ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Dong, Xiaoge

Working Paper

Willful ignorance in legal contexts: A mechanism design approach

Center for Mathematical Economics Working Papers, No. 752

Provided in Cooperation with:
Center for Mathematical Economics (IMW), Bielefeld University

Suggested Citation: Dong, Xiaoge (2025) : Willful ignorance in legal contexts: A mechanism design approach, Center for Mathematical Economics Working Papers, No. 752, Bielefeld University, Center for Mathematical Economics (IMW), Bielefeld, https://nbn-resolving.de/urn:nbn:de:0070-pub-30070577

This Version is available at:
https://hdl.handle.net/10419/333505

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.


https://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Center for Mathematical Economics
Working Papers

752

September 2025

Willful Ignorance in Legal Contexts: A Mechanism Design Approach

Xiaoge Dong

Center for Mathematical Economics (IMW)
Bielefeld University
Universitätsstraße 25
D-33615 Bielefeld · Germany

e-mail: imw@uni-bielefeld.de
uni-bielefeld.de/zwe/imw/research/working-papers
ISSN: 0931-6558

Unless otherwise noted, this work is licensed under a Creative Commons Attribution 4.0 International (CC BY) license. Further information:

https://creativecommons.org/licenses/by/4.0/deed.en

https://creativecommons.org/licenses/by/4.0/legalcode.en

Willful Ignorance in Legal Contexts: A Mechanism Design Approach

Xiaoge Dong*

September 24, 2025

Abstract

We study "willful ignorance" - choosing not to learn whether a task is illegal - in a lawmaker-principal-agent game and characterize the penalty policies that implement welfare-maximizing behavior. The model delivers an *implementability frontier*: which equilibrium behaviors can exist and be selected by penalties. With perfect inquiry, this frontier is aligned with the welfare ordering, so the lawmaker can make the welfare-maximizing behavior both exist and be preferred by all parties. With imperfect inquiry, noise breaks that alignment and produces two failures: inquiry that is socially desirable may be infeasible at any penalty, and inquiry that is socially undesirable may persist because it cannot be switched off. We compare harm-based, compliance-based, and dual-penalty rules: harm-based rules preserve control but tightens feasibility; compliance-based rules relax feasibility but sacrifices control; dual penalty rules recover both levers subject to simple bounds. The framework yields practical guidance for calibrating penalties to harm, inquiry accuracy, and inquiry costs. It also implies that ignorance cannot serve as a shield: the absence of knowing crime in equilibrium is driven by incentives rather than morality, making non-inquiry the true strategic margin of liability design.

JEL Codes: K14, K42, D82, D86, H23.

Keywords: willful ignorance, ostrich instruction, law and economics, asymmetric information.

*Zeppelin Universität Friedrichshafen, Fallenbrunnen 3, 88045 Germany.

I thank Niels Boissonnet, Yves Breitmoser, Herbert David, Manuel Förster, Ruveyda Gozen, Martina Miotto, Frank Riedel, and Gerald Willmann for valuable comments and suggestions. Earlier versions of this paper were presented at the BiGSEM colloquium at Bielefeld University, and I thank the participants for their helpful feedback. Financial support from the DFG through the project RUTHLESS is gratefully acknowledged. All remaining errors are my own.

1 Introduction

Willful ignorance - also called "willful blindness" or "deliberate ignorance" - arises when a defendant intentionally avoids learning facts that would trigger legal liability (Kirfel and Hannikainen 2023). Since *United States v. Jewell* (1976), courts have widely adopted the willful-ignorance doctrine, treating failure to inquire as tantamount to knowledge (Charlow 1991; A. Sarch 2018; Hellman 2009; Luban 1998). Yet debate persists about its foundations and the proper design of penalties; definitions vary across jurisdictions and enforcement practice is uneven. Simons (2021) urges caution until clearer, broadly accepted standards emerge. Against this background, the policy question we take up is: *how should penalties be structured so that equilibrium behavior aligns with social welfare when actors may strategically avoid information?*

We study willful ignorance in a lawmaker-principal-agent game. The *lawmaker* sets a penalty rule ex ante. Nature then draws the principal's type (good or bad). The principal's type determines the kind of task he offers: a good principal offers a legal task, while a bad principal offers an illegal one. The *principal* strategically proposes a *transfer* for the task to the agent to induce performance, anticipating the agent's responses and the legal rule. The *agent* decides whether to *inquire* into the legality of the task (at a cost) and whether to perform it; if an illegal task is performed, a *penalty* is imposed. This framework applies directly to compliance domains such as anti-money-laundering due diligence, export controls, and product-safety regulation, where inquiry duties are central and penalties vary in structure. Our objective is to characterize the *"implementability frontier"*-which equilibrium behaviors can both *exist* and be *selected* by penalties-and to map that frontier into welfare and policy guidance.

