NBER WORKING PAPER SERIES

LE BUREAU DES LÉGENDES: A DYNAMIC THEORY OF DOUBLE AGENTS

Sebastian Galiani
Franco Mettola La Giglia

Working Paper 35085
http://www.nber.org/papers/w35085

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

We used generative artificial intelligence tools (Refine and Claude by Anthropic) for coding assistance, literature verification, mathematical verification, and editorial support. All analytical content and substantive conclusions are the authors' own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Sebastian Galiani and Franco Mettola La Giglia. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Le Bureau des Légendes: A Dynamic Theory of Double Agents
Sebastian Galiani and Franco Mettola La Giglia
NBER Working Paper No. 35085
April 2026
JEL No. C70

ABSTRACT

This paper models double agents—individuals coerced into simultaneously serving two rival intelligence agencies—as a finite-horizon trilateral game between Agency A, Organization B, and the Mole M. The double agent persists within a corridor of survival: a set of bilateral beliefs under which both principals continue the relationship. Under a linear-quadratic-Gaussian benchmark, we obtain a closed-form characterization of the corridor geometry, analytical upper bounds on expected duration, and three main results: (i) existence of a double-agent equilibrium, (ii) structural transitoriness— belief updating, terminal unraveling, and compounding survival risk ensure inevitable collapse, and (iii) comparative statics linking monitoring technology, punishment severity, and protection costs to expected duration. The key mechanism is self-destructive experimentation: B learns about the mole's type through both the type-dependence of effort and the traceability channel , which amplifies this learning. Extensions establish existence under general specifications and show that duration decays exponentially in the number of rival agencies. Predictions are consistent with historical patterns from Kim Philby to the Cold War mole hunts. The setting—bilateral coercion, existential participation constraints, and Bayesian learning in a finite-horizon trilateral structure— defines what we term antagonistic common agency.

Sebastian Galiani
University of Maryland, College Park
Department of Economics
and NBER
sgaliani@umd.edu

Franco Mettola La Giglia
Universidad de San Andres
fmettolalagiglia@udesa.edu.ar

Le Bureau des Légendes: A Dynamic Theory of Double Agents

Sebastian Galiani* Franco Mettola La Giglia†

Abstract

This paper models double agents—individuals coerced into simultaneously serving two rival intelligence agencies—as a finite-horizon trilateral game between Agency A, Organization B, and the Mole M. The double agent persists within a *corridor of survival*: a set of bilateral beliefs under which both principals continue the relationship. Under a linear-quadratic-Gaussian benchmark, we obtain a closed-form characterization of the corridor geometry, analytical upper bounds on expected duration, and three main results: (i) existence of a double-agent equilibrium, (ii) structural transitoriness—belief updating, terminal unraveling, and compounding survival risk ensure inevitable collapse, and (iii) comparative statics linking monitoring technology, punishment severity, and protection costs to expected duration. The key mechanism is *self-destructive experimentation*: B learns about the mole's type through both the type-dependence of effort and the traceability channel γ, which amplifies this learning. Extensions establish existence under general specifications and show that duration decays exponentially in the number of rival agencies. Predictions are consistent with historical patterns from Kim Philby to the Cold War mole hunts. The setting—bilateral coercion, existential participation constraints, and Bayesian learning in a finite-horizon trilateral structure—defines what we term *antagonistic common agency*.

Keywords: Espionage, double agents, credible threats, imperfect monitoring, Bayesian equilibrium, coercion, common agency.

JEL Codes: C73, D74, D82, D83.

1 Introduction

Kim Philby served as a senior officer in the British Secret Intelligence Service for nearly three decades while operating as a Soviet agent. His survival inside MI6 owed much to institutional trust, class-based deference, and monitoring systems too noisy to resolve the ambiguity about

*Department of Economics, University of Maryland. Email: sgaliani@umd.edu.

†Universidad de San Andrés. Email: fmettolalagiglia@udesa.edu.ar.

We used generative artificial intelligence tools (Refine and Claude by Anthropic) for coding assistance, literature verification, mathematical verification, and editorial support. All analytical content and substantive conclusions are the authors' own.

his loyalties (Macintyre, 2014). At the opposite extreme, Oleg Penkovsky—a colonel in Soviet military intelligence who passed critical secrets to the CIA and MI6 during the early 1960s—was detected and executed within roughly eighteen months (Schecter and Deriabin, 1992). His behavioral inconsistencies eroded Soviet confidence quickly enough to cross the threshold for elimination.

These two cases bracket the central puzzle: the same underlying strategic structure—an individual under simultaneous coercion from two rival organizations, each monitoring his behavior through imperfect signals—produces outcomes that differ by an order of magnitude in duration. A satisfactory theory must explain not only *why* double agents can exist as a sustained arrangement, but also *why* they persist for vastly different lengths of time and *why* the arrangement invariably collapses.[1]

Existing literature on espionage has addressed organizational design under infiltration risk (Baccara and Bar-Isaac, 2008) and the strategic dimensions of intelligence in adversarial settings. The specific mechanism by which a double agent is sustained as an equilibrium outcome—and the structural forces driving its dissolution—remains unformalized. The standard principal-agent framework is ill-suited to this problem: the agent's outside option is not a reservation wage; it is death. The agent also serves two principals with diametrically opposed objectives—a structure distinct from the standard common agency model of Bernheim and Whinston (1986).

We study a finite-horizon game among three players: Agency A (the recruiting intelligence service), Organization B (M's parent organization), and the Mole M, coerced by A and required to maintain his standing with B. The roles are defined by function, not identity: for Philby, A is the KGB and B is MI6; for Penkovsky, A is the CIA/MI6 and B is the GRU. In each period, M allocates effort between the two sides (effort toward A delivers intelligence; effort toward B maintains cover), while A and B observe noisy signals and update their beliefs about his loyalty. M's overriding objective is survival—his own and that of his family. We term this setting *antagonistic common agency*. The label is analogical: unlike the contracting framework of Bernheim and Whinston (1986), where principals offer competing contracts, here neither principal offers a contract—each observes noisy signals and decides whether to terminate. The common feature is the trilateral structure, but the interaction is a dynamic bilateral reputation game under coercion.

Our main results are threefold. First, we characterize a *corridor of survival*—a region in bilateral belief space (θ_A, θ_B) within which both A and B continue engaging M; a double-agent

[1]The term "double agent" is frequently conflated with "undercover agent" or "spy." A double agent is an individual recruited or coerced by one intelligence service and subsequently turned—or appearing turned—by another. The simultaneity of two coercive relationships, not merely the act of espionage, defines the phenomenon. See Andrew and Mitrokhin (1999) for a systematic treatment.

equilibrium exists inside this corridor. Second, the equilibrium is *structurally transitory*: Bayesian updating compounds small behavioral inconsistencies over time, and combined with last-period effects and the extreme payoff asymmetry, beliefs eventually exit the corridor. Third, comparative statics link observable features of intelligence operations to model primitives—monitoring improvements reduce expected duration, the punishment comparative static operates through shifts in the continuation threshold (not the shape of the elimination function), and imperfect family protection generates distinctive fragility. We work throughout under a linear-quadratic-Gaussian (LQG) benchmark—normal signals, quadratic costs, linear elimination cutoffs—under which the corridor geometry and equilibrium effort admit closed-form characterization; the duration bound is analytically rigorous for the Gaussian belief dynamics but uses a certainty-equivalence approximation near corridor boundaries (Section 7.3).

The model builds on the theory of repeated games with imperfect monitoring (Green and Porter, 1984; Abreu et al., 1990), adapted to a trilateral setting with coercion. It draws on Schelling (1960)'s analysis of credible threats as the conceptual foundation for the corridor mechanism. The common agency framework of Bernheim and Whinston (1986) informs the multi-principal structure;the information design perspective of Kamenica and Gentzkow (2011) and the career concerns tradition of Holmström (1999) illuminate M's strategic management of information flows (the mole's effort affects future beliefs about his type, as in career concerns, but under existential rather than pecuniary stakes); and the treatment of coercion connects to the literature on labor coercion (Acemoglu and Wolitzky, 2011).

Beyond espionage, the structure applies to any setting with simultaneous coercion from two principals under imperfect monitoring—corporate whistleblowers, undercover law enforcement, confidential informants, and witness protection all share the trilateral coercion structure, though formal application requires adapting the signal structure to each domain.

Section 2 reviews the related literature. Section 3 presents the model. Section 4 characterizes the equilibrium and derives comparative statics. Section 5 establishes robustness beyond the benchmark. Section 6 interprets the findings through historical cases. Section 7 covers the N-agency fragility result, applications, and limitations. Section 8 concludes. Proofs are in the Appendix.

2 Related Literature

The paper builds on the literature on repeated games with imperfect monitoring. Green and Porter (1984) showed that cooperative outcomes can be sustained through trigger strategies

under noisy public signals, and Abreu et al. (1990) characterized the full set of equilibrium payoffs. Our model departs from this tradition in three ways: the game is trilateral with asymmetric roles, the horizon is finite (creating backward-induction pressure absent in discounted infinite-horizon games), and one player faces existential rather than pecuniary consequences. The reputation literature—Kreps and Wilson (1982) and Milgrom and Roberts (1982) showed that small amounts of incomplete information sustain cooperation in finite games; Fudenberg and Levine (1989) extended these results to infinite horizons—is equally central. The double-agent equilibrium relies on both principals maintaining uncertainty about M's type; Sobel (1985)'s analysis of how credibility is built and destroyed maps directly onto the evolution of bilateral beliefs. Cripps et al. (2004) show reputations are impermanent under imperfect monitoring; Mailath and Samuelson (2006) provide the comprehensive treatment. Our transitoriness result parallels these findings in a finite-horizon bilateral setting. The strategic experimentation literature—particularly Keller et al. (2005) on exponential bandits with free-riding—provides the closest precedent for our self-destructive experimentation mechanism, though the mole's experimentation is involuntary and self-defeating rather than strategically delayed. Board and ter Vehn (2013) study reputation for quality in continuous time with Brownian signals and endogenous exit—the closest methodological antecedent to the corridor structure—but their single-principal setting lacks bilateral coercion and antagonistic monitoring. Faingold and Sannikov (2011) characterize reputation effects in continuous-time games using techniques that a continuous-time extension would naturally employ. Ely and Välimäki (2003) show that a long-run player may distort actions to signal type, making reputation harmful in equilibrium. The mole faces a related tension: his two reputations operate on separate belief processes monitored by adversarial principals, so reputation with B is simultaneously a survival asset and a target of hostile inference—a bilateral liability with no counterpart in the single-principal bad-reputation literature.

A second strand concerns strategic information transmission. Crawford and Sobel (1982) study cheap talk with misaligned interests; here M sends costly signals to two receivers simultaneously, with existential consequences. The information design framework of Kamenica and Gentzkow (2011) illuminates how M manages what each side learns; Doval and Ely (2020) extend this to sequential information design with a sequence of receivers, paralleling M's period-by-period disclosure to two principals, though M acts under duress rather than committing to a signal structure ex ante. Pavan et al. (2014) develop the dynamic mechanism design framework for evolving private information; our general existence result addresses a related dynamic allocation problem under coercion rather than voluntary participation. On multi-principal design: Bernheim and Whinston (1986) establish the foundational contracting framework; Martimort and Stole (2002) characterize revelation and delegation principles.

The double-agent problem departs from standard common agency on all three dimensions—principals have opposed objectives, participation is involuntary, and the outside option is elimination—motivating the term *antagonistic common agency*.

The third strand is the economics of coercion and credible threats. Schelling (1960) showed that threat power lies in credibility, not severity; the corridor formalizes this as a belief-space region sustained by mutual deterrence. Acemoglu and Wolitzky (2011) model labor coercion; we extend this to bilateral coercion under antagonistic monitoring, where each side's threat credibility depends on the other's monitoring capacity.

3 The Model

3.1 Players and Environment

Three players interact over $T < \infty$ periods: **Agency** A, **Organization** B, and **the Mole** M.

M was originally a member of B. Agency A has compromised M—through capture, blackmail, or ideological recruitment—and coerced him into serving as an informant. M must continue operating within B to maintain the access that makes him valuable to A, while appearing loyal to B to avoid detection and elimination.

Assumption 1 (Coercion)**.** *The mole M does not voluntarily choose to enter the relationship with A. His participation is sustained by credible threats from both sides. M's primary objective is survival—his own and that of his family.*

3.2 Actions

In each period t, the players choose the following actions (timing is specified in Section 3.5):

- **The Mole** M chooses effort levels $e_{A,t} \in [0,1]$ and $e_{B,t} \in [0,1]$, representing the quality of intelligence delivered to A and B. Higher effort toward one side raises detection risk by the other.
- **Agency** A chooses monitoring intensity $m_{A,t} \in [0, \bar{m}_A]$ and protection level $p_t \in [0, \bar{p}]$ for M's family. Monitoring generates informative signals about M's loyalty at a cost.
- **Organization** B chooses monitoring intensity $m_{B,t} \in [0, \bar{m}_B]$ and eliminates M if it detects betrayal with sufficient confidence.

