The Stability of Folk Theorem Equilibria with Imperfect Costly Monitoring

Soham Pattnaik

July 12, 2026

Abstract

We study the stability of Folk Theorem outcomes in indefinitely repeated Prisoner's Dilemma games in which monitoring is imperfect and, critically, *costly and endogenous*: players choose not only a stage-game action but also an information-acquisition intensity that controls the precision of the private signal they subsequently receive about their opponent's behavior. We show that a version of the Folk Theorem survives this extension: for any feasible, individually rational payoff vector and any admissible cost function, there exists a discount factor threshold above which the payoff vector is sustainable as a subgame-perfect equilibrium, with monitoring intensity chosen optimally along the equilibrium path. Our principal theoretical contribution is a stability analysis of the equilibrium payoff correspondence $V(\sigma^2, c)$: we prove that this correspondence is upper hemicontinuous in the baseline noise variance σ^2 and the cost function c, establish a comparative-statics result showing that the correspondence is monotonically shrinking (in the Hausdorff sense) in both arguments, and derive an explicit threshold cost level above which sustaining near-efficient outcomes requires strictly greater patience. We complement the analytical results with an evolutionary agent-based simulation in which a heterogeneous population of monitoring strategies – blind free-riders, fixed-overhead monitors, and trust-adjusted adaptive monitors – competes under mutation and selection, alongside a persistent adversarial subpopulation designed to probe for exploitable blind spots. The simulation corroborates the theoretical prediction that endogenous, trust-contingent monitoring dominates both blind cooperation and static monitoring once an adversarial threat is present, while illustrating that the presence of such a threat introduces persistent volatility into aggregate welfare even after the population has converged on the theoretically optimal strategy class.

1 Introduction

Repeated games are the canonical framework for modeling long-run strategic relationships: the same set of players meets period after period, each period's payoffs depend on that period's action profile, and players carry information about the past into future decisions. When a player's lifetime payoff is a discounted sum of per-period payoffs, the discount factor $\delta \in (0, 1)$ captures the weight placed on future consequences relative to immediate gain. This

repeated structure is qualitatively different from a one-shot interaction. In a single stage of the Prisoner's Dilemma, mutual defection is the unique Nash equilibrium, even though mutual cooperation is Pareto superior. The repetition of the game reopens the possibility of cooperation, because today's action can be conditioned on tomorrow's consequences: a player who defects today can be identified and punished tomorrow, and the credible threat of punishment can render defection unprofitable in present-value terms.

This logic is formalized by the *Folk Theorem*, one of the foundational results of repeated-game theory. Under perfect monitoring, where each player observes, without error, the entire history of actions taken by every other player, the Folk Theorem states that essentially any feasible and individually rational payoff vector can be supported as a subgame-perfect equilibrium (SPE) once players are sufficiently patient. The result is remarkable precisely because it holds for such a broad class of payoffs: efficiency, fairness, and a wide range of intermediate outcomes are all simultaneously attainable, with the specific outcome pinned down not by the primitives of the stage game but by the self-fulfilling expectations of the players.

The perfect-monitoring assumption, however, is rarely satisfied outside of stylized settings. Firms cannot observe their rivals' internal decisions directly; they observe market prices, sales volumes, or leaked signals that are correlated with, but not necessarily identical to, the rival's true action. Business partners cannot observe each other's effort directly; they observe noisy proxies such as output, revenue, or third-party reports. Tax authorities cannot observe true income; they observe filings and can conduct costly audits to sharpen their inference. In each of these settings, two features of monitoring interact: first, the signals available to a player are *noisy*, so that punishing on the basis of a bad signal risks punishing an innocent cooperator (a false alarm) or failing to punish a genuine defector (a missed detection); second, the *precision* of the signal is not exogenously fixed but can be improved by the player at a cost – through auditing, monitoring technology, data collection, or surveillance effort.

The literature on repeated games with imperfect monitoring, beginning with the foundational contributions of Fudenberg, Levine, and Maskin [FLM94] and Abreu, Pearce, and Stacchetti [APS90], established that the Folk Theorem survives the introduction of noisy *public* signals, provided those signals satisfy suitable statistical richness conditions (informally, that the signal distribution varies sufficiently across action profiles that deviations can eventually be statistically detected). A parallel literature has asked what happens when the *precision* of monitoring itself is a choice variable subject to cost. Miyagawa, Miyahara, and Sekiguchi [MMS08] showed that the Folk Theorem continues to hold when players must pay a cost to observe each other's actions, so long as perfect monitoring remains *feasible* at some (possibly large) cost. Related work on games played at high frequency has emphasized that what matters for sustaining cooperation is not the level of monitoring cost in isolation, but its size *relative to* the stage-game payoffs and the frequency of interaction.

This paper sits at the intersection of these two branches. We study a repeated Prisoner's Dilemma in which, in every period, each player simultaneously chooses (i) a stage-game

action, cooperate or defect, and (ii) a monitoring intensity that determines the variance of the noisy private signal that player subsequently receives about the opponent's action. Higher monitoring intensity strictly reduces signal variance but is strictly costly, and the cost is borne every period regardless of the realized signal. Our contribution is threefold.

First, we verify that a Folk Theorem-type result survives this extension of the environment: for any feasible, individually rational target payoff and any admissible (increasing, convex) monitoring cost function, sufficiently patient players can sustain an SPE whose average payoff is arbitrarily close to the target, with monitoring chosen optimally by each player along the equilibrium path (Theorem 3.1).

Second, and this is the paper's central concern, we ask how *stable* this conclusion is. Concretely, we study the correspondence $V(\sigma^2, c)$ mapping a baseline noise level σ^2 and a cost function c to the set of payoffs sustainable in SPE, and we characterize how this set responds to perturbations of its arguments. We show that $V(\sigma^2, c)$ is upper hemicontinuous (Theorem 4.1) and that it shrinks, in the Hausdorff sense, as either baseline noise or monitoring cost increases (Corollary 4.2). We further identify an explicit critical cost threshold $\bar{c}(\delta)$ below which near-efficient cooperation remains attainable and above which sustaining any given level of cooperation requires strictly greater patience (Corollary 4.3).

Third, we complement the analytical treatment with an evolutionary agent-based simulation. Because the theoretical results are asymptotic in δ and rely on equilibrium refinements that are difficult to visualize directly, the simulation lets us examine *disequilibrium* dynamics: what happens when a heterogeneous population of boundedly sophisticated monitoring strategies competes, mutates, and is selected on the basis of realized fitness, in the presence of an adversarial subpopulation that actively searches for exploitable weaknesses in the monitoring technology. The simulation shows that trust-adjusted adaptive monitoring, which invests heavily in information when trust is low and economizes on monitoring cost once trust is established, comes to dominate the population, driving both blind cooperation and static (fixed-cost) monitoring to extinction. At the same time, the presence of a persistent adversarial threat introduces volatility into aggregate welfare that does not disappear even after the population converges on the theoretically preferred strategy class, illustrating a gap between the existence of a stable equilibrium *class* and the stability of realized *payoffs* along any given evolutionary path.

The remainder of the paper is organized as follows. Section 2 lays out the model: the stage game, the endogenous monitoring technology, the strategy space, and the equilibrium concept. Section 3 states and proves the baseline Folk Theorem result with costly endogenous monitoring. Section 4 contains the paper's main theoretical contribution, the stability analysis of the equilibrium payoff correspondence. Section 5 describes the computational simulation and reports its results. Section 6 discusses robustness, extensions, and applications, and Section 7 concludes.

2 The Model

2.1 Repeated Games: Preliminaries

We first recall the standard repeated-game framework before introducing the monitoring extension.

Definition 2.1 (Repeated Game)**.** A repeated game is played by a finite set of players $N = \{1, \ldots, n\}$ over discrete periods $t = 1, 2, \ldots$. In each period, players simultaneously choose actions from a finite (or, in our extension, compact) action set, receive a stage payoff determined by the joint action profile, and receive some (potentially imperfect) information about the actions chosen by others. Play then proceeds to the next period, with all players able to condition future choices on their private history.