Our main results can be summarized as follows. *1) With perfect inquiry, the lawmaker can always implement the welfare-maximizing behavior.* Penalties rule out uninformed action; whether the market remains active then depends only on whether the good type's surplus covers the cost of inquiry. If so, the agent performs only legal tasks; if not, the market shuts down. In this setting, the agent never knowingly performs an illegal task, so transfers cannot signal types and no separating equilibrium arises. Penalties determine market composition, while inquiry cost determines whether valuable activity survives-an *alignment* that guarantees welfare optimization. The absence of knowing crime is thus not evidence of higher morality but an endogenous outcome of incentives: all harmful conduct flows through deliberate ignorance, underscoring that ignorance cannot serve as a shield in liability design.

2) With imperfect inquiry, welfare and implementability can diverge. Noise generates false positives and negatives, lowering the value of screening and producing two failures: (i) inquiry may be welfare-maximizing but infeasible, if good types cannot bear inquiry costs plus residual risk; (ii) screening may persist even when shut-down would be better, since penalties cannot fully switch it off. How false positives are treated becomes pivotal. Under *harm-based* rules, agents remain liable after inquiry, preserving leverage to deter screening when it is welfare dominated, but this can also make desirable screening infeasible. Under *compliance-based* rules, meeting the inquiry standard eliminates residual risk, sustaining screening, but removes the lawmaker's off-switch. A *dual-penalty scheme* sets one penalty for non-inquiry and another for post-inquiry exposure, combining the advantages of both approaches. Equalizing the two recovers harm-based rules; setting the post-inquiry penalty to zero recovers compliance-based rules. Dual penalties thus restore control, allowing the lawmaker both to shut off screening when exclusion is optimal and

to sustain it when inquiry is welfare-maximizing.

Relation to the literature. This paper contributes to the legal debate on willful ignorance and the "ostrich instruction" by providing a tractable economic model that complements normative and jurisprudential analyses. Additionally, it relates to the literature on strategic ignorance in economics, but differs by focusing on penalty design and implementability rather than on social preferences or self-image. In the end, it speaks to the optimal-deterrence tradition in law and economics, extending standard prescriptions to environments where ignorance itself is the strategic margin. A full discussion of related work appears in Section 2.

Contribution. This paper makes three contributions. First, it develops the first tractable lawmaker-principal-agent framework for willful ignorance, embedding the inquiry decision in a mechanism-design setting where penalties are chosen ex ante. In doing so, it formalizes the doctrine that deliberate ignorance may be treated like knowledge (the "ostrich instruction") and shows how this affects implementability and welfare. Second, it characterizes the lawmaker's *implementability frontier*, identifying when the welfare-maximizing behavior can be both sustained and selected, and when feasibility and desirability diverge. This extends the literature on strategic ignorance by showing how liability rules-harm-based, compliance-based, and dual-map into equilibrium outcomes. Third, it translates the analysis into operational guidance: calibrate penalties directly to the cutoff conditions that sustain the socially desirable behavior; use harm-based rules when inquiry cannot be verified; reserve compliance-based exemptions for environments with auditable inquiry and tolerance for persistent screening; and deploy dual penalties where flexibility is needed to deter non-inquiry while keeping screening feasible. When enforcement absorbs real resources, implement the target behavior with the *minimal expected penalty* and avoid on-path sanctions. Together, these contributions extend the optimal-deterrence tradition to environments where ignorance itself is the strategic margin, and offer a practical menu for legal design.

Roadmap. Section 2 reviews the literature. Section 3 presents the model. Section 4 develops the results for perfect and imperfect inquiry and compares liability designs. Section 5 discusses applications in legal practice. Section 6 collects extensions. Section 7 concludes. All proofs are presented in the Appendix.

