate the probability that her opponent in her current interaction will cooperate, given that the current interaction has reached some period t with 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 analogy classes, and then calculates the empirical frequency of C and D actions in the record for each analogy class. She then attaches this probability to every history in the analogy class and assumes these probabilities describe the behavior of her current opponent. We defer until Section 4 the question of how these analogy classes are determined.

Formally, each player i is endowed with an analogy partition An_i, that is a partition of H. We require that if two histories h and $\tilde{h}$ agree in the sequence of actions, then h and $\tilde{h}$ necessarily belong to the same analogy class in An_i. This reflects our assumption that the record reports actions but not signals. However, a single analogy class may also contain histories with different sequences of actions, presumably reflecting a view that these differences are relatively unimportant. We let $\mathfrak{a}_i$ denote a typical analogy class in An_i, and for every history h we refer to $\mathfrak{a}_i(h)$ as the analogy class in An_i to which h belongs. For each analogy class $\mathfrak{a}_i$, player i identifies all histories in the record corresponding to $\mathfrak{a}_i$ after which an action is taken (histories after which the game ends with no further actions are irrelevant), and then calculates the proportion of cooperation at these histories. An analogy-based expectation for player i is denoted by $\beta_i : An_i \to \Delta\{C, D\}$, where $\beta_i(\mathfrak{a}_i)$ is the frequency of cooperation observed at

histories in $\mathfrak{a}_i$ (in the record of interactions).[5] These expectations constitute player i's model of her opponent.

For example, suppose that An_i contains just two analogy classes, the first including histories in which there has never been a defection ($\mathfrak{a}_1$), and the second including histories in which there has been a defection ($\mathfrak{a}_2$). To keep the illustration simple, suppose the record consists of just the following two observations of past plays of the game:

Observation 1 :	Player 1 :	$CCCC$
	Player 2 :	$CCDD$
Observation 2 :	Player 1 :	DD
	Player 2 :	DD

The first observation comes from a game that lasted four periods, while the second comes from a game that lasted for only two periods. The following table lists the nonterminal histories that appear in these two observations, the observation from which each history is drawn, the analogy class containing the history, and the actions taken at that history:

History	Observation	AnalogyClass	Actions
$\emptyset$	1	$\mathfrak{a}_1$	$\begin{smallmatrix} C \\ C \end{smallmatrix}$
$\begin{smallmatrix} C \\ C \end{smallmatrix}$	1	$\mathfrak{a}_1$	$\begin{smallmatrix} C \\ C \end{smallmatrix}$
$\begin{smallmatrix} CC \\ CC \end{smallmatrix}$	1	$\mathfrak{a}_1$	$\begin{smallmatrix} C \\ D \end{smallmatrix}$
$\begin{smallmatrix} CCC \\ CCD \end{smallmatrix}$	1	$\mathfrak{a}_2$	$\begin{smallmatrix} C \\ D \end{smallmatrix}$
$\emptyset$	2	$\mathfrak{a}_1$	$\begin{smallmatrix} D \\ D \end{smallmatrix}$
$\begin{smallmatrix} D \\ D \end{smallmatrix}$	2	$\mathfrak{a}_2$	$\begin{smallmatrix} D \\ D \end{smallmatrix}$

Notice that a given observation contributes multiple histories. A game that lasts four periods is informative about what actions are taken after the null history as well as histories of lengths one, two and three. The histories $\begin{smallmatrix} CCCC \\ CCDD \end{smallmatrix}$ and $\begin{smallmatrix} DD \\ DD \end{smallmatrix}$ do not appear on this list because these are terminal histories after which no actions occur, and hence are not relevant in estimating frequencies of play.

Four of the histories in this record fall into analogy class $\mathfrak{a}_1$. There are five observations of C after these histories and three observations of D, and so player i's estimate $\beta_i(\mathfrak{a}_1)$ of the incidence of cooperation after histories in $\mathfrak{a}_1$ is $5/8$. Two histories fall into analogy class $\mathfrak{a}_2$. There is one observation of C after these histories and three observations of D, and so player i's estimate $\beta_i(\mathfrak{a}_2)$ of the incidence of cooperation after histories in $\mathfrak{a}_2$ is $1/4$.

[5] It is possible that the record contains no histories corresponding to some analogy class $\mathfrak{a}_i$. In this case, we place no restrictions on the belief $\beta_i(\mathfrak{a}_i)$. Jehiel [22] explains how one could instead use trembles to discipline such beliefs. Such analogy classes do not arise in our analysis.

2.4 The Equilibrium Concept

The analogy-based expectations equilibrium concept, introduced by (Jehiel [21]), requires that the players' actions are best responses to their beliefs, and that these beliefs match the frequencies contained in an *infinite* number of draws from the equilibrium strategies. In practice, the record will be finite. Indeed, this was part of our motivation for restricting attention to a small number of analogy classes. As a result, in practice the players' beliefs will be perturbed by estimation error. This estimation error disappears with an infinite record, and the analogy-based expectations equilibrium concept thus isolates the implications of assuming players' beliefs are given by the empirical frequencies of play in a limited number of analogy classes, without the confounding effects of estimation error.

For each strategy profile σ and history h, we let $P^\sigma(h)$ denote the probability of reaching history h when players play according to σ (given the monitoring technology and realization of δ). We say that player i's belief β_i is *consistent* with the strategy profile σ if for every $\mathfrak{a}_i \in An_i$ that is reached with positive probability (i.e., such that there exists $h \in \mathfrak{a}_i$ with $P^\sigma(h) > 0$), we have:

$$\beta_i(\mathfrak{a}_i) = \frac{\sum_{h \in \mathfrak{a}_i} P^\sigma(h)\sigma_j(h)}{\sum_{h \in \mathfrak{a}_i} P^\sigma(h)}, \tag{3}$$

where, of course, $\sigma_j(h)$ is identified with $\sigma_j(h_j)$, with h_j being player j's private history associated to h.

The belief β_i for player i induces a β_i-perceived strategy of player j, denoted by $\sigma_j^{\beta_i}$ and defined by[6]

$$\sigma_j^{\beta_i}(h) = \beta_i(\mathfrak{a}_i(h)) \text{ for every } h \in H.$$

Let $P^{\sigma_i,\sigma_j^{\beta_i}}(h)$ be the (subjective) probability that player i attaches to reaching history h (for each $h \in H$) under the strategy profile $(\sigma_i, \sigma_j^{\beta_i})$. Then we say that σ_i is a *best-response* to the expectation β_i if σ_i is a best-response after every private history for player i, i.e., if for each private history of length $\tau \in \{0, 1, \ldots\}$ that arises with positive probability under $(\sigma_i, \sigma_j^{\beta_i})$, the strategy σ_i maximizes

$$\mathbb{E}_{P^{\sigma_i,\sigma_j^{\beta_i}}} \left\{ \sum_{t=\tau}^{\infty} u_i(a_i(t), a_j(t))\delta^{t-\tau} \mid h_i \right\},$$

where $a_i(t) \in \{C, D\}$ is the action taken by player i in period t and u_i is player i's stage-game payoff function. Notice that given the probability measure $P^{\sigma_i,\sigma_j^{\beta_i}}$, player i's maximization problem is identical to that of a conventional repeated

[6] Observe that $\sigma_j^{\beta_i}$ is not in general an admissible strategy for player j since j's strategy has to be measurable with respect to H_j and $\sigma_j^{\beta_i}$ need not be so.

game. The distinctive features of the analogy-based expectations equilibrium all appear in the formation of beliefs.[7]

Definition 1 *A strategy profile σ is an analogy-based expectation equilibrium given a profile of analogy partitions An if and only if there exist a profile β of analogy-based expectations such that for every player i*

1) β_i is consistent with σ, and

2) σ_i is a best-response to β_i.

We think of an analogy-based expectation equilibrium as a steady state of a learning process involving populations of players who would have access from previous play to the frequencies of behaviors in each of their analogy classes.

Remark 1 Player i is assumed to know his own monitoring technology (i.e., the statistical link between j's action and what signal i observes), his own payoff structure, and the continuation probability. However, because player i draws all of her inferences about j's behavior from the record, player i need not be aware of her opponent j's payoff structure or of the monitoring technology of player j (what j observes about the actions of i).