Assumption 2 (Effort Trade-off)**.** *The effort pair $(e_{A,t}, e_{B,t})$ is constrained by the cost structure (Assumption 4(iii)): higher effort toward one side raises the marginal cost of effort toward the other (substitutability). The signal structure is asymmetric: B's signal is degraded*

by effort toward A (through the traceability parameter $\gamma > 0$), but A's signal depends only on e_A and does not directly reflect e_B. The effort trade-off for A operates indirectly through cost substitution, not through signal leakage.

3.3 Information Structure

Information is imperfect and asymmetric: neither A nor B observes the other's signals, monitoring choices, or M's full action profile.

- Agency A observes a noisy signal $s_{A,t}$ of M's effort toward A:

$$s_{A,t} = e_{A,t} + \varepsilon_{A,t}, \qquad \varepsilon_{A,t} \sim \mathcal{N}(0, \sigma_A^2(m_{A,t})),$$

  where $\sigma_A^2(\cdot)$ is decreasing in $m_{A,t}$: more monitoring yields more precise signals.

- Organization B observes a noisy signal $s_{B,t}$ of M's behavior within B:

$$s_{B,t} = e_{B,t} - \gamma \cdot e_{A,t} + \varepsilon_{B,t}, \qquad \varepsilon_{B,t} \sim \mathcal{N}(0, \sigma_B^2(m_{B,t})),$$

  where $\gamma > 0$ is the *traceability parameter*: it captures how much of M's effort toward A leaves detectable traces in B's observation. Traceability is a structural feature of the operational environment (the inherent visibility of espionage activity), distinct from B's monitoring precision $\sigma_B^2(m_{B,t})$, which is B's active investment in counterintelligence. In the Kalman filter, γ amplifies B's signal loading (from $h_B = b$ at $\gamma = 0$ to $h_B = b(1+\gamma)$, where b is the type-sensitivity of effort derived in Section 3.6), while σ_B^2 controls the noise floor.

The mole's *loyalty type* $\omega \sim \mathcal{N}(\mu_0, P_0)$ is his intrinsic propensity toward cooperation; higher ω means a more cooperative agent.[2] Each side maintains a *loyalty belief*—the posterior mean of ω (or $1-\omega$) given signals through period t:

$$\theta_{A,t} \equiv \mathbb{E}_A\big[\omega \mid \mathcal{H}_{A,t}\big], \tag{1}$$

$$\theta_{B,t} \equiv \mathbb{E}_B\big[1-\omega \mid \mathcal{H}_{B,t}\big] = 1 - \mathbb{E}_B\big[\omega \mid \mathcal{H}_{B,t}\big], \tag{2}$$

where $\mathcal{H}_{j,t}$ denotes j's signal history through period t. $\theta_{A,t}$ measures A's confidence in M's cooperation: a higher value means A believes M is delivering better intelligence, while a lower value means A suspects withholding or fabricating. Symmetrically, $\theta_{B,t}$ measures B's confidence in M's loyalty: a higher value means B has less reason to suspect betrayal, while

[2] For the N-agency extension (Theorem 22), we additionally assume compact type support $\omega \in [\underline{\omega}, \bar{\omega}]$ to ensure a uniform lower bound on the per-period exit probability; all other results hold under the Gaussian specification.

a lower value means B's counterintelligence assessment points toward compromise. Under the Gaussian prior and signal structure, beliefs update via the Kalman filter exactly, and $\theta_{j,t}$ is a sufficient statistic for j's belief.

3.4 Payoffs and Functional Form Properties

Payoffs reflect the fundamental asymmetry between M and the agencies.

Assumption 3 (Elimination and Abandonment Technologies)**.** *The functions $d_B : \mathbb{R} \to [0,1]$ and $d_A : \mathbb{R} \to [0,1]$ satisfy:*

(i) $d_B(\theta_B)$ is twice continuously differentiable on $(0, \underline{\theta}_B)$, non-increasing, and convex in θ_B, with $d_B(1) = 0$ and $d_B(0) = 1$. There exists a threshold $\underline{\theta}_B \in (0,1)$ such that $d_B(\theta_B) = 0$ if and only if $\theta_B \geq \underline{\theta}_B$. For $\theta_B > 1$, $d_B(\theta_B) = 0$; for $\theta_B < 0$, $d_B(\theta_B) = 1$.

(ii) $d_A(\theta_A)$ is twice continuously differentiable on $(0, \underline{\theta}_A)$, non-increasing, and convex in θ_A, with $d_A(1) = 0$ and $d_A(0) = 1$. There exists a threshold $\underline{\theta}_A \in (0,1)$ such that $d_A(\theta_A) = 0$ if and only if $\theta_A \geq \underline{\theta}_A$. For $\theta_A > 1$, $d_A(\theta_A) = 0$; for $\theta_A < 0$, $d_A(\theta_A) = 1$.

Assumption 4 (Regularity)**.** *(i) The intelligence value function $v_A : [0,1] \to \mathbb{R}_+$ is twice continuously differentiable, strictly increasing, and concave, with $v_A(0) = 0$.*

(ii) The leakage loss $\ell : [0,1] \to \mathbb{R}_+$ is twice continuously differentiable, strictly increasing, and convex, with $\ell(0) = 0$.

(iii) The effort cost function $c : [0,1] \times [0,1] \times \mathbb{R} \to \mathbb{R}$ depends on the mole's type ω:

$$c(e_A, e_B; \omega) = \tilde{c}(e_A, e_B) + \eta\big[(1-\omega)\, e_A + \omega\, e_B\big],$$

where $\tilde{c}$ is twice continuously differentiable, jointly convex, strictly increasing in each argument, with $\tilde{c}(0,0) = 0$, $D\tilde{c}(0,0) = 0$, $D^2\tilde{c}$ positive definite on $(0,1)^2$, and $\partial^2\tilde{c}/\partial e_A \partial e_B \geq 0$. The parameter $\eta > 0$ governs type-dependent cost: a more cooperative type (higher ω) finds effort toward A cheaper and effort toward B costlier, making equilibrium effort type-dependent and signals informative about ω.

(iv) The signal precision functions $\sigma_j^2 : [0, \bar{m}_j] \to \mathbb{R}_{++}$ for $j \in \{A, B\}$ are continuously differentiable and strictly decreasing, with $\sigma_j^2(0) = \bar{\sigma}_j^2 < \infty$ and $\lim_{m \to \bar{m}_j} \sigma_j^2(m) = \underline{\sigma}_j^2 > 0$.

(v) The family protection function $q : [0, \bar{p}] \to (0,1)$ is continuously differentiable, strictly increasing, and concave, with $q(0) = \underline{q} > 0$ and $q(\bar{p}) < 1$.

(vi) For each principal $j \in \{A, B\}$, monitoring costs $\kappa_j : [0, \bar{m}_j] \to \mathbb{R}_+$ are twice continuously differentiable, strictly increasing, and strictly convex, with $\kappa_j(0) = 0$. The family protection cost $\phi : [0, \bar{p}] \to \mathbb{R}_+$—the monetary cost to A of safeguarding M's family, a

distinctive feature of coerced agency with no counterpart in standard principal-agent models—is twice continuously differentiable, strictly increasing, and strictly convex, with $\phi(0) = 0$.

The Mole M**.** Let $\lambda_t \in \{0, 1\}$ denote M's survival status (1 = alive). The survival probability depends on both sides' actions and beliefs:[3]

$$\mathbb{P}(\lambda_{t+1} = 1 \mid \lambda_t = 1) = \underbrace{\big(1 - d_B(\theta_{B,t})\big)}_{\text{not detected by } B} \cdot \underbrace{\big(1 - d_A(\theta_{A,t})\big)}_{\text{not abandoned by } A} \cdot \underbrace{q(p_t)}_{\text{family protection}} . \tag{3}$$

The multiplicative $q(p_t)$ reflects the operational reality that protection failure ends the arrangement: a mole whose family is seized by B will defect, surrender, or become operationally useless regardless of either principal's beliefs.[4]

The mole M's preferences are:

$$U_M = \sum_{t=1}^{T} \delta^{t-1} \lambda_t \Big[\alpha - c(e_{A,t}, e_{B,t}; \omega)\Big], \tag{4}$$

where ω is M's privately known loyalty type (Section 3.6). The parameter $\alpha > 0$ is the per-period value of survival—physical safety, family security, and continued freedom. We require $\alpha > \bar{\alpha}(T)$, ensuring the value of survival is large enough to sustain positive effort in the first period.[5] The discount factor $\delta \in (0, 1)$ captures M's time preference. The type-dependent cost creates separation: higher-ω types find cooperation with A cheaper, so they choose higher e_A^*, making signals informative about ω and enabling Bayesian learning.

The finite horizon T captures natural endpoints of intelligence relationships (retirement, reassignment, regime change); its anticipation drives the unraveling dynamics of Section 4.

Agency A**.** Agency A's payoff trades off intelligence value against operational costs:

$$U_A = \sum_{t=1}^{T} \delta^{t-1} \lambda_t \Big[v_A(e_{A,t}) - \kappa_A(m_{A,t}) - \phi(p_t)\Big]. \tag{5}$$

[3]The beliefs $\theta_{j,t}$ in this expression are the post-signal beliefs from step 5 of the timing protocol, consistent with the termination decision at step 6.

[4]Threats to family members are among the most effective levers for turning or neutralizing agents. The model treats family compromise as functionally equivalent to detection: once M's family cover is blown, the relationship collapses.

[5]The threshold $\bar{\alpha}(T)$ is characterized in Proposition 9 (Section 5).

Organization B. Organization B receives flow benefit $v_B > 0$ from retaining M (institutional knowledge, operational continuity, network value) but suffers leakage costs and monitoring costs:

$$U_B = \sum_{t=1}^{T} \delta^{t-1} \lambda_t \Big[v_B - \ell(e_{A,t}) - \kappa_B(m_{B,t}) \Big]. \tag{6}$$

3.5 Timing

Within each period t, the sequence of play is:

1. M observes information from his position within B.
2. A sets monitoring $m_{A,t}$ and protection p_t; B sets monitoring $m_{B,t}$.
3. M chooses effort levels $e_{A,t}$ and $e_{B,t}$, without observing the monitoring levels set in step 2.
4. Signals $s_{A,t}$ and $s_{B,t}$ are realized and privately observed.
5. A and B update beliefs $\theta_{A,t}$ and $\theta_{B,t}$.
6. A and B each decide whether to continue (C) or terminate (D) the arrangement. If either terminates, the game ends.

Belief timing convention. $\theta_{j,t}$ denotes the post-signal belief from step 5; monitoring (step 2) and effort (step 3) condition on the entering-period belief $\theta_{j,t-1}$.[6]

3.6 Maintained Parametrization

The following specification is maintained throughout Sections 4–6. The Gaussian signal structure guarantees Bayes' rule is well-defined at every information set. The LQG structure ensures Bayesian updating reduces to the Kalman filter exactly, the corridor is a fixed rectangle, and the mole's optimization is a standard concave dynamic program.

Assumption 5 (LQG Benchmark). *(i) Elimination and abandonment follow linear cutoffs:*

$$d_B(\theta_B) = \max\left\{0,\ 1 - \frac{\theta_B}{\underline{\theta}_B}\right\}, \qquad d_A(\theta_A) = \max\left\{0,\ 1 - \frac{\theta_A}{\underline{\theta}_A}\right\}.$$

(ii) Effort costs are quadratic with cross-substitution and type-dependent linear term:

$$c(e_A, e_B;\, \omega) = \tfrac{1}{2}\big(e_A^2 + e_B^2 + 2\rho\, e_A\, e_B\big) + \eta\big[(1-\omega)\, e_A + \omega\, e_B\big], \qquad \rho \in [0,1),\ \eta > 0.$$

The parameter $\rho \in [0,1)$ governs effort substitutability.

(iii) Signal variances: $\sigma_A^2(m_A) = \bar{\sigma}_A^2/(1+m_A)$ and $\sigma_B^2(m_B) = \bar{\sigma}_B^2/(1+m_B)$.

[6] The value functions $V_t^A(\theta_A)$ and $V_t^B(\theta_B)$ are written as functions of $\theta_{j,t}$ with the understanding that principals anticipate how their step-2 choices affect $\theta_{j,t}$ through the signal structure.