Lifetime payoffs are evaluated using discounted utility. For player i,

$$U_i = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} u_i(a^t), \tag{1}$$

where $\delta \in (0, 1)$ is the common discount factor and $u_i(a^t)$ is player i's realized stage-game payoff in period t. The normalizing factor $(1 - \delta)$ rescales the discounted sum onto the same units as a single stage payoff, which is standard and convenient for stating Folk Theorem results, since it allows average equilibrium payoffs to be compared directly against the stage-game payoff matrix.

Although it was not formalized and published until the 1970s, the Folk Theorem takes its name from the fact that the basic insight was circulated informally as "folk wisdom" among game theorists for decades beforehand. It seems intuitive that when cooperation has a higher payoff than playing individually, it is possible for players to maintain cooperation.

Theorem 2.2 (The Folk Theorem, perfect monitoring)**.** *Let G be a finite, simultaneous-move game of complete information, let $(u_1^*, \ldots, u_n^*)$ denote the minimax (individually rational) payoffs of G, and let $(u_1, \ldots, u_n)$ be any feasible payoff vector of G satisfying $u_i > u_i^*$ for every $i \in N$. Then there exists $\bar{\delta} < 1$ such that for every $\delta \in (\bar{\delta}, 1)$, the infinitely repeated game $G(\delta)$ admits a subgame-perfect equilibrium whose average payoff is arbitrarily close to $(u_1, \ldots, u_n)$.*

Theorem 2.2 presumes that deviations are perfectly and publicly observed, so that punishment can be triggered with certainty exactly when, and only when, a deviation occurs. Removing this assumption is the point of departure for the present paper.

2.2 The Stage Game

We specialize to the two-player Prisoner's Dilemma. In each period, players simultaneously choose to Cooperate (C) or Defect (D). Stage payoffs follow the standard Temptation–Reward–Punishment–Sucker parameterization, summarized in Table 1, with $T = 5$, $R = 3$,

$P = 1$, $S = 0$ (so that $T > R > P > S$ and $2R > T + S$, the two conditions required for the Prisoner's Dilemma to have the usual efficiency properties).

	C	D
C	$(3, 3)$	$(0, 5)$
D	$(5, 0)$	$(1, 1)$

Table 1: Prisoner's Dilemma stage-game payoffs.

The unique stage-game Nash equilibrium is mutual defection, yielding the minimax payoff $u_i^* = 1$ for each player, while the symmetric cooperative payoff $(3, 3)$ is both feasible and strictly individually rational, making the pair a natural target for Theorem 2.2 under perfect monitoring.

2.3 Endogenous Monitoring Technology

The key departure from the classical model is that actions are not directly observed. Instead, each player i receives, at the end of every period, a noisy continuous private signal $s_i^t \in [-0.5, 1.5]$ about the opponent j's action $a_j^t \in \{0, 1\}$ (with 0 coded as Cooperation and 1 as Defection). The signal is generated by

$$s_i^t = \text{clip}\big(a_j^t + \varepsilon_i^t,\, -0.5,\, 1.5\big)\,, \qquad \varepsilon_i^t \sim \mathcal{N}\big(0,\, \sigma_i^2(m_i^t)\big)\,, \tag{2}$$

where $m_i^t \in [0, 1)$ is player i's chosen monitoring intensity in period t, and the variance of the observation noise is governed by

$$\sigma_i^2(m_i^t) = \frac{\sigma_0^2}{1 + \alpha \cdot (m_i^t \cdot \beta)}, \tag{3}$$

with baseline ambient noise $\sigma_0^2 = 4.0$, information-precision sensitivity $\alpha = 35.0$, and monitoring scale multiplier $\beta = 2.5$. Equation (3) embeds two natural features of a monitoring technology: at $m_i^t = 0$, signal quality reduces to the ambient baseline σ_0^2, and $\sigma_i^2(\cdot)$ is strictly decreasing and convex in m_i^t, so that monitoring investment exhibits diminishing marginal returns to precision: an early unit of monitoring effort buys much more clarity than a later one.

Monitoring is costly independent of its outcome: the cost is paid in every period regardless of the signal realized. We work with a general cost function $c_i : M_i \to \mathbb{R}_{\geq 0}$, assumed strictly increasing and strictly convex, of which the two functional forms used in our simulation, a constant (fixed) cost $\bar{c}_i$ and a trust-contingent cost obtained by linearly interpolating between $\bar{c}_i$ and a negligible floor, are special cases.

2.4 Strategies

Definition 2.3 (Private Histories and Strategies)**.** Let $H_i^t = \prod_{\tau=1}^{t-1}(A_i \times M_i \times S_i)$ denote the set of player i's private histories through period t, with $H_i^1 = \{\varnothing\}$. A behavioral strategy for player i is a sequence of maps $\sigma_i = (\sigma_i^t)_{t=1}^{\infty}$, where

$$\sigma_i^t : H_i^t \to \Delta(A_i \times M_i) \tag{4}$$

assigns to each private history a (possibly correlated) probability distribution over the current-period action $a_i^t \in A_i$ and monitoring intensity $m_i^t \in M_i$.

The instantaneous net payoff to player i in period t, before discounting, is

$$\pi_i^t = u_i(a_i^t, a_j^t) - c_i(m_i^t), \tag{5}$$

where u_i is the Prisoner's Dilemma payoff from Table 1. This is the object whose discounted sum defines lifetime utility U_i as in equation (1), with $u_i(a^t)$ there understood to already net out the period's monitoring cost.

To fix intuition and to connect the theory to the simulation of Section 5, it is useful to describe three natural strategic archetypes, each of which specifies both a monitoring policy and an action rule.

- **Blind free-rider.** Sets $m_i^t \equiv 0$ in every period, avoiding all information-acquisition cost but facing maximal signal noise $\sigma_i^2 = \sigma_0^2$ in every period.
- **Fixed-overhead monitor.** Sets $m_i^t \equiv \bar{c}_i \in [0.3, 0.6]$ regardless of history, purchasing a constant, history-independent improvement in signal precision.
- **Trust-adjusted adaptive monitor.** Maintains a scalar trust state $\mathcal{T}_i^t \in [0, 1]$, initialized at $\mathcal{T}_i^1 = 0.5$, that summarizes the player's posterior confidence that the opponent is cooperating. Trust evolves according to

$$\mathcal{T}_i^t = \mathrm{clip}\Big(\mathcal{T}_i^{t-1} + \omega_i \cdot \Big[\big(1 - \mathrm{clip}(s_i^{t-1}, 0, 1)\big) - 0.5\Big],\ 0,\ 1\Big), \tag{6}$$

  for an individual wariness parameter $\omega_i > 0$, and monitoring intensity is chosen by linear interpolation between a high-cost, high-precision regime and a negligible-cost regime as trust rises:

$$m_i^t = \mathrm{interp}\big(\mathcal{T}_i^t;\ [0, 1],\ [\bar{c}_i,\ 0.01]\big). \tag{7}$$

  Actions follow a trust-modulated Tit-for-Tat rule with stochastic forgiveness rate ϕ_i:

$$a_i^t = \begin{cases} 0 & \text{if } s_i^{t-1} < 1 - \mathcal{T}_i^t, \text{ or with probability } \phi_i \\ 1 & \text{otherwise.} \end{cases} \tag{8}$$

The adaptive archetype is designed so that monitoring intensity is itself an equilibrium-relevant state variable: a player who has recently received signals consistent with cooperation lets trust rise, at which point equation (7) economizes on the (otherwise wasted) cost of further monitoring, while a player who receives signals suggestive of defection lets trust fall, which simultaneously tightens the cutoff in the action rule (8) and raises monitoring intensity, restoring precision exactly when it is most valuable for detecting continued defection.

2.5 Equilibrium Concept

We look for subgame-perfect equilibria of the repeated game with private monitoring.

