Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

E-Commerce Claim Verification Under Generative AI: Deterrence and Evidence Selection to Minimize Platform Loss

Jianjun Li [1,2,*], Lin Cao [1], Yu Yang [1,2] and Shuo Yang [1]

[1] School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China; caoliny2024@s.hrbcu.edu.cn (L.C.); yangyu@hrbcu.edu.cn (Y.Y.); yangshuoy2024@s.hrbcu.edu.cn (S.Y.)
[2] Public Policy and Modern Service Industry Innovation Think Tank, Harbin 150028, China
* Correspondence: lijianjun@hrbcu.edu.cn

Abstract

We study how an e-commerce platform should allocate verification effort and choose between claimant-generated and transaction-linked evidence as manipulation becomes cheaper. In a sequential game for filed disputes, platform loss includes fraud, verification expenditure, wrongful rejection of legitimate claims and the additional burden of transaction-linked checks. Under the benchmark commitment and deterioration path, incumbent verification rises to maintain zero incremental manipulation and then falls as the platform accommodates manipulation; an initial correction-supported plateau can precede this pattern. Switching evidence bases can minimize loss before or after incremental manipulation begins, with the ordering governed by implementation burden, final-error rates and prevention ceilings. The decision rule compares the additional burden with savings in verification expenditure, residual-fraud loss and legitimate-claim error loss. Joint changes and alternative timing can reverse benchmark selection patterns. Synthetic calculations illustrate the boundaries, while monetary examples show how the per-dispute burden depends on allocable cost and case volume.

Keywords: generative artificial intelligence; e-commerce disputes; fraudulent claims; costly verification; evidence manipulation; platform governance

1. Introduction

E-commerce platforms resolve disputes over refunds, delivery failures and product conditions. Platform-governance research treats rules, control and coordination as governance mechanisms [1–3], and recent JTAER work identifies complaint handling as a governance interface [4]. China's Measures for the Supervision and Administration of Online Trading Platform Rules, effective from 1 February 2026, require dispute-resolution mechanisms, fair evidentiary responsibilities and safeguards against abuse [5]. Generative AI adds a verification problem by making claimant-generated photographs, videos, screenshots and text easier to manipulate.

Authentication remains imperfect: studies document difficulties distinguishing AI-generated and authentic content [6–8]. A 2026 arXiv field investigation based on interviews with 17 merchants and 13 platform workers reports GenAI-enabled refund fraud in Chinese e-commerce, including fabricated product-defect evidence and responses such as video requests and traceability [9]. Related JTAER research examines governance of AIGC misuse [10]. These observations motivate a comparison of evidence sources that differ in claimant control, verification cost and legitimate-claim errors.

Check for updates

Academic Editor: Qingfeng Zeng

Received: 5 September 2026
Revised: 3 November 2026
Accepted: 4 November 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Consider a buyer requesting a refund for a damaged product. A photograph may be easy to alter, while order and delivery records may reveal little about the product's condition. The platform can inspect the photograph more deeply or change which records anchor its decision. The latter choice brings integration costs and may disadvantage a genuine claimant whose complaint leaves no transaction-record trace.

We ask how verification within an evidence basis adjusts as manipulation becomes cheaper and when switching to transaction-linked evidence reduces expected platform loss. The comparison includes verification resources, successful fraud, final wrongful rejection of legitimate claims and the additional burden of the alternative basis.

We use an established verification–manipulation contest in which the platform commits to a dominant evidence basis and effort before fraudulent claimants choose additional manipulation. Claim type is unobserved. Deterrence denotes zero incremental manipulation after filing. The two bases can differ in resistance, marginal resource cost, prevention ceiling and legitimate-claim errors.

Three findings answer the research question. Along the benchmark path, incumbent verification rises while zero manipulation is maintained and falls once the optimum moves to active manipulation; legitimate-claim correction can support an initial plateau. Evidence selection follows a separate boundary, allowing a cost-minimizing switch before or after manipulation begins. Where both bases use positive verification and suppress incremental manipulation, Equation (25) gives the additional-burden threshold from their resource, residual-fraud and final-error cost differences. The contribution is to characterize these boundaries when both bases are reoptimized and to show how joint changes and effort timing can reverse their benchmark ordering.

Section 2 positions the model; Sections 3–5 define and solve it. Section 6 examines the extensions that change the main conclusions. Section 7 discusses theory, governance and limitations; Section 8 concludes. Appendix A contains the proofs and supporting derivations.

2. Literature Review

2.1. E-Commerce Disputes, Misconduct, and Platform Governance

Platform architecture and governance arrangements coevolve [11,12]. Socio-technical and ecosystem research emphasizes interdependence among participants [13,14], while studies of platform-dependent entrepreneurs document power asymmetries [15]. These perspectives place dispute procedures and evidentiary requirements within platform governance.

Lenient return policies may induce opportunistic returns [16], and research on fraudulent returns documents abuse and retailer countermeasures [17]. Promotional reviews, review fraud and organized fake-review markets provide related examples of strategic information manipulation [18–20]. These studies motivate verification within platform governance; our model focuses on evidence used in filed disputes.

2.2. Costly Verification, Strategic Misrepresentation, and Evidence

Costly state verification and audit design are established foundations [21,22]. With endogenous monitoring costs, privately costly evasion can raise verification costs [23]. Efficient insurance contracts can nevertheless retain claims inflation [24]. Our fixed refund and procedure exclude contract design. Our comparison focuses on the basis–intensity choice.

Audit credibility affects honest policyholder welfare [25], while fraud indicators can determine investigative referrals [26]. We instead impose class-level effort commitment. The economics of crime provides a foundation for detection and sanctions [27]; Section 6.4 separates their post-filing effects from participation.

Contest success functions and strategic effort commitment provide the established technical foundations for the verification–manipulation game [28–30].

One theoretical foundation is costly misrepresentation. Strategic communication models show that lying costs shape both the extent of distortion and the amount of information ultimately revealed [31,32]. We use the same reduced-form logic for manipulation resistance: lower resistance makes manipulation less costly without assuming that manipulated evidence necessarily escapes detection.

A second foundation is endogenous monitoring and verification. In costly monitoring and verification models, inspection is a principal's choice rather than an exogenous accuracy parameter [33,34]. Work on hard evidence, partial state verifiability, and costly verification in collective decisions also shows how verifiable information changes incentives and allocation [35–37]. This motivates treating platform verification intensity as endogenous.

Research on evidence formalizes certifiable information [38], surveys its role in games and mechanisms [39], and studies endogenous acquisition [40]. Selling with certifiable evidence [41] and probabilistic verification [42] provide further foundations. Commitment-free implementation is established for a specified class of evidence mechanisms [38]; the effort commitment used here is defined in Section 3.1.

2.3. Generative AI and Manipulable Digital Evidence

Research on AI-generated disinformation identifies information-integrity and governance risks [43,44]. Surveys document challenges in deepfake detection and generalization [45,46], and experiments show context-dependent limits to identifying political deepfakes and AI-generated images [47,48].

These findings motivate changes in manipulation resistance for claimant-controlled evidence. Organizational responses to AI-enabled mis/disinformation also extend beyond technical detection [49], supporting a platform-governance perspective on evidence verification.

2.4. Research Gap and Positioning

The closest benchmarks establish distinct choice mechanisms. Allocation with costly report verification can be optimized without transfers [50]. Transparent cutoff-with-appeal mechanisms allow a private principal signal [51]. Other models jointly choose incentive contracts and monitoring categories [52], design tests against covert falsification with endogenous approval probabilities [53], and compare evidence procedures with costly distortion and adjudication error [54]. Anti-counterfeiting design affects counterfeiter entry and consumer channel choice, including nonmonotone design effects [55]. Signal-distribution design can precede allocation with costly verification [56]. Building on these mechanisms, we hold candidate technologies and adjudication rules fixed and compare independently optimized basis-specific platform losses.

A concealment–detection contest between taxpayers and an authority studies auditing arrangements that reduce wasteful expenditures [57]. Our comparison applies established contest incentives to independently optimized evidence bases, with legitimate-claim correction included in platform loss.

Related work examines lie detection following a sender's claim [58] and AI-driven fabrication, verification and targeting costs across two markets [59]. Table 1 compares their choices and mechanisms with the present basis–intensity decision.

Table 1. Positioning against related theory.

Research Strand and Representative Study	Established Research Problem	Distinction of the Present Model
Classical costly verification and insurance fraud [21–24]	Costly evasion [23]; efficient contracts may retain claims inflation [24].	Fixed refunds and procedures; optimize manipulation, basis and effort.
Credible and informative claim auditing [25,26]	Audit credibility and honest policyholder welfare [25]; fraud-based referrals [26].	Class-level effort commitment; no posterior-contingent audit policy.
Sanctions and contests [27–30]	Detection incentives, effort competition and commitment.	Conditional post-filing sanctions; maintained ratio technology.
Costly report verification [50]	Allocation with costly checking and no transfers.	Filed claims fixed; optimize basis and intensity, not allocation rules.
Transparency and appeals [51]	Transparency and appeal-triggered verification with a principal signal.	Maintained correction procedure; no optimal appeals design.
Manipulable tests [53]	Optimal tests with falsifiable inputs and endogenous errors.	Fixed bases and rules; compare reoptimized losses when fraudulent claimants manipulate.
Verification and information design [56]	Signal design followed by costly verification allocation.	No signal distribution or allocation mechanism is chosen.
Strategic lie detection [58]	Inspection follows the sender's claim.	Effort precedes manipulation; simultaneous continuation tested separately.
GenAI-induced fraud reallocation [59]	AI changes fabrication, verification and targeting across two markets.	No entry or cross-market targeting; basis choice for filed disputes.
Endogenous monitoring [52]	Joint incentive-contract and monitoring-category choice.	Fixed technologies isolate correction costs, effort regimes and selection.
Evidence procedures [54]	Costly distortion and adjudication error across procedures.	Fixed procedure; final-error loss enters the basis comparison.
Anti-counterfeiting [55]	Design affects counterfeiter entry and channel choice.	Filed fraud fixed; compare deterrence and selection boundaries.
This study	Verification–manipulation contest with legitimate-claim error and two fixed bases.	Conditional effort and selection boundaries; joint-change counterexample; platform-loss objective.

3. Model Setup

3.1. Verification Bases and Timing

Consider a filed post-transaction claim of an unobserved type. An exogenous fraction $\pi \in (0,1)$ is fraudulent; legitimate claimants do not manipulate. The platform minimizes expected successful-fraud loss, verification expenditure and final wrongful-rejection loss. Throughout, "preferable" means lower optimized platform loss.

The platform chooses between two stylized dominant verification bases. The claimant-generated basis (D) relies primarily on photographs, videos, screenshots, and text supplied by the claimant. The transaction-linked basis (V) relies more on order and logistics records, platform logs, traceability data, and cross-source checks. Real platforms may combine these sources; D and V denote the dominant basis rather than mutually exclusive technologies.

Before supplementary evidence is prepared, the platform commits to a dominant basis $j \in \{D, V\}$ and verification intensity $g \geq 0$. The fraudulent claimant observes this policy and chooses incremental manipulation $e \geq 0$. Commitment requires an enforceable resource arrangement or auditable protocol that fixes effective effort for the relevant class of filed disputes and makes later deviation costly or infeasible [25]. Section 6.3 examines an alternative in which the basis is chosen first and effort is simultaneous. Figure 1 summarizes the decision sequence and final adjudication outcomes.

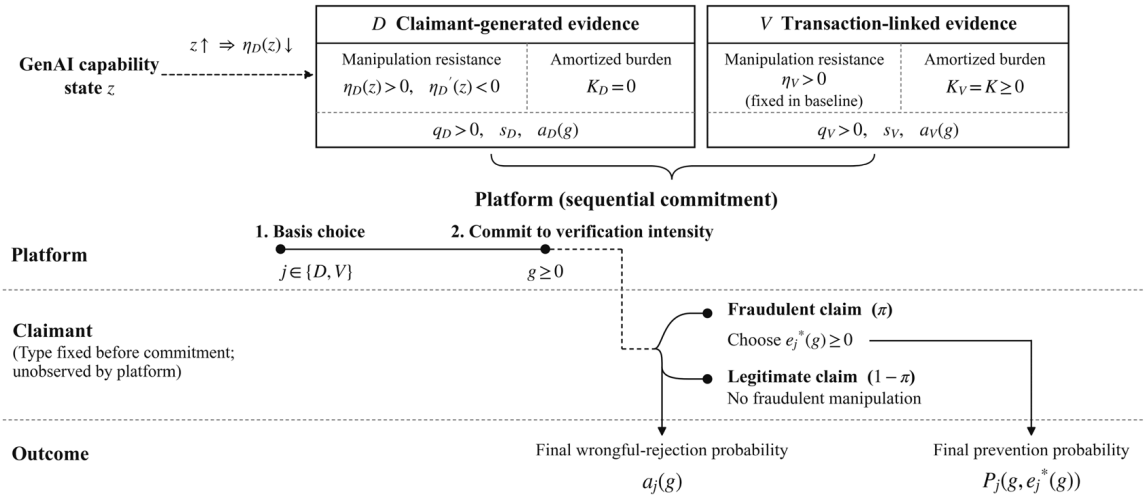


Figure 1. Decision timing and final adjudication outcomes. The diagram shows the sequence for a filed claim: basis and effort commitment, incremental manipulation, and final adjudication by true claim type. The arrows indicate the sequence of decisions and the outcome paths for the two claim types.