(iv) Intelligence value and leakage loss are linear: $v_A(e_A) = \nu \cdot e_A$ *and* $\ell(e_A) = \ell_0 \cdot e_A$.

Under this specification, the equilibrium effort strategy is affine in type: $e^*_{A,t}(\omega) = a_t + b_t\omega$ with $b_t > 0$. Under quadratic cost, the slope $b_t = \eta/(1-\rho)$ is constant across periods and strictly positive for $\rho \in [0, 1)$.[7] This affine dependence ensures the Kalman filter applies exactly, $\theta_{j,t}$ is a sufficient statistic for j's belief, and the signal loadings $h_A \equiv b$ and $h_B \equiv b(1+\gamma)$ are time-invariant (Section 5.3). We write b without subscript hereafter. For non-LQG specifications, the extended Kalman filter provides a first-order approximation; Section 5 develops the general case under a certainty equivalence assumption.

3.7 Principals' Optimization

Under the benchmark, the continuation thresholds have closed forms. Agency A continues when expected intelligence value $\nu(a_t + b\theta_A)$ exceeds monitoring and protection costs $\kappa^*_A + \phi^*$, yielding

$$\underline{\theta}_A = \frac{\kappa^*_A + \phi^* - \nu a_t}{\nu b},$$

where $\kappa^*_A \equiv \kappa_A(m^*_A)$ and $\phi^* \equiv \phi(p^*)$. Organization B continues when expected leakage $\ell_0 \cdot \mathbb{E}_B[e^*_{A,t}] = \ell_0(a_t + b(1-\theta_B))$ is tolerable. Defining $\bar{\ell} \equiv v_B - \kappa_B - \ell_0 a_t$ (B's maximum tolerable leakage rate), B's continuation condition $\ell_0 b(1-\theta_B) \leq \bar{\ell}$ yields

$$\underline{\theta}_B = 1 - \frac{\bar{\ell}}{\ell_0\, b}.$$

We impose the stationarity condition $a_t = a$ for all $t < T$, which holds in the infinite-horizon limit and approximates well when T is large relative to the unraveling zone near $t = T$. Both thresholds are then constants, making the corridor a fixed rectangle. Near $t = T$, terminal unraveling drives $a_t \to 0$ and thresholds become time-varying; the fully endogenous thresholds are developed in Section 5.1.

[7] Writing $b = \partial e^*_A/\partial\omega$ and $d = \partial e^*_B/\partial\omega$, differentiating the FOC system with respect to ω gives $b + \rho d = \eta$ and $d + \rho b = -\eta$; solving yields $b = \eta/(1-\rho)$. The intercept a_t is time-varying (it depends on the continuation value); the slope b is not, because the type-dependent term enters linearly and the quadratic Hessian $D^2\tilde{c}$ is type-independent.

4 Equilibrium Analysis

The equilibrium concept is **Perfect Bayesian Equilibrium** (PBE): strategies are sequentially rational given beliefs, and beliefs update via Bayes' rule wherever possible. Off-path beliefs are set to $\theta_{j,t}^{\text{off}} = 0$, triggering termination.[8]

4.1 The Corridor of Survival

The central object is the set of belief pairs under which the double-agent arrangement persists.

Definition 1 (Corridor of Survival)**.** *The* corridor of survival *at period t is the set*

$$\mathcal{C}_t \equiv \left\{ (\theta_{A,t}, \theta_{B,t}) \in \mathbb{R}^2 \;\middle|\; \theta_{A,t} \geq \underline{\theta}_{A,t} \;\; \textit{and} \;\; \theta_{B,t} \geq \underline{\theta}_{B,t} \right\},$$

where $\underline{\theta}_{A,t}$ is the minimum belief at which A prefers to continue rather than terminate M, and $\underline{\theta}_{B,t}$ is the minimum belief at which B does not eliminate M.

Under the LQG benchmark, both thresholds are constants (Section 3.7), so the corridor is a fixed rectangle:

Proposition 2 (Corridor Geometry)**.** *Under Assumptions 1–5, and treating the effort intercept a_t and principals' monitoring levels as approximately constant (valid away from the terminal period; see Section 7.3), the corridor is $\mathcal{C} \approx [\underline{\theta}_A, 1] \times [\underline{\theta}_B, 1]$, with comparative statics:*

(i) $\underline{\theta}_A$ is increasing in κ_A and ϕ (higher costs tighten A's threshold).

(ii) $\underline{\theta}_B$ is increasing in ℓ_0 (leakage sensitivity) and independent of A's monitoring parameters (though it depends on A's parameters through the equilibrium effort intercept a).

(iii) Higher γ accelerates belief erosion toward the corridor boundary, reducing expected duration (the static corridor boundaries are γ-invariant).

(iv) Higher ρ raises $b = \eta/(1-\rho)$, which tightens $\underline{\theta}_B$ upward (narrowing the corridor) and accelerates belief erosion, reducing expected duration.

The proof exploits the linearity of the cutoffs and the quadratic cost structure.[9] Figure 1 illustrates a typical belief trajectory within the corridor under the benchmark parametrization.

[8] This is a belief specification, not implied by Bayes' rule (which does not apply off-path). The Gaussian signal structure has full support, so all signal realizations are on-path; the off-path specification applies to unilateral effort deviations that alter the signal distribution.

[9] See Appendix E for the full derivation.

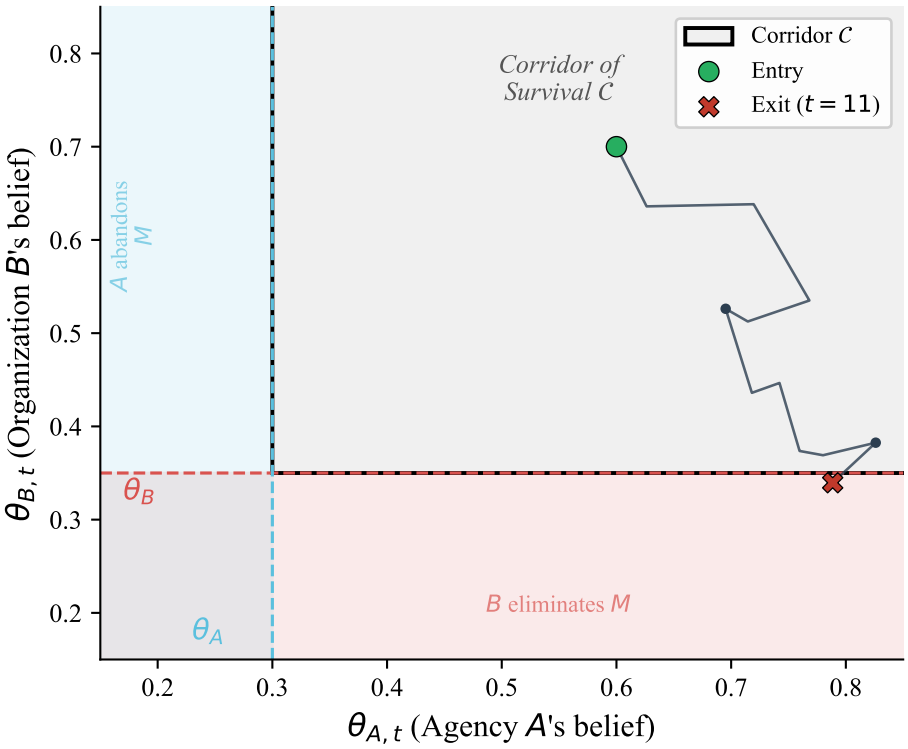


Figure 1: A sample belief trajectory in the corridor of survival. The green circle marks entry at $(\theta_{A,0}, \theta_{B,0}) = (0.6, 0.7)$; the red cross marks exit through B's threshold $\underline{\theta}_B$ after 11 periods. Small dots along the path mark every fifth period, showing the pace of belief movement. The trajectory illustrates the self-destructive experimentation channel: θ_B erodes faster than θ_A because B's signal loading $h_B = b(1+\gamma) = 0.27$ exceeds A's loading $h_A = b = 0.15$, due to the combined direct and traceability channels. The corridor $\mathcal{C}$ (gray rectangle) is bounded below by $\underline{\theta}_A = 0.3$ (dashed blue) and $\underline{\theta}_B = 0.35$ (dashed red). The shaded regions below each threshold indicate the termination zones. Benchmark parameters: $\sigma_A^2 = \sigma_B^2 = 0.04$, $b = \eta/(1-\rho) = 0.15$, $T = 50$.

4.2 Existence

Proposition 3 (Existence of a Double-Agent Equilibrium)**.** *Under Assumptions 1–5, suppose:*

(i) Initial beliefs satisfy $(\theta_{A,0}, \theta_{B,0}) \in \mathcal{C}$.

(ii) The horizon T *is sufficiently long that the information value to* A *exceeds the costs of the arrangement for at least one period.*

Then there exists a PBE in which M *operates as a double agent for at least one period.*

The construction follows backward induction (Appendix B): at $t = T$ effort collapses to zero (Proposition 10), while the positive continuation value sustains interior effort for $t < T$. Under the benchmark, the FOC system has a unique interior solution for generic parameters.

4.3 Transitoriness

Proposition 4 (Structural Transitoriness)**.** *In any PBE of the finite game, the probability that the double-agent arrangement terminates strictly before the horizon T converges to 1 as T grows. Three forces drive collapse:*

(i) ***Belief erosion.*** *Beliefs are martingales, but their strictly positive variance (Proposition 14) combined with exit boundaries ensures eventual corridor exit. The corridor has exit boundaries at $\underline{\theta}_A$ and $\underline{\theta}_B$: each period carries a strictly positive probability of corridor exit, so the probability of surviving all T periods vanishes as T grows.*

(ii) ***Last-period unraveling.*** *As $t \to T$, the continuation value V_{t+1} shrinks, reducing M's incentive to exert costly effort and degrading the intelligence M delivers, causing A's belief $\theta_{A,t}$ to decline. Backward induction propagates this unraveling to earlier periods, shrinking the effective corridor.*

(iii) ***Compounding survival risk.*** *Even inside the corridor ($d_A = d_B = 0$), the family protection probability $q(p_t) < 1$ introduces per-period survival risk. The cumulative survival probability $\prod_{s=1}^{t} q(p_s)$ converges to zero as t grows, making the arrangement increasingly fragile with age. The primary drivers are the belief-erosion channels (i)–(ii); compounding survival risk (iii) is an additional accelerating force.*

The key mechanism behind (i) is *self-destructive experimentation*: B learns about M's type through two compounding channels. The *direct channel* operates through the type-dependence of e_B: since $e_B^*(\omega)$ has slope $-b$, B's signal carries information about ω even at $\gamma = 0$, with loading $h_B = b$. The *traceability channel* amplifies this: effort toward A leaks into B's signal with coefficient $-\gamma b$, raising the total loading to $h_B = b(1 + \gamma)$. The parameter γ amplifies an already-active learning process; at $\gamma = 0$, B still learns from e_B's type-dependence. For A, only the direct channel operates ($h_A = b$), so A's belief erodes more slowly. The formal proof is in Appendix C; Figure 2 illustrates the effect of γ on belief trajectories.

4.4 Comparative Statics

The model yields several testable predictions.[10]

Corollary 5 (Monitoring Technology)**.** *An improvement in monitoring technology (reduction in σ_j^2 for $j \in \{A, B\}$) reduces expected duration. Under the benchmark, corridor boundaries are independent of signal precision, but more precise signals increase belief volatility, accelerating convergence to the boundary.*

[10] Proofs are in Appendix D.