Definition 2.4 (Feasibility and Individual Rationality)**.** A payoff vector $v = (v_1, \dots, v_n)$ is *feasible* if it lies in the convex hull of the stage-game payoff vectors $\{u(a) : a \in A\}$. It is *individually rational* if $v_i \geq u_i^*$ for every i, where u_i^* is player i's minimax payoff, i.e., the greatest payoff i can guarantee against an adversarial best response from the remaining players.

We write V for the set of feasible, individually rational payoff vectors, and $V(\sigma^2, c)$ for the (possibly strict) subset of V that is sustainable as an SPE payoff of the repeated game with baseline noise variance σ^2 and monitoring cost function c.

3 The Folk Theorem with Costly Endogenous Monitoring

We now state and prove the extension of Theorem 2.2 to the environment of Section 2.

Theorem 3.1 (Folk Theorem with Costly Endogenous Monitoring)**.** *Consider any finite two-player stage game satisfying the standard full-dimensionality condition (the set of feasible payoffs has nonempty interior relative to the individually rational payoff set). Suppose the private signal generating process satisfies the* individual full rank *condition: for each player i, the family of signal distributions $\{F_i(\cdot \mid a_i, m_i, a_j)\}_{a_j \in A_j}$ is linearly independent across a_j whenever m_i is bounded away from zero. Then for any admissible, strictly increasing and strictly convex cost function $c_i(\cdot) \geq 0$ and any feasible, individually rational payoff vector $v \in V$, there exists $\bar{\delta}(v, c) < 1$ such that for every $\delta \in (\bar{\delta}, 1)$ there is a subgame-perfect equilibrium of the repeated game whose average discounted payoff is arbitrarily close to v, and in which every player's monitoring intensity along the equilibrium path is chosen optimally given the continuation value function.*

Proof. The proof proceeds in four steps, following the self-generation methodology of Abreu, Pearce, and Stacchetti [APS90] as extended to endogenous monitoring choices.

Step 1 (Statistical detectability). Fix a target payoff $v \in V$ in the interior of the individually rational payoff set (the boundary case follows by a standard limiting argument). By the

individual full-rank condition, for any monitoring intensity m_i bounded away from 0, the signal distributions induced by j's two possible actions, $F_i(\cdot \mid m_i, a_j = 0)$ and $F_i(\cdot \mid m_i, a_j = 1)$, are statistically distinguishable: there exists a bounded statistic (a likelihood-ratio test on the realized signal s_i^t) whose expectation strictly differs across $a_j \in \{0, 1\}$. Because the noise in (2)–(3) is Gaussian with strictly positive, finite variance for every $m_i < 1$, this condition holds for every monitoring intensity in the admissible range so long as m_i is not driven to its degenerate upper bound; a positive but arbitrarily small level of monitoring is enough to make j's two actions statistically distinguishable in the limit of many observations, and a stationary monitoring intensity chosen once and for all makes them distinguishable within any finite window with probability that can be pushed arbitrarily close to one by lengthening the review phase.

Step 2 (Review-phase punishment construction). Construct a candidate equilibrium in *review-strategy* form, dividing time into review phases of L periods. Within a review phase, both players are prescribed the action profile that supports v in the corresponding stage game (e.g., mutual cooperation if v is the cooperative payoff) together with a constant monitoring intensity $m_i^\dagger$ chosen so as to satisfy $m_i^\dagger \in \arg\max_{m_i} \left[\Pr(\text{correct detection} \mid m_i) - c_i(m_i)/\lambda \right]$ for a shadow price λ to be determined; strict convexity of c_i guarantees this maximizer exists and is unique. At the end of a review phase, each player computes a sufficient statistic of the L realized signals about the opponent (for instance, the sample average of s_i^t over the phase). If the statistic falls below a threshold consistent with the opponent having defected with objectively low probability, play continues into the next review phase at the cooperative profile; otherwise, players switch to a finite punishment phase in which the deviator's continuation payoff is driven down toward the minimax payoff u_i^* via mutual defection with minimal monitoring (since punishment does not require fine detection, $c_i(m_i) = 0$ is optimal during punishment).

Step 3 (Incentive compatibility of the action choice). By Step 1, as $L \to \infty$ the review statistic converges in probability to its true mean under either action, so the probability of a false trigger (punishing a truthful cooperator) and the probability of a missed detection (failing to punish a genuine defector) can both be made arbitrarily small simultaneously, for any fixed, strictly positive $m_i^\dagger$. Standard arguments (as in Fudenberg, Levine, and Maskin [FLM94]) then show that for δ sufficiently close to 1, the one-period temptation payoff $T - R$ from a unilateral deviation to defection is outweighed by the discounted loss in continuation value from triggering punishment with probability approaching one, so that following the prescribed action is optimal at every private history consistent with the equilibrium path.

Step 4 (Incentive compatibility of the monitoring choice). It remains to verify that no player wishes to deviate from $m_i^\dagger$ to some $m_i' \neq m_i^\dagger$, either upward (over-monitoring) or downward (under-monitoring, i.e., "skimping"). Because $c_i(\cdot)$ is strictly convex and the detection probability is concave and bounded above by 1 as a function of m_i (a consequence of (3), under which additional monitoring intensity yields strictly diminishing reductions in signal variance), the objective in Step 2 is strictly concave in m_i, so $m_i^\dagger$ is the unique unconstrained optimum of a player who is indifferent about the informational externality on the opponent's

continuation value. A deviation to lower monitoring ($m_i' < m_i^\dagger$) saves the deviator a cost of at most $c_i(m_i^\dagger) - c_i(m_i')$ per period but, because the deviator's own detection of the opponent's behavior is what triggers the deviator's own switch into a favorable continuation path when the opponent has in fact cooperated, under-monitoring strictly raises the probability that the deviator mistakenly triggers punishment against an innocent opponent and forfeits the surplus of continued cooperation; for δ sufficiently close to 1 the forfeited continuation surplus dominates the one-period cost saving, by the same discounting argument as in Step 3. A symmetric argument, using strict convexity of c_i, rules out profitable deviations to $m_i' > m_i^\dagger$: the marginal cost of additional precision exceeds its marginal contribution to continuation value once $m_i^\dagger$ already satisfies the first-order condition of Step 2.

Step 5 (Self-generation and convergence to v). Steps 2–4 exhibit, for each δ sufficiently large, a self-generating set of payoff vectors in the sense of Abreu, Pearce, and Stacchetti [APS90]: a compact set $W(\delta)$ of payoffs such that every $w \in W(\delta)$ can be decomposed into a current action/monitoring profile and a continuation payoff also lying in $W(\delta)$, consistent with the incentive constraints verified above. Since $W(\delta)$ is self-generating, $W(\delta) \subseteq E(\delta)$, the set of SPE payoffs of the repeated game at discount factor δ. As the review-phase length L is increased (which is feasible for any fixed δ sufficiently close to 1, since the discounted cost of lengthening the review phase before the first punishment decision vanishes as $\delta \to 1$), the detection errors in Step 3 shrink to zero, and the achievable payoff within $W(\delta)$ converges to the target stage-game payoff v net of the equilibrium monitoring cost $c_i(m_i^\dagger)$, which itself can be made arbitrarily small by choosing $m_i^\dagger$ close to (but bounded away from) zero while still satisfying Step 1's detectability requirement, since detectability only requires strictly positive monitoring, not monitoring bounded away from zero uniformly in L. Combining these limits, for every $\eta > 0$ there exists $\bar{\delta}$ such that for $\delta > \bar{\delta}$ the constructed equilibrium delivers an average payoff within η of v, which establishes the claim. □

Remark 3.2. Theorem 3.1 generalizes Miyagawa, Miyahara, and Sekiguchi [MMS08] in two respects: the cost of monitoring is paid continuously, in every period, rather than only when perfect observation is purchased outright, and the informativeness of the signal is itself a continuous, endogenously chosen function of cost via (3), rather than a binary choice between observing and not observing. The proof strategy, however, remains within the classical self-generation paradigm of [APS90] together with the statistical detectability arguments of [FLM94]; the innovation lies in Step 4, where convexity of the cost function is used to pin down a unique, incentive-compatible monitoring intensity rather than simply assuming monitoring is exogenously fixed once purchased.