Remark 2 In some applications, data from past play would be anonymous, reporting profiles of actions but not which player chose which actions, forcing the players to work with analogy partitions that bundle histories that can be obtained from one another by permuting the roles of players i and j. In such anonymous feedback scenarios, we would also have to modify the definition of consistency and replace $\sigma_j(h)$ by $\frac{\sigma_1(h)+\sigma_2(h)}{2}$ in (3) so as to reflect that the behaviors of both players (not just j) would contribute to the aggregate frequency observed by player i. When symmetric strategies are considered (as we do in our analysis), the two notions of consistency are the same, but not otherwise. ■

Throughout the course of the interaction player i uses her signals to update her beliefs about which analogy class contains the current history, and hence her beliefs about her opponent's behavior, and plays best responses to these beliefs. Player i thus conditions her current actions on her past actions and past signals, while modeling opponents' actions as depending only on analogy classes, which contain no signals but possibly the past actions of both players. In some sense, player i appears to believe that she is more sophisticated than her opponent. But, our preferred interpretation here is that this reflects not inconsistency but ignorance. Player i uses all of the information at her disposal when choosing actions, and assumes that opponents do the same. However, i may not know what information j has available, entertaining a wide range

[7] Best-responses are defined only at private histories h_i arising with positive probability, as otherwise it is not clear with which distribution one should define the expectation operator. Beliefs are defined after every history, ensuring sequential rationality. As is familiar in the literature, adding trembles would allow us to deal with all histories. The resulting refinement is irrelevant for our purpose.

of possible information structures for player j, involving various signals and actions. A committed Bayesian would endow player i with a belief over the possible information structures of player j (and similarly for j), and solve for an equilibrium of the consequent game of incomplete information. We view this as taking us beyond the small worlds (Savage [37]) in which such an analysis is appropriate. Instead, we model player i as constructing the best empirical model of j allowed by the record, and then best responding to this model.

Similar ideas appear in other equilibrium concepts that capture various aspects of bounded rationality. Player i in a cursed equilibrium (Eyster and Rabin [12] fail to recognize that j's actions depend on j's information, even as i conditions her actions on her information. Player i in a behavioral equilibrium Esponda [10]) or the Bayesian Network personal equilibrium (Spiegler [39]) has a model of opponent's behavior that is somewhat more sophisticated, but still less sophisticated that i's own behavior. Each agent in a Berk-Nash equilibrium (Esponda and Pouzo [11]) entertains a model of her opponent's behavior that may be inconsistent with her own. Every player above level zero in a level-k equilibrium (Nagel [33], Stahl and Wilson [40]) believes they are more sophisticated than others in the game.

3 Equilibrium

3.1 Equilibrium Strategies

3.1.1 The Candidate Equilibrium

The equilibrium concept itself provides no guidance as to how many analogy classes a player is likely to use in examining the data, nor how these classes are to be determined. Our intuition is that the number of analogy classes is likely to be small, reflecting either limitations of the historical record or parsimony in the players' reasoning. Toward that end, we suppose that players arrange histories into two analogy classes, 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 2.3 contains an illustration of such a calculation.

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 (even if the record provides such information). 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.

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 and the probabilities p and q. Once this probability drops below a threshold, player i switches to defecting. Once player i defects, i views every subsequent history as being dirty (and hence the opponent as cooperating with probability q), and defects thereafter.

The estimated probabilities p and q are equilibrium phenomena—the estimated probabilities must match the empirical frequencies of behavior, which in turn must be optimal given the estimated probabilities. Even before solving the fixed-point problem, we can infer that 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 clean histories in a high-continuation-probability exhibiting a first defection. It may then 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 $q < 1$.

3.1.2 Restless Bandits

We first fix probabilities p and q and examine an individual player's problem. As we have noted, a player who has once defected will thereafter always defect. Each player must then solve a stopping problem, determining how long to cooperate before switching to defection. We formulate this stopping problem as a restless bandit problem. There are two arms, a C arm (corresponding to cooperating) and a D arm (corresponding to defecting).

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[8]

$$\phi(z,c) = \frac{zp(1-\varepsilon)}{zp(1-\varepsilon) + z(1-p)\varepsilon + (1-z)[q(1-\varepsilon) + (1-q)\varepsilon]} \tag{4}$$

with probability $zp(1-\varepsilon) + z(1-p)\varepsilon + (1-z)[q(1-\varepsilon) + (1-q)\varepsilon]$; while a d signal is observed and both arms move to state

$$\phi(z,d) = \frac{zp\varepsilon}{zp\varepsilon + z(1-p)(1-\varepsilon) + (1-z)[q\varepsilon + (1-q)(1-\varepsilon)]} \tag{5}$$

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 history 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

$$\begin{aligned} W(z) &= (1-\overline{\delta})[(zp + (1-z)q)(1+k)] + \overline{\delta}q(1+k) \\ &= (1-\overline{\delta})z(p-q)(1+k) + q(1+k). \end{aligned} \tag{6}$$

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 $(1-\overline{\delta})z(p-q)(1+k)$ 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.

[8] This is the probability player i attaches to the event that the period-$t+1$ history is clean, given probability z that the period-t history is clean and given that i played C and observed signal c, and is readily constructed from the following accounting of outcomes:

Current history	Probability	Signal/Next history	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	$(1-z)[(1-q)(1-\varepsilon) + q\varepsilon]$

.

3.1.3 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 $\overline{z}$.*

Proof The existence of an optimal policy is standard, having been established by Whittle [44], 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 (6), that

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

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

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

It follows from (4) that $\phi(z, \cdot)$ is increasing in z, and $V(z)$ is increasing in z (since an agent with belief z' can mimic the actions of a player with belief $z < z'$, with the former securing a higher expected payoff), and so $\mathbb{E}(V(\phi(z, \cdot)))$ 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. ■

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 repeated play of the consequent solution of the bandit problem would induce values $\hat{p}$ and $\hat{q}$ in the record, 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 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 include only observations of the form

$$\begin{array}{c} DDDDDD \\ DDDDDD \end{array} \quad .$$

The length of the observations will vary, but all observations will exhibit mutual defection in every period. There only clean histories in the record will then be null histories, since these are the only histories which exhibit no instances of defection. The only actions observed after such histories are defections, and so players observing this record will estimate $p = 0$. All other histories are dirty, and again the only actions observed after such histories are defections, and so players observing this record will estimate $q = 0$. Player i thus believes that her opponent will defect after every history, regardless of i's actions. Player i's best response is then similarly to always defect, giving an equilibrium featuring relentless defection. Notice that this argument would apply no matter what the analogy classes.

3.2.2 Cooperation

We turn to the existence of nontrivial (i.e., exhibiting at least some cooperation) equilibria. 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 restless bandit problem from pulling the C arm, the first pull of the D arm, and subsequent pulls of the D arm, as

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

We can now focus attention on the key aspects of these payoffs by subtracting $q(1+k)$ from each payoff, yielding an equivalent restless bandit problem with payoffs

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

If $p \leq q$, then any strategy in which the player chooses C at least once is dominated by the strategy of always choosing D, ensuring the equilibrium is

trivial.[9] ■

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.[10]

Proposition 1 *Let $\varepsilon \in (0, 1/2)$ and $\alpha \in (0, 1)$.*

[1.1] Suppose

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

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

[1.3] Suppose

$$k < \frac{\overline{\delta}}{2 - \overline{\delta}}.$$

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

Appendix 5.1 contains the proof.

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

$$\frac{k}{1+k} \leq \overline{\delta} p(p - q). \tag{7}$$

Notice that if $p = 1$ and $q = 0$, 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 $\overline{\delta}$ can decrease the estimate of q extracted from the record, making defecting less attractive.

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

[10] 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.

3.2.3 The Value of Cooperation

Proposition 1 ensures the existence of an equilibrium with some cooperation, but makes no statement as to how much cooperation we can expect. 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 Nash reversion equilibrium in a game of perfect monitoring. Appendix 5.2 proves:

Proposition 2

[2.1] Suppose $\overline{\delta}$ *satisfies*

$$\overline{\delta} > \frac{k}{1+k}. \tag{8}$$

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 $\overline{\delta}$ *satisfies*

$$\overline{\delta} > \frac{2k}{1+k}. \tag{9}$$

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 the history 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 cooperation to defection. 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 Nash reversion 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 Nash reversion 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, ensuring that adverse signals are sufficiently likely to reflect defection, 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. Notice that this is just the opposite of the reasoning exploited by the first condition, and the latter reasoning gives rise to precisely the type of inference that scuttles cooperation in standard repeated games of private monitoring. If adverse signals are interpreted as quirks of the noisy monitoring, how do incentives to cooperate arise? The argument in the current setting relies on the misspecification in the players' models of their interaction to verify the incentives to cooperate are not disrupted in the process. We return to this in Section 4.1.

3.3 Examples

We illustrate the results with two examples. To keep the notation uncluttered, we set $\underline{\delta}=0$ and denote $\overline{\delta}$ simply by δ.