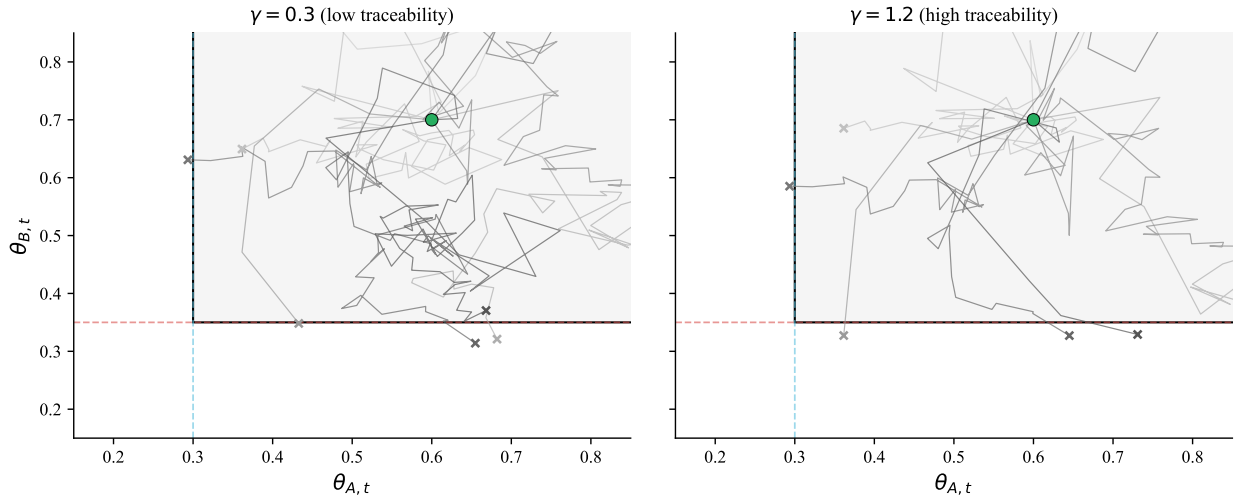


Figure 2: Simulated belief trajectories under low traceability ($\gamma = 0.3$, left) and high traceability ($\gamma = 1.2$, right). Each panel shows 8 trajectories from the same initial beliefs. Under high γ, the signal loading $h_B = b(1+\gamma)$ is larger, amplifying the variance of $\theta_{B,t}$ increments and causing faster absorption at the corridor boundary. Most high-γ trajectories exit through B's threshold (bottom boundary), consistent with the self-destructive experimentation mechanism.

Remark 6 (Punishment Severity)**.** *The shape of d_B below the corridor threshold $\underline{\theta}_B$ does not affect equilibrium behavior, since the continuation value is zero outside the corridor (Section 5.2). The relevant comparative static is a shift in the threshold $\underline{\theta}_B$ itself: a lower threshold (more tolerant B) widens the corridor and extends expected duration, while a higher threshold (less tolerant) narrows it. Once $\underline{\theta}_B$ is fixed, M's equilibrium effort and the corridor dynamics are determined entirely by the belief process inside the corridor, where $d_B = 0$.*

Corollary 7 (Protection Costs)**.** *An increase in the cost of family protection $\phi(p)$ weakly shrinks the corridor from A's side: $\underline{\theta}_{A,t}$ increases, requiring M to deliver higher-quality intelligence to justify A's continued investment. This accelerates collapse.*

4.5 Expected Duration

Expected duration is bounded above. Define $\tau = \min\{t : (\theta_{A,t}, \theta_{B,t}) \notin \mathcal{C}\}$ and $\tau_B = \min\{t : \theta_{B,t} \leq \underline{\theta}_B\}$. Since $\tau \leq \tau_B$, any upper bound on $\mathbb{E}[\tau_B]$ bounds $\mathbb{E}[\tau]$. Since $\theta_{B,t}$ is a martingale with per-period conditional variance $v_{B,t} = h_B^2 P_{B,t}^2/(h_B^2 P_{B,t} + \sigma_B^2)$ where $h_B = b(1+\gamma)$, the optional stopping theorem gives (since $\tau_B \leq T$ is bounded, the OST applies without a uniform integrability condition)

$$\mathbb{E}[\tau] \leq \mathbb{E}[\tau_B] \leq \frac{(\theta_{B,0} - \underline{\theta}_B)^2}{\underline{v}},$$

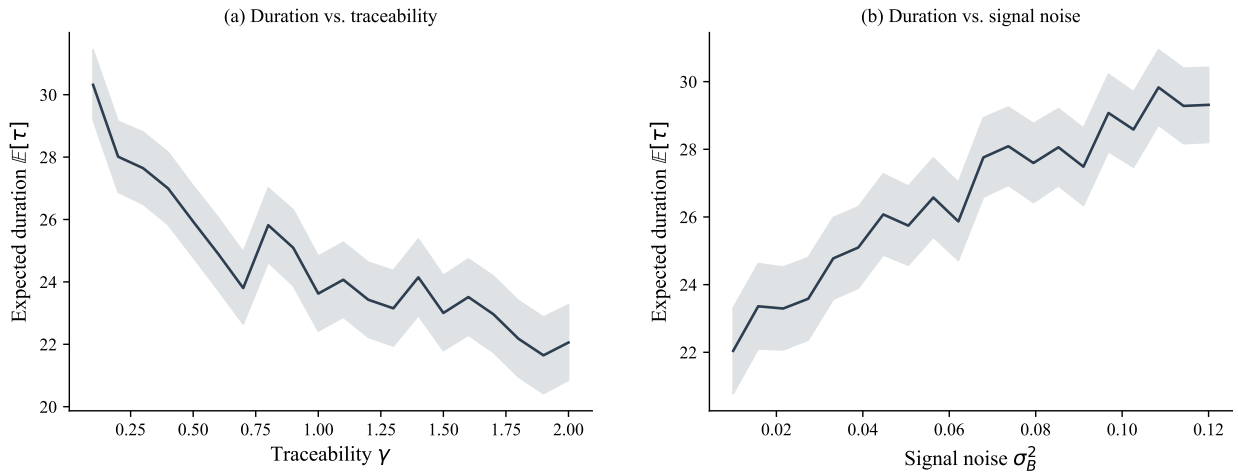


Figure 3: Expected duration as a function of (a) traceability γ and (b) signal noise σ_B^2. Each point is the mean of 1,000 simulated trajectories; shaded bands show 95% confidence intervals. Panel (a) confirms Proposition 8(i): duration is decreasing in γ once the traceability channel dominates (for very small γ, B's signal loading is negligible and the binding constraint is the horizon or A's side). Panel (b) confirms Proposition 8(ii): noisier signals slow belief updating and extend duration.

where $\underline{v} = \inf_t v_{B,t}$ is the minimum per-period variance, attained at the last period since $P_{B,t}$ is decreasing. The bound is most informative for moderate horizons; for very large T, it may exceed T and become vacuous. Under the benchmark, $\underline{v}$ is decreasing in σ_B^2 and increasing in γ.

An analogous A-side bound holds with $h_A = b$; since $h_B = b(1+\gamma) > h_A$ for all $\gamma > 0$, the B-side bound is tighter given equal initial slacks.

The following comparative statics sharpen the duration predictions.

Proposition 8 (Duration Comparative Statics)**.** *Under the benchmark parametrization:*

(i) $\mathbb{E}[\tau]$ *is decreasing in* γ*, the traceability parameter.*

(ii) $\mathbb{E}[\tau]$ *is increasing in* σ_B^2*, the noise in* B*'s signal.*

(iii) $\mathbb{E}[\tau]$ *is increasing in* $(\theta_{B,0} - \underline{\theta}_B)$*, the initial slack in* B*'s continuation condition.*

Proof. All three follow from the duration bound $\mathbb{E}[\tau] \leq (\theta_{B,0} - \underline{\theta}_B)^2/\underline{v}$. (i) An increase in γ raises $h_B = b(1+\gamma)$, which increases $v_{B,t} = h_B^2 P_{B,t}^2/(h_B^2 P_{B,t} + \sigma_B^2)$ and hence $\underline{v}$, reducing the bound. (ii) An increase in σ_B^2 decreases $v_{B,t}$ (for fixed $P_{B,t}$), reducing $\underline{v}$ and increasing the bound. (iii) The bound is quadratic in $(\theta_{B,0} - \underline{\theta}_B)$, so increasing this slack increases expected duration directly. Figure 3 confirms these predictions via Monte Carlo simulation under the benchmark parametrization. □

5 Robustness Beyond the Benchmark

The key qualitative findings from the LQG benchmark—existence, transitoriness, and comparative statics—are robust to relaxing the functional form assumptions. We replace Assumption 5 with two weaker conditions and show that the corridor structure, equilibrium existence, and welfare results carry over. Assumptions 1–4 are maintained throughout. Proofs are in Appendices F–I.

Assumption 6 (Absolute Continuity)**.** *For each $j \in \{A, B\}$ and each $m_j > 0$, the signal distribution under any effort pair is absolutely continuous with respect to the distribution under any alternative effort pair. This is immediately satisfied by the Gaussian signal structure.*

Assumption 6 guarantees Bayes' rule is well-defined at every information set reached with positive probability.

5.1 Principals' Value Functions

Under general specifications, the continuation thresholds and monitoring levels are endogenous equilibrium objects. The recursive value functions implicit in the benchmark analysis are:

Agency A's value function at period t is

$$V_t^A(\theta_A) = \max_{m_A, p} \Big\{ v_A(e^*_{A,t}) - \kappa_A(m_A) - \phi(p) + \delta \cdot \Pi_t \cdot W^A_{t+1}(\theta_A, m_A, p) \Big\}, \tag{7}$$

where W^A_{t+1} is A's expected continuation value conditional on survival. The first-order conditions are:

$$\kappa'_A(m^*_{A,t}) = -\delta \cdot \frac{\partial W^A_{t+1}}{\partial P_{A,t+1}} \cdot \frac{\partial P_{A,t+1}}{\partial \sigma^2_A} \cdot (\sigma^2_A)'(m^*_{A,t}), \tag{8}$$

$$\phi'(p^*_t) = \delta \cdot (1 - d_B(\theta_{B,t}))(1 - d_A(\theta_{A,t})) \cdot q'(p^*_t) \cdot W^A_{t+1}. \tag{9}$$

A continues when $V_t^A(\theta_A) \geq 0$, giving threshold $\underline{\theta}_{A,t} = \min\{\theta_A : V_t^A(\theta_A) \geq 0\}$.

Organization B's value function is

$$V_t^B(\theta_B) = \max_{m_B} \Big\{ v_B - \ell(e^*_{A,t}) - \kappa_B(m_B) + \delta \cdot \Pi^B_t \cdot W^B_{t+1}(\theta_B, m_B) \Big\}, \tag{10}$$

where Π^B_t is the continuation probability from B's perspective. The monitoring FOC is

$$\kappa'_B(m^*_{B,t}) = -\delta \cdot \Pi^B_t \cdot \frac{\partial W^B_{t+1}}{\partial P_{B,t+1}} \cdot \frac{\partial P_{B,t+1}}{\partial \sigma^2_B} \cdot (\sigma^2_B)'(m^*_{B,t}). \tag{11}$$

B continues when $V_t^B(\theta_B) \geq 0$, yielding threshold $\underline{\theta}_{B,t} = \min\{\theta_B : V_t^B(\theta_B) \geq 0\}$.

5.2 The Mole's Dynamic Optimization

The mole knows his own type ω and, under equilibrium play, can track both principals' beliefs (θ_A, θ_B) using his knowledge of the equilibrium strategy profile and the signal structure.[11] The terminal condition is $V_{T+1}(\omega, \theta_A, \theta_B) \equiv 0$. For $t \leq T$ and $(\theta_A, \theta_B) \in \mathcal{C}_t$, the Bellman equation is

$$V_t(\omega, \theta_A, \theta_B) = \max_{(e_A, e_B) \in [0,1]^2} \Big\{ \alpha - c(e_A, e_B;\, \omega) + \delta \cdot \Pi_t(\theta_A, \theta_B) \cdot W_{t+1}(e_A, e_B;\, \omega, \theta_A, \theta_B) \Big\}, \quad (12)$$

where $\Pi_t(\theta_A, \theta_B) = (1 - d_B(\theta_B))(1 - d_A(\theta_A)) \cdot q(p_t^*(\theta_A))$ and W_{t+1} is the expected continuation value conditional on survival within the corridor. The protection level p_t^* is chosen at step 2 (before signals) and thus conditions on the entering-period belief $\theta_{A,t-1}$; by the certainty-equivalence assumption and the deterministic evolution of the Kalman variance, p_t^* can equivalently be expressed as a function of the current-period post-signal belief θ_A in the recursive formulation.[12] For $(\theta_A, \theta_B) \notin \mathcal{C}_t$, $V_t(\theta_A, \theta_B) = 0$.

V_t is continuous on $\mathcal{C}_t$ (Maximum Theorem), weakly increasing in each belief, jointly concave under the benchmark, and zero on the corridor boundary.[13]

The mole's first-order conditions are:

$$\delta \cdot \Pi_t \cdot \frac{\partial W_{t+1}}{\partial e_A} = \frac{\partial c}{\partial e_A}, \quad (13)$$

$$\delta \cdot \Pi_t \cdot \frac{\partial W_{t+1}}{\partial e_B} = \frac{\partial c}{\partial e_B}. \quad (14)$$

Under the benchmark, $\partial W_{t+1}/\partial e_j > 0$ for moderate γ (Proposition 9), ensuring an interior solution.

Proposition 9 (Interior Solution)**.** *For $\alpha > \bar{\alpha}(T)$, the optimal effort $(e_{A,t}^*, e_{B,t}^*)$ is interior for all $t < T$ whenever $(\theta_A, \theta_B) \in \text{int}(\mathcal{C}_t)$.* (Proof in Appendix F.)