4 Stability of the Equilibrium Payoff Set

Theorem 3.1 establishes existence: cooperative payoffs remain sustainable once monitoring is costly and endogenous. It does not, however, address how sensitive this conclusion is to the two structural parameters that govern the monitoring environment – the baseline ambient

noise level σ^2 and the cost function c. This is the paper's central question, and we address it by studying the correspondence $V(\sigma^2, c)$ introduced in Section 2.

4.1 Setup

Let $G = (N, (A_i)_{i\in N}, (u_i)_{i\in N})$ be the finite stage game of Section 2, and consider the infinitely repeated version played at discrete periods $t \in \{1, 2, \dots\}$. At each period, players simultaneously choose an action $a_i^t \in A_i$ and a monitoring intensity $m_i^t \in M_i \equiv [0, 1)$; write $A = \prod_i A_i$ and $M = \prod_i M_i$. The true action profile $a^t \in A$ is never directly observed. Instead, each player i receives a private continuous signal $s_i^t \in S_i \subseteq \mathbb{R}$, and the joint signal vector $s^t = (s_1^t, \dots, s_n^t) \in S = \prod_i S_i$ is distributed according to an absolutely continuous probability measure $F(\cdot \mid a^t, m^t)$ with density $f(\cdot \mid a^t, m^t)$ with respect to Lebesgue measure.

Let $c_i : M_i \to \mathbb{R}_{\geq 0}$ be player i's information-acquisition cost function, assumed strictly increasing and C^2 with $\partial^2 c_i/\partial m_i^2 > 0$. Lifetime utility is

$$U_i = (1-\delta)\,\mathbb{E}_F\left[\sum_{t=1}^{\infty} \delta^{t-1}\big(u_i(a^t) - c_i(m_i^t)\big)\right]. \tag{9}$$

We are interested in how the set of SPE payoffs varies as (σ^2, c) varies, where σ^2 enters through the noise variance function (3) and $c = (c_1, c_2)$ ranges over the space of admissible cost functions equipped with the C^2 topology.

4.2 Upper Hemicontinuity

Theorem 4.1 (Upper Hemicontinuity of the Equilibrium Payoff Correspondence)**.** *The subgame-perfect equilibrium payoff correspondence $V(\sigma^2, c) : \mathbb{R}_{++} \times C^2 \rightrightarrows \mathbb{R}^n$ is upper hemicontinuous in the baseline noise variance σ^2 and the cost function c, under the Hausdorff metric topology on subsets of $\mathbb{R}^n$.*

Proof. Let $\big(\sigma_k^2, c_k\big)_{k=1}^{\infty}$ be a sequence of structural parameters converging to a limit environment (σ^2, c), in the sense that $\sigma_k^2 \to \sigma^2$ in $\mathbb{R}_{++}$ and $c_k \to c$ in the C^2 topology (i.e., c_k, c_k', and c_k'' converge uniformly on compact subsets of M_i). Let $v_k \in V(\sigma_k^2, c_k)$ be an arbitrary associated sequence of sustainable equilibrium payoff vectors. We must show that every convergent subsequence of (v_k) has its limit in $V(\sigma^2, c)$; since the Hausdorff topology is metrizable, this is equivalent to upper hemicontinuity.

Step 1 (Compactness and existence of a convergent subsequence). For every k, $V(\sigma_k^2, c_k) \subseteq V$, the feasible and individually rational payoff set of the underlying stage game, which does not depend on k. Since V is the intersection of the (compact) convex hull of a finite set of stage-game payoff vectors with a finite collection of closed half-spaces (the individual-rationality constraints), V is compact. Hence $(v_k)_{k=1}^{\infty} \subset V$ is a bounded sequence in $\mathbb{R}^n$, and by the Bolzano–Weierstrass theorem it admits a convergent subsequence, which without loss of generality we continue to denote (v_k), with limit $v = \lim_{k\to\infty} v_k \in V$.

Step 2 (Convergence of the signal-generating process). Fix any measurable, bounded continuation-value functional of the form used in the self-generation construction of Theorem 3.1 – that is, any bounded function of finite private histories that a candidate equilibrium strategy profile can condition on. Because $\sigma_k^2 \to \sigma^2$ and the noise term in (2) is Gaussian with variance given by (3), the conditional signal densities $f_k(\cdot \mid a, m)$ converge pointwise, and hence (by Scheffé's lemma, since each $f_k(\cdot \mid a, m)$ is a proper density) in L^1 and therefore in total variation, to the limiting density $f(\cdot \mid a, m)$, uniformly over the compact action and monitoring spaces $A \times M$. Consequently, for any bounded measurable payoff-relevant statistic ψ of the signal,

$$\lim_{k\to\infty} \mathbb{E}_{F_k}\big[\psi(s) \mid a, m\big] = \mathbb{E}_{F}\big[\psi(s) \mid a, m\big], \tag{10}$$

by the Dominated Convergence Theorem, since ψ is bounded and $F_k \to F$ in total variation. In particular, taking $\psi = u_i(a) - c_{i,k}(m_i)$ and using $c_{i,k} \to c_i$ uniformly on the compact set M_i,

$$\lim_{k\to\infty} \mathbb{E}_{F_k}\big[u_i(a) - c_{i,k}(m_i)\big] = \mathbb{E}_{F}\big[u_i(a) - c_i(m_i)\big]. \tag{11}$$

Step 3 (Preservation of incentive constraints in the limit). For each k, the payoff vector v_k is supported by an SPE strategy profile ζ_k. Incentive compatibility requires that at every private history h^t reached with positive probability under ζ_k, no one-period deviation to (a_i', m_i') raises player i's continuation payoff:

$$(1-\delta)\big(u_i(a_i^t, a_{-i}^t) - c_{i,k}(m_i^t)\big) + \delta\, \mathbb{E}_{F_k}\big[V_i(h^{t+1})\big] \;\geq\; (1-\delta)\big(u_i(a_i', a_{-i}^t) - c_{i,k}(m_i')\big) + \delta\, \mathbb{E}_{F_k}\Big[V_i(h^{t+1}_{a_i', m_i'})\Big], \tag{12}$$

where $V_i(\cdot)$ denotes the continuation-value function implied by ζ_k. Every term in (12) is of the form covered by Step 2: it is either a bounded stage payoff net of monitoring cost, or a conditional expectation of a bounded continuation value under $F_k(\cdot \mid a, m)$. By Step 2, each such term converges as $k \to \infty$ to its counterpart evaluated at the limit environment (σ^2, c). Suppose, toward a contradiction, that the limit payoff v together with the corresponding limit of the strategy profiles ζ_k (which, since strategies are bounded functions on a fixed countable history space, admit a further subsequence converging pointwise by a diagonal argument) failed to satisfy the limiting incentive constraint at some history h^t – that is, suppose some deviation (a_i', m_i') yielded a strictly higher payoff in the limit environment. By continuity of each term in (12) established in Step 2, this strict profitability would have to appear already at some finite k along the sequence, for k large enough that the gap between the limiting and k-th environment is smaller than the profitability margin of the deviation. This contradicts the assumption that ζ_k is incentive compatible (and hence that $v_k \in V(\sigma_k^2, c_k)$) for every k. Hence no such profitable deviation can exist in the limit, and the limiting strategy profile is incentive compatible at every history, i.e., it constitutes an SPE of the limit environment (σ^2, c).