3.3.1 Example 1: Perfect Monitoring

If $\varepsilon=0$, so that monitoring is perfect, 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 low continuation probability games. Each player will estimate $q=0$, observing dirty histories only in low continuation probability games whose players routinely defect. We see here the motivation for including low continuation probability games in the analysis, since otherwise there would be no observations on which to form this estimate.

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

$$p=\frac{\alpha\sum_{t=0}^{\infty}\delta^t}{\alpha\sum_{t=0}^{\infty}\delta^t+(1-\alpha)}=\frac{\frac{\alpha}{1-\delta}}{\frac{\alpha}{1-\delta}+1-\alpha}=\frac{\alpha}{\alpha+(1-\alpha)(1-\delta)}. \tag{10}$$

The numerator calculates the frequency of clean histories in the record after which a player cooperates. The denominator calculates the frequency of all

clean histories.[11]

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

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

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

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

or, using our solution for p and rearranging,

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

This inequality holds for sufficiently large δ, 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.

3.3.2 Example 2: Imperfect Monitoring

This example, returning to imperfect monitoring, illustrates the forces behind Proposition 2.1.

[11] Given the proposed equilibrium behavior, the distribution of *observations* in the record will be the following:

Observation	Probability
$\emptyset^D_D$	$(1-\alpha)$

Observation	Probability
$\emptyset^C_C$	$\alpha(1-\delta)$
$\emptyset^{CC}_{CC}$	$\alpha\delta(1-\delta)$
$\emptyset^{CCC}_{CCC}$	$\alpha\delta^2(1-\delta)$
$\vdots$	$\vdots$

.

For example, with probability $(1-\alpha)$, a low continuation probability is drawn, in which case both players defect and then the game ends, giving observation $\emptyset^D_D$. (Low continuation probability interactions would also contribute longer interactions had we not simplified by setting $\underline{\delta} = 0$.) With probability $\alpha\delta^2(1-\delta)$, for example, a high continuation probability is drawn (probability α), the players cooperate in the first period and the game continues two additional periods (probability δ^2) during which they also cooperate, and then the game ends (probability $1-\delta$). Now we calculate the proportion of cooperation at clean *histories*. The total incidence of clean histories (the denominator in (10) is given by 1 (every observation contains the clean history $\emptyset$) plus $\alpha\delta$ (proportion $\alpha\delta$ of the observations contain the clean history $\emptyset^C_C$) plus $\alpha\delta^2$ (proportion $\alpha\delta^2$ of the observations contain the clean history $\emptyset^{CC}_{CC}$) plus .... Rearranging this sum gives the denominator of the intermediate expression in (10). All but $1-\alpha$ of these these clean histories exhibit only cooperation (with the $1-\alpha$ remainder, corresponding to the clean history $\emptyset$ in low continuation probability interactions, exhibiting only defection), and subtracting $1-\alpha$ from the denominator in (10) gives the numerator.

The posterior probability that the history is clean, given that i has hitherto always cooperated, that i has a prior probability z that the history is clean, and that i observes a c signal, is denoted by $\phi(z,c)$ and given by (4). 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 the history 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. 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)

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

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 that the the history is clean, 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 denoted by $\phi(z,d)$ and given by (5). 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 show:

Proposition 3 *For all sufficiently large $\alpha < 1$ and δ, there exists $\varepsilon^*(\alpha)$ such that for all $\varepsilon < \varepsilon^*(\alpha)$, there exists an equilibrium in which high-continuation-probability agents initially cooperate, and continue to cooperate until receiving their first d signal, and then thereafter defect. As $\alpha \to 1$, the required lower bound on δ approaches $k/(1+k)$.*

This gives us an equilibrium of the type described in Proposition 2.1, with a value approaching one, the value of permanent cooperation.

Remark 3 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. ■

The proof of Proposition 3 makes the nature of our equilibrium construction clear, and so we present it in the remainder of this section.

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 the posterior z^- 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 $q \leq \varepsilon/2$. 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 defect in the next period with probability 1.

We can calculate p, obtaining:[12]

$$p = \frac{\alpha + (1-\varepsilon)\alpha\delta\sum_{n=0}^{\infty}(\delta(1-\varepsilon)^2)^n}{1 + \alpha\delta\sum_{n=0}^{\infty}(\delta(1-\varepsilon)^2)^n}.$$

We can simplify to

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

[12] Following the logic of footnote (11), 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	$\alpha\delta$	$(1-\varepsilon)$
2	$\alpha\delta^2(1-\varepsilon)^2$	$(1-\varepsilon)$
3	$\alpha\delta^3(1-\varepsilon)^4$	$(1-\varepsilon)$
	$\vdots$	

To 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 $\alpha\delta$, 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. Subsequent terms are analogous.

The key characteristic we will use is that p goes to $\frac{\alpha}{1-\delta+\alpha\delta}$ 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 latter calculation reflects the fact that as ε approach zero, a d signal is overwhelmingly likely to have come from a defection rather than an erroneous signal. This alone is not enough to ensure that z^- approaches zero—if p approaches 1, defections may themselves be yet more overwhelmingly unlikely than erroneous signals. However, p is approaching $\frac{\alpha}{1-\delta+\alpha\delta}$, ensuring that a d signal is interpreted as a defection, and hence that z^- is arbitrarily small.

It remains to confirm incentives. 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, and so we 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

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

recalling that $\phi(z,c)$ 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 $W(\phi(z,d))$ 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

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

From (6), we have $W(z) = (1-\delta)z(p-q)(1+k) + q(1+k)$. The incentive constraints for equilibrium are

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

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 $V(z^-) = (1-\delta)(-k)$, 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 $p + (1-p)(-k) \geq (1-\delta)p(1+k)$, then becomes

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

which, using our limiting expression for p, becomes

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

If we were to now let α approach one, then we would recover the limit $\delta \geq \frac{k}{1+k}$ from the traditional repeated game of perfect monitoring.

4 Discussion

4.1 The Importance of Misspecified Models

Equilibrium cooperation rests on three pillars. First, player i must believe that player j will (at least sometimes) 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.

The basic difficulty is a tension between conditions one and two. Under a Nash reversion equilibrium hypothesis, j's interpretation of a first-period d signal is that i cooperated and the signal is erroneous. Given this, j will continue 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 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. The proofs of Propositions 1 and 2 complete the argument by 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. Nonetheless, i observes that histories with a defection exhibit more subsequent defection than do histories without, and this empirical regularity leads player i to overestimate the potential of an initial defection to induce opponents to defect. This overestimation is the key to Proposition 2.2. We can illustrate the mechanics of this

misspecification in a particularly stark setting, in the process making it clear that players in our setting can support cooperation under conditions that would ordinarily consign them to persistent defection.

Let $\varepsilon = 1/2$, so that signals carry no information. To keep things simple, we again let $\underline{\delta} = 0$ and denote $\overline{\delta}$ simply as δ. As before, each player believes that the opponent cooperates with probability p after clean histories and probability q after dirty histories, and 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, each player will cooperate, 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.

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 as t increases, the probability that j has not yet defected declines.

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

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

The numerator calculates the frequency of clean histories after which a player cooperates.[13] The denominator calculates the frequency of all clean histories.

As T grows from 0 to ∞, the value of $p(T)$ grows from $\alpha/(1+\delta\alpha)$ to $\alpha/(1-\delta+\delta\alpha)$. 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 clean histories when $T=\infty$) or δ (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 $p(T)$ satisfy the incentive constraints. The incentive constraints for these periods will ensure the incentive constraints hold in other periods. The value of cooperation

[13] 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 δ we also get a period-1 history added to the list, with probability δ^2 we get a period-2 history, and so on, through probability δ^T that we get a period-T history. After that, there is no more cooperation.

in period T is $z_T[p(1+\delta p(1+k))+(1-p)(-k)]+(1-z_T)(-k)$. 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 $z_T[p(-k+\delta p(1+k))+(1-p)(-k)]+(1-z_T)(-k)\geq 0$, or

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

The equilibrium condition is then

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

We can use (11) and rearrange to obtain

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

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 δ 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 δ approaches one, this initial cooperation fades into insignificance, with payoffs approaching zero ($\delta^T \to 1$), or whether are payoffs bounded away from zero ($\delta^T < 1$). The latter is the case.[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

[14]To see this, we first note that

$$p^T \approx \frac{k}{1+k} \tag{14}$$

as δ 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, at 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

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

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

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

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 δ^T.

that for large T, we have $\lim_{\alpha \to 1}(p(T))^{T+2} = 1$. Hence, as α approaches 1, as long as $k/(1+k)\delta < 1$, 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.