3.2. Capability State and Assumptions

Let $z \geq 0$ be the GenAI capability index and η_j the marginal claimant cost per unit of manipulation effort under basis j. Lower η_j represents cheaper creation or adaptation of manipulated evidence. Section 6.5 discusses potential measurement and identification.

Verification g and manipulation e use a fixed common effective-input scale. On this scale, q_j is platform currency per verification unit and η_j is claimant currency per manipulation unit; B, R, F, K_j are monetary quantities with the conditional or per-dispute meanings defined below. Comparing η_D and η_V requires the same manipulation normalization.

Assumption 1 (Baseline deterioration). *Claimant-generated evidence becomes easier to manipulate as the capability state increases:*

$$\eta_D(z) > 0, \qquad \eta_D \in C^1, \qquad \eta_D'(z) < 0. \tag{1}$$

The transaction-linked basis has constant manipulation resistance η_V in the baseline. Section 6.1 allows it to deteriorate as well. GenAI, therefore, enters the benchmark through claimant-side manipulation costs.

Assumption 2 (resources and final legitimate-claim errors). *The platform incurs resource cost $q_j g$, with $q_j > 0$. A legitimate claim is finally rejected with probability $a_j(g) \in [0, 1]$. We study correction-oriented procedures: a_j is continuous at zero, twice continuously differentiable on $(0, \infty)$, non-increasing and convex, with finite right derivative at zero. Let*

$$\beta = \pi B > 0, \qquad A_j(g) = q_j g + (1 - \pi) F a_j(g). \tag{2}$$

Assume $F \geq 0$ and a unique minimizer $h_j \geq 0$ of A_j. Uniqueness is imposed separately because convexity permits flat minima; bounded error probabilities and positive resource prices ensure coercivity. Here $a_j(g)$ is final wrongful rejection after the standing review rule and any supplementary checks or reconsideration. At $g = 0$, the standing rule still operates, so $a_j(0)$ may be positive. Increasing verification is assumed to correct misleading adverse evidence while keeping admissibility rules and claimant proof requirements fixed. Assumption 2 describes this correction-oriented procedure class.

Article 19 of China's platform-rule measures [5] motivates the distinction between initial processing and final disposition: it requires accessible appeals and timely re-examination, including human adjudication on request when a matter has been handled solely by AI. The final-error response to supplementary verification remains a model assumption.

A communication model with privately informed speakers combines costly investigation of imperfect reliability with losses from erroneous verdicts [60], providing a conceptual comparison for type-dependent error outcomes.

Assumption 3 (state-independent additional burden of the transaction-linked basis). The basis-specific burden satisfies

$$K_D = 0, \qquad K_V = K \geq 0. \tag{3}$$

K allocates the transaction-linked basis's incremental integration, data access, coordination and operating burden per filed dispute. The allocation is held constant for the baseline selection classification in Section 5.2. Its independence from effort leaves the within-basis optimizer unchanged and shifts the relative optimized loss.

Manipulation resistance, prevention ceilings and final legitimate-claim errors are separate primitives: a record-poor genuine claim may face a higher rejection risk under V even when V is harder to manipulate. Policy-independent procedure costs enter the relative intercept; common constants cancel. The normalization $K_D = 0$ and $K_V = K \geq 0$ assumes V's net intercept is at least D's. F measures platform-borne consequences of final error.

3.3. Detection and Payoffs

The probability that a fraudulent claim is finally prevented from obtaining the improper benefit is

$$P_j(g,e) = \begin{cases} s_j g/(g+e), & g+e>0, \\ 0, & g=e=0, \end{cases} \qquad 0 < s_j \leq 1. \tag{4}$$

We use the ratio-form contest technology [61] with general prevention ceiling s_j. The benchmark has $s_j = 1$, and the main robustness scenario allows residual non-detection. At zero incremental manipulation, any positive verification yields the ceiling; the origin is treated separately below.

Classical contest theory motivates the ratio technology [28,29] and the commitment comparison [30]. The recent move-order study [61] is supplementary; these sources do not establish the imposed residual ceiling.

Here $g = 0$ means no supplementary verification; the standing evidentiary rule still operates. For example, a filed-claim class may contain fraudulent submissions that satisfy formal document requirements and genuine claims that lack a required record. Without an authenticity check, the former may succeed while the latter are rejected. This illustrates the maintained origin convention. If the feasible zero-review rule instead accepts every claim, $a_j(0) = 0$ and the assumed non-increase imply $a_j \equiv 0$, eliminating the correction-supported regime.

Under verification basis j, the fraudulent claimant's expected payoff is

$$U_{F,j}(g,e) = R\big[1 - P_j(g,e)\big] - \eta_j e, \qquad R > 0. \tag{5}$$

Expected loss per filed dispute is

$$L_j(g,e) = \beta\big[1 - P_j(g,e)\big] + A_j(g) + K_j. \tag{6}$$

R is the claimant's conditional benefit and B the platform's conditional fraud exposure. The population-averaged resource cost $q_j g$ is charged once per filed dispute. Legitimate-claim loss is $(1-\pi)Fa_j(g)$, and K_j excludes resources already charged through q_j. Backward induction first optimizes effort within each basis and then compares the optimized losses.

4. Equilibrium Analysis

4.1. Claimant Response and the Verification Trade-Off

At positive verification, the claimant balances the gain from escaping detection against the marginal cost of manipulation. Strict concavity gives the unique best response:

$$e^*(g;\eta) = \max\Big\{\sqrt{sRg/\eta} - g, 0\Big\}. \tag{7}$$

Lemma 1. *Positive manipulation occurs precisely when* $0 < g < t = sR/\eta$. *At* $g \geq t$, *incremental manipulation is zero, but fraud still succeeds with probability* $1 - s$ *if* $s < 1$. *At* $g = 0$, *the unique response is* $e = 0$ *and fraud succeeds with certainty. This corner is not a successful deterrence outcome.*

Figure 2 illustrates this response with $s = R = 1$. Small increases in verification initially elicit more manipulation; beyond the response peak, additional verification reduces manipulation until $g \geq t$ makes it unprofitable. The platform anticipates this response when choosing its own intensity.

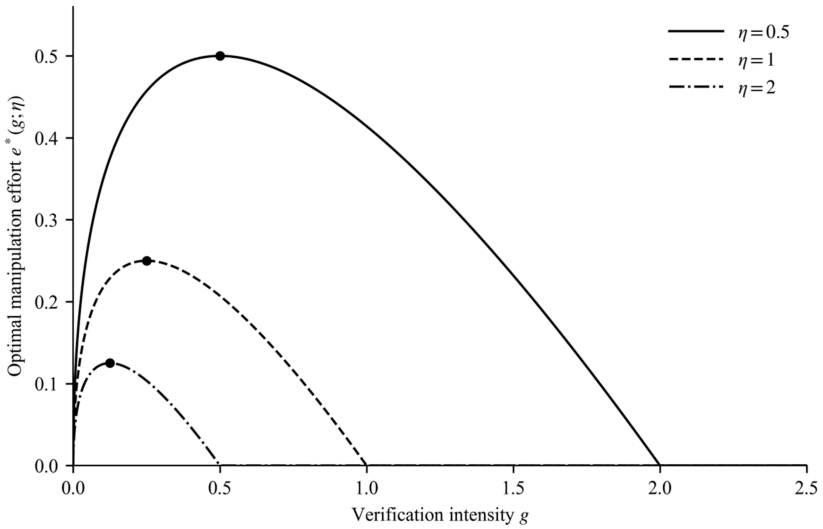


Figure 2. Fraudulent claimant's best-response geometry. Under $s = R = 1$, the curve is $e^*(g;\eta) = \max\Big\{\sqrt{g/\eta} - g, 0\Big\}$. The peak is at $g = e^* = 1/(4\eta)$ and manipulation is zero for $g \geq 1/\eta$. This conditional response is distinct from the platform's optimized paths in the subsequent platform analysis.

Substitution gives the platform's reduced objective, excluding K:

$$\Phi(g;\eta) = \begin{cases} \beta + A(0), & g = 0, \\ \beta\Big[1 - \sqrt{s\eta g/R}\Big] + A(g), & 0 < g < t, \\ \beta(1-s) + A(g), & g \geq t. \end{cases} \tag{8}$$

The active objective is strictly convex. The deterrence objective is minimized at $\max\{h, t\}$, rather than always at t. Define $x > h$ by

$$xA'(x) = \beta s/2, \qquad \eta^T = sR/x; \quad \eta^H = sR/h \text{ if } h > 0. \tag{9}$$

The monotonicity of $gA'(g)$ for $g > h$ ensures the unique root. Appendix A proves global optimality, including the zero-intensity corner.

4.2. Fixed-Basis Equilibrium

Proposition 1 (fixed-basis equilibrium). *Under Assumption 2, the platform's unique optimal intensity is*

$$g^*(\eta) = \begin{cases} h, & h > 0,\ t \le h, \\ t, & h < t \le x, \\ g_A, & t > x, \end{cases} \qquad A'(g_A) = \frac{\beta}{2}\sqrt{\frac{s\eta}{Rg_A}}, \quad h < g_A < t. \tag{10}$$

The optimum balances fraud prevention with resources and legitimate-claim correction. When the correction incentive alone supports sufficient review, verification stays at h. Otherwise, the platform maintains zero incremental manipulation at t until that policy becomes too costly; it then chooses the interior active-manipulation intensity g_A. The formulas agree at the shared boundaries, and the additive basis burden leaves the within-basis optimizer unchanged.

Under Assumption 2, $h_j > 0$ exactly when $A_j'(0+) = q_j + (1-\pi)Fa_j'(0+) < 0$: the initial reduction in final-error loss exceeds marginal review expenditure. Convexity, a finite right derivative and the unique minimizer suffice. If $h = 0$, the correction-supported regime is absent. A positive-length initial plateau additionally requires $t_j(z_{\text{start}}) < h_j$; equality starts at its endpoint. Appendix A proves global optimality.

Let $W(\eta) = \Phi\big(g^*(\eta); \eta\big)$. The optimized loss is

$$W(\eta) = \begin{cases} \beta(1-s) + A(h), & t \le h, \\ \beta(1-s) + A(t), & h < t \le x, \\ \beta\big[1 - \sqrt{s\eta g_A/R}\big] + A(g_A), & t > x. \end{cases} \tag{11}$$

For an absent first regime its row is simply omitted. W is continuously differentiable across both regime boundaries, with

$$W_\eta = \begin{cases} 0, & t \le h, \\ -A'(t)sR/\eta^2, & h < t \le x, \\ -\frac{\beta}{2}\sqrt{\frac{sg_A}{R\eta}}, & t > x. \end{cases} \tag{12}$$

Thus W is non-increasing in resistance, with strict decline outside the correction plateau. As $\eta \downarrow 0, g^* \to h$ and $W \to \beta + A(h)$. The original normalized closed forms are recovered only by setting $s = 1$, $A(g) = qg$ and $\beta = B$. Appendix A records that restricted case.

Table 2 summarizes the resulting equilibrium regimes. Section 5 applies this fixed-basis solution to changing manipulation resistance and then compares the two optimized evidence bases.

Table 2. Equilibrium regimes and feasible burden regions.

Panel (**A**). Within-basis equilibrium (Proposition 1)					
Equilibrium egime	**Parameter condition**	**Optimal verification, g^***	**Optimal manipulation, e^***	**Prevention probability, P^***	**Optimized platform loss, W**
Correction-supported verification	$h > 0, t \le h$	h	0	s	$\beta(1-s) + A(h)$
Minimum deterrence	$h < t \le x$	t	0	s	$\beta(1-s) + A(t)$
Active manipulation	$t > x$	$g_A \in (h,t)$, uniquely defined by (10)	$\sqrt{sRg_A/\eta} - g_A$	$\sqrt{s\eta g_A/R}$	$\beta\left[1 - \sqrt{s\eta g_A/R}\right] + A(g_A)$
Panel (**B**). Fixed-V selection with nonnegative burden (Proposition 3)					
Feasible burden regime	**Preferred basis and equality states**				
$0 \le K < K_0$	V for all $z \ge 0$. Feasible only if $K_0 > 0$.				
$K = K_0$	Tie at $z = 0$ and on any initial correction plateau; V thereafter.				
$K_0 < K < K_T$	Unique tie : $0 < z_S < z_T$; D before z_S, V after.				
$K = K_T$	Unique tie : $z_S = z_T$; D before it, V after.				
$K_T < K < K_\infty$	Unique finite tie : $z_S > z_T$; D before it, V after.				
$K = K_\infty$	D at every finite z; tie only as $z \to \infty$.				
$K > K_\infty$	D at every finite z; V's limiting net advantage is negative.				