Step 4 (Feasibility and individual rationality of the limit, and conclusion). Since $v_k \to v$ and $v_k \in V$ for every k, and V is closed (being compact, as shown in Step 1), the limit v lies

in V: it is feasible and individually rational. Combined with Step 3, v is an SPE payoff of the environment (σ^2, c), i.e., $v \in V(\sigma^2, c)$. Since (v_k) was an arbitrary sequence of equilibrium payoffs along an arbitrary convergent sequence of environments, this establishes that every limit point of every such sequence lies in $V(\sigma^2, c)$, which is precisely the definition of upper hemicontinuity of $V(\cdot, \cdot)$ at (σ^2, c). Since (σ^2, c) was arbitrary, V is upper hemicontinuous everywhere on its domain. □

Corollary 4.2 (Monotone Shrinkage of the Equilibrium Set)**.** *Fix δ. If $\sigma_2^2 \geq \sigma_1^2$ (holding c fixed) or $c_2 \geq c_1$ pointwise (holding σ^2 fixed), then $V(\sigma_2^2, c) \subseteq_{\mathrm{H}} V(\sigma_1^2, c)$ and $V(\sigma^2, c_2) \subseteq_{\mathrm{H}} V(\sigma^2, c_1)$ in the Hausdorff sense: the equilibrium payoff set at the higher noise level or higher cost function is contained, up to a vanishing perturbation as the parameters converge, within the equilibrium payoff set evaluated at the lower noise level or lower cost function.*

Proof. Higher σ^2 strictly increases $\sigma_i^2(m_i)$ pointwise in m_i for every $m_i \in [0, 1)$ by (3), which strictly reduces the informativeness of the signal (in the sense of Blackwell) at every monitoring intensity, and hence strictly weakens the detectability condition used in Step 1 of the proof of Theorem 3.1: the review-phase length L required to achieve any target detection error now must increase. Because the discounted cost of a longer review phase before the first possible punishment is strictly increasing in L for fixed $\delta < 1$, the set of δ for which any given target payoff remains achievable shrinks, so the threshold $\bar{\delta}(v, c)$ in Theorem 3.1 is weakly increasing in σ^2; equivalently, at fixed δ, the set of achievable v weakly shrinks. An analogous argument applies to c: a pointwise increase in c_i raises the equilibrium monitoring cost $c_i(m_i^\dagger)$ incurred along the path constructed in Theorem 3.1 for any fixed target detectability, which strictly lowers the best achievable net payoff and, by the strict convexity argument of Step 4 in that proof, can also force a reduction in the equilibrium $m_i^\dagger$ itself, further weakening detectability. Both channels operate in the same direction, establishing the stated containments; continuity in the Hausdorff metric as $\sigma_2^2 \to \sigma_1^2$ or $c_2 \to c_1$ then follows directly from Theorem 4.1. □

Corollary 4.3 (Threshold on Monitoring Costs)**.** *For every fixed discount factor $\delta \in (0, 1)$, there exists a critical cost level $\bar{c}(\delta)$, strictly increasing in δ, such that: (i) if the cost function c satisfies $c_i(m_i) \leq \bar{c}(\delta)$ for all m_i in the relevant monitoring range and all i, then payoffs within η of the fully efficient payoff (R, R) remain sustainable for some η that can be driven to 0 as $\delta \to 1$; and (ii) if c exceeds $\bar{c}(\delta)$, sustaining any fixed target payoff v strictly interior to V requires a strictly higher discount factor $\delta' > \delta$ than would be required under $\bar{c}(\delta)$.*

Proof. This follows directly from Corollary 4.2 together with the explicit threshold construction in Step 5 of the proof of Theorem 3.1. Define $\bar{\delta}(v, c)$ as the infimum discount factor for which the review-phase construction of Theorem 3.1 sustains a payoff within η of v; monotone shrinkage (Corollary 4.2) shows $\bar{\delta}(v, c)$ is weakly increasing in c. Fixing δ and inverting this monotone relationship in c, holding v and η fixed, defines $\bar{c}(\delta)$ as the largest cost function (in the pointwise partial order) for which $\bar{\delta}(v, c) \leq \delta$; strict monotonicity of $\bar{\delta}(\cdot, c)$ in c (established via the strict convexity argument in Step 4 of Theorem 3.1's proof,

which shows every increase in c strictly raises the equilibrium monitoring cost paid along the path) implies $\bar{c}(\cdot)$ is itself strictly increasing in δ. Statement (i) is the content of Theorem 3.1 evaluated at $c = \bar{c}(\delta)$ and v approaching (R, R). Statement (ii) is the contrapositive of the monotonicity just established. □

4.3 Robustness and Breakdown

Corollaries 4.2 and 4.3 together characterize the conditions under which cooperation is robust to the frictions introduced by imperfect, costly monitoring, and the conditions under which it is not.

Cooperation is robust when three conditions hold jointly: players are sufficiently patient (δ close to 1); punishment strategies remain credible and severe enough to deter deviation even accounting for the residual detection noise that a costly monitoring technology cannot fully eliminate; and monitoring investment is itself incentive-compatible, in the sense verified in Step 4 of Theorem 3.1's proof – that the marginal value of better information during a punishment or review phase outweighs its marginal cost. A grim-trigger strategy adapted to this setting captures the intuition simply: cooperate and monitor at a moderate intensity as long as signals remain consistent with cooperation; upon observing a signal realization inconsistent with cooperation by more than a statistically calibrated margin, switch to mutual defection and (since detection is no longer required) monitor minimally during the punishment phase.

Cooperation breaks down, in the sense of Corollary 4.3, along two limiting paths. As the cost function $c \to \infty$ pointwise, players are driven to under-invest in monitoring regardless of δ, since Corollary 4.2's mechanism – rising equilibrium cost forcing down $m_i^\dagger$ – eventually drives $m_i^\dagger \to 0$; in the limit, only the stage-game Nash payoff (P, P) remains sustainable, exactly as in a one-shot game. Symmetrically, as the baseline noise variance $\sigma^2 \to \infty$, no finite level of monitoring investment can restore informativeness (since $\sigma_i^2(m_i) \geq \sigma_0^2/(1+\alpha\beta)$ is bounded below by a term proportional to σ_0^2 for any $m_i < 1$), so the signal becomes uninformative regardless of cost incurred, and the repeated game collapses to one with effectively no monitoring, where – by the classical unraveling argument – only the stage-game Nash payoff can be sustained in the limit.

5 Computational Simulation

The theoretical results of Sections 3–4 are asymptotic statements about the discount-factor threshold required to sustain a *given* equilibrium construction. They say little about which strategy classes would actually emerge from a boundedly rational selection process, nor about the transient dynamics of aggregate welfare as a heterogeneous population adjusts. We address these questions computationally, using an evolutionary agent-based model calibrated to the environment of Section 2.

5.1 Simulation Design

The simulation implements the stage game of Table 1 and the signal-generating process (2)–(3), with $\sigma_0^2 = 4$, and matches between pairs of agents lasting 100 rounds. The evolving population consists of 120 agents, divided at initialization roughly evenly into the three archetypes of Section 2: blind free-riders ($m_i \equiv 0$); fixed-overhead monitors, with base cost $\bar{c}_i$ drawn uniformly from $[0.3, 0.6]$; and trust-adjusted adaptive monitors, with $\bar{c}_i \sim U[0.3, 0.6]$, wariness $\omega_i \sim U[0.08, 0.20]$, and forgiveness rate $\phi_i \sim U[0.04, 0.15]$.

In addition to the evolving population, the simulation maintains a fixed background cohort of 40 adversarial agents – "testers" – that never mutate or die and are excluded from the fitness ranking used for selection. Each tester probes a newly matched opponent with three consecutive periods of unconditional defection under (in this idealized adversarial benchmark) perfect, costless monitoring of the true opponent action, rather than the noisy signal available to ordinary players. After the probe, the tester audits whether the opponent retaliated in at least two of the three probe rounds; if not, the tester classifies the opponent as exploitable and continues unconditional defection for the remainder of the match, while if the opponent did retaliate, the tester reverts to playing a flawless, noiseless Tit-for-Tat. The tester cohort is designed to represent a persistent, adversarial background risk – akin to a population of bad-faith counterparties, fraudulent actors, or adversarial market participants – against which the evolving population must remain robust, and it is held fixed in size and behavior across generations so that the theoretical predictions of Section 4 can be assessed against a constant, rather than co-evolving, threat level.