It is obviously impossible for i's actions to affect j's behavior in this case. Nonetheless, i interprets the regularity in the record that defection tends to be followed by increased defection as indicating a link between a defection by i and j's subsequent behavior, prompting i to initially cooperate. While cases in which players receive no information about opponents' actions are likely to be exceptional, the point is that analogical reasoning can support cooperation in noisy monitoring situations that standard equilibria could not.

4.2 What Difference Does an Analogy Make?

What determines the players' analogy classes? We explore two possibilities. To keep things simple, we continue to let $\underline{\delta} = 0$ and to denote $\overline{\delta}$ simply as δ.

4.2.1 More Analogy Classes

One possibility is that the analogy classes are shaped by the nature of the information contained in the record. For example, the record in our two-analogy-class example may report profiles of actions while not distinguishing which player took which action. The players may eschew some analogy classes because the data are too thin to provide reliable estimates. Other analogy classes may be combined because they yield similar probabilities of cooperation. In general, the analogy classes in this interpretation reflect primarily characteristics of the data.

In this section, we illustrate the issues that arise when comparing different configurations of analogy classes by examining a model with three analogy classes. These analogy classes refine the two-class partition of Sections 2–3 by allowing players to take account of which participant in a dirty history has defected.

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 player i 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.

Player i's model of player j is that j is initially healthy, and that j cooperates with probability p when healthy, cooperates with probability q when exposed, and cooperates with probability r when infected. As long as i continues to cooperate, i will view j as being either healthy or infected, and will update the probability i attaches 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.

The candidate equilibrium strategies are that each player begins by cooperating. 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. 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 Appendix 5.3. We can then show that we have a nontrivial equilibrium as long as players are sufficiently patient and the monitoring is sufficiently precise. Appendix 5.3 proves:

Proposition 4 *Suppose*

$$k < \frac{\delta\alpha^2}{4 - \overline{\delta}\alpha^2} \tag{17}$$

Then there exists $\overline{\varepsilon}$ such that for all $\varepsilon < \overline{\varepsilon}$, 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

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

The next step is to place some bounds on the values of p and q. 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).

Now we fix the continuation probability δ and show that, as the monitoring structure becomes increasingly precise and the proportion of impatient players shrinks, then there exists an equilibrium with payoff approaching 1, the payoff of the Nash reversion equilibrium in a game of perfect monitoring.

Proposition 5 *Fix $\delta \geq \frac{k}{1+k}$. 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 $\delta \geq \frac{k}{1+k}$ on the continuation probability is precisely the bound for the Nash reversion 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 establishing the sufficient condition that, for small ε and large α, we have

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

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

The sufficient conditions for supporting cooperation are more demanding in the case of three analogy classes. Proposition 4 requires ε small while Proposition 1 does not, and Proposition 5 requires an order of limits while Proposition 2 does not. This is initially unexpected. Under two analogy classes, the punishment consists of receiving $q(1+k)$ forever, while under three analogy classes, $q(1+k)$ is received only temporarily, until the opponent switches from exposed to infected, giving a seemingly more severe punishment.

It may well be that three analogy classes gives rise to a more severe punishment that more readily supports cooperation. However, the reverse may also occur. The proof of Proposition 1 shows that a sufficient condition for a player to cooperate in the two-analogy-class case is given by (7):

$$\frac{k}{1+k} \leq \delta p(p-q),$$

while the counterpart (18) of this condition appearing in the proof of Proposition 5 shows that the with three analogy classes a *necessary* condition for cooperation is:

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

Since $1/(1-q\delta) > 1$, the second condition appears less demanding, but this misses the fact that the two-analogy-class system gives a smaller estimate of q than does three analogy classes. With three analogy classes, 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 analogy classes, 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, giving a smaller value of q and hence a more intimidating punishment. One can then construct examples in which cooperation is possible under two but not three analogy classes.[15]

4.2.2 Information Design

We have thus far taken the analogy classes to be exogenously fixed, focusing attention on the extent to which such analogy classes can support cooperation in an analogy-based expectation equilibrium. This may be appropriate if the

[15] For example, fixing $\alpha \in (0,1)$ and $\varepsilon \in (0,1/2)$, letting $\delta \to 1$ ensures that $q \to 0$ in the two-analogy-class case. For sufficiently large δ, the necessary condition under three analogy classes will be more demanding than the sufficient condition under two analogy classes if

$$p_2^2 > \frac{p_3(p_3-q_3)}{1-q_3},$$

where p_2 is the equilibrium value of p from the two-analogy-class case, and so on.

specification of the analogy classes is fixed by the nature of the information in the record or by cognitive limitations of the players.

An alternative is that a designer chooses the analogy classes and the attendant equilibrium in order to maximize the players' payoffs. The designer can perhaps be viewed as a third party in charge of recording and disclosing feedback from past interactions. The designer's objective is to maximize the players' payoffs,[16] while the designer faces the constraint that the players form their models of their opponents' behavior in an empirical fashion as formalized in the analogy-based expectation equilibrium.

We continue to assume that signals are not accessible from past interactions, which puts constraints on the type of analogy partitions that can be considered by the designer. Depending on the context, additional constraints may arise. For example, the record may keep track of actions but not the identity of the players taking those actions, restricting the designer to analogy classes similar to those examined in Sections 2–3. There may also be constraints on the number of analogy classes if it is more costly for the designer to disclose (or difficult for players to absorb) a larger number of aggregate statistics. But beyond these constraints, the designer is free to consider the analogy partitions of her choice.

This perspective has some similarity with the information design perspective developed in the Bayesian persuasion literature pioneered by Kamenica and Gentzkow [24], but with some notable differences. Most importantly, the sender in a game of Bayesian persuasion shapes the monitoring technology determining the information that is transmitted to the receiver in the game that the sender plays with the receiver. In contrast, the monitoring technology remains the same in our game regardless of the analogy partitions. Instead, the information design we are considering concerns the feedback from past interactions that determines how the new players in the current game model their opponents. The analogy-based expectation equilibrium is the tool we use to model the effect of such feedback in the steady state, in contrast to the Bayes-Nash equilibrium in the Bayesian persuasion literature. Nonetheless, one can view the perspective suggested here as extending the question of information design to a repeated-game setting.

Our perspective also bears some similarity with the community enforcement perspective pioneered by Kandori [25], to the extent that it seeks the feedback to be passed to new players that will most effectively induce cooperation. Our formulation differs from the previous community enforcement literature in several key dimensions. First, we consider repeated interactions with the same players while the community enforcement literature has focused on frequent rematching. Second, the feedback in our setting is not about the past history of the current partner but about the collective play of past teams in the community, which shapes expectations in the current match. As is the case for the Bayesian persuasion literature, the misspecification present in our approach has no counterpart in the community enforcement literature.

When analogy partitions can differentiate the actions of the two players (i.e.,

[16] We keep things simple by assuming the designer entertains only symmetric equilibria.

actions are not recorded in an anonymous way), as in Section 4.2.1, then the optimal design problem has a straightforward solution. The following proposition identifies an optimal analogy partition for α and δ large enough.

Proposition 6 *Consider the analogy partition specifying that each player i's history is dirty if i has ever defected, and is clean otherwise. There exist δ^* and α^* such that for all $\delta > \delta^*$ and $\alpha > \alpha^*$, it is an analogy-based expectation equilibrium that players in high continuation probability interactions cooperate after clean histories and defect after dirty histories, while players always defect in low probability interactions. This leads to a payoff no smaller than the one attained in any analogy-based expectation equilibrium whatever the analogy partitions.*

Proof The equilibrium path induced by these strategies is that players in high continuation payoff interactions always cooperate, while those with low continuation payoffs always defect. The players will then estimate that q, the probability of cooperating after a dirty history, is zero, since all such histories appear in low continuation payoff interactions. The probability of cooperation after a clean history will be

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

The incentive constraint that cooperation is optimal at clean histories for players with high continuation probabilities is

$$p-(1-p)k \geq (1-\delta)p(1+k),$$

which will obviously hold if α and δ are sufficiently large. It is clear that independently of the analogy partitions, no analogy-based expectation equilibrium can deliver a higher payoff, establishing the final part of the proposition. ■

We leave for future research the study of the best analogy partitions when other considerations preclude the construction in Proposition 6, such as when the record does not report the identities of the players taking the various actions and thus analogy partitions must be anonymous.

4.3 Relationship to the Literature

Escaping Matsushima. As noted in Section 1.2, any attempt to support cooperation in the face of private monitoring must break free of Matsushima's [30] three conditions (independent signals, pure strategies, strategies measurable with respect to beliefs about opponent histories). Like the bulk of the literature that followed Matsushima [30], we retain the independent monitoring (with Sugaya [41] pioneering the extension to arbitrary signal structures), while relaxing the restriction to pure strategies. The mixtures in conventional models ensure that adverse signals are informative about behavior, and hence can prompt incentive-creating reactions, rather than be interpreted as quirks of

the noisy monitoring. Analogously, the inclusion of low continuation probability interactions in the record ensures that no player expects an opponent to cooperate with probability one, again allowing signals to be interpreted as conveying information about behavior that prompts incentive-creating responses.