2 Related Literature

The willful-ignorance doctrine addresses a practical tactic: defendants who strategically avoid learning incriminating facts. Since *United States v. Jewell*, 532 F.2d 697 (9th Cir. 1976), courts have permitted juries to treat deliberate ignorance as knowledge-the so-called "ostrich instruction." Legal scholarship has debated the legitimacy of this doctrine (Sarch 2014; Simons 2021; Hellman 2009; Luban 1998), but existing analyses remain largely normative. Our model provides the first tractable mechanism-design framework to analyze willful ignorance as a policy instrument, showing how treating ignorance as culpable affects implementability and welfare, and thereby extending the jurisprudential debate into a formal economic analysis.

Closest to our analysis is Yaffe (2018), who offers a normative defense of willful ignorance and shows that failure to inquire indicates disregard for others' interests. Yaffe's model builds in social-preference concerns, treating ignorance as culpable because it reflects deficient regard for others. Our approach differs in primitives and question. We keep players fully rational and self-interested, and show that willful ignorance arises as an *equilibrium response* to incentives—transfers, penalties, and inquiry costs—rather than from diminished concern for others. This shift allows us to characterize implementability and to analyze how liability rules map into welfare. Social concern enters in an extension in Section 6, as a parameter that substitutes for legal penalties under pooling.

Empirical work speaks to perceptions rather than design. Kirfel and Hannikainen (2023) show that willfully ignorant actors are judged more antisocial than unsuspecting ones but less than knowing violators (see also Alter et al. 2007). These findings inform the legitimacy and likely acceptance of ostrich instructions, but they are orthogonal to the penalty-design problem we study. Our analysis takes incentives as primary and, in extensions, allows for non-material inquiry burdens and social concern as distinct welfare primitives.

Our paper is also related to economic models of strategic ignorance. Kartik et al. (2007) develop a theoretical model of motivated information avoidance, where agents prefer to remain uninformed in order to preserve plausible deniability in communication. Grossman and Van Der Weele (2017) study willful ignorance in social decisions, showing that individuals avoid information about the consequences of their actions to protect their self-image. Related experimental work by Dana et al. (2007) demonstrates how principals deliberately remain ignorant about payoffs to excuse exploitative choices. These papers highlight the behavioral logic of ignorance but do not study optimal penalty design. We differ in focusing squarely on legal enforcement: the lawmaker sets penalties ex ante, and ignorance is treated as a strategic margin in equilibrium. This mechanism-design perspective allows us to derive implementability conditions and to compare alternative liability rules, which is absent in the existing information-acquisition literature.

Finally, our enforcement results connect to the optimal-deterrence tradition (see Polinsky and Shavell 2000). When severity is resource costly, welfare favors the *minimal* penalty that implements the desired behavior; when enforcement intensity is costly, welfare favors the *minimal* intensity. Under imperfect inquiry, these prescriptions interact with residual legal risk after inquiry, explaining why parity can render inquiry infeasible when it is desirable, why exemption can make inquiry hard to switch off when exclusion is better, and how dual penalties can restore control by separating the levers for uninformed action and post-inquiry exposure.

3 Model

The game. We study a market environment with three parties: the lawmaker ($\mathcal{L}$), the principal ($\mathcal{P}$), and the agent ($\mathcal{A}$). The lawmaker drafts and announces a penalty rule: if an *illegal* task is performed, the agent will be convicted and penalized at level $T \in \mathbb{R}_+$. Nature ($\mathcal{N}$) then draws the principal's type $\omega \in \{G, B\}$, with $\Pr(G) = \theta \in (0, 1)$ common knowledge.[1] The principal privately observes ω and offers the agent a corresponding task

1. We assume $0 < \theta < 1$, so that both good and bad types lie in the support of beliefs. This captures the doctrinal idea of "substantial suspicion": the agent attaches positive probability to illegality, so that non-inquiry is a strategic choice rather than innocent ignorance. Legal scholarship diverges on

with transfer $X \in \mathbb{R}_+$. By a slight abuse of notation, we use G (resp. B) to denote both the good (resp. bad) principal and the legal (resp. illegal) task he offers.

The agent does not directly observe legality but can conduct an inquiry at cost $k > 0$. We denote her *belief* that the task is legal by $\mu \in [0, 1]$. After observing X and T, she chooses two strategies in sequence: i) an *inquiry strategy* $\sigma_A^i(X) \in \{i, ni\}$, where i denotes inquiry and ni denotes no inquiry; and ii) an *action strategy* $\sigma_A^a(X, \cdot) \in \{a, na\}$, where a denotes acceptance of the task and na non-acceptance. In the baseline game, belief is fully determined by the information structure: without inquiry, $\mu = \theta$, reflecting the prior probability that the task is legal; with perfect inquiry, $\mu \in \{0, 1\}$, as the inquiry reveals type with certainty.