5. Strategic Verification Under Generative AI

5.1. Verification Adaptation Under the Incumbent Basis

Suppose $\eta_D(z)$ is continuously differentiable, strictly decreases from $\eta_D(0) > \eta_D^T$, and tends to zero. There is a unique finite z_T with $\eta_D(z_T) = \eta_D^T$.

Table 3 gives the synthetic parameters and their effective input and monetary scales. The resistance path is $\eta_D(z) = 4e^{-z}$ with fixed $\eta_V = 5$. Appendix A records the separately normalized linear-cost benchmark. Computed results are rounded for display, usually to four decimal places; additional digits are retained for small differences or boundary comparisons. Calculations and comparisons use the unrounded S1 values.

Table 3. Reproducible synthetic scenarios and computed boundaries.

Quantity	**Benchmark (Perfect Fraud Prevention at Deterrence)**	**Positive Residual Non-Detection**
π, B, R, F	0.20, 5, 1, 10	0.20, 5, 1, 10
$\beta = \pi B$, $(1-\pi)F$	1, 8	1, 8
q_D, q_V	0.20, 0.30	0.20, 0.30
$a_D(g)$, $a_V(g)$	Equation (17)	Equation (17)
$\eta_D(z)$, η_V	$4e^{-z}$, 5	$4e^{-z}$, 5
s_D, s_V	1, 1	0.90, 0.95
h_D, x_D	1, 3.2229	1, 3
z_H, z_T	1.3863, 2.5566	1.4917, 2.5903

Table 3. *Cont.*

Quantity	Benchmark (Perfect Fraud Prevention at Deterrence)	Positive Residual Non-Detection
W_V (excluding K)	0.6467	0.6959
K_0, K_T, K_∞	0.1133, 0.3474, 1.1133	0.1641, 0.3641, 1.0641
K_L, K_C, K_H	0.20, K_T, 0.50	0.20, K_T, 0.50
z_S at K_L	2.1622	2.0197
z_S at K_C	2.5566	2.5903
z_S at K_H	2.9035	2.9320

For $s_D = s_V = 1$, the D correction plateau ends at $z_H = 1.3863$, deterrence ends at $z_T = 2.5566$, and $W_V = 0.6467$. The burden thresholds are $(K_0, K_T, K_\infty) = (0.1133, 0.3474, 1.1133)$. The cases $K = 0.20, K_T, 0.50$ give $z_S = 2.1622, 2.5566, 2.9035$, respectively. These values establish nonempty pre-, coincident and post-deterrence examples under the unified cost function.

Proposition 2 (verification adaptation along the specified resistance path). *Along this path, g_D^* is initially constant at h_D where the correction regime is visited, rises as $s_D R/\eta_D$ in the minimum-deterrence regime, reaches its unique maximum x_D at z_T, and then decreases toward h_D in the active regime. Optimized incumbent loss is non-decreasing and is strictly increasing after any correction plateau. Final legitimate-claim rejection weakly falls while verification rises, and weakly rises when verification subsequently falls.*

On the active branch,

$$\frac{dg_A}{d\eta} = \frac{A'(g_A)/(2\eta)}{A''(g_A) + A'(g_A)/(2g_A)} > 0. \tag{13}$$

The loss derivative follows (12) and the chain rule. During minimum deterrence, prevention stays at s_D while resources rise and final legitimate-claim errors fall. Effective cost nevertheless increases because effort is above its correction-only minimizer. The complete plateau–rise–decline sequence requires $\eta_D(z_{\text{start}}) > \eta_D^H$; starting at equality gives only the plateau endpoint, while $h_D < t_D(z_{\text{start}}) < x_D$ yields the rise and decline without an initial plateau. Figure 3 illustrates these incumbent-basis paths for the perfect-prevention scenario in Table 3.

5.2. Verification-Basis Selection

With fixed V parameters and fixed $K \geq 0$, define

$$G(z) = W_D(\eta_D(z)) - W_V(\eta_V), \qquad \Delta(z) = G(z) - K. \tag{14}$$

The platform selects V if $\Delta > 0$, D if $\Delta < 0$, and either basis at equality. Basis-specific costs and error functions mean that $\eta_V > \eta_D(0)$ alone no longer guarantees an initially positive gross advantage, or even that V is on its deterrence branch. We verify V's branch from its own h_V, x_V, t_V. For Table 3 V is on the minimum-deterrence branch throughout the baseline because its primitives are fixed; hence its optimized within-basis loss is constant.

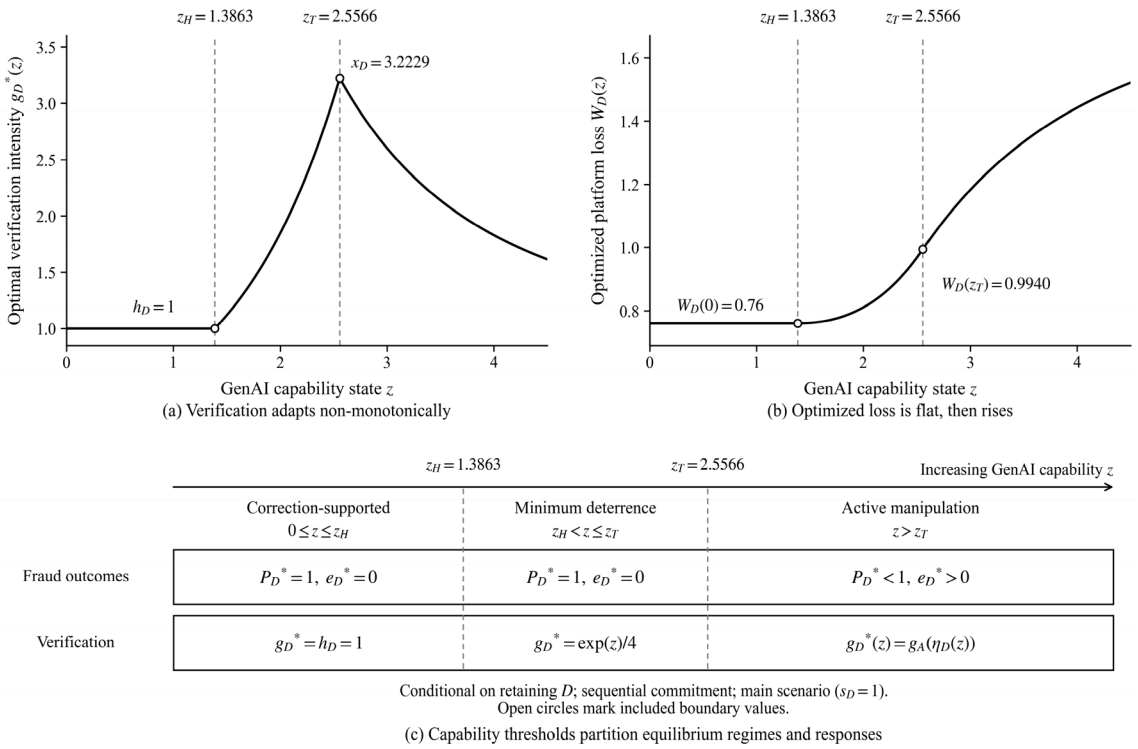


Figure 3. Verification adaptation under the incumbent basis. Table 3's perfect-prevention scenario supplies the three panels: incumbent verification, optimized loss and regime classification. These are conditional D outcomes; the residual case is analyzed separately. Markers include the boundary values.

Proposition 3 (evidence-basis selection under sequential commitment). *Under Proposition 2's path assumptions, G is continuous and non-decreasing, initially constant if D is on a correction plateau, and strictly increasing thereafter. Define*

$$K_0 = G(0), \quad K_T = G(z_T), \quad K_\infty = \beta + A_D(h_D) - W_V. \tag{15}$$

Then $K_0 < K_T < K_\infty$, but any of these numbers can be negative. The seven feasible burden cases are in Table 2. In particular, a unique strictly positive pre-deterrence equality state exists if $K_0 < K < K_T$ and $K \ge 0$. At $K = K_0$, equality can persist until the end of an initial plateau. At $K = K_\infty$, equality is only asymptotic. No finite boundary exists above it. If $A_j(g) = \bar{A}_j(g) + b_j$ with b_j independent of g, then $W_j = \bar{W}_j + b_j$ and the within-basis optimizer is unchanged. Holding all other primitives fixed, each selection-burden threshold shifts by $b_D - b_V$; a common additive component cancels. Intensity-dependent correction gains can change both the within-basis regimes and the reoptimized selection boundary. Joint changes require Proposition 4.

Corollary 1. *Whenever the relevant burden intervals have nonempty intersections with* $[0, \infty)$*, zero manipulation under D need not imply that D is economically preferred, and active manipulation need not imply that V is preferred. Specifically, a pre-deterrence crossing is feasible if* $K_T > max\{0, K_0\}$*, and a strictly post-deterrence finite crossing is feasible if* $K_\infty > max\{0, K_T\}$*. The numerical cases in Table 3 demonstrate both.*

Early substitution also occurs in the no-correction special case (Appendix A). Legitimate-claim correction changes effective marginal verification cost, the effort regimes and the

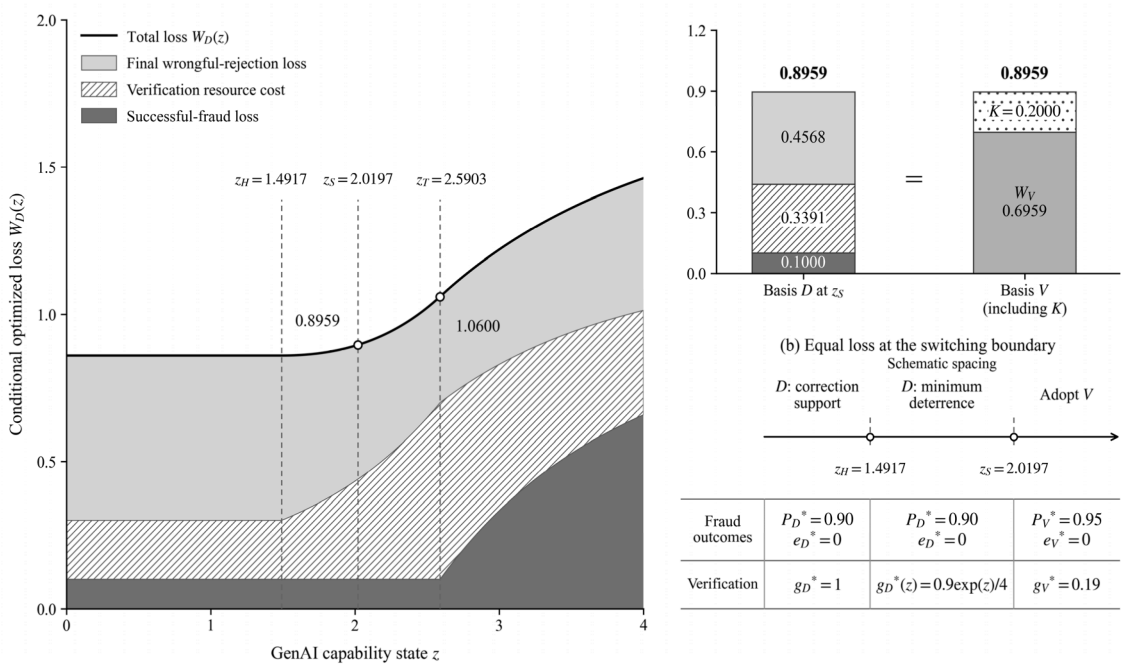
selection boundary. Corollary 2 isolates this channel by varying the final-error cost weight while holding other primitives fixed.

Corollary 2 (correction incentives and evidence-basis substitution) *Let* $0 < \pi < 1$, $q_j > 0$, and let the maintained effective-cost assumptions hold for every F *in a considered interval, with all other primitives fixed. For each basis* j, *the correction-cost minimizer* $h_j(F)$ *and the deterrence-boundary intensity* $x_j(F)$ *are weakly increasing in* F. *At an interior switching state* $z_S(F)$ *where* G *is continuously differentiable in a neighborhood of* $(z_S(F), F)$, $G(z_S(F), F) = K$, *and* $G_z(z_S(F), F) > 0$,

$$\frac{dz_S}{dF} = -\frac{(1-\pi)[a_D(g_D^*) - a_V(g_V^*)]}{G_z(z_S(F), F)}. \tag{16}$$

The optimized wrongful-rejection probabilities are evaluated at the switching state. Higher wrongful-rejection loss therefore locally advances selection of V when $a_D(g_D^*) > a_V(g_V^*)$ and delays it under the reverse inequality. The directional conclusion does not cover plateau ties, disappearing crossings or simultaneous changes in K and other primitives. The proof is in Appendix A.