Belief-Based Equilibria. Our analysis then proceeds on two fronts. Proposition 2.1 exploits reasoning similar to that in Sekiguchi [38]. Players in Sekiguchi's analysis initially mix between always defecting and playing a counterpart of the Nash reversion strategy. Each agent i who chooses the latter continually updates her beliefs until the probability that the opponent both initially chose the Nash reversion strategy and is still cooperating becomes low enough that i switches to defecting. Each player's behavior thus consists of a string of cooperation, until making a permanent switch to defection.

The initial probability attached to always defecting in Sekiguchi's equilibrium is the functional equivalent of the low-continuation-probability interactions in our analysis, and the subsequent beliefs are the counterpart of our agents' beliefs that the history is still clean. The equilibrium described in Proposition 2.1 again yields an initial string of cooperation for each player, but both players ultimately switch to perpetual defection. Conceptually, our formulation differs in that prior beliefs emerge from an empirical interpretation of the data and are then processed via a misspecified model, rather than coming to life as part of the equilibrium concept in a correctly-specified model. The resulting behavior is qualitatively similar, though we have effectively purified Sekiguchi's initial mixture.

Belief-Free Equilibria. In a belief-free equilibrium, each player i is in each period indifferent between playing C and D. Players cooperate in the first period and in each subsequent period each player i chooses a mixture that depends only on i's previous action and signal. Each player is more likely to cooperate after having received a c signal, with this difference in the opponent's future behavior calibrated to make a player indifferent between playing C and D in each period.[17] The behavior thus exhibits a form of stationarity, with the

[17] Letting π_{Xy} be the probability attached to cooperate after having played $X \in \{C, D\}$ and received signal $y \in \{c, d\}$, we have

$$\begin{aligned}
\pi_{Cc} &= \frac{(V_C - \delta V_D)(1-2\varepsilon) + (1-\delta)[k + \varepsilon - (1-\varepsilon)(1+k)]}{\delta(1-2\varepsilon)(V_C - V_D)} \\
\pi_{Cd} &= \frac{(V_C - \delta V_D)(1-2\varepsilon) + (1-\delta)[\varepsilon - (1-\varepsilon)(1+k)]}{\delta(1-2\varepsilon)(V_C - V_D)} \\
\pi_{Dc} &= \frac{(1-\delta)[k + V_D(1-2\varepsilon) - k\varepsilon]}{\delta(1-2\varepsilon)(V_C - V_D)} \\
\pi_{Dd} &= \frac{V_D[(1-2\varepsilon) - k\varepsilon]}{\delta(1-2\varepsilon)(V_C - V_D)},
\end{aligned}$$

where V_C and V_D are the expected continuation to player i when player j plays C and when j plays D. The values V_C and V_D must be chosen so that these are all probabilities, and can both be chosen arbitrarily close to one for sufficiently large δ and small ε.

previous-period outcome determining current-period behavior, no matter what the current period. Supporting payoffs close to those of perpetual cooperation is then a matter of verifying that as players become patient and the monitoring precise, the mixtures can be chosen so as to place relatively little weight on defection, allowing high payoffs to be sustained throughout the interaction.

Proposition 2.2 follows Ely and Välimäki [9] in exploiting the dependence of i's current action on i's previous action. However, the mechanisms are quite different. The continually adjusted actions in a belief-free equilibrium (as seen in footnote 17) induce a stationary pattern of behavior, while players in our construction typically have strict incentives and play pure strategies, beginning with a persistent string of cooperation and culminating in persistent defection. Payoffs close to those of perpetual cooperation in a belief-free equilibrium can be obtained (under appropriate conditions) by tuning the mixtures appropriately. Supporting such payoffs in our setting is a matter of verifying that the switch to cooperation can be postponed so long as to have a negligible effect on payoffs. Unlike the belief-free equilibrium, this does not require that monitoring become close to perfect.[18] The indifferences supporting the mixtures in a belief-free equilibrium reflect correct beliefs about the strategies of their opponents. In the equilibrium of Proposition 2.2, agent i's preference to defect after having once defected reflects the fact that i has estimated an average of j's behavior across the the histories in the analogy class dirty, causing i to overestimate the subsequent probability that j will defect.

Misspecification. Compte and Postlewaite [7] examine a model of cooperation in the repeated prisoners' dilemma that shares with our analysis the classification of histories into categories. Their interpretation is that players resort to categorization not because they must estimate opponents' play from a limited record, but because the set of feasible strategies is limited by psychological considerations and cognitive limitations. Their analysis focuses on cases in which players can be in one of two mental states, with a strategy induced by attaching an action to each state and specifying rules for how the players move between states. The mental states are reminiscent of our analogy classes, but because the restriction is placed directly on the strategies of both players, the players are effectively solving a game with limited strategy space with the usual mutual best-response requirement.

As we noted in Section 2.4, our model joins a collection of misspecification-based equilibrium concepts, while applying such reasoning to the question of cooperation in repeated games.[19] Interestingly, Hansen, Mishra and Pai [18] report that oligopolistic firms often use machine learning algorithms to investigate their demand curves. In doing so, each firm i typically assumes firm j's pricing behavior is fixed and unresponsive to i's behavior. In our terms, each firm i

[18] Section 1.2 notes that Matsushima [31] and Yamamoto [45] relax the nearly-perfect-monitoring requirement by building review phases into the equilibrium. Our equilibria have no counterpart of review phases, with an obvious indication of the difference being that the switch to defection is permanent in our equilibria.

[19] Jehiel and Samuelson [23] apply similar ideas in a model of reputation building.

puts all histories into a single analogy class. Hansen, Mishra and Pai identify conditions under which the firms' algorithms (unknowingly) induce correlated price experiments, causing each firm to underestimate the price sensitivity it faces and leading the firms to set high, effectively collusive prices. One can view the machine learning as the counterpart of our sampling from a record, the assumption that opposing firms do not react to one's own pricing as the counterpart of our analogy classes, and the collusive behavior as the counterpart of our cooperative outcomes.

Monitoring. The equilibria we have constructed in general feature only temporary cooperation, though the expected duration of cooperation may be long relative to the expected length of the relationship. Observational studies of cooperation stress the importance of improving the monitoring in making lasting cooperation possible.

Marshall and Marx [29] emphasize that monitoring in oligopoly interactions is inherently noisy and private—firms have no way of ascertaining how well their own actions are tracked by others, or whether the seemingly abnormal quantities of others reflect deliberate actions or the capriciousness of the market. As a result, Marshall and Marx 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.

Porter [36] and Ulen [42, 43] report that the Joint Executive Committee, the governing body of a railroad cartel operating in the 1880s, 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. In a similar vein, Levenstein and Suslow [26] and Harrington and Skrzypacz [19, 20] stress the importance cartels place on disseminating public sales information to their members, perhaps through the creation of joint sales agencies or trade associations.

Elinor Ostrom (e.g., [34]; see also Ellickson [8]) takes a similar view of cooperation in the management of common resources. Ostrom again emphasizes the importance of continual communication and the creation of informal mechanisms or explicit agreements specifying how the participants are to monitor and verify other's behavior. Blomquist, Schlager, Tang and Ostrom [5] identify features that are common to a large number of cases in which cooperation has been sustained in the use of common-pool resources, including organized procedures for monitoring actions, making deviations known, and assessing and enforcing sanctions. These monitoring arrangements were often backed by formal institutions.

Our interpretation of this literature is that one cannot typically expect to sustain *permanent* cooperation without converting private monitoring into pub-

lic monitoring.[20] 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. Relations between countries, where institutions to provide monitoring are sparse, are one 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 led to the demise of the empire) while in the British empire, cooperation persevered but the empire withered away.

Experiments. We can interpret the results of two experiments in light of our model. First, 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 also 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. They 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. This is consistent with our finding that seemingly temporary cooperation can be quite valuable.

Second, Section 4.1 explains how equilibria supporting cooperation can arise when signals are arbitrarily noisy, perhaps even uninformative. Experiments on the repeated prisoners' dilemma with private monitoring have naturally focused on relatively informative signals. However, Aoyagi and Frechette [2] included some treatments of the repeated prisoners' dilemma in which subjects received no feedback as to their opponents' actions. The players in these interactions typically do not come close to sustaining full cooperation. However, consistent with the equilibria sketched in Section 4.1, the players do sometimes cooperate. The average incidence of cooperation was .31.[21] Figure 1 shows the average

[20] Perfect monitoring may be literally impossible, so the goal may better be described as making the monitoring pubic and sufficiently close to perfect that the results of Mailath and Morris [27, 28] apply.