[11] M does not observe the principals' signals or monitoring levels directly, but since the Kalman gain and equilibrium effort are common knowledge in a PBE, M can compute the belief dynamics from his known type and the equilibrium mapping.

[12] This equivalence holds exactly under the benchmark because the protection FOC depends on beliefs only through the continuation value, which under certainty equivalence is evaluated at the posterior mean.

[13] Formal statements and proofs are in Appendix F. Under the LQG benchmark, concavity ensures existence via backward induction; for the general case, the Kakutani-Fan-Glicksberg argument requires only upper hemicontinuity.

Proposition 10 (Terminal-Period Unraveling). *At $t = T$, the unique optimal effort is* $(e^*_{A,T}, e^*_{B,T}) = (0, 0)$.[14] (Proof in Appendix B.)

Proposition 11 (Effort Monotonicity in Beliefs). *$e^*_{A,t}$ is weakly increasing in θ_B and $e^*_{B,t}$ is weakly decreasing in θ_B. For θ_A: $e^*_{B,t}$ is weakly increasing in θ_A (safety on A's side frees resources for B), while the sign of $\partial e^*_{A,t}/\partial\theta_A$ depends on the relative strength of the direct continuation-value effect versus the cross-cost rebalancing effect.* (Proof in Appendix F.)

5.3 Belief Dynamics

Assumption 7 (Certainty Equivalence). *The mole optimizes as if each principal's belief equals its conditional expectation given M's history. Under the benchmark parametrization, this is exact for the flow payoff (quadratic cost) and belief updating (Kalman filter), but it is an approximation for the continuation value: the survival probability Π_t involves nonlinear cutoffs d_j and corridor exit boundaries, which break exact certainty equivalence. The approximation is accurate when beliefs are in the interior of the corridor (where $d_j = 0$ and the stopping constraint is slack). For general specifications, this is a maintained modeling assumption; its accuracy improves as posterior variance decreases (Lemma 12). Under the benchmark, the following results are exact: the effort slope $b = \eta/(1-\rho)$, the Kalman recursion, and the martingale property of beliefs. The duration bound, the effort intercept a_t, and the welfare comparison depend on certainty equivalence and are therefore approximate near the corridor boundaries.*

Under the Gaussian signal structure, substituting the affine effort strategies $e^*_A(\omega) = a_t + b\omega$ and $e^*_B(\omega) = a'_t - b\omega$,[15] the signal loadings on ω are:

- A's signal: $\partial s_A/\partial\omega = b$, so $h_A = b$.
- B's signal: $\partial s_B/\partial\omega = d - \gamma b = -b - \gamma b = -b(1+\gamma)$, so $|h_B| = b(1+\gamma)$.

Since $\theta_{B,t} = \mathbb{E}_B[1-\omega]$, the loading on $1-\omega$ is $h_B = b(1+\gamma) > 0$. The posterior variance evolves via

$$P_{j,t+1} = \frac{P_{j,t} \cdot \sigma_j^2(m^*_{j,t})}{h_{j,t}^2 P_{j,t} + \sigma_j^2(m^*_{j,t})}, \tag{15}$$

with Kalman gain $K_{j,t} = h_{j,t} P_{j,t}/(h_{j,t}^2 P_{j,t} + \sigma_j^2)$.

Lemma 12 (Variance Strict Monotonicity). *$P_{j,t+1} < P_{j,t}$ for all t. This follows directly from the recursion: $P_{j,t+1}/P_{j,t} = \sigma_j^2/(h_{j,t}^2 P_{j,t} + \sigma_j^2) < 1$ whenever $h_{j,t} > 0$ and $P_{j,t} > 0$.*

[14] The analog of backward-induction unraveling in finitely repeated games (Benoit and Krishna, 1985); Bayesian learning sustains cooperation in earlier periods through a mechanism distinct from Kreps et al. (1982).

[15] The slope of e^*_B with respect to ω is $d = -b = -\eta/(1-\rho)$; see footnote in Section 3.6.

Proposition 13 (Martingale Characterization). *Under the equilibrium strategy profile, $\{\theta_{j,t}\}_{t=1}^{T}$ is a martingale: $\mathbb{E}[\theta_{j,t+1} \mid \mathcal{F}_t^j] = \theta_{j,t}$ for all t.*

Proposition 14 (Conditional Variance of Belief Increments). *The conditional variance of belief increments is*

$$v_{j,t} \equiv \mathrm{Var}(\theta_{j,t+1} - \theta_{j,t} \mid \mathcal{F}_t^j) = \frac{h_{j,t}^2 P_{j,t}^2}{h_{j,t}^2 P_{j,t} + \sigma_j^2(m_{j,t}^*)} = P_{j,t} - P_{j,t+1}, \tag{16}$$

where $h_A = b$ and $h_B = b(1+\gamma)$. This variance is strictly positive, equals the per-period reduction in posterior variance, and is strictly decreasing in t.

Proposition 15 (Joint Dynamics of (θ_A, θ_B)). *The bilateral belief process satisfies:*

(i) Each marginal $\{\theta_{j,t}\}$ is a martingale (Proposition 13).

(ii) The signal noises $\varepsilon_{A,t}$ and $\varepsilon_{B,t}$ are conditionally independent given principals' information. Unconditionally, *belief increments are correlated because both signals share exposure to the common type through $e_{A,t}^*(\omega)$.*

(iii) Under the benchmark parametrization, the unconditional covariance (integrating over the unknown type) is

$$\mathrm{Cov}(\Delta\theta_{A,t}, \Delta\theta_{B,t}) = -K_{A,t} \cdot K_{B,t} \cdot (1+\gamma) \cdot b^2 \cdot P_0, \tag{17}$$

where $b > 0$ and $P_0 = \mathrm{Var}(\omega)$ is the prior type variance. Since principals do not share signals, the relevant object is the unconditional P_0, not either principal's private posterior.[16] *The magnitude $|K_{A,t}K_{B,t}(1+\gamma)b^2P_0|$ declines over time as Kalman gains shrink. The covariance is strictly negative: when ω is high, A's signal is on average more favorable while B's is less favorable through the γ-leakage channel.*

5.4 Equilibrium Characterization (General Case)

Off-path deviations are attributed to the worst type: $\theta_{j,t}^{\mathrm{off}} = 0$, triggering termination.[17]

Theorem 16 (Existence of Equilibrium). *Under Assumptions 1–4 and 6–7, suppose:*

(i) Both A and B can credibly commit to punishment.

(ii) Initial beliefs satisfy $(\theta_{A,0}, \theta_{B,0}) \in \mathcal{C}_0$.

(iii) The horizon T is finite and $\delta \in (0,1)$.

[16] This covariance is a model-level property from the perspective of an external observer; it is not decision-relevant to either A or B.

[17] The standard revelation principle fails in this adversarial common agency setting; cf. (Peters, 2001; Epstein and Peters, 1999).

(iv) There exists $(e_A, e_B) \in (0,1)^2$ *such that* M *can keep both beliefs above threshold in expectation for at least one period.*

Then there exists a PBE (σ^*, μ^*) *in which* M *operates as a double agent for at least one period.*

Under the LQG benchmark, existence follows by backward induction on a concave dynamic program. For the general case, we apply the Kakutani-Fan-Glicksberg fixed point theorem to the joint best-response correspondence.[18]

Theorem 17 (Local Uniqueness of FOC Solutions)**.** *For generic parameter vectors*

$$(\alpha, \gamma, \delta, \bar{\sigma}_A^2, \bar{\sigma}_B^2) \in \mathbb{R}^5_{++}$$

(all except a closed set of Lebesgue measure zero), the FOC system has locally unique interior solutions. Under the LQG benchmark, this yields a unique interior PBE effort profile; for general specifications, multiple equilibria may coexist under extreme parameters (Proposition 18).

The result follows from the parametric transversality theorem applied to the smooth FOC system.[19]

Proposition 18 (Multiplicity Under Extreme Parameters)**.** *When* γ *is sufficiently large and* α *is near* $\bar{\alpha}(T)$*, multiple equilibria coexist: an "aggressive informing" equilibrium (high* e_A^**, low* e_B^**) and a "cautious dual-loyalty" equilibrium (low* e_A^**, high* e_B^**).* (Proof in Appendix H.)

5.5 Welfare Analysis

Let $\tau = \min\{t : (\theta_{A,t}, \theta_{B,t}) \notin \mathcal{C}_t\}$ denote the first exit time from the corridor.

[18] The proof in Appendix H equips the strategy space with the weak-$*$ topology (compact by Banach-Alaoglu), verifies best responses are non-empty, convex-valued, and upper hemicontinuous, and constructs beliefs via Bayes' rule.

[19] The FOC system defines $F : (0,1)^2 \times \mathbb{R}^5_{++} \to \mathbb{R}^2$; by the transversality theorem (Mas-Colell, 1985, Chapter 8), Debreu (1970), solutions are generically isolated. See Appendix H.

Proposition 19 (Expected Equilibrium Payoffs)**.** *The expected equilibrium payoffs are:*

$$\mathbb{E}[U_M(\sigma^*)] = (\alpha - c^*) \cdot \frac{1 - \mathbb{E}[\delta^\tau]}{1 - \delta}, \tag{18}$$

$$\mathbb{E}[U_A(\sigma^*)] = \sum_{t=1}^{T} \delta^{t-1} \Big[\big(v_A(e^*_{A,t}) - \kappa_A(m^*_{A,t}) - \phi(p^*_t) \big) \cdot \mathbb{P}(\tau > t) \Big], \tag{19}$$

$$\mathbb{E}[U_B(\sigma^*)] = \sum_{t=1}^{T} \delta^{t-1} \Big[\big(v_B - \ell(e^*_{A,t}) - \kappa_B(m^*_{B,t}) \big) \cdot \mathbb{P}(\tau > t) \Big], \tag{20}$$

where $c^* \approx c(e^*_A, e^*_B; \omega)$ *is the (approximately constant) equilibrium effort cost, and* $\mathbb{P}(\tau > t) = \mathbb{E}[\lambda_t] = \mathbb{E}\big[\prod_{s=1}^{t-1} \Pi_s\big]$ *accounts for belief-contingent termination and protection failure risk. The general formula is* $\mathbb{E}[U_M] = \sum_{t=1}^{T} \delta^{t-1} \mathbb{E}[\lambda_t(\alpha - c(e^*_{A,t}, e^*_{B,t}; \omega))]$*; equation (18) is its benchmark simplification.* (Proof in Appendix I.)

Remark 20 (Deadweight Monitoring Cost)**.** *Imperfect observability of effort generates positive monitoring expenditure:* $\kappa_A(m^*_{A,t}) + \kappa_B(m^*_{B,t}) > 0 = \kappa_A(0) + \kappa_B(0)$*, where* $(m^{FB}_A, m^{FB}_B) = (0, 0)$ *under observable effort.*

Remark 21 (Institutional Cost of Coercion)**.** *Consider a voluntary benchmark in which A achieves the same effort profile through incentive payments—M delivers intelligence voluntarily, eliminating family protection (*$\phi = 0$*) and reducing monitoring to observable-effort levels. Holding the effort profile fixed, the primary surplus terms cancel and the welfare loss is*

$$\Delta S = S^{vol} - S(\sigma^*) = \mathbb{E}\Bigg[\sum_{t=1}^{T} \delta^{t-1} \lambda_t \cdot \big[\phi(p^*_t) + \kappa_A(m^*_{A,t}) - \kappa_A(m^{vol}_{A,t}) + \kappa_B(m^*_{B,t}) - \kappa_B(m^{vol}_{B,t}) \big] \Bigg] > 0. \tag{21}$$

This is a partial-equilibrium comparison holding effort fixed; see Section 7.3 for the scope of this exercise.

6 Historical Interpretation

The theoretical results map onto documented patterns in intelligence history; a tabulation appears in Appendix A.

Kim Philby's three decades inside MI6 illustrate the wide-corridor regime. Institutional trust reduced effective monitoring (σ^2_B was large), class-based deference biased prior beliefs

($\theta_{B,0}$ was high), and bureaucratic inertia slowed updating (Macintyre, 2014). Aldrich Ames exhibits a parallel pattern within the CIA (Andrew and Mitrokhin, 1999).

Oleg Penkovsky illustrates the narrow-corridor regime. His behavioral inconsistencies accumulated quickly enough that Soviet counterintelligence beliefs fell below $\underline{\theta}_B$ within roughly eighteen months (Schecter and Deriabin, 1992). High traceability (γ was large) and intense Soviet monitoring made belief erosion swift.[20] Philby's decades versus Penkovsky's months maps directly onto these primitives: Philby operated under low γ and high $\bar{\sigma}_B^2$, while Penkovsky faced high γ and intense surveillance.