Each generation, every agent in the combined pool (evolving population plus tester cohort) plays six matches against randomly drawn opponents from the combined pool, and fitness is defined as mean total payoff across those matches. Selection operates only within the evolving population: the top half of the evolving population by fitness survives, and the population is replenished to its original size by cloning surviving parents with small mutations to $\bar{c}_i$, ω_i, and ϕ_i (Gaussian perturbations applied with independent probability to each parameter). The simulation is run for 40 generations.

5.2 Results

Figure 1 reports the two headline outputs of the simulation: the evolving composition of the population by archetype (left panel) and the mean fitness of the evolving population (right panel), both plotted against generation number.

Three qualitative patterns are apparent, each of which maps directly onto the theoretical results of Section 4.

Fixed-overhead monitoring is driven to extinction almost immediately. The fixed-cost archetype pays its monitoring cost $\bar{c}_i \in [0.3, 0.6]$ every period, in every match, regardless of whether the opponent is cooperative or adversarial, and regardless of whether informative monitoring is actually needed at that moment. Against a population containing a substantial cooperative subgroup, this constant overhead is a pure fitness cost relative to agents who

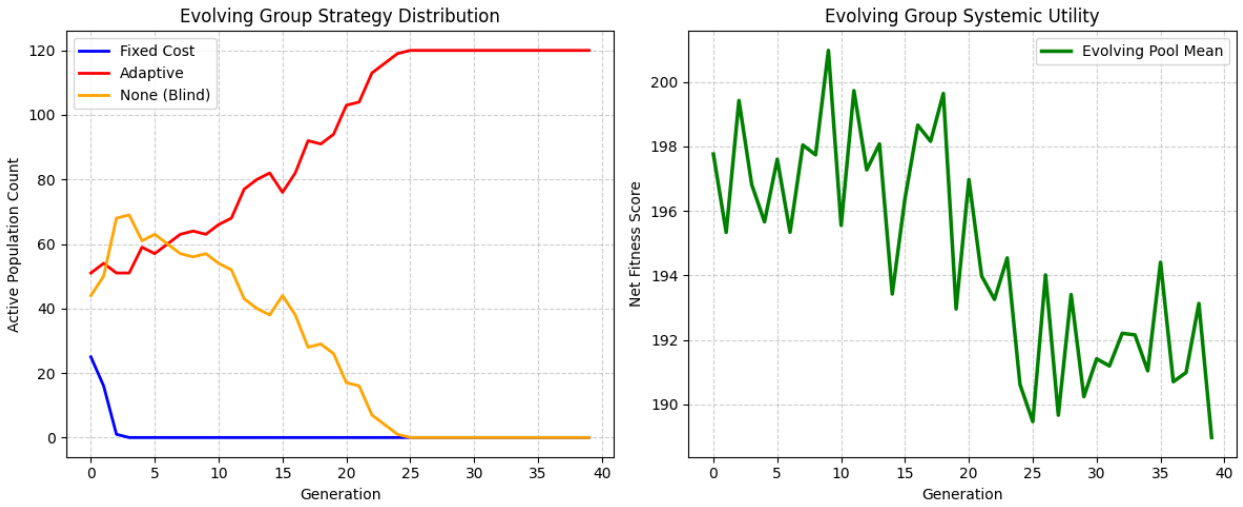

Figure 1: Left: population counts of the three evolving archetypes (fixed-overhead, adaptive, and blind) over 40 generations, against a constant background of 40 adversarial testers. Right: mean net fitness (discounted-equivalent total payoff) of the evolving population over the same horizon.

can economize on monitoring once trust is established; the fixed-overhead population count falls from roughly a quarter of the evolving population at initialization to effectively zero by generation 3, consistent with Corollary 4.3: bearing cost $\bar{c}_i$ unconditionally is dominated once a state-contingent alternative is available in the same strategy space.

Blind free-riding survives temporarily but is eventually eliminated. Unlike fixed monitors, blind agents pay no monitoring cost, and especially early in the simulation while the population still contains many fixed-cost agents whose cost burden is even less efficient, blind agents are competitive, and their population share briefly rises. However, the blind agents have no ability to distinguish an ordinary cooperative opponent from a tester in its exploitative phase, since they never receive any signal at all; they are therefore maximally vulnerable to sustained exploitation by the tester cohort. As adaptive monitors come to dominate the rest of the population, blind agents are increasingly likely to be matched against either testers or, indirectly, against a population environment in which the absence of any monitoring is simply first-order dominated, and their count decays smoothly to zero by roughly generation 25.

Trust-adjusted adaptive monitoring comes to dominate the population, converging to fixation. The adaptive archetype's defining feature, coupling monitoring intensity to a trust state that falls swiftly upon receiving signals suggestive of defection, per (6)–(7), allows it to pay the fixed monitors' cost only when trust is low (i.e., exactly when detection value is highest, as in the early rounds of a match against an untested opponent or a tester's probe phase) while economizing on cost once trust is established against a genuinely cooperative counterpart. This is precisely the state-contingent monitoring policy shown to be incentive-compatible in Step 4 of Theorem 3.1's proof. By generation 25, the adaptive archetype has

dominated the population space at 120 agents.

Aggregate welfare fluctuates persistently even after fixation. The right panel of Figure 1 shows that mean fitness of the evolving population does not converge to a stable value once the adaptive archetype reaches fixation; instead, it continues to oscillate, with visible downward as well as upward swings, over the remaining generations. This is a direct consequence of the fixed, non-co-evolving tester cohort: because 40 of the 160 agents in the combined pool are adversarial testers using perfect, costless monitoring, roughly one match in four for any evolving agent is against an opponent that cannot itself be reformed by selection, and the ex-post realized payoff of any given generation's random match assignment against this threat is itself a stochastic quantity, independent of any further improvement within the now-homogeneous adaptive population. This illustrates a distinction not visible in the asymptotic theory of Sections 3–4: convergence of the equilibrium *strategy class* is not the same as convergence, or even stabilization, of realized *payoffs*, when a persistent adversarial subpopulation of fixed size is present. The theoretical machinery of Theorem 4.1 describes how the sustainable payoff *set* responds to changes in the noise and cost environment; it is silent on the higher-frequency volatility induced by finite-sample matching against a fixed adversarial share, which is exactly what the simulation surfaces.

5.3 Discussion of the Simulation Relative to the Theory

The simulation should be read as simply an illustration of the theoretical results and not a proof, since the evolutionary dynamics implement boundedly rational strategies (fixed archetypes evolving continuous parameters) rather than the fully optimized, sophisticated equilibrium strategies constructed in the proof of Theorem 3.1. With that caveat, the qualitative correspondence is close: Corollary 4.2 predicts that the sustainable payoff set shrinks as monitoring becomes less efficiently targeted, and the simulation shows exactly this inefficiency being competed away, as fixed-overhead monitoring (an unconditional, non-state-contingent cost) is eliminated in favor of trust-contingent monitoring (a state-contingent cost that approximates the incentive-compatible $m_i^\dagger$ of Theorem 3.1's construction). The persistent volatility in aggregate payoff documented above is not a contradiction of the stability results of Section 4, which characterize the equilibrium payoff *set* as a function of the noise and cost primitives, holding those primitives fixed; it is instead a reminder that a fixed adversarial share is itself best modeled as a component of the individual-rationality constraint (via the minimax payoff u_i^*) rather than as noise to be averaged away, a point we return to in Section 6.

6 Discussion and Extensions

Several extensions of the model are natural and, we believe, tractable within the framework developed here.

Co-evolving adversaries. The simulation of Section 5 fixes the tester cohort's size and

strategy across generations. A natural extension allows the adversarial share itself to evolve, or allows testers to adapt their probing strategy in response to the evolving population's monitoring intensity. This would connect the evolutionary simulation more tightly to the minimax-based individual-rationality constraint of Definition 4.4, since a rational adversary's best response to a population with monitoring intensity m is itself a well-defined optimization problem.