Figure 4 illustrates Proposition 3 with the positive-residual Table 3 parameters and Equation (17).



(a) Conditional loss decomposition under basis D (c) Schematic policy order; computed outcomes

Figure 4. Early basis substitution with positive residual non-detection. The positive-residual scenario in Table 3 is used. (**a**) Conditional D losses and their components, including the D continuation beyond the switch. (**b**) Reoptimized losses are equal at the switching boundary, including V's burden K. (**c**) The arrow and horizontal spacing show the policy order schematically; the threshold labels and equilibrium entries are computed model values. D is retained at equality.

5.3. Analytical Illustrations and Comparative Statics

The illustrative probability functions are

$$a_D(g) = 0.02 + \frac{0.10}{1+g}, \qquad a_V(g) = 0.04 + \frac{0.04}{1+g}. \tag{17}$$

Equation (17) specifies bounded synthetic final-error curves, with a persistent component and a correctable component that decreases with verification. The coefficients and functional forms are illustrative assumptions.

5.3.1. Final-Error-Loss Sensitivity and Legitimate-Claim Outcomes

Both bases are reoptimized for each F, holding $K = 0.20$, $s_D = s_V = 1$ and all other benchmark primitives in Table 3 fixed. For $F = (0, 2, 5, 10, 20, 40)$, the equal-loss states are respectively $z_S = (1.6487, 1.7172, 1.8540, 2.1622, 2.7274, 3.6045)$.

For the same F values, final-error probabilities under D are (0.0635, 0.0618, 0.0585, 0.0515, 0.0408, 0.0393); under V they are (0.0733, 0.0733, 0.0733, 0.0733, 0.0674, 0.0594).

The switching state increases with F in this scenario. At $F = 10$, $g_D^* = 2.1725$ and $g_V^* = 0.20$. Legitimate-claim final wrongful-rejection risk is approximately 5.152% under D and 7.333% under V, a 2.181 percentage-point difference; the all-claim incidence difference is 1.745 points when $\pi = 0.20$. Thus, stronger concern for legitimate-claim errors can favor more verification within a basis while delaying V adoption. Equation (16) gives the local sign condition.

5.3.2. Selection Boundaries and Residual-Fraud Losses

Figure 5 translates Proposition 3 into a selection map. With Table 3's residual parameters, the boundary starts at 0.1641, reaches 0.3641 at the deterrence endpoint and approaches 1.0641. A higher burden delays the switch when a finite crossing exists. Thus $K = 0.20$ supports an early switch, whereas $K_H = 0.50$ supports a switch after manipulation begins.

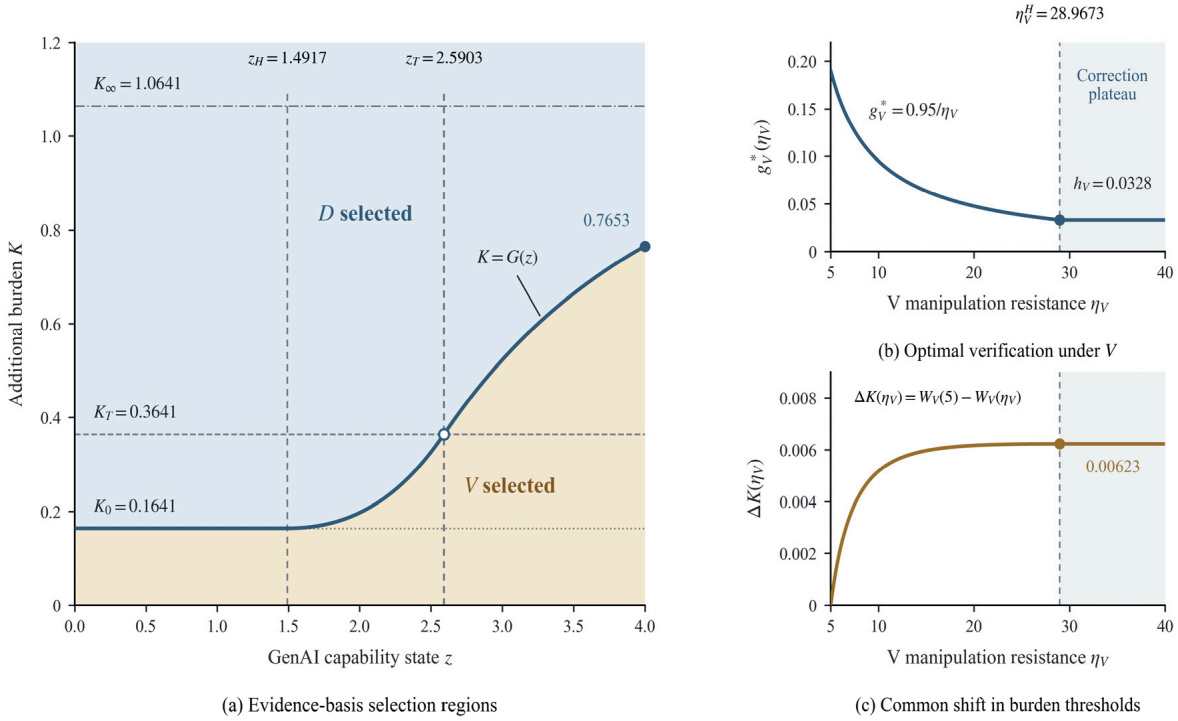


Figure 5. Basis-selection regions and V-resistance comparative statics. The positive-residual scenario in Table 3 is used. (**a**) Selection regions at $\eta_V = 5$; equality holds on $K = G(z)$. (**b**) Reoptimized V verification as resistance varies. (**c**) The burden thresholds shift by $\Delta K(\eta_V) = W_V(5) - W_V(\eta_V)$ for $r \in \{0, T, \infty\}$, with D primitives fixed.

Increasing alternative resistance reduces the review needed to prevent incremental manipulation. Once correction of legitimate-claim errors determines V's effort, further resistance gains leave its optimized loss unchanged. In Figure 5b,c, the common increase in

the three burden thresholds therefore saturates at approximately 0.00623. The comparison identifies the source and the limit of the gain from harder-to-manipulate evidence.

The residual case also separates behavioral deterrence from successful fraud prevention. Figure 6 retains the verification hump, but final prevention stays at 0.90 while incremental manipulation is zero and falls after the deterrence endpoint. The platform therefore incurs fraud loss even in the minimum-deterrence region. All panels show the conditional D continuation, including states in which switching to V would minimize total loss.

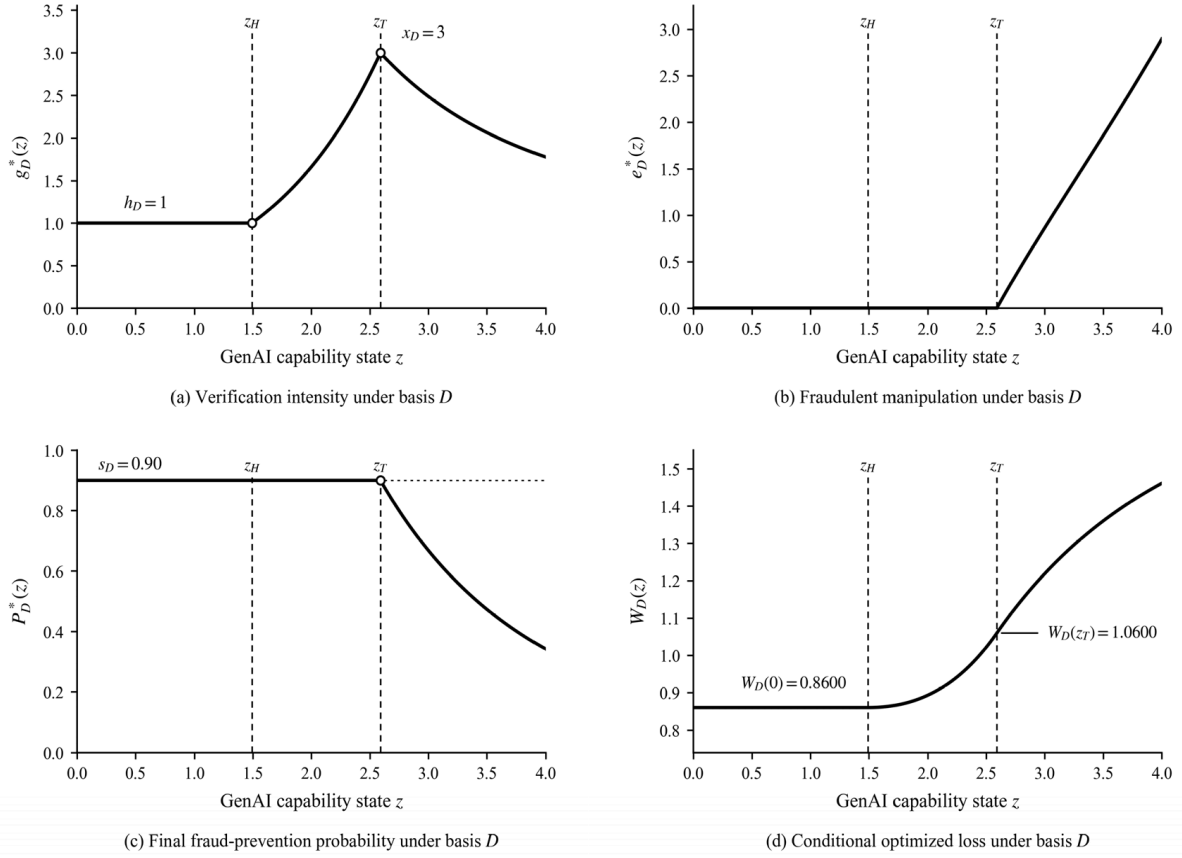
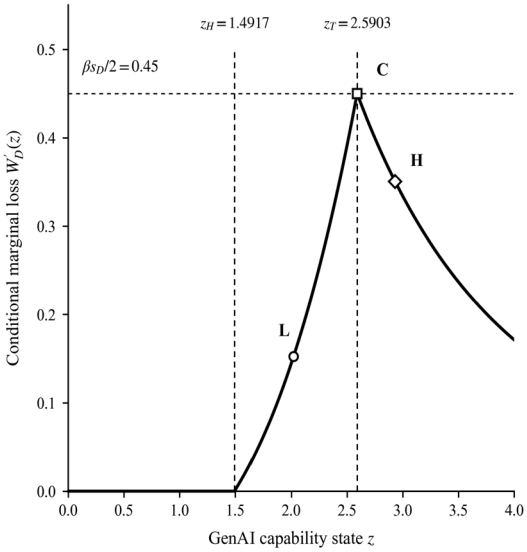


Figure 6. Equilibrium outcomes with positive residual non-detection. The residual parameters in Table 3 use $s_D = 0.90$ and $s_V = 0.95$. Panels report D verification, manipulation, final prevention and optimized within-basis loss.

The rising loss in Figure 6d can be decomposed into successful-fraud, review-resource and final wrongful-rejection costs. For the unit-rate exponential resistance path, the capability derivative is

$$\frac{dW_D}{dz} = \begin{cases} 0, & t \le h_D, \\ tA_D{}'(t), & h_D < t \le x_D, \\ \beta P_D^*/2, & t > x_D. \end{cases} \tag{18}$$

On the correction plateau, the derivative is zero. During minimum deterrence, higher review expenditure is partly offset by lower final-error loss. At z_T, the signed contributions change across regimes, but their sum is continuous and equals $\beta s_D/2 = 0.45$. Figure 7 reports this decomposition at the early, coincident, and late switching states. After the endpoint, lower verification raises both successful-fraud and final-error losses while saving review resources. The chain-rule factor must be restored for other resistance paths.



(a) Marginal loss at switching positions

Signed contributions to $W_D'(z)$

State	Fraud loss	Review resources	Final-error loss	Net loss
L	0.00000	0.33914	−0.18668	0.15246
C^-	0.00000	0.60000	−0.15000	0.45000
C^+	0.67500	−0.30000	0.07500	0.45000
H	0.50102	−0.22007	0.06946	0.35041

C^- and C^+: one-sided limits at z_T

Net = fraud + resources + final error

(b) Signed marginal-loss contributions

Figure 7. Marginal incumbent loss and its signed channels. The positive-residual scenario is used throughout. L and H are the switching states for K = 0.20 and K = 0.50; C is the coincident case at the deterrence boundary. Panel (**a**) reports the total derivative. Panel (**b**) tabulates the successful-fraud, verification-resource and final wrongful-rejection contributions and their signed sum. C^- and C^+ are one-sided limits at the boundary; entries are rounded to five decimals.