Timing. The sequence of moves is summarized in Figure 1.

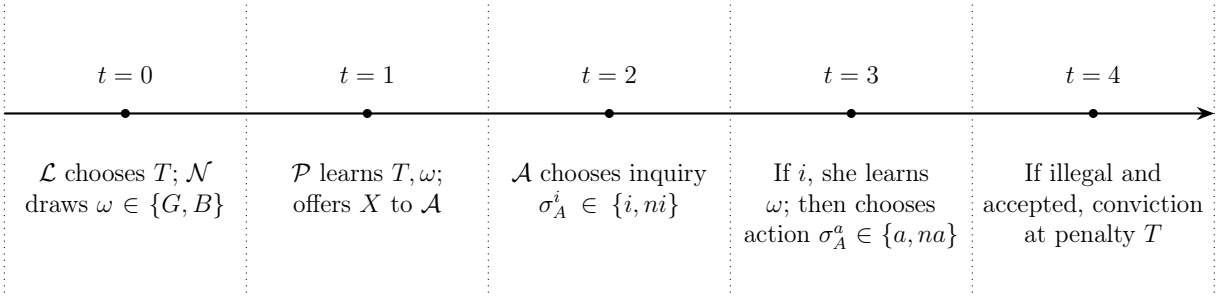


Figure 1: Timing of the game

Notes: The figure summarizes the sequence of moves in the lawmaker-principal-agent game. At $t = 0$, the lawmaker sets the penalty T and nature chooses the principal's type. At $t = 1$, the principal offers a task; at $t = 2$, the agent decides whether to inquire; at $t = 3$, the agent decides whether to perform the task; and at $t = 4$, conviction occurs if the task is illegal.

Payoffs. If the legal task is *performed*, the good principal gets $y_G - X$ and the agent gets $X - k \cdot \mathbf{1}\{i\}$. If the illegal task is *performed* and *illegal*, the bad principal gets $y_B - X$ and the agent gets $X - T - k \cdot \mathbf{1}\{i\}$; society bears harm H. If the task is *not performed*, both principal and agent get 0, and the agent pays the inquiry cost iff she inquired. Transfers and penalties are pure redistributions between principal and agent.

As reference, we formalize the interaction in the game illustrated in Figure A.1, which we refer to as the $\mathcal{LPA}$ game. This game models offenses involving inculpatory propositions of the form: "...where the underlying action would not be independently wrongful absent the defendant's knowledge of the inculpatory proposition." The delegated task-such as transporting a substance or developing software-is not inherently illicit but becomes so when performed with knowledge that the substance is contraband or the software facilitates money laundering. We assume perfect detection of illegal acts, but only the agent bears the legal consequences. We relax these assumptions in section 6.

how demanding this threshold should be. Some argue that any positive probability suffices to ground culpability (e.g. Luban 1998; Hellman 2009), while others emphasize awareness of a high probability of wrongdoing as the correct doctrinal test (e.g. Charlow 1991; A. F. Sarch 2014; Simons 2021). Our baseline collapses the threshold to its minimal value, effectively $s = 0$. One could alternatively introduce a suspicion threshold $s \in (0, 1)$ and treat ignorance as culpable only if $1 - \tilde{\theta} \geq s$ given the agent's subjective belief $\tilde{\theta}$. A higher s enlarges the range in which ignorance is excused, narrowing implementability, but the core logic of penalty design remains unchanged.

Equilibrium concept. Given a fixed penalty T, we analyze the *continuation game* induced by the lawmaker's move. We refer to any perfect Bayesian equilibrium of this subgame as a *continuation equilibrium*. Such an equilibrium satisfies *sequential rationality* and *Bayesian consistency*. In the analysis that follows, we assume that if the agent receives an off-equilibrium offer X', she infers it comes from a bad type principal, refrains from inquiry, and accepts the task only if the legal penalty is fully offset. For off-equilibrium offers X', we denote her belief by $\hat{\mu}$. Formally, this implies:

$$\hat{\mu}(G \mid X') = 0, \quad \sigma_A^i(X') = ni, \quad \text{and} \quad \sigma_A^a(X', ni) = a \iff X' \geq T.$$