Network monitoring. The model as presented is dyadic: each player monitors a single counterparty. Many applications, like reputation systems, decentralized markets, and multilateral collusion all involve monitoring across a network of counterparties, where a player's signal about one counterparty may be informative about others (for instance, through shared reputation scores). Extending Theorem 4.1 to a network setting would require replacing the pairwise signal-generating process (2) with a joint process over the network and re-deriving the statistical detectability condition of Step 1 in Theorem 3.1's proof at the network level.

Continuous time. High-frequency trading, algorithmic collusion, and other applications naturally suggest a continuous-time limit of the model, in which the discrete review-phase construction of Theorem 3.1 is replaced by a continuous monitoring flow. The threshold cost result of Corollary 4.3 would then be expected to take the form of a bound relating the monitoring cost *rate* to the discount *rate*, in the spirit of existing results on repeated games played at high frequency.

Applications. The threshold structure identified in Corollary 4.3 has direct interpretations across the applications mentioned in Section 3. In antitrust settings, it suggests that collusive outcomes among firms are most vulnerable to erosion precisely when auditing or market-intelligence costs rise relative to the discount rate implied by firms' cost of capital – consistent with the empirical observation that cartels tend to be fragile in industries with high monitoring costs (e.g., differentiated products with opaque pricing) relative to industries with low ones (e.g., commodity markets with public benchmark prices). In organizational economics, it implies that the erosion of internal oversight budgets should be expected to produce not a gradual decline in cooperative effort but a comparatively sharp transition once monitoring costs cross the threshold $\bar{c}(\delta)$, an implication that could in principle be tested against panel data on internal audit spending and measures of employee or subsidiary compliance.

7 Conclusion

This paper studied the stability of Folk Theorem equilibria in a repeated Prisoner's Dilemma when monitoring is imperfect, costly, and endogenously chosen. We showed that a Folk Theorem-type existence result survives this extension (Theorem 3.1), extending the observation-cost literature of Miyagawa, Miyahara, and Sekiguchi [MMS08] to a setting in which signal precision, rather than the binary choice to observe or not, is the object of costly choice. Our principal contribution is the stability analysis of Section 4: the equilibrium payoff correspondence $V(\sigma^2, c)$ is upper hemicontinuous in the baseline noise level and cost structure

(Theorem 4.1), shrinks monotonically as either parameter worsens (Corollary 4.2), and admits an explicit, discount-factor-dependent cost threshold governing whether near-efficient cooperation remains attainable (Corollary 4.3). The accompanying evolutionary simulation of Section 5 illustrated these results in a setting with boundedly rational, competing strategy archetypes and a persistent adversarial threat, and additionally surfaced a phenomenon not visible in the asymptotic theory: convergence of the population to the theoretically dominant strategy class does not eliminate volatility in realized payoffs when a fixed adversarial share is present, pointing toward the co-evolving-adversary extension discussed in Section 6 as a natural next step. Taken together, the theoretical and computational results suggest that the classical Folk Theorem's central insight of patience and punishment is what sustains cooperation, and this survives the introduction of realistic informational frictions, but does so only within limits that are governed jointly by how patient players are and by how expensive it is, at the margin, for them to find out what their counterparties are actually doing.

A Simulation Code

The following is the complete Python implementation used to generate the results reported in Section 5 and Figure 1.