The same boundary determines the signed gap between switching and the onset of manipulation. Let $\Omega = z_T - z_S$ for a unique finite crossing with $G_z(z_S) > 0$. Holding D primitives and the burden fixed,

$$\frac{\partial \Omega}{\partial \eta_V} = \frac{-W_{V,\eta}(\eta_V)}{G_z(z_S)} \geq 0. \tag{19}$$

Greater V resistance weakly increases this gap because it reduces V's optimized loss. The effect becomes zero on V's correction plateau. The derivative requires a finite regular crossing; an initial tie interval or an absent crossing must be classified directly using Proposition 3. Appendix A gives the separately normalized linear-cost sensitivity.

6. Extensions

6.1. Deterioration of Both Verification Bases

Write $\tilde{W}_j(\eta, z) = \min_g \Phi_j(g; \eta, z)$ and $W_j(z) = \tilde{W}_j\big(\eta_j(z), z\big)$, allowing effective costs and K to vary with z. Hold β, R, s_j fixed. Let $M_j = -\partial \tilde{W}_j / \partial \eta \geq 0$, evaluated at the current policy.

Proposition 4 (joint changes). *Where value functions are differentiable,*

$$\Delta'(z) = M_D(-\eta_D{}') - M_V(-\eta_V{}') + A_{D,z}(g_D^*, z) - A_{V,z}(g_V^*, z) - K'(z). \tag{20}$$

The direct cost derivatives hold g fixed; no optimal-effort derivative is added. If Δ is continuous and strictly increasing over an interval and has opposite signs at its endpoints, exactly one D-to-V preference switch occurs there. A strictly positive derivative throughout is sufficient under the usual piecewise smoothness but is not necessary. At nondifferentiable boundaries, use the global monotonicity condition rather than dividing marginal values. M_V can be zero on its correction plateau. If s_j, R or β also change, (20) needs their direct derivatives.

A counterexample uses $\beta = R = 1$, $s_D = s_V = 0.9$, $\eta_D = 4e^{-z}$, $\eta_V = 6e^{-z}$, $K(z) = 0.05 + 0.02(1 - e^{-z})$, $q_D(z) = 0.5 + 0.03(1 - e^{-z})$, $q_V(z) = 0.6 + 0.02(1 - e^{-z})$, $a_D(g) =$

$0.0025 + 0.01/(1+g)$ and $a_V(g) = 0.00375 + 0.01/(1+g)$, with $\pi = 0.2, B = 5, F = 10$. Independent optimization gives net-loss roots 1.1768 and 1.8395, with D, then V, then D strictly preferred. Both cost paths, both resistance paths, the burden and nonconstant legitimate-claim costs are included. V remains harder to manipulate throughout. Figure 8 illustrates this reversal under the specified joint changes.

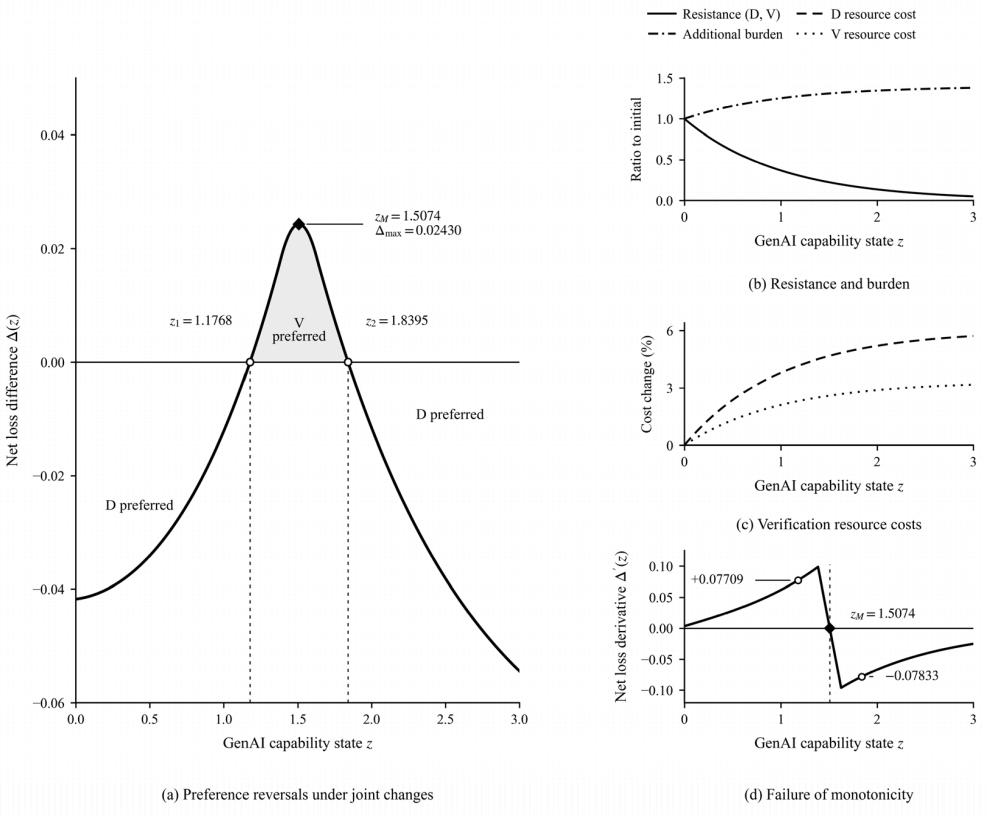


Figure 8. Preference reversals under joint changes. The panels use the dedicated joint-change counterexample in Section 6.1, varying both resistance paths, both resource prices and additional burden. (**a**) $\Delta = W_D - W_V - K$ is positive where V is preferred and negative where D is preferred. (**b**) Resistance and burden ratios to initial levels; coincident resistance ratios do not imply equal resistance levels. (**c**) Resource-price changes, $100\left[q_j(z)/q_j(0) - 1\right]$ percent. (**d**) The derivative from (20) changes sign, so global single crossing fails. Circles mark the same equal-loss states in (**a**,**d**); diamonds identify the maximum of Δ and its zero derivative. The curves are synthetic illustrations.

Falling platform review prices can offset cheaper manipulation and make the net advantage nonmonotone. Reoptimizing both bases under the declining-price paths moves the illustrative crossing from 2.0197 to 3.7092. Proposition 4, rather than the fixed-cost classification, governs this comparison.

An exogenously rising fraud share changes both fraud exposure and legitimate-error weights. In the separate composition sensitivity, reoptimization advances selection from 2.0197 to 1.5505.

6.2. Residual Non-Detection

Costly falsification and inexact verification can coexist in an optimal insurance contract [62]. Here Equations (7)–(12) apply with $s_j < 1$ after recomputing $x_j A_j'\left(x_j\right) = \beta s_j/2$ and $t_j = s_j R/\eta_j$. Table 3 reports the resulting thresholds, and Figures 6 and 7 show the positive-residual outcomes.

Basis-specific resource costs and prevention ceilings. For each evidence basis j, the platform independently minimizes its within-basis loss with marginal verification cost

q_j and prevention ceiling s_j. At a differentiable optimum, the envelope-theorem effect of increasing q_j on minimized loss is the optimal verification intensity ${g_j}^*$.

Changing only the fixed alternative's prevention ceiling leaves the incumbent path and Proposition 3's monotonicity proof unchanged. Lowering that ceiling weakly raises V's optimized loss, so all three burden thresholds fall by the same amount. Intersect the recalculated intervals with nonnegative burdens; the feasible pre-endpoint interval can disappear. Changes in D's ceiling or joint technology paths do not have this shift-only implication.

In the reverse-ceiling example, reducing the alternative ceiling to 0.40 makes the endpoint burden threshold negative (-0.1803); no nonnegative burden then permits a strictly pre-endpoint crossing. The lower prevention ceiling outweighs V's resistance advantage in this case.

6.3. Simultaneous Effort with Ex Ante Evidence-Basis Selection

The alternative timing has two stages. The platform first selects its dominant basis $j \in \{D, V\}$. Conditional on that basis, the platform and fraudulent claimant simultaneously choose verification and manipulation; the legitimate type does not manipulate and the individual type remains unobserved. Stage one compares the induced Nash losses, including each basis's burden.

For each fixed basis, the platform and claimant use (5) and (6) with exactly (4)'s zero-input convention; the constant K_j does not affect the effort continuation. For $g+e > 0$, within-basis platform loss excluding K_j is $\beta[1 - sg/(g + e)]+A(g)$. Under Assumption 2's convexity, differentiability and unique-minimizer conditions, when $h > 0$,$(h, 0)$ is a Nash equilibrium if and only if $\eta \geq sR/h$. Otherwise, the unique interior effort equilibrium satisfies

$$A'(g_N) + \frac{\beta\eta}{R} = \beta\sqrt{\frac{s\eta}{Rg_N}}, \qquad e_N = \sqrt{\frac{sRg_N}{\eta}} - g_N > 0. \tag{21}$$

When $h = 0$, there is no zero-manipulation Nash equilibrium at finite positive resistance under the specified zero-input convention. A correction-supported zero-manipulation region can remain when $h > 0$, but the commitment model's minimum-deterrence maintenance branch does not generally exist. Holding A, β, R, s fixed as resistance varies, the interior effort peak is characterized by $yA'(y) = \beta s/4$ and $\eta = sR/(4y)$. Appendix A supplies the global-best-response argument, including the zero-effort corner; these conditions must be checked for each specified cost class.

For a fixed state, define the induced within-basis Nash loss

$$N_j = \beta\left[1 - P_j\left(g_{Nj}, e_{Nj}\right)\right] + A_j\left(g_{Nj}\right).$$

The first-stage policy is $j_N^* \in \operatorname{argmin}_{j\in\{D,V\}}\left\{N_j + K_j\right\}$; V is strictly selected when $N_D - N_V - K > 0$. Both bases are evaluated using their induced Nash continuations, including the continuation after an unselected basis. Equal first-stage losses permit either basis.

In the residual scenario with K = 0.20, commitment selects V before D manipulation begins. With simultaneous effort, manipulation begins at z = 1.4917 and selection follows at 1.7817, reversing the benchmark ordering.

6.4. Detected Fraud Sanctions

Let $\tau \geq 0$ be a monetary-equivalent sanction paid by a fraudulent claimant when detected. Filing and withdrawal are fixed, and sanction receipts are excluded from platform revenue. The claimant payoff is

$$U_j = R - (R + \tau)P_j - \eta_j e. \tag{22}$$

At positive verification, the payoff is strictly concave in manipulation, giving the response and minimum intensity:

$$e_j^{\tau*} = max\left\{\sqrt{\frac{s_j(R+\tau)g}{\eta_j}} - g, 0\right\}, \quad t_j^{\tau} = \frac{s_j(R+\tau)}{\eta_j}. \tag{23}$$

Replace the claimant stake R by R + τ in Proposition 1, keeping platform exposure β = πB and effective cost unchanged. Peak intensity x is unchanged. On Table 3's unit-rate exponential path, the endpoint moves earlier by log(1 + τ/R); with decay rate α > 0, divide that shift by α. A negative endpoint means manipulation is active from the reference state. At a feasible endpoint, incumbent loss is independent of τ while V's minimized loss weakly rises, yielding

$$\eta_j^{T,\tau} = \frac{s_j(R+\tau)}{x_j}, \quad K_T^{\tau} = \beta(1-s_D) + A_D(x_D) - W_V^{\tau}. \tag{24}$$

For sanctions 0, 0.5 and 1 in the positive-residual Table 3 scenario, endpoint states are 2.5903, 2.1848 and 1.8971; selection states are 2.0197, 1.6672 and 1.4405. Endpoint burden thresholds fall from 0.3641 to 0.3555 and 0.3441. An earlier selection state therefore does not imply a larger feasible pre-endpoint burden region. At fixed capability, the relative value effect is generally ambiguous.

At positive verification with zero manipulation, claimant utility is R(1 − s) − τs. Under perfect prevention and a positive sanction, this is negative, so allowing an outside option would change filing participation. Appendix A proves the effort result for the fixed filed-claim population.

6.5. Monetary Magnitudes and Volume Sensitivity

This illustration translates the model into currency units and break-even volumes. The card-chargeback populations in [63,64] differ from platform refund claims, and the merchant tracking subscription in [65] covers a data-access component. The values serve as scale anchors. Applying the comparison to a large platform requires its own allocable integration costs, dispute volumes, exposure and final-error costs.

Mastercard/Datos report a USD 84 average retail chargeback [63]. The Visa Acceptance Solutions/Verifi/MRC report gives USD 78 as a trimmed all-related first-party-misuse resolution cost, with estimates capped at USD 299 [64]. We use USD 84 only as a monetary scale; the USD 78 measure is not a marginal verification-cost estimate.