This specification of off-path beliefs maximizes the equilibrium set and is adopted without loss of generality.[2] More optimistic conventions ($\hat{\mu} > 0$) shrink the existence windows but do not alter the qualitative welfare rankings, so our main results are robust to nearby belief specifications. We adopt standard tie-breaking in favor of the preceding mover: indifferent agents accept; indifferent principals offer the smallest transfer that induces acceptance. Alternative tie-breaking rules shift only knife-edge boundaries and do not affect the welfare rankings reported below. Given a continuation equilibrium, we classify the structure of play into four types:

i. *Pooling*: Both types of principals offer the same transfer; the agent does not inquire and performs the task. Beliefs are not updated. Formally put: $X_B = X_G$, $\sigma_A^i(\cdot) = ni$, $\sigma_A^a(\cdot) = a$.

ii. *Semi-pooling*: Both types offer the same transfer; the agent inquires before deciding. Beliefs are updated. Formally put: $X_B = X_G$, $\sigma_A^i(\cdot) = i$, $\sigma_A^a(\cdot) \in \{a, na\}$. Because the key feature of this equilibrium is that the agent pays the inquiry cost and conditions her action on the inquiry result, we will refer to this equilibrium type as *screening* in the analysis that follows. "*Semi-pooling*" and "*screening*" are thus interchangeable terms in what follows.

iii. *Separating*: Types separate via transfer; the agent does not inquire and decides based on the offer. Beliefs are not updated. Formally put: $X_B > X_G$, $\sigma_A^i(\cdot) = ni$, $\sigma_A^a(\cdot) \in \{a, na\}$.

iv. *Inactive*: Both types offer below-penalty transfer; the agent does not inquire and refuses the task. Beliefs are not updated. Formally put: $X_B, X_G \in [0, T)$, $\sigma_A^i(\cdot) = ni$, $\sigma_A^a(\cdot) = na$.

Nontrivial mixed-strategy equilibria do not exist in our environment; equilibrium behavior is exhausted by the pure types we analyze.[3]

Social welfare and benchmarks. We use a utilitarian welfare measure: the sum of expected payoffs across players minus expected social harm. We include the bad principal's payoff with a normative weight $\tau \in [0, 1]$ to capture different policy views: τ near 0 fits predatory or cross-border crimes where the offender's gain is not valued; τ near 1 fits local externalities (e.g., pollution abatement failures) where the principal's output still counts. Once a behavior is fixed, transfers and penalties are pure redistributions and do not affect SW (penalties are resource-costless; see Section 6 for costly enforcement).

2. Formal bounds for general $\hat{\mu} \in [0, 1]$ are provided in Appendix C.2.

3. See Appendix C.3 for a formal argument.

We will compare our results to two benchmarks: (i) the *Perfect Information Equilibrium* (PIE), where the principal's type is common knowledge; and (ii) the *No Inquiry Equilibrium* (NIE), where type is unknown and players never inquire.

Lemma 3.1 (Benchmark welfare under PIE and NIE)**.** *Under PIE,*

$$\mathrm{SW}_{PIE} = \begin{cases} \theta\, y_G + (1-\theta)(\tau y_B - H), & \text{if } \tau y_B - H \geq 0, \\ \theta\, y_G, & \text{if } \tau y_B - H < 0, \end{cases}$$

and under NIE,

$$\mathrm{SW}_{NIE} = \begin{cases} \theta\, y_G + (1-\theta)(\tau y_B - H), & \text{if } \theta y_G + (1-\theta)(\tau y_B - H) \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Moreover, $\mathrm{SW}_{NIE} \leq \mathrm{SW}_{PIE}$*, with strict inequality whenever* $\tau y_B - H < 0$*.*

Proof. All proofs are relegated to the Appendix. □

Intuition. If the bad type's net contribution is nonnegative ($\tau y_B - H \geq 0$), both PIE and NIE feature both types, so welfare coincides. If it is negative, PIE can exclude the bad type while NIE shuts the market down (welfare $0 < \theta y_G = \mathrm{SW}_{\mathrm{PIE}}$). A full proof appears in Appendix B.1.