[21] The average incidence of cooperation is the number of times the action C was observed divided by the total number of actions. The results are complied from the original data from Aoyagi and Frechette's [2] experiment, available at

incidence of cooperation in each period (aggregated over all games) and the average incidence of cooperation in each of the games (aggregated over periods) in the ten-game sequence of repeated games. The former hints at a slight tendency for cooperation to increase as the play within a game proceeds, while the latter suggests that cooperation decreases as participants gain experience with the game.

The incidence of cooperation in these games is noticeably lower than that of comparable experiments with monitoring (Aoyagi and Frechette [2, Table 2, p. 1146]), but noticeably more cooperation than in one-shot prisoners' dilemma games (Dal Bó [6, Table 5, p. 1599, lines 1 and 4]). These results suggest that cooperation in a repeated prisoners' dilemma without monitoring is conceivable, while it is clear the the equilibrium presented in Section 4.1 is far from a perfect match for the data.[22]

4.4 Beyond the Prisoners' Dilemma

Our equilibrium construction relies heavily on the fact that the prisoners' dilemma has only two actions. Can we extend the analysis to more general games?

While we leave for future research the study of general repeated games, we suggest how our two analogy class construction can be extended to the simplest instance of the setting that provided much of the early motivation for examining collusion, the Cournot duopoly. In the stage game firms 1 and 2 choose quantities x_1 and x_2. Prices are then given by

$$\begin{aligned} p_1 &= f(x_1, x_2) + \varepsilon_1 \\ p_1 &= f(x_2, x_1) + \varepsilon_2, \end{aligned}$$

where $f : \mathbb{R}^2_+ \to \mathbb{R}$ is decreasing in both arguments and ε_1 and ε_2 are identically-distributed random variable with zero means and full support on $\mathbb{R}$. Profits are then given by

$$\begin{aligned} \pi_1(x_1, x_2) &= x_1[f(x_1, x_2) + \varepsilon_1] \\ \pi_1(x_1, x_2) &= x_2[f(x_2, x_1) + \varepsilon_2]. \end{aligned}$$

Our interpretation is that the firms are selling differentiated substitutes. The function f describes demand conditions in the market, with $f(x_i, x_j)$ giving the expected price received by firm i when i sets quantity x_i and firm j sets quantity x_j. The expected price of i is decreasing in firm i's quantity, as expected, and is also decreasing in firm j's quantity, as is expected in the case of substitutes. The random variables ε_1 and ε_2 capture idiosyncratic shocks to the firms' demands. We keep things simple by building symmetry into the function f determining expected prices as well as into the demand shocks, and by assuming that marginal costs are constant, at which point one can incorporate the marginal costs into the function f.

http://people.cess.fas.nyu.edu/frechette/data/Aoyagi_2009a_data.txt.

[22] A tighter experimental test of our approach would provide subjects with summary statistics of behavior observed in similar interactions, consistent with our analogy classes.

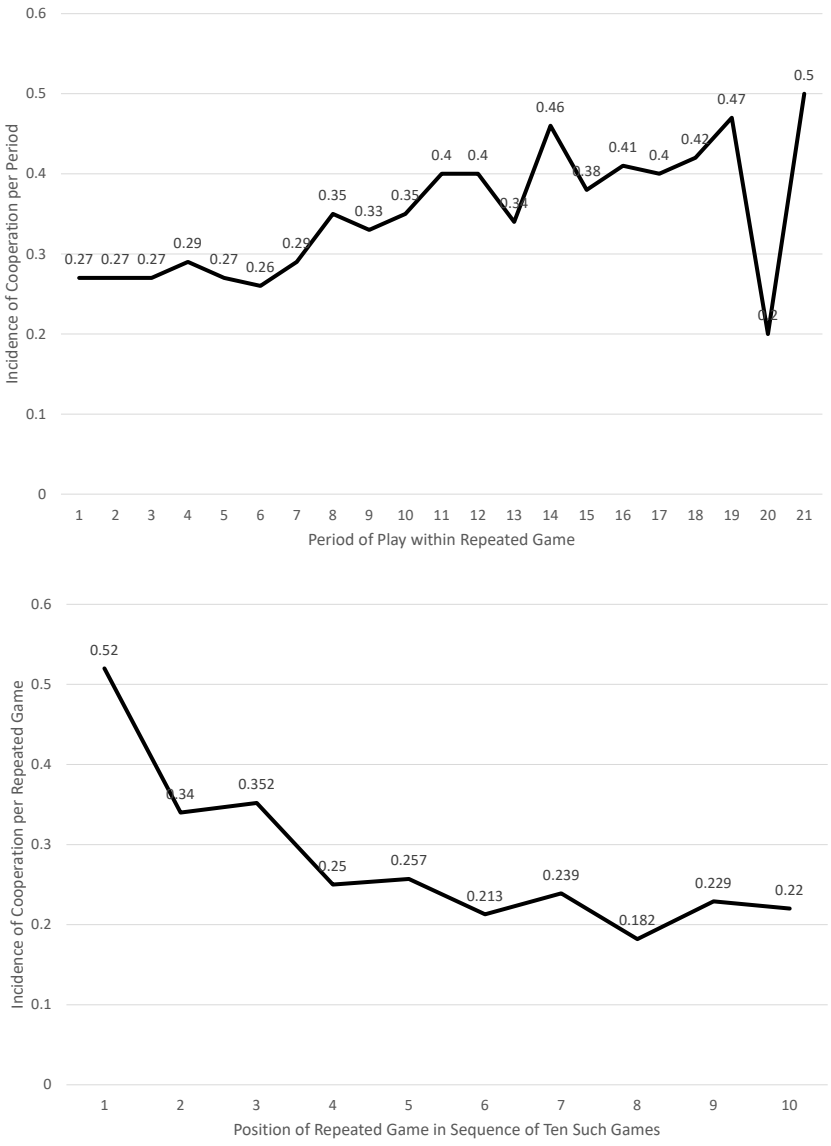

Figure 1: Summary of data from the no-monitoring treatment of Aoyagi and Frechette [2]. Participants played ten repeated-prisoners'-dilemma games, each of random length. The top panel shows the incidence of cooperation in each period, aggregated over all games. The sample sizes for later periods become small, as few games lasted that long. The bottom panel shows the average incidence of cooperation, aggregated over periods in a game and aggregated over all games in each position, for positions one through ten in the sequence of ten repeated games.

In the stage game, the firms simultaneously choose quantities x_1 and x_2, and then the random variables ε_1 and ε_2 are drawn and prices are realized. Firm 1 (and similarly firm 2) chooses its quantity x_1 to maximize its expected profit given x_2, and hence solves $\max_{x_1} x_1 f(x_1, x_2)$. We assume that $x_i f(x_i, x_j)$ is strictly concave in x_i for all x_j, so that best-response functions are downward sloping, and that the stage game has a unique Nash equilibrium, denoted by (x^N, x^N).[23] Let (x^*, x^*) denote the profile that maximizes the sum of the players' payoffs in the stage game, and note the familiar result that $x^* < x^N$.

When players are matched to play the repeated game, with probability α they are drawn to have a zero continuation probability, in which case they play (x^N, x^N). With the complementary probability they are drawn to have a higher continuation probability δ.[24] In the repeated game, we assume that firms observe their own quantities and prices, but not those of the opponent. Firm i's quantity and price allow i to draw inferences about j's quantity.

The record consists of sequences of quantities but not prices. Analogously to our treatment of the prisoners' dilemma in Sections 2–3, suppose players arrange histories into two analogy classes, clean and dirty. There exists a quantity $\overline{x} \in [x^*, x^N)$, described in more detail presently, such that a history is clean if it contains only quantities smaller than $\overline{x}$, and is otherwise dirty.

As in the prisoners' dilemma, there will always exist an equilibrium in which every player chooses x^N after every history. We are interested in conditions under which the following candidate equilibrium is indeed an equilibrium. The equilibrium is characterized by the quantity $\overline{x}$, a cutoff probability $\overline{z}$, a quantity $\underline{x} > \overline{x}$ and a function $\hat{x}(z)$ that maps beliefs into quantities larger than $\overline{x}$. Player i plays $\overline{x}$ as long as the posterior probability z that the history is clean remains above $\overline{z}$. If z dips below $\overline{z}$ and player i has hitherto played only $\overline{x}$, then player i plays an action $\hat{x}(z) > \overline{x}$ that depends on the belief z . Once player i has played such an action $\hat{x}$, then i plays the action $\underline{x}$ in every subsequent period.