Robert Hanssen's case at the FBI (1979–2001) ended when advances in counterintelligence methods accelerated belief updating; the Webster Commission report (Commission for the Review of FBI Security Programs, 2002) documents how monitoring improvements increased belief volatility and shortened the expected time to corridor exit.

Juan Pujol García ("Garbo") actively managed the information flowing to each side, fabricating a network of fictitious sub-agents to control the signal structure (Talty, 2012). He effectively altered the signal structure itself—beyond the model's formal action space—but the effort choice (e_A, e_B) captures the core trade-off he navigated.

Oleg Gordievsky was exfiltrated by MI6 when the corridor shrank to near-zero after a tip from Aldrich Ames (Gordievsky, 1995). His case suggests an extension: when beliefs approach the corridor boundary, A may find it optimal to extract M rather than allow collapse, introducing an endogenous exit option.[21]

7 Discussion

7.1 General Observations

The bilateral model captures the core mechanism in its simplest form. Recruitment mode systematically affects initial conditions and structural parameters, generating testable predictions about duration and collapse mode (Section 6).

[20] Penkovsky approached Western intelligence voluntarily. We treat ideological recruitment as initiation into Agency A's service: once he began passing intelligence, the operational structure became functionally identical to a coerced arrangement. The key difference is that $\theta_{A,0}$ is high, widening the slack on A's side while leaving B's detection channel unchanged.

[21] Ashraf Marwan ("The Angel") illustrates a further subtlety: the label "double agent" can be endogenous to beliefs. Whether Marwan fed disinformation to Israel or was an Israeli asset remains disputed (Bar-Joseph, 2016)—consistent with the model's prediction that when the corridor is wide, outside observers cannot determine which side M truly serves.

Adding agencies rapidly destroys the arrangement: each contributes an independent detection channel, and corridor volume decays exponentially (Theorem 22), explaining why double agents almost always serve exactly two sides.

Replace the single agency A with $N \geq 2$ rival agencies $A_1, \ldots, A_N$. M chooses effort vector $\mathbf{e}_t = (e_{A_1,t}, \ldots, e_{A_N,t}, e_{B,t}) \in [0,1]^{N+1}$. Organization B observes

$$s_{B,t} = e_{B,t} - \sum_{i=1}^{N} \gamma_i \cdot e_{A_i,t} + \varepsilon_{B,t}, \tag{22}$$

and the corridor becomes a hyperrectangle $\mathcal{C}_t^N = \{(\theta_{A_1}, \ldots, \theta_{A_N}, \theta_B) : \theta_{A_i} \geq \underline{\theta}_{A_i,t},\ \theta_B \geq \underline{\theta}_{B,t}\}$ with survival probability $\Pi_t^N = (1 - d_B(\theta_B)) \prod_{i=1}^{N} \left[(1 - d_{A_i}(\theta_{A_i})) \cdot q_i(p_{i,t})\right]$.

Theorem 22 (Fragility Scaling)**.** *Under symmetry ($\gamma_i = \gamma$, $\sigma_{A_i}^2 = \sigma_A^2$, $\underline{\theta}_{A_i} = \underline{\theta}_A$ for all i), independence ($\varepsilon_{A_i,t} \perp\!\!\!\perp \varepsilon_{A_j,t}$ for $i \neq j$), and compact type support $\omega \in [\underline{\omega}, \bar{\omega}]$:*

(i) The corridor measure (Lebesgue volume) decays exponentially: $\mu_L(\mathcal{C}_t^N) = (1 - \underline{\theta}_A)^N \cdot (1 - \underline{\theta}_B)$.

(ii) The expected duration decays exponentially:

$$\mathbb{E}[\tau^N] - 1 \leq (T-1) \cdot e^{-N p_{\min}}, \tag{23}$$

where $p_{\min} = \inf_\omega \mathbb{P}(\tau_1 \leq 1 \mid \omega) > 0$ is the worst-case (over types) first-period exit probability for a single agency. In particular, $\mathbb{E}[\tau^N] \to 1$ as $N \to \infty$.

Proof sketch. Conditioning on the common type ω, the exit times τ_{A_i} are conditionally independent (since the signal noises are independent). For any ω, the per-period exit probability $p_1(\omega) > 0$ is bounded below by $p_{\min} > 0$. Then $\mathbb{P}(\tau^N > t \mid \omega) \leq [1 - p_{\min}]^N$, and integrating over ω gives $\mathbb{P}(\tau^N > t) \leq [1 - p_{\min}]^N$. Summing: $\mathbb{E}[\tau^N] \leq 1 + (T-1)e^{-N p_{\min}}$. Full proof in Appendix H. □

Remark 23. *The exponential scaling breaks down under correlated signals: if agencies share intelligence, additional constraints are not independent and the corridor shrinks more slowly.*

7.2 Applications Beyond Espionage

The trilateral structure extends to any setting where an individual serves two principals with conflicting objectives under imperfect monitoring and existential stakes. A corporate whistleblower approached by a regulator faces the same structure: the regulator (A) coerces cooperation through the threat of prosecution, while the firm (B) threatens termination and retaliation. The whistleblower must gather evidence while appearing loyal—a corridor

sustained by imperfect monitoring on both sides. The Dodd-Frank Act's whistleblower provisions widen this corridor by reducing A's abandonment threshold and improving $q(p)$. Undercover law enforcement, confidential informants, and protected witnesses share the same dynamics: bilateral monitoring, survival-contingent payoffs, and the tension between serving one side and maintaining cover with the other. In each domain, duration increases in signal noise and initial trust and decreases in monitoring technology and traceability. Formal application requires adapting the signal structure and participation constraints to each institutional context.

7.3 Limitations

Several modeling choices constrain the scope of the results. The certainty-equivalence assumption is exact for the flow payoff and belief updating under the LQG benchmark but is an approximation for the continuation value near the corridor boundaries, where nonlinear cutoffs break exact certainty equivalence. The approximation is accurate in the corridor interior and improves as posterior variance decreases (Lemma 12).

The benchmark corridor is a fixed rectangle under the stationarity condition $a_t = a$, which holds away from the terminal period. Near $t = T$, terminal unraveling drives $a_t \to 0$ and thresholds become time-varying; the fully endogenous thresholds are in Section 5.1.

The duration bound $(\theta_{B,0} - \underline{\theta}_B)^2/\underline{v}$ is most informative for moderate horizons; it may exceed T for very long horizons and become vacuous.

The coercion assumption requires M's participation to be involuntary, fitting cases of blackmail, capture, or family threats. For ideologically motivated agents such as Penkovsky, the predictions apply primarily through B's detection channel; the comparative statics on γ and σ_B^2 remain relevant even when A's abandonment threshold is effectively zero.

The model assumes a one-dimensional loyalty type ω. A multi-dimensional type would require a multivariate Kalman filter and a higher-dimensional corridor; we conjecture the exit-boundary dynamics are qualitatively similar, but verification requires extending the martingale exit-time argument to multiple correlated belief processes. The welfare comparison (Remark 21) holds the effort profile fixed; a full comparison would require specifying the voluntary alternative explicitly.

8 Conclusion

This paper formalizes the double-agent problem as a trilateral game of antagonistic common agency under bilateral coercion. Three results emerge from the LQG benchmark. First,

the *corridor of survival*—a fixed rectangle in bilateral belief space—captures how double agents persist: both principals continue only while their beliefs remain above endogenous thresholds. Second, the equilibrium is *structurally transitory*: Bayesian updating, terminal-period unraveling, and compounding survival risk ensure eventual collapse. Third, the *self-destructive experimentation* dynamic gives the model its distinctive predictive content.

The comparative statics are consistent with historical patterns: duration increases with signal noise and initial trust and decreases with monitoring technology and traceability. The linear-quadratic-Gaussian specification provides closed-form characterizations of the corridor geometry and optimal effort allocation, as well as analytical upper bounds on expected duration. Robustness results confirm that equilibrium existence, local uniqueness, and positive welfare losses from coercion all extend beyond the benchmark. The N-agency extension shows expected duration decays exponentially in the number of rival agencies.

Several directions merit future investigation. Formalizing endogenous recruitment would enrich the model's predictive content. Strategic exit (exfiltration as an alternative to collapse) modifies the equilibrium near the corridor boundary. A continuous-time formulation, connecting to Sannikov (2007), Sannikov (2008), and Faingold and Sannikov (2011), would yield sharper characterizations of belief dynamics. Calibrating the model to declassified operational data remains an open empirical challenge.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti, "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 1990, *58* (5), 1041–1063.

Acemoglu, Daron and Alexander Wolitzky, "The Economics of Labor Coercion," *Econometrica*, 2011, *79* (2), 555–600.

Aliprantis, Charalambos D. and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed., Berlin: Springer, 2006.

Andrew, Christopher and Vasili Mitrokhin, *The Sword and the Shield: The Mitrokhin Archive and the Secret History of the KGB*, New York: Basic Books, 1999.

Baccara, Mariagiovanna and Heski Bar-Isaac, "How to Organize Crime," *Review of Economic Studies*, 2008, *75* (4), 1039–1067.

Bar-Joseph, Uri, *The Angel: The Egyptian Spy Who Saved Israel*, New York: Harper, 2016.

Benoit, Jean-Pierre and Vijay Krishna, "Finitely Repeated Games," *Econometrica*, 1985, *53* (4), 905–922.

Bernheim, B. Douglas and Michael D. Whinston, "Common Agency," *Econometrica*, 1986, *54* (4), 923–942.

Board, Simon and Moritz Meyer ter Vehn, "Reputation for Quality," *Econometrica*, 2013, *81* (6), 2381–2462.

Commission for the Review of FBI Security Programs, *A Review of FBI Security Programs*, Washington, DC: U.S. Department of Justice, 2002. (Webster Commission Report).

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Cripps, Martin W., George J. Mailath, and Larry Samuelson, "Imperfect Monitoring and Impermanent Reputations," *Econometrica*, 2004, *72* (2), 407–432.

Debreu, Gerard, "Economies with a Finite Set of Equilibria," *Econometrica*, 1970, *38* (3), 387–392.

Doval, Laura and Jeffrey C. Ely, "Sequential Information Design," *Econometrica*, 2020, *88* (6), 2575–2608.

Ely, Jeffrey C. and Juuso Välimäki, "Bad Reputation," *Quarterly Journal of Economics*, 2003, *118* (3), 785–814.

Epstein, Larry G. and Michael Peters, "A Revelation Principle for Competing Mechanisms," *Journal of Economic Theory*, 1999, *88* (1), 119–160.

Faingold, Eduardo and Yuliy Sannikov, "Reputation in Continuous-Time Games," *Econometrica*, 2011, *79* (3), 773–876.

Fudenberg, Drew and David K. Levine, "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, 1989, *57* (4), 759–778.

Gordievsky, Oleg, *Next Stop Execution: The Autobiography of Oleg Gordievsky*, London: Macmillan, 1995.

Green, Edward J. and Robert H. Porter, "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 1984, *52* (1), 87–100.

Holmström, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 1999, *66* (1), 169–182.

Kamenica, Emir and Matthew Gentzkow, "Bayesian Persuasion," *American Economic Review*, 2011, *101* (6), 2590–2615.

Keller, Godfrey, Sven Rady, and Martin Cripps, "Strategic Experimentation with Exponential Bandits," *Econometrica*, 2005, *73* (1), 39–68.

Kreps, David M. and Robert Wilson, "Reputation and Imperfect Information," *Journal of Economic Theory*, 1982, *27* (2), 253–279.

_ , Paul Milgrom, John Roberts, and Robert Wilson, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 1982, *27* (2), 245–252.

Macintyre, Ben, *A Spy Among Friends: Kim Philby and the Great Betrayal*, New York: Crown, 2014.

Mailath, George J. and Larry Samuelson, *Repeated Games and Reputations: Long-Run Relationships*, New York: Oxford University Press, 2006.

Martimort, David and Lars A. Stole, "The Revelation and Delegation Principles in Common Agency Games," *Econometrica*, 2002, *70* (4), 1659–1673.

Mas-Colell, Andreu, *The Theory of General Economic Equilibrium: A Differentiable Approach*, Cambridge: Cambridge University Press, 1985.

Milgrom, Paul and Chris Shannon, "Monotone Comparative Statics," *Econometrica*, 1994, *62* (1), 157–180.

_ **and Ilya Segal**, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 2002, *70* (2), 583–601.

_ **and John Roberts**, "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory*, 1982, *27* (2), 280–312.

Pavan, Alessandro, Ilya Segal, and Juuso Toikka, "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 2014, *82* (2), 601–653.

Peskir, Goran and Albert Shiryaev, *Optimal Stopping and Free-Boundary Problems*, Basel: Birkhäuser, 2006.