Implementation objective. Let $\mathcal{E} = \{\text{pooling}, \text{semi-pooling}, \text{inactive}\}$ denote the set of equilibrium behaviors, and write $\mathcal{E}(T)$ for the behaviors sustained under penalty T. For each $e \in \mathcal{E}$, let $\mathrm{SW}(e)$ denote the associated level of social welfare. We call an equilibrium behavior *socially desirable* if it maximizes welfare,

$$e^\star \in \arg\max_{e \in \mathcal{E}} \mathrm{SW}(e).$$

The lawmaker chooses T to *implement the socially desirable behavior*, i.e. to ensure that $e^\star \in \mathcal{E}(T)$. When $\mathcal{E}(T)$ contains multiple behaviors, we adopt a minimal *implementation convention*: whenever possible, we pick T in the *interior of the penalty region* that sustains $e^\star$ to secure uniqueness; if uniqueness cannot be forced, we credit $e^\star$ as implemented only when it is *uniformly preferred (Pareto-dominant)* among coexisting behaviors (every player weakly prefers $e^\star$ to any alternative and at least one player strictly prefers it). This convention is not an additional instrument for the lawmaker, but merely a tie-breaking device for exposition. When Pareto dominance does not hold, we adopt the conservative stance and treat the desirable behavior as non-implementable.

4 Analysis

Equilibrium preliminaries. Two observations will be used throughout. First, with perfect inquiry ($\alpha = 1$), no equilibrium separates in transfers: if $X_B \neq X_G$, the agent would skip inquiry, infer type from the offer, and the bad type would profitably mimic the good type. Second, whenever an illegal task is performed with positive probability on the path, it occurs only without inquiry (willful ignorance): if the agent inquired and still performed the illegal task, she would also perform the legal one after inquiry, so inquiry

would be payoff-irrelevant and strictly dominated by not inquiring. Consequently, under $\alpha = 1$ social welfare equals $\theta y_G - k$ in semi-pooling, $\theta y_G + (1-\theta)(\tau y_B - H)$ in pooling, and 0 in inactive. (Proofs are in Appendix B.2.)

Benchmark $T = 0$. When $T = 0$ (for any $\alpha \in (1/2, 1)$), there is a pooling equilibrium in which both types offer $X = 0$, the agent does not inquire, and all tasks are performed; welfare is $\theta y_G + (1-\theta)(\tau y_B - H)$. This coincides with the no-inquiry equilibrium (NIE) used above. If the bad type's task is socially beneficial ($\tau y_B - H > 0$), this benchmark is welfare maximizing and no legal intervention is needed. Henceforth we focus on the relevant case $\tau y_B - H < 0$.

4.1 Baseline - Perfect Inquiry

We start with the baseline game and use the implementation objective stated in Section 3: choose T to implement the socially desirable behavior.

Proposition 4.1 (Perfect inquiry)**.** *In the baseline with perfect inquiry, the lawmaker can always induce socially desirable behavior through choice of T. Consequently, maximal welfare under perfect inquiry weakly exceeds the No Inquiry Equilibrium (NIE) benchmark.*

Intuition. At low harm-and thus low penalties-both principal types remain and the agent acts without inquiry (pooling). Raising T makes action without inquiry unattractive and pushes out the bad type. What remains depends only on the good type's ability to fund inquiry: if $\theta y_G > k$, he pays for inquiry and the agent performs only the legal task; if $\theta y_G \leq k$, even the good type cannot support inquiry and the market shuts down. Under perfect inquiry, the same inequality $\theta y_G \gtrless k$ governs both feasibility and welfare, so the lawmaker can pick T to implement exactly the welfare-maximizing behavior. Since the lawmaker can use T as leverage to eliminate uninformed illegal performance while preserving legal activity whenever $\theta y_G > k$, maximal attainable social welfare is weakly higher than that of NIE benchmark.

Numerical Example(perfect inquiry). For the parameters in Figure 2, we have $\theta y_G - k = 0.4 \cdot 4 - 0.5 = 1.1 > 0$. With perfect inquiry, the lawmaker can implement welfare $\max\{\theta y_G - k,\ \theta y_G + (1-\theta)(\tau y_B - H)\}$. Under the no inquiry equilibrium, the guarantee is $\max\{0,\ \theta y_G + (1-\theta)(\tau y_B - H)\}$. Hence whenever $-k > (1-\theta)(\tau y_B - H)$ (i.e., $\tau y_B - H < -\frac{5}{6}$ here), perfect inquiry delivers strictly higher welfare.

Ignorance as a strategic margin. One major implication from our analysis is that ignorance is not a shield. Choosing not to inquire is itself a rational, strategic choice: the agent balances the private cost of inquiry against the expected penalty from remaining ignorant, and may optimally prefer ignorance whenever the latter is lower. In this sense, ignorance is not a deficiency of knowledge but a calculative decision. If the law were to exempt ignorance from liability, every harmful task would be undertaken without inquiry, and penalties would lose all deterrent power. Liability design must therefore focus on deterring ignorance, not merely punishing knowing violations, by ensuring that acting without inquiry is privately unattractive.