Table 4 reports monetary scenarios M1 and M2, retaining the positive-residual error curves and effective-input scales. Both set the claimant benefit to USD 84 and scale resource prices and resistance accordingly. Exposure-to-refund and final-error-to-refund ratios are scenario assumptions.

The reported AfterShip Tracking Premium price starts at USD 59 per month, with 6000 shipments per year and API/webhook access [65]. Annualizing gives USD 708. Table 4 allocates this small-merchant access component over hypothetical eligible dispute volumes. It is not an estimate of large-platform adjudication integration.

Each volume cutoff divides the assumed annual USD 708 component by the endpoint burden threshold. Equality uses unrounded values. M2 at 120 disputes lies just USD 0.000664 below its threshold, illustrating sensitivity near a break-even boundary.

Only incremental allocable costs enter the additional burden. Shared or sunk expenditures and resources already included in variable verification costs are excluded.

Table 4. Monetary scenarios and volume-dependent selection.

Quantity or Allocable-Cost Assumption	M1: Exposure/Refund = 5; Error/Refund = 10	M2: Both Loss Ratios = 1
Refund benefit; exposure; final-error loss (USD)	84; 420; 840	84; 84; 84
Initial; endpoint; limiting burden (USD)	13.783765; 30.583765; 89.383765	1.714891; 5.900664; 14.289176
Incumbent endpoint (capability)	2.590267	0.877568
600 eligible disputes/year; burden USD 1.18	V from reference state	V from reference state
120/year; burden USD 5.90	V from reference state	Pre-endpoint selection: 0.877480
20/year; burden USD 35.40	Post-endpoint selection: 2.724142	No finite selection
Volume cutoff for endpoint equality (per year)	23.149537	119.986492

Table 5 maps the primitives to potential measurements and identification sources. Effective verification and manipulation inputs require a fixed common normalization; cost records, validated outcomes and experimental variation play different roles in estimating the parameters.

Table 5. Potential measurement and identification of model primitives.

Primitive	Unit	Observable Proxy and Possible Source	Identification Status and Boundary
η_j	Claimant currency/e unit	Manipulation resources in controlled red-team tasks and tool-use records.	Structural or experimental identification after fixing the e scale; not directly observed in dispute logs.
s_j	Probability (0–1)	Final prevention in validated fraudulent cases with zero incremental manipulation; positive-review tests.	Maintained ceiling until validated; flags or maximum observed detection rates do not identify it.
q_j	Platform currency/g unit	Incremental reviewer, compute and data costs; workload and accounting records.	Marginal-cost estimate after normalizing g and adjusting workload mix; not average total resolution cost.
K	Platform currency/filed dispute	Incremental allocable integration, access and coordination costs/eligible dispute volume.	Records inform costs and volume; allocation horizon/shares remain assumptions. Exclude sunk/common costs and resources in q_j.
F	Platform currency/final wrongful rejection	Attributable redress, remediation and retention losses; validated cases linked to cost/customer records.	Direct costs from records; causal retention/reputation effects need experimental or quasi-experimental identification, or assumptions.

Table 5. *Cont.*

Primitive	Unit	Observable Proxy and Possible Source	Identification Status and Boundary
π	Share of filed claims (0–1)	Fraud share in a representative, independently validated audit sample of filed disputes.	Estimable with sampling and validation; flagged cases are selected. Imperfect validation requires correction or bounds.
B	Platform currency/successful fraudulent claim	Net platform exposure after recovery/reimbursement; outcome and accounting records.	Record-based estimate with attribution rules; need not equal the refund or R.
R	Claimant currency/successful fraudulent claim	Improper benefit retained; payment, recovery and account records.	Transfers inform monetary benefit; private valuation/nonmonetary gains require assumptions or experimental identification.
$a_j(g)$	Final-error probability (0–1)	Final rejection among validated legitimate claims by review input; adjudication audits.	Effort effects need exogenous variation or a credible design; decreasing convexity remains an assumption here.

6.6. Scope Conditions and Extensions

Table 6 collects the conditions under which the benchmark results and extensions apply. Corollary 2 holds the other primitives and the additional burden fixed when varying the final-error weight.

Table 6. Applicability of the benchmark results.

Maintained Comparison	Propositions and Path Result	Selection Implication and Limit
Fixed V and burden; only D resistance falls	1–3 under their assumptions; correction plateau may be absent	Unique finite crossing only in Proposition 3's feasible intervals
Only fixed V ceiling or resistance changes	1–3 after reoptimization	All thresholds shift equally; feasible nonnegative intervals may vanish
Joint prices, resistances or burden changes	1 pointwise under its assumptions; 2–3 need not hold along the path	Use 4; monotonicity plus endpoint signs suffice; reversals possible
Rising filed-fraud share	Recompute fraud and legitimate-error weights	Add direct prevalence effects to 4; no filing equilibrium
Fixed detected-fraud sanction	1 with stake $R + \tau$; platform exposure unchanged. Path conditions still required	Endpoint burden weakly falls under fixed V; fixed-state and filing effects differ
Quadratic cost or conditional fee	Separate branch checks; a fee may cause nonattainment	Use S1's derivations and global checks; 1–3 do not transfer automatically
Simultaneous effort, segmentation or service cost	Nash continuations, nested policies or reoptimized costs	No invariant ordering, universal mean-parameter policy or welfare ranking

Across these comparisons, the objective is expected platform loss for filed claims under fixed adjudication procedures. Technology paths and fraud prevalence are exogenous. The numerical illustrations establish the reported cases; platform-specific implementation requires the measurements in Table 5.

7. Discussion

7.1. Comparison with Auditing and Contest Theory

Classical auditing models allow evasion to raise monitoring costs [23] and efficient insurance contracts to retain claims inflation [24]. Our active-manipulation region shares this accommodation mechanism, while the basis for comparison holds refunds and procedures fixed. Credibility affects honest policyholder welfare [25], and scoring models guide fraud-based investigative referrals [26]. The present analysis evaluates class-level effort and evidence-basis choice through optimized platform loss.

The verification hump follows from increasing effort to maintain zero manipulation and subsequent accommodation under the specified ratio technology and deterioration path. Established contest foundations [28–30] explain the roles of effort competition and commitment. Joint changes and the alternative timing show how the basis-selection ordering can change; Table 6 summarizes the relevant conditions.

7.2. Platform Governance Implications

During minimum deterrence, manipulation remains zero and final prevention stays at s_j while verification expenditure rises. Part of the added effort also corrects legitimate-claim errors. Stable fraud outcomes can therefore conceal changing review costs and error losses. Workload, case duration, evidence requests and escalations can help a platform assess these changes.

When both bases use positive verification and zero incremental manipulation, V minimizes platform loss exactly when its additional burden is below the following full cost difference:

$$K < q_D {g_D}^* - q_V {g_V}^* + \pi B(s_V - s_D) + (1 - \pi)F[a_D({g_D}^*) - a_V({g_V}^*)]. \quad (25)$$

The rule reduces to review-spending savings when the optimized bases have equal prevention ceilings and equal final-error losses. Costs are per filed dispute. At a zero-verification corner, use the full objective; a procedural service loss adds its separate cost difference.

For delivery or order disputes, a platform can pilot transaction-linked checks and measure the relevant cost and error differences. Product-condition disputes additionally require validated condition evidence. Comparing these outcomes by dispute category makes the burden threshold operational and helps identify genuine claimants who lack transaction-linked support.

Evidence of discrepancies between declared governance and seller experiences on Amazon Marketplace [66] reinforces the need to verify how a proposed protocol is implemented.

7.3. Limitations and Future Research

The game begins with filed disputes and an exogenous fraud share. Endogenizing submission, withdrawal, seller responses and appeals would connect verification policy to the composition of claims. The sanction result concerns effort within the fixed filed-claim population.

The two dominant bases represent fixed adjudication procedures, and capability and productivity paths are exogenous. Future work could study evidence mixtures, endoge-

nous technology investment and procedures derived from signals. The correction channel requires verification to reduce final legitimate-claim errors; procedures with the opposite response need a different specification.

Empirical application requires platform-specific costs and validated claim outcomes. Table 5 outlines potential identification routes. The card-dispute and merchant-subscription anchors illustrate monetary units and break-even logic; they do not estimate platform adoption thresholds or the frequency of the model's selection regions.

8. Conclusions

This paper asks how an e-commerce platform should adjust verification and choose its dominant evidence basis as manipulation becomes cheaper. Under the benchmark commitment and deterioration path, verification rises to maintain zero incremental manipulation and falls once accommodation becomes optimal; legitimate-claim correction can support an initial plateau. The contribution is to characterize evidence-basis selection after both bases are reoptimized with legitimate-claim error costs. Its boundary can precede or follow the onset of manipulation, depending on the additional burden and the relative cost and accuracy of the two bases. Joint changes and alternative effort timing identify limits to that benchmark ordering.

For platform managers, the comparison is between complete optimized losses. Where both bases use positive verification and suppress incremental manipulation, Equation (25) sets the additional-burden threshold using review expenditure, residual-fraud exposure and final wrongful-rejection loss. A harder-to-manipulate evidence source can still perform poorly for legitimate claims or product-condition disputes. A useful pilot therefore measures these differences and allocable cost per eligible dispute before changing the dominant evidence basis. The monetary illustrations show how to apply the accounting; the parameter map identifies what a platform would need to measure.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jtaer21110414/s1. Supplementary File S1 contains the notation glossary, parameters, executable code, numerical outputs, figure scripts and additional technical analyses of nonlinear costs, tool adoption, segmentation and service burden.

Author Contributions: J.L. contributed to study conceptualization and outline; L.C. contributed to writing, modeling, formal analysis, and analytical illustrations; Y.Y. contributed to writing guidance; S.Y. contributed to editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the 2022 Heilongjiang Philosophy and Social Science Research Planning Project (grant number 22SHE416); the 2026 Higher Education Teaching Reform Research Projects of Harbin University of Commerce (grant numbers HSDJY202637 and HSDJY202661); and the 2026 Culture and Tourism Research Projects of Heilongjiang Province (grant numbers 2026WL075 and 2026WL002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The synthetic parameters, code and generated numerical outputs are provided in Supplementary File S1. Published sources used for the monetary illustration are cited in Section 6.5.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Proofs and Supporting Derivations

A.1. Claimant Response and the Zero-Input Corner

For $g > 0$, $U_e = sRg/(g+e)^2 - \eta$ and $U_{ee} = -2sRg/(g+e)^3 < 0$. The first-order condition and the derivative at zero give (7) and the threshold t. At g = 0, the convention gives $U = R - \eta e$, uniquely maximized at zero. This establishes Lemma 1, including the distinct economic meaning of the origin.

A.2. Global Platform Solution

On $0 < g < t$, (8) has derivative $A'(g) - (\beta/2)\sqrt{s\eta/(Rg)}$ and strictly positive second derivative $A''(g) + (\beta/4)\sqrt{s\eta/R}\,g^{-3/2}$. The derivative tends to minus infinity at zero. At t from below it equals $A'(t) - \beta s/(2t)$; from above it equals $A'(t)$. There is an upward derivative jump. Together with convexity of A and continuity at t, the reduced objective is convex on the positive half-line and strictly convex on the active interval.

Because A is convex, coercive and uniquely minimized at h, $A'(g) > 0$ for $g > h$. The function $gA'(g)$ is strictly increasing there, with derivative $A'(g) + gA''(g) > 0$, and tends to infinity; hence (9) has a unique $x > h$. If $t \le h$, the active objective decreases up to t and the deterrence objective is minimized at h. If $h < t \le x$, the active derivative at t is nonpositive and A increases to its right, so t is optimal. If $t > x$, the active derivative crosses zero once at $g_A \in (h, t)$ and the objective increases thereafter. These comparisons establish (10) using only feasible candidates.

The corner at zero has loss $\beta + A(0)$. At a sufficiently small positive g, its difference from this loss is $-\beta\sqrt{s\eta/R}\sqrt{g} + A'(0+)g + o(g) < 0$. Thus zero is strictly dominated for finite positive resistance, under the finite-derivative assumption. The technological discontinuity has been treated explicitly rather than assumed away. Unique positive minimizers connect at t = h and t = x. This proves Proposition 1 and (11). To verify the endpoint criterion, if $A'(0+) < 0$, small positive effort lowers A and its minimizer cannot be zero. Conversely, if the unique minimizer satisfies $h > 0$, then $A(h) < A(0)$ and convexity implies $A'(0+) \le [A(h) - A(0)]/h < 0$. Thus the equivalence holds under Assumption 2 without strengthening it to strict convexity.

A.3. Comparative Statics and Limits

Implicit differentiation gives (13). Its numerator and denominator are positive since $g_A > h$, so verification increases with resistance on the active branch. On the minimum-deterrence branch, $dg/d\eta = -sR/\eta^2$; on the correction plateau it is zero. The unique maximum along the stated path is x at z_T. Substitution into (8) and the envelope theorem give (12). At $t = x$, the two nonzero derivative formulas equal $-\beta\, s/(2\eta^T)$; at $t = h$, the deterrence derivative tends to zero. This proves Proposition 2 and the stated smooth joining of the value function.