Peters, Michael, "Common Agency and the Revelation Principle," *Econometrica*, 2001, *69* (5), 1349–1372.

Sannikov, Yuliy, "Games with Imperfectly Observable Actions in Continuous Time," *Econometrica*, 2007, *75* (5), 1285–1329.

_ , "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies*, 2008, *75* (3), 957–984.

Schecter, Jerrold L. and Peter S. Deriabin, *The Spy Who Saved the World: How a Soviet Colonel Changed the Course of the Cold War*, New York: Charles Scribner's Sons, 1992.

Schelling, Thomas C., *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960.

Sobel, Joel, "A Theory of Credibility," *Review of Economic Studies*, 1985, *52* (4), 557–573.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott, *Recursive Methods in Economic Dynamics*, Cambridge, MA: Harvard University Press, 1989.

Talty, Stephan, *Agent Garbo: The Brilliant, Eccentric Secret Agent Who Tricked Hitler and Saved D-Day*, Boston: Houghton Mifflin Harcourt, 2012.

Wise, David, *Spy: The Inside Story of How the FBI's Robert Hanssen Betrayed America*, New York: Random House, 2002.

A Historical Double-Agent Operations

Agent	Period	Handler / Target	Duration	Collapse Mode	Model Reading
Kim Philby	1934–1963	KGB / MI6	∼29 yrs	Defection under suspicion	Wide corridor: high σ_B^2 (noisy monitoring), high $\theta_{B,0}$, institutional trust
Aldrich Ames	1985–1994	KGB / CIA	∼9 yrs	Arrested (mole hunt)	Corridor narrowed by improved CI methods; $\sigma_B^2 \downarrow$ over time
Robert Hanssen	1979–2001	KGB-SVR / FBI	∼22 yrs	Arrested (surveillance)	Long duration from compartmentalized access; eventual monitoring breakthrough
Oleg Penkovsky	1961–1962	CIA, MI6 / GRU	∼18 mo	Arrested, executed	Narrow corridor: high γ, intense Soviet monitoring; rapid belief erosion
Juan Pujol García ("Garbo")	1941–1944	MI5 / Abwehr	∼3 yrs	War ended	Active signal management; M shaped (s_A, s_B) through fabricated sub-agents
Oleg Gordievsky	1974–1985	MI6 / KGB	∼11 yrs	Exfiltrated by MI6	Corridor collapse $\rightarrow$ extraction as endogenous exit option

Agent	Period	Handler / Target	Duration	Collapse Mode	Model Reading
Ashraf Marwan ("The Angel")	1970–?	Mossad? / Egypt?	Disputed	Died 2007 (disputed)	Label endogenous to beliefs: each side claims M as their own

Sources: Macintyre (2014), Andrew and Mitrokhin (1999), Wise (2002), Schecter and Deriabin (1992), Talty (2012), Gordievsky (1995), Bar-Joseph (2016), Commission for the Review of FBI Security Programs (2002).

B Proof of Proposition 3 (Existence)

We construct a PBE by backward induction on $t = T, T-1, \ldots, 1$.

Terminal period ($t = T$). With $V_{T+1} \equiv 0$, M's problem reduces to

$$\max_{(e_A, e_B) \in [0,1]^2} \alpha - c(e_A, e_B).$$

The survival probability depends on current beliefs through d_A and d_B, but not on M's effort—effort affects only future beliefs, and there is no future at $t = T$. Since $c(0, 0; \omega) = 0$ and c is strictly increasing in each argument, the unique optimal effort is

$$(e^*_{A,T}, e^*_{B,T}) = (0, 0). \tag{24}$$

This is the backward-induction unraveling analog (Benoit and Krishna, 1985): without continuation value, effort is purely costly. For A: $\max_{m_A, p}\ v_A(0) - \kappa_A(m_A) - \phi(p)$ yields $m^*_{A,T} = 0$ and $p^*_T = 0$; similarly $m^*_{B,T} = 0$. Both continue if beliefs remain in the corridor (condition (ii)).

Inductive step ($t < T$). Suppose equilibrium strategies and beliefs have been constructed for periods $t+1, \ldots, T$ with $V_{t+1} > 0$ on the corridor interior. In period t, M's Bellman equation is

$$V_t(\theta_A, \theta_B) = \max_{(e_A, e_B) \in [0,1]^2} \alpha - c(e_A, e_B) + \delta \cdot \Pi_t(\theta_A, \theta_B) \cdot W_{t+1}(e_A, e_B;\, \theta_A, \theta_B),$$

where $\Pi_t = (1 - d_B(\theta_B))(1 - d_A(\theta_A)) \cdot q(p_t^*)$. The first-order conditions are

$$\delta \cdot \Pi_t \cdot \frac{\partial W_{t+1}}{\partial e_A} = \frac{\partial c}{\partial e_A}, \tag{25}$$

$$\delta \cdot \Pi_t \cdot \frac{\partial W_{t+1}}{\partial e_B} = \frac{\partial c}{\partial e_B}. \tag{26}$$

At $(0,0)$, the marginal cost $\partial c/\partial e_A|_{(0,0)} = \eta(1-\omega)$ is positive but small for high-ω types, while the marginal benefit $\delta\Pi_t \cdot (\partial W_{t+1}/\partial e_A)$ is strictly positive since $V_{t+1} > 0$. For $\alpha > \bar{\alpha}(T)$, the marginal benefit exceeds $\eta(1-\omega)$ at the origin for all types. At $e_A = 1$, the marginal cost dominates by convexity. The intermediate value theorem delivers an interior solution. Continuity of V_{t+1} (Maximum Theorem) and positive definiteness of $D^2\tilde{c}$ yield a unique optimal pair $(e_{A,t}^*, e_{B,t}^*)$.

Beliefs are martingales on the equilibrium path (Proposition 13). The positive continuation value deters off-path deviations via the worst-type belief $\theta_{j,t}^{\text{off}} = 0$. The linear cutoff structure makes $\underline{\theta}_A$ and $\underline{\theta}_B$ constants; for $\theta_{j,0} > \underline{\theta}_j$, the positive martingale variance and positive continuation value sustain the equilibrium for at least one period. A and B's best responses are as in the terminal period; both continue if beliefs remain in the corridor.

Belief consistency. The type-dependent cost makes M's optimal effort affine in ω: $e_{A,t}^*(\omega) = a_t + b\omega$, so $s_{A,t} = a_t + b\omega + \varepsilon_{A,t}$ is a Gaussian signal about ω with known coefficients. The Kalman filter applies exactly and $\theta_{j,t}$ is a sufficient statistic for j's belief. Off path, beliefs satisfy sequential rationality.

This completes the construction. □

C Proof of Proposition 4 (Structural Transitoriness)

We prove each part separately.

Part (i): Belief erosion. Fix $j \in \{A, B\}$. Under Bayesian updating, $\theta_{j,t+1} = \theta_{j,t} + \beta_{j,t}(s_{j,t} - \mathbb{E}[s_{j,t} \mid \theta_{j,t}])$, where $\beta_{j,t} > 0$ is the Kalman gain. The innovation has mean zero (since j correctly anticipates M's equilibrium effort), so $\{\theta_{j,t}\}$ is a martingale.

The corridor imposes exit boundaries: if $\theta_{j,t}$ falls below $\underline{\theta}_{j,t}$, the game terminates. Since beliefs are martingales with strictly positive per-period variance $v_{j,s} > 0$ (Proposition 14), each period carries a positive probability of corridor exit. Define $p_s = \mathbb{P}(\theta_{j,s+1} \leq \underline{\theta}_j \mid$

$\theta_{j,s} > \underline{\theta}_j) > 0$. Then $\mathbb{P}(\tau_j > T) = \prod_{s=1}^{T}(1-p_s) \to 0$ as $T \to \infty$. The OST applied to $M_t = (\theta_{j,\tau_j \wedge t} - \underline{\theta}_j)^2 - \sum_{s<\tau_j \wedge t} v_{j,s}$ gives $\mathbb{E}[\tau_j] \cdot \underline{v} \leq (\theta_{j,0} - \underline{\theta}_j)^2$.[22]

The *self-destructive experimentation* dynamic: B's signal loading $h_B = b(1+\gamma)$ combines the direct effort channel (e_B depends on ω with slope $-b$) and the traceability channel (e_A leaks into s_B with coefficient $-\gamma b$). Higher γ amplifies h_B, increasing per-period variance of $\theta_{B,t}$ and accelerating absorption. See Section 4.5 for the duration bound.

Part (ii): Last-period unraveling. At $t = T$, $V_{T+1} = 0$, so $(e^*_{A,T}, e^*_{B,T}) = (0,0)$. For $t = T-1$, the positive $V_T > 0$ generates interior effort; but zero effort at T reduces A's expected signal there, causing A to reduce protection at $T-1$ and further lowering M's survival probability. By backward induction, this propagates to earlier periods. The implicit function theorem on the FOCs gives $\partial e^*_{A,t}/\partial V_{t+1} > 0$; since V_{t+1} decreases as $t \to T$, effort declines as the horizon approaches.

Part (iii): Compounding survival risk. Inside the corridor, $d_B = d_A = 0$, so termination risk from beliefs is zero. But $q(p_s) < 1$ for any finite protection level, so the cumulative survival probability $\prod_{s=1}^{t-1} q(p_s)$ converges to zero as $t \to \infty$. Near the corridor boundary, the additional risk from $d_j > 0$ further accelerates collapse. □

D Proofs of Corollary 5, Remark 6, and Corollary 7

Corollary 5 (Monitoring Technology). A reduction in σ_j^2 increases the Kalman gain, causing more aggressive belief updating. Since $\partial v_{j,t}/\partial \sigma_j^2 < 0$, lower σ_j^2 increases belief volatility; corridor boundaries are σ_j^2-independent (Proposition 2), so expected duration decreases. □

Remark 6 (Punishment Severity). Inside the corridor, $d_B = 0$, so the shape of d_B below $\underline{\theta}_B$ does not enter W_{t+1}. The relevant comparative static is a shift in $\underline{\theta}_B$ itself. □

Corollary 7 (Protection Costs). An increase in ϕ raises the cost of A's continuation. The threshold satisfies $\mathbb{E}[v_A(e_A) \mid \underline{\theta}_{A,t}] = \kappa_A(m^*_A) + \phi(p^*)$, so the implicit function theorem gives $\partial \underline{\theta}_{A,t}/\partial \phi > 0$: A's continuation condition becomes more stringent, shrinking the corridor and accelerating collapse. □

[22] The total quadratic variation $\sum_{s=1}^{T} v_{j,s} \leq P_{j,0}$ is bounded by the prior variance, so the "divergent quadratic variation" argument does not apply. Transitoriness relies instead on the per-period exit probability argument and the compounding protection failure channel.

E Benchmark Corridor and Computations

Proof of Proposition 2

Under Assumption 5, the FOCs become

$$\delta \cdot \Pi_t \cdot \frac{\partial W_{t+1}}{\partial e_A} = e_A + \rho\, e_B + \eta(1-\omega),$$
$$\delta \cdot \Pi_t \cdot \frac{\partial W_{t+1}}{\partial e_B} = e_B + \rho\, e_A + \eta\omega.$$

With linear cutoffs and quadratic costs, the system yields closed-form effort expressions with key features:

(i) e_A^* is increasing in ν (intelligence value) and decreasing in ℓ_0 (leakage loss) and γ (traceability).

(ii) e_B^* is decreasing in e_A^*: higher effort toward A necessarily reduces apparent effort toward B.

The corridor boundaries: A continues when $\nu(a+b\theta_A) \geq \kappa_A^* + \phi^*$, giving $\underline{\theta}_A = (\kappa_A^* + \phi^* - \nu a)/(\nu b)$; B continues when $\ell_0 b(1-\theta_B) \leq \bar{\ell}$, giving $\underline{\theta}_B = 1 - \bar{\ell}/(\ell_0 b)$. Both are constants under the benchmark, so $\mathcal{C}$ is a fixed rectangle. Comparative statics by inspection:

- $\underline{\theta}_A$ is increasing in κ_A and ϕ (higher costs tighten A's threshold).
- $\underline{\theta}_B$ is increasing in ℓ_0 (more sensitive to leakage $\Rightarrow$ lower tolerance).
- Signal precision (higher $m_j/\bar{\sigma}_j^2$) does not affect the static corridor boundaries but increases the speed of belief updating, reducing expected duration.
- Higher γ raises the signal-to-noise ratio in B's monitoring, accelerating belief erosion and reducing expected duration.
- Higher ρ raises $b = \eta/(1-\rho)$, which tightens $\underline{\theta}_B$ upward (narrowing the corridor) and accelerates belief erosion, reducing expected duration.