Meanwhile, the absence of knowing crime in equilibrium is endogenous, not moral. One might be tempted to interpret the non-existence of knowing crime as evidence of a moral constraint, or to hard-wire an "exclusion of knowing crime" principle into the model.

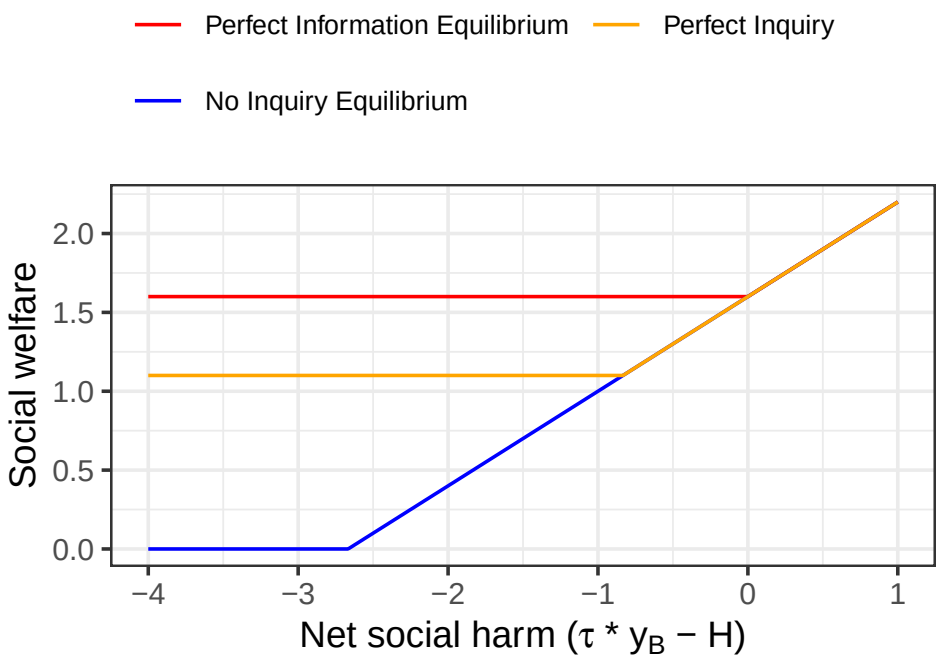


Figure 2: Social welfare under perfect inquiry

Notes: Parameters are $\theta = 0.4$, $y_G = 4$, $k = 0.5$, $\tau = 1$, and $H = 8$. The x-axis plots net social harm, $\tau y_B - H$; the y-axis plots social welfare. Curves compare Perfect Information Equilibrium (PIE), No Inquiry Equilibrium (NIE), and Perfect Inquiry (PIq). For $\tau y_B - H \geq 0$, pooling is efficient and all curves coincide. As net harm turns negative, PIE excludes the bad task and yields θy_G. PIq switches from pooling to screening when $\theta y_G \geq k$, delivering $\theta y_G - k$; if $\theta y_G < k$, PIq becomes inactive (zero). NIE tracks pooling while $\theta y_G + (1-\theta)(\tau y_B - H) \geq 0$ and otherwise falls to zero.

Yet doing so would not alter outcomes: knowing crime is already ruled out by rational incentives. Conversely, the fact that agents refrain from knowingly illegal behavior is not proof of a higher moral standard-it simply reflects that inquiry is privately dominated when illegality is certain. From a policy perspective, this underscores that the true strategic margin is the choice to remain ignorant, and that liability rules must target this margin if socially harmful behavior is to be deterred.

Penalty design and conventional benchmarks.