Since $h < g_A < x$ and (13) is positive, g_A has a finite limit as $\eta \to 0$. The active first-order condition forces that limit to minimize A; uniqueness gives h. Consequently $P^* = \sqrt{s\eta\, g_A/R} \to 0$ and $W \to \beta + A(h)$. This also covers $h = 0$ and does not require a strictly decreasing final-error curve.

A.4. Selection and Feasible Burden Cases

Equation (12) implies that G is constant on any initial correction plateau and strictly increases afterward. Under $\eta_D(0) > \eta_D^T$ there is a nonempty strictly increasing minimum-deterrence segment before z_T, so $K_0 < K_T$. On the active segment W strictly increases toward its limit and remains below that limit at every finite z, so $K_T < K_\infty$. No sign restriction on the three levels follows from a resistance ordering when basis costs differ.

https://doi.org/10.3390/jtaer21110414

If $K < K_0$, V is strictly preferred throughout. At $K = K_0$, equality holds initially and over any initial plateau, with V strictly preferred afterward. For $K_0 < K < K_T$, continuity and strict increase after the plateau give one pre-deterrence equality. The same argument gives equality at z_T for $K = K_T$, and one post-deterrence equality for $K_T < K < K_\infty$. At $K = K_\infty$, $G < K$ at every finite state; a larger burden also implies D throughout. Intersect each case with $K \geq 0$. This proves Proposition 3 and Table 2. The feasible intervals in Corollary 1 and numerical cases in Table 3 distinguish economic preference from behavioral deterrence.

On a finite equality branch, differentiating $W_D(z_S) - W_V(\eta_V) - K = 0$ gives $dz_S/d\eta_V = W_{V,\eta}/G_z$. Since z_T is fixed under that perturbation, (19) follows. It does not apply to a plateau tie or an absent finite crossing.

Proof of Corollary 2. Fix a basis and abbreviate $A(g;F) = qg + (1-\pi)Fa(g)$. Let $F_2 > F_1 \geq 0$ and h_i minimize $A(\cdot;F_i)$. Optimality at both parameters implies

$$0 \geq (1-\pi)(F_2 - F_1)[a(h_2) - a(h_1)].$$

Thus $a(h_2) \leq a(h_1)$. If $h_2 < h_1$, monotonicity of a implies $a(h_2) \geq a(h_1)$, hence equality. But then $qh_2 < qh_1$ with the same error probability, contradicting the optimality of h_1 since $q > 0$. Therefore $h_2 \geq h_1$, including zero-effort corners.

For each F, $x(F) > h(F)$ uniquely solves $\phi(g,F) = gA_g(g;F) = \beta s/2 > 0$. At any fixed positive g, $\phi(g,F_2) \leq \phi(g,F_1)$ since $a'(g) \leq 0$. On $g > h(F)$,$\phi_g = A_g + gA_{gg} > 0$. If $x(F_2) < x(F_1)$, then, because $x(F_2) > h(F_2) \geq h(F_1)$, monotonicity of $\phi(\cdot,F_1)$ on that interval would imply $\phi(x(F_2),F_1) < \phi(x(F_1),F_1) = \beta s/2$, while $\phi(x(F_2),F_1) \geq \phi(x(F_2),F_2) = \beta s/2$, a contradiction. Hence $x(F_2) \geq x(F_1)$.

For fixed capability state z, the optimized within-basis loss is $W_j(z,F) = \min_{g\geq 0}\{\beta[1 - P_j(g, e_j^*(g))] + q_j g + (1-\pi)Fa_j(g)\}$. The maintained unique optimum and regularity conditions yield the envelope derivative $\partial_F W_j = (1-\pi)a_j(g_j^*)$. Thus $G_F = (1-\pi)[a_D(g_D^*) - a_V(g_V^*)]$ for $G = W_D - W_V$. At an interior equality $G(z_S(F),F) = K$, suppose G is continuously differentiable in a neighborhood of $(z_S(F),F)$, $G_z > 0$, and K is fixed. The implicit-function theorem then ensures a locally unique differentiable equality branch and gives $z_S'(F) = -G_F/G_z$, which is (16). This derivative does not apply when the crossing is absent or nonregular. □

A.5. Joint Changes and Residual Non-Detection

At an interior active optimum, the envelope derivative of W with respect to η is (12) and its direct z derivative is $A_z(g^*,z)$ when β, R and s are fixed. In the minimum-deterrence regime, explicit differentiation of $\beta(1-s) + A(sR/\eta,z)$ gives the same statement. In the correction regime, the A-minimizer first-order condition (or the boundary optimum when h = 0) removes the derivative of h. Applying the chain rule and subtracting K′ yields (20). A strictly increasing continuous Delta with opposite endpoint signs has a root by continuity and at most one by strict monotonicity. This is a sufficient condition; the counterexample violates it. No division by a possibly zero marginal loss is used.

The claimant calculation and the positive-half-line convexity proof already include s. Hence the correct residual boundary is (9). For a pure-linear total cost $A = qg + b$, $g_A = \beta^2 s\eta/(4q^2R)$, $x = \beta s/(2q)$ and $\eta^T = 2qR/\beta$. Here $W = \beta(1-s) + qsR/\eta + b$ in deterrence and $W = \beta - \beta^2 s\eta/(4qR) + b$ in active manipulation. A nonzero b is not itself multiplied by s. With a nonlinear correction component, this affine scaling is unavailable. The numerical discrepancy is recorded in Supplementary File S1, Section S1.3.

A.6. Simultaneous Global Best Responses

At $(0,0)$, the platform improves with sufficiently small positive g, so the origin is not Nash. If $e = 0$, its positive-g objective is $\beta(1-s) + A(g)$. For $h > 0$, the unique positive best response is h, which dominates zero since $A(h) \leq A(0)$ and $\beta s > 0$. The claimant chooses zero manipulation exactly when $h \geq sR/\eta$. If $h = 0$, the positive-g infimum is not attained because of the jump at zero, so there is no zero-manipulation equilibrium.

For e > 0 the platform objective is strictly convex: second derivative $2\beta se/(g+e)^3 + A''(g) > 0$. Combining its first-order condition $A'(g) = \beta se/(g+e)^2$ with the claimant condition $sRg/(g+e)^2 = \eta$ gives (21). The left side minus the right side of (21) is strictly increasing in g, negative at h (or tends to minus infinity at zero) and positive at t when t > h. Thus, exactly one interior solution exists in (h,t). Its two first-order conditions are global best responses by convexity and concavity. At t = h it joins (h,0).

Holding A, β, R and s fixed, differentiating (21) yields a zero intensity slope exactly when g = t/4. Substitution gives $yA'(y) = \beta s/4$, with a unique y > h. The unique interior intensity peak is attained at $\eta = sR/(4y)$. When h > 0, it is distinct from the zero-manipulation endpoint $\eta = sR/h$; when h = 0, there is no finite zero-manipulation endpoint to compare. The comparison under Table 3 parameters therefore legitimately changes boundary ordering. At every state, the preceding argument supplies a unique pure effort continuation after each possible basis choice under Assumption 2. Let its within-basis loss be N_j as defined in Section 6.3. Backward induction then gives $j_N^* \in \operatorname{argmin}_j\left(N_j + K_j\right)$ at stage one, with both bases admissible at a tie. The resulting basis-first solution uses Nash continuations for both the selected and the unselected basis. This completes the two-stage policy comparison; it does not identify N_j with the commitment value W_j.

A.7. Restricted Recovery of the Original Results

Extending the mixture model to its boundary $\pi = 1$, setting $s = 1$ and $q_D = q_V = q$ removes legitimate-claim costs and gives $\beta = B$. Then $h = 0$, $x = B/(2q)$, $\eta^T = 2qR/B$, $g^* = R/\eta$ in deterrence and $B^2\eta/\left(4q^2R\right)$ in active manipulation. Optimized loss is qR/η or $B - B^2\eta/(4qR)$, respectively. These recover the original fixed-basis equations exactly. $\pi = 1$ is a boundary special case, not a value in the maintained interior mixture assumption. Setting $F = 0$ with $\pi < 1$ instead recovers the same algebra with B replaced by β, not the unchanged original B.

The no-correction possibility also holds at an interior filed-claim mixture. Set $F = 0$, $s = 1$ and $q_D = q_V = q$, retaining $\pi \in (0,1)$ and $\beta = \pi B$. If $\eta_V > \eta_D(0) > 2qR/\beta$, then

$$K_0 = qR[1/\eta_D(0) - 1/\eta_V], \qquad K_T = \beta/2 - qR/\eta_V$$

satisfy $0 < K_0 < K_T$. Proposition 3 therefore yields a unique pre-deterrence crossing for every $K \in (K_0, K_T)$. Early substitution does not require legitimate-claim correction costs; this is the nested linear-cost case, with no additional numerical scenario.

For that restricted model only, normalize $\tilde{\eta} = \eta/(2qR/B)$, $\tilde{g} = 2qg/B$, $\tilde{e} = 2qe/B$, $v = W/B$, $k = K/B$ and $\xi = \log[\eta_D(0)/\eta_D(z)]$. With initial normalized D resistance 2.5 and V resistance 3, the original $\xi_T = 0.9163$, $(k_0, k_T, k_\infty) = (1/30, 1/3, 5/6)$ and L/C/H crossings $(0.4595, 0.9163, 1.1820)$ are correct. The original finite-branch sensitivity at the coincident threshold is 1/9. They can be retained as a compact analytical special case, not relabeled as outcomes of (17).

The normalized burdens producing those L/C/H states are $k_L = 0.15$, $k_C = 1/3$ and $k_H = 0.45$. They are separate from the unnormalized Table 3 burdens $K_L = 0.20$ and $K_H = 0.50$.

A.8. Sanctions and Reverse Prevention Ordering

At positive verification, (22) has marginal payoff $s(R+\tau)g/(g+e)^2 - \eta$ and negative second derivative, proving the global response (23). The platform's active branch is $\beta[1 - \sqrt{(s\eta g/(R+\tau))}] + A(g)$, with boundary derivative $A'(t) - \beta s/(2t)$. Branch matching yields $xA'(x) = \beta s/2$ independently of τ and proves (24). Appendix A's argument also establishes dominance over the zero corner.

A larger sanction weakly increases manipulation and fraud success at every fixed positive verification input, so each basis's minimized platform loss weakly increases. The incumbent transition loss $\beta(1 - s_D) + A_D(x_D)$ is constant in τ; subtracting the alternative value proves that K_T is nonincreasing. Solving $\eta_D(z) = s_D(R+\tau)/x_D$ under the unit-rate path used in Table 3 gives the stated endpoint shift; for exponential decay rate α the shift is divided by α. A positive endpoint requires the incumbent's initial resistance to exceed its sanction-adjusted threshold; otherwise use Proposition 1 from the actual initial regime. Filing remains fixed and sanction receipts are excluded.

At fixed positive input, optimized prevention is $\min\{s, \sqrt{(s\eta g/R)}\}$, nondecreasing in s. Loss is therefore nonincreasing in s pointwise and after minimization; the zero-input value is unchanged. Changing only the fixed alternative ceiling shifts all three Proposition 3 thresholds by the negative of its optimized-value change. Their relative ordering is preserved, but intersection with nonnegative burdens can become empty, irrespective of the ranking of the two ceilings.