Listing 1: Evolutionary simulation of monitoring strategies under the costly-signal environment of Section 2.

```
# -*- coding: utf-8 -*-
"""Untitled1.ipynb

Automatically generated by Colab.

Original file is located at
    https://colab.research.google.com/drive/1H7xX8V2D2xdlX0xKMOlGyzoJEkzIPH-a
"""

import numpy as np
import matplotlib.pyplot as plt
import random
from collections import Counter
from typing import Tuple

# Prisoner's Dilemma Matrix: 0 = Cooperate (C), 1 = Defect (D)
# T=5 (Temptation), R=3 (Reward), P=1 (Punishment), S=0 (Sucker)
PAYOFFS = np.array([[[3, 3], [0, 5]],
                    [[5, 0], [1, 1]]])

class Player:
    def __init__(self, player_type: str, base_cost: float = 0.5,
                 wariness: float = 0.15, forgiveness_rate: float = 0.08):
        self.player_type = player_type
        self.base_cost = base_cost
```

```
        self.wariness = wariness
        self.forgiveness_rate = forgiveness_rate
        self.history = []
        self.total_payoff = 0.0
        self.trust = 0.5 # Start neutral
        self.tester_exploiting = False

    def reset(self):
        self.history = []
        self.total_payoff = 0.0
        self.trust = 0.5
        self.tester_exploiting = False

    def choose_monitoring_cost(self, round_num: int) -> float:
        if self.player_type in ["none", "evil_tester"]:
            return 0.0
        elif self.player_type == "fixed":
            return self.base_cost
        elif self.player_type == "adaptive":
            if self.history:
                last_signal = self.history[-1][1]
                signal_quality = 1.0 - np.clip(last_signal, 0.0, 1.0)
                self.trust += self.wariness * (signal_quality - 0.5)
                self.trust = np.clip(self.trust, 0.0, 1.0)

            # Efficient dynamic mapping: High trust -> negligible information tax
            cost = np.interp(self.trust, [0.0, 1.0], [self.base_cost, 0.01])
            return cost
        return 0.0

    def choose_action(self, round_num: int, opp_real_last_action: int = None) -> int:
        # --- PERFECT MONITORING MULTI-PROBE TESTER ---
        if self.player_type == "evil_tester":
            PROBE_ROUNDS = 3
            if round_num < PROBE_ROUNDS:
                return 1 # Continuous early defection pressure

            if round_num == PROBE_ROUNDS:
                # Audit rounds 1, 2, 3 to check for opponent's response to our probes
                # Elements: (my_action, observed_signal, cost, true_opp_action,
                    net_payoff)
                retaliations = sum([h[3] for h in self.history[1:PROBE_ROUNDS+1] if h[3]
                     is not None])
                if retaliations < 2:
                    self.tester_exploiting = True # Target is blind or weak, lock
                        execution
                else:
                    self.tester_exploiting = False # Target is secure, ease up

            if self.tester_exploiting:
```

```
                return 1 # Ruthless continuous culling
            else:
                # Ease up and execute flawless noiseless Tit-for-Tat
                return opp_real_last_action if opp_real_last_action is not None else 0

        # --- STANDARD STRATEGIES ---
        if round_num == 0 or not self.history:
            return 0

        last_signal = self.history[-1][1]
        cutoff = 1.0 - self.trust if self.player_type == "adaptive" else 0.5

        if last_signal < cutoff:
            return 0
        else:
            if random.random() < self.forgiveness_rate:
                return 0
            return 1

    def observe_signal(self, true_opp_action: int, m: float, base_variance: float =
        4.0) -> float:
        if self.player_type == "evil_tester":
            return float(true_opp_action) # Flawless perfect monitoring channel
        variance = base_variance / (1.0 + 40.0 * m)
        noise = np.random.normal(0, np.sqrt(variance))
        return np.clip(true_opp_action + noise, -0.5, 1.5)

def simulate_match(p1: Player, p2: Player, num_rounds: int = 100) -> Tuple[float,
    float]:
    p1.reset()
    p2.reset()
    a1_last, a2_last = None, None
    for r in range(num_rounds):
        a1 = p1.choose_action(r, a2_last)
        m_cost1 = p1.choose_monitoring_cost(r)
        a2 = p2.choose_action(r, a1_last)
        m_cost2 = p2.choose_monitoring_cost(r)

        net1 = PAYOFFS[a1][a2][0] - m_cost1
        net2 = PAYOFFS[a1][a2][1] - m_cost2

        sig1 = p1.observe_signal(a2, m_cost1 * 2.5)
        sig2 = p2.observe_signal(a1, m_cost2 * 2.5)

        # Structure payload history: (my_action, my_signal, my_cost, opp_true_action,
            my_net)
        p1.history.append((a1, sig1, m_cost1, a2, net1))
        p2.history.append((a2, sig2, m_cost2, a1, net2))

        p1.total_payoff += net1
```

```
        p2.total_payoff += net2
        a1_last, a2_last = a1, a2
    return p1.total_payoff, p2.total_payoff

def run_evolution(num_generations: int = 40, pop_size: int = 160, rounds_per_match: int = 100, matches_per_agent: int = 6):
    # Establish isolated background threat cohort (1/4 of total system volume)
    num_testers = pop_size // 4 # 40 permanent testers
    testers_pool = [Player('evil_tester') for _ in range(num_testers)]

    # The remaining 120 slots belong strictly to the evolving pool
    evolving_size = pop_size - num_testers
    population = []

    for _ in range(evolving_size // 3):
        population.append(Player('adaptive', base_cost=random.uniform(0.3, 0.6), wariness=random.uniform(0.08, 0.20), forgiveness_rate=random.uniform(0.04, 0.15)))
    for _ in range(evolving_size // 3):
        population.append(Player('fixed', base_cost=random.uniform(0.3, 0.6)))
    for _ in range(evolving_size - (2 * (evolving_size // 3))):
        population.append(Player('none'))

    history = {'fixed': [], 'adaptive': [], 'none': [], 'avg_payoff': []}

    for gen in range(num_generations):
        full_pool = population + testers_pool
        fitness = [0.0] * len(full_pool)

        for i in range(len(full_pool)):
            total_score = 0.0
            for _ in range(matches_per_agent):
                opp_idx = random.randint(0, len(full_pool)-1)
                while opp_idx == i:
                    opp_idx = random.randint(0, len(full_pool)-1)
                score1, _ = simulate_match(full_pool[i], full_pool[opp_idx])
                total_score += score1
            fitness[i] = total_score / matches_per_agent

        # Strict Extraction: Calculate metrics strictly using the evolving genotypes
        evolving_fitness = fitness[:evolving_size]
        avg_evolving_payoff = np.mean(evolving_fitness)

        # Truncation selection executed strictly within the evolving subset
        sorted_indices = np.argsort(evolving_fitness)[::-1]
        survivors = [population[i] for i in sorted_indices[:evolving_size // 2]]

        new_evolving_pop = list(survivors)
        while len(new_evolving_pop) < evolving_size:
            parent = random.choice(survivors)
```

```
            child = Player(parent.player_type, base_cost=parent.base_cost, wariness=
                getattr(parent, 'wariness', 0.15), forgiveness_rate=parent.
                forgiveness_rate)

            if random.random() < 0.20:
                child.base_cost = max(0.1, child.base_cost + random.gauss(0, 0.04))
            if random.random() < 0.15 and child.player_type == 'adaptive':
                child.wariness = max(0.02, min(0.35, child.wariness + random.gauss(0,
                    0.02)))
            if random.random() < 0.15:
                child.forgiveness_rate = max(0.0, min(0.35, child.forgiveness_rate +
                    random.gauss(0, 0.01)))

            new_evolving_pop.append(child)

        population = new_evolving_pop[:evolving_size]

        counts = {'fixed': 0, 'adaptive': 0, 'none': 0}
        for p in population:
            counts[p.player_type] += 1

        for k in history.keys():
            if k in counts: history[k].append(counts[k])
        history['avg_payoff'].append(avg_evolving_payoff)

        print(f"Gen {gen:2d}: Fixed={counts['fixed']:3d} | Adaptive={counts['adaptive
            ']:3d} | None={counts['none']:3d} | [Testers Background={num_testers}] |
            Avg Evolving Payoff={avg_evolving_payoff:.1f}")

        # ================================================================================
    # DIAGNOSTIC TRAIT METRIC CHECK FOR SURVIVING AGENTS (FINAL GEN)
    # ================================================================================
    print("\n" + "="*60)
    print(" CHARACTERISTICS OF THE BEST SURVIVING AGENTS (FINAL GEN) ")
    print("="*60)

    adaptive_survivors = [p for p in population if p.player_type == 'adaptive']

    if adaptive_survivors:
        base_costs = [p.base_cost for p in adaptive_survivors]
        wariness_factors = [p.wariness for p in adaptive_survivors]
        forgiveness_rates = [p.forgiveness_rate for p in adaptive_survivors]

        mean_cost = np.mean(base_costs)
        mean_wariness = np.mean(wariness_factors)
        mean_forgiveness = np.mean(forgiveness_rates)

        std_cost = np.std(base_costs)
        std_wariness = np.std(wariness_factors)
        std_forgiveness = np.std(forgiveness_rates)
```

```
211
212         print(f"Total Surviving Adaptive Profiles Evaluated: {len(adaptive_survivors)}"
                )
213         print(f"1. Base Cost Ceiling (base_cost) : Mean = {mean_cost:.4f} | Std Dev = {
                std_cost:.4f}")
214         print(f"2. Systemic Wariness Factor (wariness): Mean = {mean_wariness:.4f} |
                Std Dev = {std_wariness:.4f}")
215         print(f"3. Stochastic Forgiveness (forgiveness): Mean = {mean_forgiveness:.4f}
                | Std Dev = {std_forgiveness:.4f}")
216
217         # Identify the single highest-performing elite agent vector based on total
                accumulated payoff
218         sorted_elites = sorted(adaptive_survivors, key=lambda x: x.total_payoff,
                reverse=True)
219         elite = sorted_elites[0]
220         print("-"*60)
221         print(" Trait Blueprint of the Single Highest Performing Elite Agent:")
222         print(f" -> Optimized Base Cost Override: {elite.base_cost:.4f}")
223         print(f" -> Optimized Wariness Scalar : {elite.wariness:.4f}")
224         print(f" -> Optimized Forgiveness Floor : {elite.forgiveness_rate:.4f}")
225     else:
226         print("No Adaptive strategies survived the evolutionary timeline.")
227     print("="*60 + "\n")
228
229
230     # Plot
231     plt.figure(figsize=(12, 5))
232     gens = range(num_generations)
233
234     plt.subplot(1, 2, 1)
235     plt.plot(gens, history['fixed'], label='Fixed Cost', color='blue', lw=2)
236     plt.plot(gens, history['adaptive'], label='Adaptive', color='red', lw=2)
237     plt.plot(gens, history['none'], label='None (Blind)', color='orange', lw=2)
238     plt.xlabel('Generation')
239     plt.ylabel('Active Population Count')
240     plt.title('Evolving Group Strategy Distribution')
241     plt.legend()
242     plt.grid(True, linestyle='--', alpha=0.6)
243
244     plt.subplot(1, 2, 2)
245     plt.plot(gens, history['avg_payoff'], color='green', lw=2.5, label='Evolving Pool
            Mean')
246     plt.xlabel('Generation')
247     plt.ylabel('Net Fitness Score')
248     plt.title('Evolving Group Systemic Utility')
249     plt.legend()
250     plt.grid(True, linestyle='--', alpha=0.6)
251     plt.tight_layout()
252     plt.show()
253
```

```
254 if __name__ == '__main__':
255     run_evolution()
```

References

[Abr88] D. Abreu. On the Theory of Infinitely Repeated Games with Discounting. *Econometrica*, 56(2):383–396, 1988.

[APS90] D. Abreu, D. Pearce, and E. Stacchetti. Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica*, 58(5):1041–1063, 1990.

[FLM94] D. Fudenberg, D. K. Levine, and E. Maskin. The Folk Theorem with Imperfect Public Monitoring. *Econometrica*, 62(5):997–1039, 1994.

[FM86] D. Fudenberg and E. Maskin. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica*, 54(3):533–554, 1986.

[MMS08] E. Miyagawa, Y. Miyahara, and T. Sekiguchi. The Folk Theorem for Repeated Games with Observation Costs. *Journal of Economic Theory*, 139(1):192–221, 2008.

[Sel75] R. Selten. Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games. *International Journal of Game Theory*, 4(1):25–55, 1975.