The duration bound follows from the optional stopping theorem applied to the belief martingale; see below. □

Benchmark Parametrization

(i) Linear cutoffs: $d_B(\theta_B) = \max\{0, 1-\theta_B/\underline{\theta}_B\}$ and $d_A(\theta_A) = \max\{0, 1-\theta_A/\underline{\theta}_A\}$.

(ii) Quadratic costs with type-dependent term: $c(e_A, e_B; \omega) = \frac{1}{2}(e_A^2 + e_B^2 + 2\rho\, e_A e_B) + \eta[(1-\omega)e_A + \omega e_B]$.

(iii) Fixed variances: $\sigma_A^2(m_A) = \bar{\sigma}_A^2/(1+m_A)$ and $\sigma_B^2(m_B) = \bar{\sigma}_B^2/(1+m_B)$.

(iv) Linear intelligence value and leakage: $v_A(e_A) = \nu \cdot e_A$ and $\ell(e_A) = \ell_0 \cdot e_A$.

$T = 2$ Closed-Form Solution

Period $t = 2$ (terminal). By Proposition 10, $e^*_{A,2} = e^*_{B,2} = 0$.

Period $t = 1$ (initial). Since $V_2(\theta_A, \theta_B) = \alpha$ for $(\theta_A, \theta_B) \in \mathcal{C}_2$, the continuation value is $W_2 = \alpha \cdot \mathbb{P}((\theta_{A,2}, \theta_{B,2}) \in \mathcal{C}_2 \mid e_A, e_B)$. Under Gaussian updating and linear cutoffs:

$$W_2(e_A, e_B) \approx \alpha \cdot q(p_1^*) \cdot \Phi\Big(\frac{\theta_{A,1} - \underline{\theta}_A}{\sqrt{v_{A,1}}}\Big) \cdot \Phi\Big(\frac{\theta_{B,1} - \underline{\theta}_B}{\sqrt{v_{B,1}}}\Big),$$

where $v_{j,1}$ is the period-1 belief increment variance (Proposition 14), $q(p_1^*)$ is the equilibrium protection probability, and Φ is the standard normal CDF. The product form treats $\theta_{A,2}$ and $\theta_{B,2}$ as approximately independent; the negative correlation established in Proposition 15(iii) is second-order for moderate γ.

Expected Duration Bounds

The expected duration satisfies

$$\mathbb{E}[\tau] \le \mathbb{E}[\tau_B] \le \frac{(\theta_{B,0} - \underline{\theta}_B)^2}{\underline{v}}, \tag{27}$$

where $\underline{v} = \inf_t v_{B,t}$. The bound follows from the OST applied to $M_t = (\theta_{B,\tau_B \wedge t} - \underline{\theta}_B)^2 - \sum_{s=0}^{(\tau_B \wedge t)-1} v_{B,s}$. Since $\tau_B \le T$ is bounded, $\mathbb{E}[M_{\tau_B}] = M_0$, giving $\mathbb{E}[\sum_s v_{B,s}] \le (\theta_{B,0} - \underline{\theta}_B)^2$. Bounding $v_{B,s} \ge \underline{v}$ yields the result (Peskir and Shiryaev, 2006).

F Proofs for Section 5.2: The Mole's Dynamic Optimization

The value function properties follow standard arguments: continuity from the Maximum Theorem (Stokey et al., 1989), monotonicity from monotone comparative statics (Milgrom and Shannon, 1994), concavity by induction using the envelope theorem (Milgrom and Segal, 2002), and boundary behavior from the corridor definition.

Proof of Proposition 9 (Interior Solution)

Fix $t < T$ and $(\theta_A, \theta_B) \in \text{int}(\mathcal{C}_t)$.

At $e_A = 0$: The marginal benefit $\delta \Pi_t \cdot (\partial W_{t+1}/\partial e_A)|_{e_A=0} > 0$ while the marginal cost $\eta(1-\omega)$ is bounded ($D\tilde{c}(0,0) = 0$). For $\alpha > \bar{\alpha}(T)$, the marginal benefit exceeds $\eta(1-\omega)$

for all types. At $e_A = 1$, the marginal cost dominates by convexity. The intermediate value theorem gives an interior solution; uniqueness follows from positive definiteness of the Hessian of $-J_t$. The same argument applies to e_B. □

Proof of Proposition 11 (Effort Monotonicity)

The FOC system implicitly defines (e_A^*, e_B^*) as functions of (θ_A, θ_B). Write $F(e_A, e_B, \theta_A, \theta_B) = 0$ with $F_j = \delta\Pi_t \cdot (\partial W_{t+1}/\partial e_j) - \partial c/\partial e_j$. By the implicit function theorem ($D_{(e_A,e_B)}F$ is invertible):

$$\begin{pmatrix} \partial e_A^*/\partial\theta_B \\ \partial e_B^*/\partial\theta_B \end{pmatrix} = -\left[D_{(e_A,e_B)}F\right]^{-1} \cdot \begin{pmatrix} \partial F_1/\partial\theta_B \\ \partial F_2/\partial\theta_B \end{pmatrix}.$$

A decrease in θ_B raises $\partial W_{t+1}/\partial e_B$, shifting allocation toward e_B; joint convexity of c then raises $\partial c/\partial e_A$, reducing e_A^*. Cramer's rule gives $\partial e_B^*/\partial\theta_B \leq 0$ and $\partial e_A^*/\partial\theta_B \geq 0$. □

G Proofs for Section 5.3: Belief Dynamics

Propositions 13 and 14 follow directly from the Kalman update equation.

Proof of Proposition 15 (Joint Dynamics)

Part (i) is established in Proposition 13.

For parts (ii) and (iii): M's effort $e_{A,t}^*(\omega) = a_t + b\omega$ is determined before noise realizations. The innovations are $\xi_{A,t} = b(\omega - \theta_{A,t}) + \varepsilon_{A,t}$ and $\xi_{B,t} = -(1+\gamma)b(\omega - (1-\theta_{B,t})) + \varepsilon_{B,t}$ (the coefficient $-(1+\gamma)b$ reflects both e_B's slope $-b$ and e_A's leakage $-\gamma b$ on ω). Since $\varepsilon_{A,t} \perp\!\!\!\perp \varepsilon_{B,t}$, covariance arises solely from shared uncertainty about ω: $\mathrm{Cov}(\xi_{A,t}, \xi_{B,t}) = b \cdot (-(1+\gamma)b) \cdot P_0 = -(1+\gamma)b^2 P_0$. Multiplying by the Kalman gains yields (17). □

H Proofs for Section 5.4: Equilibrium Characterization

Proof of Theorem 16 (Existence—General Case)

We apply the Kakutani-Fan-Glicksberg fixed point theorem. The product strategy space $\Sigma = \prod_{j\in\mathcal{I}} \prod_{t=1}^{T} L^\infty(\mathcal{H}_t^j)$ is compact in the weak-$*$ topology (Banach-Alaoglu); convexity is inherited from $\Delta(\mathcal{A}_j)$. Best responses are non-empty (Weierstrass), convex-valued (action spaces are convex; objective is concave by joint convexity of c), and upper hemicontinuous (Maximum Theorem). The joint correspondence Φ satisfies the Kakutani-Fan-Glicksberg

conditions (Aliprantis and Border, 2006, Corollary 17.55), yielding $\sigma^* \in \Phi(\sigma^*)$. Beliefs μ^* are constructed via Bayes' rule on path (Assumption 6); off path, $\theta_{j,t}^{\text{off}} = 0$. □

Proof of Theorem 17 (Local Uniqueness)

Define $F : (0,1)^2 \times \mathbb{R}^5_{++} \to \mathbb{R}^2$ by the FOC system. The Jacobian $D_{(e_A,e_B)}F$ is non-singular at interior solutions (Hessian of $-J_t$ is positive definite by second-order conditions). The implicit function theorem gives local uniqueness; the parametric transversality theorem (Mas-Colell, 1985, Chapter 8) extends this to generic isolation. Intertemporal coupling is handled by backward induction from $t = T$. □

Proof of Proposition 18 (Multiplicity)

Fix α near $\bar{\alpha}(T)$ and γ large. Two fixed points of the FOC system coexist: one near $(1, 0)$ (aggressive informing) and one near $(0, 1)$ (cautious dual-loyalty). At each, the second-order conditions hold locally. Global concavity of J_t fails for large γ, creating a bimodal landscape. □

Proof of Theorem 22 (Fragility Scaling)

Part (i): $\mu_L(\mathcal{C}_t^N) = \prod_{i=1}^N (1-\underline{\theta}_{A_i,t}) \cdot (1-\underline{\theta}_{B,t})$. Each factor $(1-\underline{\theta}_{A_i,t}) < 1$, so $\mu_L(\mathcal{C}_t^N) < \mu_L(\mathcal{C}_t^{N-1})$ for all $N \geq 2$.

Part (ii): The signal noises $\varepsilon_{A_i,t}$ are independent across agencies, but the belief processes share exposure to the common type ω, creating unconditional dependence across exit times. We therefore condition on ω. Given ω, the innovations $\xi_{A_i,t} = b(\omega - \theta_{A_i,t}) + \varepsilon_{A_i,t}$ are independent across i (since the ε terms are independent and ω is fixed), making the exit times $\tau_{A_1}, \ldots, \tau_{A_N}$ conditionally independent. For any fixed ω, define $p_1(\omega) = \mathbb{P}(\tau_1 \leq 1 \mid \omega) > 0$ (strictly positive because $\theta_{A_1,1}$ is Gaussian with positive variance conditional on ω). Let $p_{\min} = \inf_{\omega \in [\underline{\omega}, \bar{\omega}]} p_1(\omega) > 0$, which is strictly positive because the type space is compact: $p_1(\omega) > 0$ for every ω (the innovation has positive variance, so $\theta_{A_1,1}$ assigns positive probability to $(-\infty, \underline{\theta}_A]$), and a continuous positive function on a compact set has a positive infimum.[23] Then:

$$\mathbb{P}(\tau^N > t \mid \omega) = \mathbb{P}(\tau_B > t \mid \omega) \cdot \prod_{i=1}^{N} \mathbb{P}(\tau_{A_i} > t \mid \omega) \leq [1 - p_{\min}]^N.$$

[23] Explicitly, $p_1(\omega) = \Phi((\underline{\theta}_A - \theta_{A,0} - K_1 b(\omega - \theta_{A,0}))/\sigma_\xi)$ where $\sigma_\xi > 0$ and Φ is the standard normal CDF. For high ω (cooperative type), $p_1(\omega)$ is small (the first-period update pushes $\theta_{A,1}$ upward, away from the boundary); for low ω, $p_1(\omega)$ is large. On the compact support $[\underline{\omega}, \bar{\omega}]$, p_1 attains its minimum at $\bar{\omega}$, giving $p_{\min} = p_1(\bar{\omega}) > 0$.

Integrating over ω: $\mathbb{P}(\tau^N > t) \leq [1 - p_{\min}]^N$. Summing:

$$\mathbb{E}[\tau^N] \leq 1 + (T-1)(1-p_{\min})^N \leq 1 + (T-1)e^{-Np_{\min}}. \quad \square$$

I Proofs for Section 5.5: Welfare Analysis

Proof of Proposition 19 (Expected Equilibrium Payoffs)

For M: $\mathbb{E}[U_M] = \sum_{t=1}^{T} \delta^{t-1}\mathbb{E}[\lambda_t(\alpha - c_t^*)]$. Under the approximation $c_t^* \approx c^*$, this reduces to $(\alpha - c^*)(1 - \mathbb{E}[\delta^\tau])/(1-\delta)$ using $\sum_{t=1}^{T} \delta^{t-1}\mathbb{P}(\tau \geq t) = (1 - \mathbb{E}[\delta^\tau])/(1-\delta)$ (since $\tau \leq T$ a.s.). For A and B: per-period flow payoffs are weighted by $\mathbb{P}(\tau > t)$, giving the formulas directly from (5)–(6). $\square$

Proof of Remark 20 (Deadweight Monitoring Cost)

Under observable effort, $m_j^{FB} = 0$ and $\kappa_j(0) = 0$. Under imperfect monitoring, both principals invest in monitoring ($m_{j,t}^* > 0$) since it reduces signal noise and extends the corridor. The inequality is strict whenever (θ_A, θ_B) is in the corridor interior. $\square$

Proof of Remark 21 (Institutional Cost of Coercion)

Under voluntary agency, $\phi = 0$ and monitoring costs are lower. The welfare loss arises from (a) protection costs $\phi(p_t^*) > 0$ and (b) excess monitoring by both principals. The inequality is strict because $q(0) = \underline{q} < 1$ implies A always provides positive protection. $\square$