References

1. Islam, H.A.; Farrell, M.; Nair, A.; Zhang, J. Understanding Transaction Platform Governance and Conflicts: A Configuration Approach. *Technol. Forecast. Soc. Change* **2023**, *189*, 122382. [CrossRef]
2. Hanisch, M.; Goldsby, C.M.; Fabian, N.E.; Oehmichen, J. Digital Governance: A Conceptual Framework and Research Agenda. *J. Bus. Res.* **2023**, *162*, 113777. [CrossRef]
3. Springer, V.; Randhawa, K.; Jovanović, M.; Ritala, P.; Piller, F.T. Platform Design and Governance in Industrial Markets: Charting the Meta-Organizational Logic. *Res. Policy* **2025**, *54*, 105236. [CrossRef]
4. Sun, C.; Li, X.; Meng, X. From Service Touchpoint to Governance Interface: Anthropomorphic AI, Complaint Severity, and Trust in C2C Platform Complaint Handling. *J. Theor. Appl. Electron. Commer. Res.* **2026**, *21*, 215. [CrossRef]
5. State Administration for Market Regulation; Cyberspace Administration of China. Measures for the Supervision and Administration of Online Trading Platform Rules, Order No. 116. 2025; (English Translation of the Chinese Title; Original in Chinese). Available online: https://www.samr.gov.cn/wljys/gzzd/art/2026/art_74b094e2fda240cbada023e223bc25bb.html (accessed on 23 October 2026).
6. Nightingale, S.J.; Farid, H. AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2120481119. [CrossRef]
7. Spitale, G.; Biller-Andorno, N.; Germani, F. AI Model GPT-3 (Dis)Informs Us Better than Humans. *Sci. Adv.* **2023**, *9*, eadh1850. [CrossRef]
8. Stockner, M.; Convertino, G.; Cambedda, S.; Mazzoni, G. Are Humans Able to Discriminate between Real and Deepfake Faces? A Systematic Review and Meta-Analysis. *Comput. Hum. Behav. Artif. Hum.* **2026**, *9*, 100332. [CrossRef]
9. Zhang, S.; He, E.; Zhan, X.; He, S.; Xiao, R.; Yi, X.; Li, H. Generative AI-Enabled Refund Fraud in Chinese E-Commerce: Investigation on Merchants and Platform Workers. *arXiv* **2026**, arXiv:2606.03215. [CrossRef]
10. Yang, Y.; Liu, J.; Huang, R.; Shuai, Q. Governing AIGC Misuse in E-Commerce Platform Content Ecosystems: A Tripartite Evolutionary Game Analysis of Platforms, Content Creators, and Users. *J. Theor. Appl. Electron. Commer. Res.* **2026**, *21*, 248. [CrossRef]
11. Tiwana, A.; Konsynski, B.; Bush, A.A. Research Commentary—Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. *Inf. Syst. Res.* **2010**, *21*, 675–687. [CrossRef]
12. Constantinides, P.; Henfridsson, O.; Parker, G.G. Introduction—Platforms and Infrastructures in the Digital Age. *Inf. Syst. Res.* **2018**, *29*, 381–400. [CrossRef]
13. Kapoor, K.; Ziaee Bigdeli, A.; Dwivedi, Y.K.; Schroeder, A.; Beltagui, A.; Baines, T. A Socio-Technical View of Platform Ecosystems: Systematic Review and Research Agenda. *J. Bus. Res.* **2021**, *128*, 94–108. [CrossRef]

14. Costabile, C. Digital Platform Ecosystem Governance of Private Companies: Building Blocks and a Research Agenda Based on a Multidisciplinary, Systematic Literature Review. *Data Inf. Manag.* **2024**, *8*, 100053. [CrossRef]
15. Cutolo, D.; Kenney, M. Platform-Dependent Entrepreneurs: Power Asymmetries, Risks, and Strategies in the Platform Economy. *Acad. Manag. Perspect.* **2021**, *35*, 584–605. [CrossRef]
16. Altug, M.S.; Aydinliyim, T.; Jain, A. Managing Opportunistic Consumer Returns in Retail Operations. *Manag. Sci.* **2021**, *67*, 5660–5678. [CrossRef]
17. Zhang, D.; Frei, R.; Senyo, P.K.; Bayer, S.; Gerding, E.; Wills, G.; Beck, A. Understanding Fraudulent Returns and Mitigation Strategies in Multichannel Retailing. *J. Retail. Consum. Serv.* **2023**, *70*, 103145. [CrossRef]
18. Mayzlin, D.; Dover, Y.; Chevalier, J. Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *Am. Econ. Rev.* **2014**, *104*, 2421–2455. [CrossRef]
19. Luca, M.; Zervas, G. Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud. *Manag. Sci.* **2016**, *62*, 3412–3427. [CrossRef]
20. He, S.; Hollenbeck, B.; Proserpio, D. The Market for Fake Reviews. *Mark. Sci.* **2022**, *41*, 896–921. [CrossRef]
21. Townsend, R.M. Optimal Contracts and Competitive Markets with Costly State Verification. *J. Econ. Theory* **1979**, *21*, 265–293. [CrossRef]
22. Mookherjee, D.; Png, I. Optimal Auditing, Insurance, and Redistribution. *Q. J. Econ.* **1989**, *104*, 399–415. [CrossRef]
23. Bond, E.W.; Crocker, K.J. Hardball and the Soft Touch: The Economics of Optimal Insurance Contracts with Costly State Verification and Endogenous Monitoring Costs. *J. Public Econ.* **1997**, *63*, 239–264. [CrossRef]
24. Crocker, K.J.; Morgan, J. Is Honesty the Best Policy? Curtailing Insurance Fraud through Optimal Incentive Contracts. *J. Polit. Econ.* **1998**, *106*, 355–375. [CrossRef]
25. Picard, P. Auditing Claims in the Insurance Market with Fraud: The Credibility Issue. *J. Public Econ.* **1996**, *63*, 27–56. [CrossRef]
26. Dionne, G.; Giuliano, F.; Picard, P. Optimal Auditing with Scoring: Theory and Application to Insurance Fraud. *Manag. Sci.* **2009**, *55*, 58–70. [CrossRef]
27. Becker, G.S. Crime and Punishment: An Economic Approach. *J. Polit. Econ.* **1968**, *76*, 169–217. [CrossRef]
28. Tullock, G. Efficient Rent-Seeking. In *Toward a Theory of the Rent-Seeking Society*; Buchanan, J.M., Tollison, R.D., Tullock, G., Eds.; Texas A&M University Press: College Station, TX, USA, 1980; pp. 97–112.
29. Skaperdas, S. Contest Success Functions. *Econ. Theory* **1996**, *7*, 283–290. [CrossRef]
30. Dixit, A. Strategic Behavior in Contests. *Am. Econ. Rev.* **1987**, *77*, 891–898.
31. Kartik, N.; Ottaviani, M.; Squintani, F. Credulity, Lies, and Costly Talk. *J. Econ. Theory* **2007**, *134*, 93–116. [CrossRef]
32. Kartik, N. Strategic Communication with Lying Costs. *Rev. Econ. Stud.* **2009**, *76*, 1359–1395. [CrossRef]
33. Antinolfi, G.; Carli, F. Costly Monitoring, Dynamic Incentives, and Default. *J. Econ. Theory* **2015**, *159*, 105–119. [CrossRef]
34. Li, Y. Mechanism Design with Costly Verification and Limited Punishments. *J. Econ. Theory* **2020**, *186*, 105000. [CrossRef]
35. Bull, J.; Watson, J. Hard Evidence and Mechanism Design. *Games Econ. Behav.* **2007**, *58*, 75–93. [CrossRef]
36. Deneckere, R.; Severinov, S. Mechanism Design with Partial State Verifiability. *Games Econ. Behav.* **2008**, *64*, 487–513. [CrossRef]
37. Erlanson, A.; Kleiner, A. Costly Verification in Collective Decisions. *Theor. Econ.* **2020**, *15*, 923–954. [CrossRef]
38. Ben-Porath, E.; Dekel, E.; Lipman, B.L. Mechanisms with Evidence: Commitment and Robustness. *Econometrica* **2019**, *87*, 529–566. [CrossRef]
39. Ben-Porath, E.; Dekel, E.; Lipman, B.L. Evidence in Games and Mechanisms. *Annu. Rev. Econ.* **2026**, *18*, 81–104. [CrossRef]
40. Ben-Porath, E.; Dekel, E.; Lipman, B.L. Mechanism Design for Acquisition of/Stochastic Evidence. *Theor. Econ.* **2026**, *21*, 99–131. [CrossRef]
41. Koessler, F.; Skreta, V. Selling with Evidence. *Theor. Econ.* **2019**, *14*, 345–371. [CrossRef]
42. Ball, I.; Kattwinkel, D. Probabilistic Verification in Mechanism Design. *Theor. Econ.* **2025**, *20*, 1247–1284. [CrossRef]
43. Feuerriegel, S.; DiResta, R.; Goldstein, J.A.; Kumar, S.; Lorenz-Spreen, P.; Tomz, M.; Pröllochs, N. Research Can Help to Tackle AI-Generated Disinformation. *Nat. Hum. Behav.* **2023**, *7*, 1818–1821. [CrossRef]
44. Feuerriegel, S.; Hartmann, J.; Janiesch, C.; Zschech, P. Generative AI. *Bus. Inf. Syst. Eng.* **2024**, *66*, 111–126. [CrossRef]
45. Mirsky, Y.; Lee, W. The Creation and Detection of Deepfakes: A Survey. *ACM Comput. Surv.* **2021**, *54*, 7. [CrossRef]
46. Kaur, A.; Noori Hoshyar, A.; Saikrishna, V.; Firmin, S.; Xia, F. Deepfake Video Detection: Challenges and Opportunities. *Artif. Intell. Rev.* **2024**, *57*, 159. [CrossRef]
47. Groh, M.; Sankaranarayanan, A.; Singh, N.; Kim, D.Y.; Lippman, A.; Picard, R. Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video. *Nat. Commun.* **2024**, *15*, 7629. [CrossRef]
48. Kramer, R.S.S.; Jones, A.L.; Fitousi, D.; Tree, J.J. AI-Generated Images of Familiar Faces Are Indistinguishable from Real Photographs. *Cogn. Res. Princ. Implic.* **2025**, *10*, 70. [CrossRef]
49. Barrett, M.; Faik, I.; Jelassi, T. Platform Governance as Institutional Custodianship: Multi-Actor Collaboration in Combating AI-Enabled Mis/Disinformation. *Inf. Organ.* **2025**, *35*, 100590. [CrossRef]

50. Ben-Porath, E.; Dekel, E.; Lipman, B.L. Optimal Allocation with Costly Verification. *Am. Econ. Rev.* **2014**, *104*, 3779–3813. [CrossRef]
51. Kattwinkel, D.; Knoepfle, J. Costless Information and Costly Verification: A Case for Transparency. *J. Polit. Econ.* **2023**, *131*, 504–548. [CrossRef]
52. Li, A.; Yang, M. Optimal Incentive Contract with Endogenous Monitoring Technology. *Theor. Econ.* **2020**, *15*, 1135–1173. [CrossRef]
53. Perez-Richet, E.; Skreta, V. Test Design Under Falsification. *Econometrica* **2022**, *90*, 1109–1142. [CrossRef]
54. Emons, W.; Fluet, C. Accuracy Versus Falsification Costs: The Optimal Amount of Evidence under Different Procedures. *J. Law Econ. Organ.* **2009**, *25*, 134–156. [CrossRef]
55. Gao, Y. On the Use of Overt Anti-Counterfeiting Technologies. *Mark. Sci.* **2018**, *37*, 403–424. [CrossRef]
56. Chen, Y.-C.; Hu, G.; Yang, X. Information Design in Allocation with Costly Verification. *Int. Econ. Rev.* **2025**, *66*, 1267–1285. [CrossRef]
57. Bayer, R.-C. A Contest with the Taxman—The Impact of Tax Rates on Tax Evasion and Wastefully Invested Resources. *Eur. Econ. Rev.* **2006**, *50*, 1071–1104. [CrossRef]
58. Sadakane, H.; Tam, Y.C. Cheap Talk and Lie Detection. *Games Econ. Behav.* **2026**, *158*, 787–814. [CrossRef]
59. Yin, X.; Wen, F. When Trust Attracts Fraud: AI and Trust Arbitrage. *Econ. Lett.* **2026**, *268*, 113234. [CrossRef]
60. Ioannidis, K.; Offerman, T.; Sloof, R. Lie Detection: A Strategic Analysis of the Verifiability Approach. *Am. Law Econ. Rev.* **2022**, *24*, 659–705. [CrossRef]
61. Gao, L.; Lu, J.; Wang, Z. Move Orders in Contests: Equilibria and Winning Chances. *Games Econ. Behav.* **2025**, *150*, 436–468. [CrossRef]
62. Hau, A. Optimal Insurance under Costly Falsification and Costly, Inexact Verification. *J. Econ. Dyn. Control* **2008**, *32*, 1680–1700. [CrossRef]
63. Mastercard. *The Chargeback Window of Opportunity: A Global View of the 2025 Chargeback Trends and How to Turn Them into Opportunities*; Mastercard: Purchase, NY, USA, 2025; Available online: https://www.mastercard.com/content/dam/mccom/shared/news-and-trends/insights/2025/2025-global-chargebacks-outlook/pdf/2025-state-of-chargebacks-report.pdf (accessed on 23 October 2026).
64. Visa Acceptance Solutions; Verifi; Merchant Risk Council. *2025 Global eCommerce Payments & Fraud Report*; Visa: San Francisco, CA, USA, 2025; Available online: https://www.cybersource.com/content/dam/documents/campaign/fraud-report/global-fraud-report-2025.pdf (accessed on 23 October 2026).
65. AfterShip. Pricing & Plans: AfterShip Tracking. Available online: https://www.aftership.com/pricing/tracking (accessed on 23 October 2026).
66. Gawer, A.; Harracá, M. Inconsistent Platform Governance and Social Contagion of Misconduct in Digital Ecosystems: A Complementors Perspective. *Res. Policy* **2025**, *54*, 105300. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.