The record will then consist of some observations in which players immediately play x^N, because they have drawn a low continuation probability. In other (high continuation probability) observations, both players will initially play $\overline{x}$. This will continue until one player (or both, if both players' posteriors dip below $\overline{z}$ for the first time in the same period) plays $\hat{x}(z)$ for a value of z just below $\overline{z}$. There will then appear some periods in which one player plays $\underline{x}$ while the other plays $\overline{x}$, with the latter player then playing $\hat{x}(z)$ (for a different value of z) and subsequently $\underline{x}$.[25]

[23] Sufficient conditions for these assumptions are straightforward and encompass the standard specifications found in the literature, such as a linear specification for f. Notice that realized prices may be negative. This common convention (see Mathews and Mirman [32] for an early example) simplifies the analysis.

[24] As in Section 2–3, the low continuation probability interactions preclude the existence of a trivial equilibrium in which all agents attach probability 1 to clean histories, clinging to this belief no matter what they observe, while being left free by the absence of any data to conjecture that deviations will bring severe punishments.

[25] Possible high continuation probability histories thus are either (i) a sequence of $\begin{smallmatrix}\overline{x}\\ \overline{x}\end{smallmatrix}$ followed by a single period of $\begin{smallmatrix}\hat{x}(z)\\ \hat{x}(z)\end{smallmatrix}$ followed by a sequence of $\begin{smallmatrix}x\\ \underline{x}\end{smallmatrix}$, or (ii) (i) a sequence of $\begin{smallmatrix}\overline{x}\\ \overline{x}\end{smallmatrix}$ followed

Player i will estimate from the record a distribution ρ describing behavior at clean histories that puts a mass of probability ρ_{x^N} on quantity x^N (from observing clean null histories in low continuation probability interactions), puts a mass $\rho_{\overline{x}}$ on quantity $\overline{x}$, and distributes the remaining probability among a collection of quantities $\hat{x}(z)$. Similarly, player i will estimate from the record a distribution θ describing behavior at dirty histories that puts masses of probability $\theta_{\overline{x}}$ and $\theta_{\underline{x}}$ on quantities $\overline{x}$ and $\underline{x}$, with the remaining probability distributed among a collection of quantities $\hat{x}(z)$. Player i's view is that the history is initially clean, with the history switching to dirty upon the first play by either player of any quantity larger than $\overline{x}$.

Given this model of the interaction, player i's best response, conditional on having always played $\overline{x}$ and conditional on taking some quantity *less* than or equal to $\overline{x}$, is to play $\overline{x}$. To see this, we note that player i's current payoff is increasing in x_i when $x_i \leq \overline{x}$ and $x_j < x^N$. Player i's view is that the nature of the history (clean or dirty) evolves independently of her quantity as long as that quantity does not exceed $\overline{x}$, and that any larger quantity renders the history dirty. Player i will thus continue playing $\overline{x}$ as long at the probability z of a clean history remains sufficiently high. Once this probability z dips below $\overline{z}$, player i will choose a best response $\hat{x}(z) > \overline{x}$ to her estimate of the current quantity produced by her opponent, maximizing her current payoff at the cost of inducing a dirty history. Player i then regards the history as dirty, and plays $\underline{x}$, a best response to the behavior described by the distribution θ.[26]

We can now pursue analogous reasoning to obtain counterparts of Lemma 1, Proposition 1 and Proposition 2. As long as $\overline{x}$ is sufficiently likely after clean histories and $\underline{x}$ sufficiently likely after dirty histories, a counterpart of Lemma 1 ensures that the proposed behavior is optimal. We can then establish conditions analogous to those of Propositions 1–2, involving some combination of precise monitoring (in the form of shrinking variance of the distributions governing ε_1 and ε_2), patience and unlikely low continuation probabilities, to ensure that some cooperation appears, and that this cooperation has significant payoff effects.

Players in this interaction have all the more reason to group histories into analogy classes, since the infinite set of possible actions multiplies the number of histories. Our example groups actions into two analogy classes, intuitively viewed as "good" or "bad." This is a simple view of the world, but not an unrealistic view. We can well imagine participants characterizing their relationship in terms of "things are OK" or "something is wrong."

---

by a single period of $\frac{\hat{x}(z)}{\overline{x}}$ followed by a (possibly empty) sequence of $\frac{\overline{x}}{\overline{x}}$ followed by a single period of $\frac{\overline{x}}{\hat{x}(x)}$ followed by a sequence of $\frac{\underline{x}}{\underline{x}}$.

[26] Hence, for player $i = \{1, 2\}$, we have

$$\begin{aligned} \hat{x}(z) &= \arg\max_{x_i} \mathbb{E}_{z\rho+(1-z)\theta} x_i f(x_i, x_j) \\ \underline{x} &= \arg\max_{x_i} \mathbb{E}_{\theta} x_i f(x_i, x_j). \end{aligned}$$

The quantities $\hat{x}(z)$ and $\underline{x}$ thus maximize the expected stage game payoff, given that the opponent's quantity is governed by $z\rho + (1-z)\theta$ and θ, respectively.

It is a familiar result that collusive agreements in Cournot duopoly are plagued by incentives to cheat by expanding output. As a result, players at clean histories in our setting produce either the largest quantity consistent with keeping the history clean ($\overline{x}$), or a quantity that renders the history dirty. Players willing to tip the history to dirty maximize their current playoff by producing $\hat{x}(z)$ (where z is their current posterior of a clean history), while players who know they face dirty histories maximize their current payoff by producing $\underline{x}$.

There will be multiple equilibria, for two reasons. First, as we have noted, there is always an equilibrium in which every player chooses x^N after every history. In addition, we have not yet tied down the value $\overline{x}$. There will typically be multiple equilibria of the type we have just described, characterized by different values of $\overline{x}$. If the players could choose between such equilibria, they would choose $\overline{x}$ to be as close to the joint profit maximizing quantity x^* as possible, effectively restraining their urge to increase output. This would also be the choice of an information designer intent on maximizing payoffs. However, equilibria also exist in where the analogy classes happen to incorporate larger values of $\overline{x}$, supporting some but not the most extreme possible collusion.

We regard this approach as being more broadly applicable. The key is that, just as the economist writes a model of a more complicated strategic interaction, so can we expect the participants in the interaction to employ models in their reasoning. We especially view modeling their opponents' strategies as challenging for the participants, prompting them to turn to information about past play for help, and in turn forcing them to organize histories into analogy classes. The formation of these analogy classes is likely to depend heavily on context, making general statements elusive, but we expect the tendency to think in terms of relatively few analogy classes, perhaps as some version of "good" and "bad," to be helpful in supporting cooperation.

4.5 Questions

Many questions remain. For example, the equilibria we have examined exhibit permanent punishment. Does analogical reasoning in private-monitoring games allow equilibria in which players recoordinate on cooperation after a punishment has been triggered, and could such a construction allow higher payoffs to be supported? We suspect not, but the question remains open.

Could one establish a folk theorem, fixing a monitoring structure and then showing that for any feasible, individually rational payoff, there is a specification of analogy classes supporting that outcome? Again, we suspect not. By expanding the number of analogy classes, one moves the game closer to a conventional game of private monitoring, raising the possibility that we could recover the constructions of Sekiguchi [38] or Ely and Välimäki [9]. However, even upon allowing our players plentiful analogy classes, they remain hampered by the absence of signals from the record. This precludes them from duplicating the understanding of opponent behavior required for conventional folk theorems.

5 Appendix: Proofs

5.1 Proof of Proposition 1

[Step 1] We first convert our restless bandit into an equivalent restless bandit with one arm whose payoff is constant. 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 period 1, the player chooses either C, for payoff $-(k/\overline{\delta}) + \mathbb{E}h(z_1|z_0)$, or chooses D, for a payoff of 0. In period 2, assuming C was chosen in period one, the player chooses either C, for payoff $-(k/\overline{\delta}) + \mathbb{E}h(z_2|z_1)$, 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 $-k/\overline{\delta} + \mathbb{E}h(z_t|z_{t-1})$. The idea is that no matter what, the player receives the period-0 bonus $h(z_0)$. Then, in the ordinary representation, the player can pay the cost k in period 0 in order to also receive a 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 $k/\overline{\delta}$ for the reward $\mathbb{E}h(z_1|z_0)$. Continuing in this way, we obtain an equivalent bandit whose D arm always gives a payoff of 0. The optimal policy is again a threshold policy that cooperates above some belief $\overline{z}$ and defects below that belief.

[Step 2] We next establish a sufficient condition under which a player will optimally pull the C arm of the modified bandit in the first period. The condition for this to be the case is that the period-1 reward from the C arm exceed that of the D arm, or

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

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

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

Remark 4 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:

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

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

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

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 3] We next identify a lower bound $\underline{z}$ on the value of $\overline{z}$, the boundary belief between cooperating and not cooperating, that applies to any equilibrium. An upper bound on the continuation payoff from cooperating and an exact calculation of the payoff from defecting are given by:

$$\begin{aligned} C: &\quad [zp + (1-z)q] + [(1 - (zp + (1-z)q))](-k) \\ D: &\quad (1-\overline{\delta})[zp + (1-z)q](1+k) + \overline{\delta}q(1+k). \end{aligned}$$

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

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

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 $\underline{z}$ on $\overline{z}$, is given by (setting $\overline{\delta} = p = 1$ and $q = 0$)

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

There are then no circumstances under which a player will cooperate when her belief that the history is clean drops below $\underline{z}$.

[STEP 4] We now constrain players to cooperate after the null history in high continuation probability interactions, while placing no other constraints on their behavior. We then construct a function 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 $(\hat{p}, \hat{q}, \hat{\overline{z}}) \in [\frac{\alpha}{2}, 1] \times [0, \frac{\alpha}{2}] \times [\underline{z}, 1]$. The function is defined as follows. First, given (p, q), a player solves for the optimal value $\hat{\overline{z}}$ in the modified bandit, Then, given this value and the induced behavior (remembering the constraint that the player cooperate after the null history in high continuation probability interactions), and working with the updating rules defined by (p, q), we construct the distribution over histories, and from this infer new values $(\hat{p}, \hat{q})$. This gives us the value $(\hat{p}, \hat{q}, \hat{\overline{z}})$. Notice that this function indeed maps into $[\frac{\alpha}{2}, 1] \times [0, \frac{\alpha}{2}] \times [\underline{z}, 1]$. 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. The probability p is bounded below by $\frac{\alpha}{2}$. To 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 $\alpha/2$. 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. The previous step has established the bound $\underline{z}$.

The argument now involves identifying conditions under which the function Φ has a fixed point, and under which any such fixed point has the property that

(19) holds, ensuring that the restriction that players cooperate after the null history in high continuation probability interactions is redundant, giving us an equilibrium. It is straightforward to confirm that Φ is continuous, ensuring the existence of a fixed point.

First, fix k, $\overline{\delta}$ and α satisfying Statement [1.1]. Using the fact that $p \geq \alpha/2$, a sufficient condition for Statement (1).1] is

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

Now fix α and $\overline{\delta}$. As ε approaches zero, so does q. In particular, 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 $\overline{\delta}$) 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. The sufficient condition then becomes

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

which rearranges to give the condition in Statement [1.1].

Second fix k, α and ε. As $\overline{\delta}$ approaches one, either p approaches 1 or q approaches 0. In particular, suppose that p is bounded below 1 as $\overline{\delta} \to 1$. 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 $\overline{\delta} \to 1$, each defection gives rise to an arbitrarily large number of dirty histories, ensuring that q converges to zero. The event that q converges to zero gives the more demanding condition, which is (substituting $\overline{\delta} = 1$, $p = \alpha/2$ and $q = 0$)

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

which is equivalent to $k \leq \frac{\alpha^2}{4-\alpha^2}$, giving the condition in Statement [1.2].

Third, Statement [1.3] is implied by Proposition 2.2, which is proven below. ■

5.2 Proof of Proposition 2

If the history is currently clean with probability z, a player who cooperates and receives a c signal forms the posterior $\phi(z, c)$ that the history is clean given by (4). 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

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

or, equivalently

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

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

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

First, let $\varepsilon \to 0$. Because (i) $\overline{\delta}$ 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 (8). 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 (9) 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 history 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 (9). The result is an equilibrium in which cooperation persists throughout virtually all interactions, as desired. ■

5.3 Details for Section 4.2.1

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[27]

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

We can solve for the value of

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

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

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

[27] 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 $q^2\delta$, the game lasts another period and j again cooperates. With probability $q^3\delta^2$, the game lasts yet another period, and j again cooperates, and so on.

Proof of Proposition 4 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):

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

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

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

Now, for a fixed δ, 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 δ approach one completes the argument, as before. In the limit as δ approaches 1, the sufficient condition is then (substituting $\delta = 1$, $p = \alpha/2$ 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] Masaki Aoyagi and Guillaume R. Frechette. Collusion as public monitoring becomes noisy: Experimental evidence. *Journal of Economic Theory*, 144:1135–1165, 2009.

[3] V. Bhaskar, George J. Mailath, and Stephen Morris. Purification in the infinitely-repeated prisoners' dilemma. *Review of Economic Dynamics*, 11(3):515–528, 2008.

[4] V. Bhaskar and Ichiro Obara. Belief-based equilibria in the repeated prisoners' dilemma with private monitoring. *Journal of Economic Theory*, 102(1):40–69, 2002.

[5] Willaim Blomquist, Edella Schlager, Shui Yan Tang, and Elinor Ostrom. Regularities from the field and possible explanations. In Elinor Ostrom, Roy Gardner, and James Walker, editors, *Rules, Games and Common-Pool Resources*, pages 301–316. University of Michigan Press, Ann Arbor, Michigan, 1994.

[6] Pedro Dal Bó. Cooperation under the shadow of the future: Experimental evidence from infinitely repeated games. *Ameican Economic Review*, 95(5):1591–1604, 2005.

[7] Olivier Compte and Andrew Postlewaite. Plausible cooperation. *Games and Economic Behavior*, 91:45–59, 2015.

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

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

[10] Ignacio Esponda. Behavioral equilibrium in economies with adverse selection. *American Economic Review*, 98(4):1269–1291, 2008.

[11] Ignacio Esponda and Demian Pouzo. Berk-Nash equilibrium: A framework for modeling agents with mispecified models. *Econometrica*, 84(3):1093–1130, 2016.

[12] Erik Eyster and Matthew Rabin. Cursed equilibrium. *Econometrica*, 73(5):1623–1672, 2005.

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

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

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

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

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

[18] Karsten T. Hansen, Kanishka Misra, and Mallesh M. Pai. Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science*, 40(1):1–12, 2021.

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

[20] Joseph E. Harrington, Jr. and Andrzej Skrzypacz. Private monitoring and communication in cartels: Explaining recent collusive practices. *American Economic Review*, 101(6):2425—-2449, 2011.

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

[22] Philippe Jehiel. Analogy-based expectation equilibrium and related concepts: Theory, applications, and beyond. Prepared for the twelth World Congress of the Econometric Society, 2020.

[23] Philippe Jehiel and Larry Samuelson. Reputation with analogical reasoning. *Quarterly Journal of Economics*, 127(4):1927–1970, 2012.

[24] Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

[25] Michihiro Kandori. The use of information in repeated games with imperfect monitoring. *Review of Economic Studies*, 59(3):591–593, 1992.

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

[27] George J. Mailath and Stephen Morris. Repeated games with almost-public monitoring. *Journal of Economic Theory*, 102(1):189–228, 2002.

[28] George J. Mailath and Stephen Morris. Coodination failure in a repeated game with almost public monitoring. *Theoretical Economics*, 1:311–340, 2006.

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

[30] Hitoshi Matsushima. On the theory of repeated games with private information: Part I: Anti-folk theorem without communication. *Economics Letters*, 35(3):253–256, 1991.

[31] Hitoshi Matsushima. Repeated games with private monitoring: Two players. *Econometrica*, 72(3):823–852, May 2004.

[32] Steven A. Matthews and Leonard J. Mirman. Equilibrium limit pricing: The effects of private information and stochastic demand. *Econometrica*, 51(4):981–996, 1983.

[33] Rosemarie Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85:1313–1326, 1995.

[34] Elinor Ostrom. *Governing the Commons*. Cambridge University Press, Cambridge, 1990.

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

[36] Robert H. Porter. A study of cartel stability: The Joint Executive Committee, 1880–1886. *Bell Journal of Economics*, 14(2):301–314, 1983.

[37] Leonard J. Savage. *The Foundations of Statistics*. Dover Publications, New York, 1972. Originally 1954.

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

[39] Ran Spiegler. Bayesian networks and boundedly rational expectations. *Quarterly Journal of Economics*, 131(3):1243–1290, 2016.

[40] Dale O. Stahl and Paul Wilson. On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10:218–254, 1995.

[41] Takuo Sugaya. Folk theorem in repeated games with private monitoringc. *Review of Economic Studies*, 2022. forthcoming.

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

[43] 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.

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

[45] Yuichi Yamamoto. Characterizing belief-free review-strategy equilibrium payoffs under conditional independence. *Journal of Economic Theory*, 2012:1998–2027, 1965.

[46] Peyton Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.