Lemma 4.1 (Calibration of T to implement $e^\star$ (perfect inquiry))**.** *Assume perfect inquiry. Let* $e^\star \in \{\text{pooling}, \text{semi}, \text{inactive}\}$ *be the socially desirable behavior (per our implementation convention). Then* T *can be placed as follows:*

(i) If $e^\star \in \{\text{semi}, \text{inactive}\}$*, choose*

$$T \;>\; \max\left\{ y_B,\; y_G + \tfrac{k}{1-\theta} \right\}.$$

(ii) If $e^\star = \text{pooling}$*, choose*

$$\begin{cases} T \;\leq\; \frac{y_G}{1-\theta} & \textit{(applies when } \theta y_G \leq k\textit{)}, \\ T \;=\; \frac{k}{\theta(1-\theta)} & \textit{(applies when } \theta y_G > k\textit{)}. \end{cases}$$

Intuition. With perfect inquiry, penalties serve one purpose: they eliminate uninformed action. Once uninformed action is ruled out, whether the market remains active depends only on the trade-off between the good type's surplus and the cost of inquiry: if the surplus covers the cost, the agent screens and undertakes only legal tasks (Semi); otherwise the market shuts down (Inactive). This is why the calibration for Semi and Inactive coincides—both first raise T to exclude pooling; which of the two results then hinges solely on whether the surplus covers the cost. For Pooling, there are two situations: when gains from inquiry are *weak*, keeping T below the acceptance cap sustains pooling; when gains are *strong*, placing T exactly at the indifference knife-edge blocks screening (which needs a strict improvement), so pooling persists. Conventional one-size penalties fail for this reason: *strict liability* ($T = H$) can overshoot the feasibility caps and deter valuable activity, while a *net-harm* penalty ($T = H - \tau y_B$) can be under the indifference threshold and leave uninformed pooling in place precisely when screening is desirable.

4.2 Imperfect Inquiry and Liability Design

In legal practice, a defendant's conduct is often judged not only by outcomes but also by whether she undertook a reasonable inquiry–for example, consulting counsel, compliance officers, or approved testing protocols–and such inquiries are rarely perfectly accurate. In *United States v. Ratzlaf*, 510 U.S. 135 (1994), the defendant structured cash withdrawals into smaller amounts to avoid triggering currency-transaction reports under the Bank Secrecy Act. Ratzlaf admitted knowledge of the reporting rule but claimed he did not know that structuring itself was illegal, pointing to regulatory complexity. The Court ruled for Ratzlaf, underscoring that both the *content* and the *reliability* of inquiry bear legal weight. To reflect this concern, we now relax perfect inquiry.

Information environment. At $t = 3$, an inquiry returns a *signal* of ω back to the agent. A signal generates *false positives* (a good signal for an illegal task) and *false negatives* (a bad signal for a legal task) with symmetric accuracy $\alpha \in (1/2, 1)$: both tasks yield a correct signal with probability α. Bayes' rule then yields posterior beliefs:

$$\mu(g) = \Pr(G \mid g) = \tfrac{\alpha\theta}{\alpha\theta+(1-\alpha)(1-\theta)}, \qquad \mu(b) = \Pr(G \mid b) = \tfrac{(1-\alpha)\theta}{(1-\alpha)\theta+\alpha(1-\theta)}.$$

We continue to assume the agent's inquiry decision and precision are verifiable.[4] The extended game form is depicted in Figure A.2 in Appendix.

Noise affects both *welfare* and *existence*. First, inquiry-based screening becomes less valuable: some illegal tasks slip through (false positives) and some legal tasks are blocked (false negatives). This lowers the welfare of semi-pooling while leaving pooling and inactive welfare unchanged. Second, the existence regions for pooling and semi-pooling shift. As a result, *desirability* and *existence* of screening can diverge: at a given penalty,

4. When we later allow choice of precision in an extension, verifiability of the chosen α and the associated cost is required.

screening may exist when market shut-down would yield higher welfare, or fail to exist when screening is socially desirable.

We now compare three liability designs, treated in parallel.

4.2.1 Harm-based liability (false positives penalized)

Legal posture. Harm-based penalty applies whether or not the agent inquired. This rule treats illegal performance identically whether the agent inquired or not: if an illegal task is performed, a penalty applies. This mirrors strict harm-based regimes (e.g., strict products liability, some environmental penalties) where penalties scale with damage regardless of inquiry accuracy.

Economic properties. Harm-based rules preserve *selection leverage*: by raising the penalty sufficiently, the lawmaker can always eliminate uninformed action (pooling). However, because the agent remains exposed to penalty risk even *after* inquiry (due to false positives), screening requires a transfer high enough to cover residual exposure and the inquiry cost, while remaining profitable for the good principal. As α falls or k rises, these feasibility constraints *tighten*.

Proposition 4.2 (Imperfect inquiry under harm-based rules)**.** *Under harm-based liability, the lawmaker can always deter inquiry when it is not socially desirable. However, for non-empty parameter ranges (lower inquiry accuracy and/or higher inquiry cost), inquiry is socially desirable yet* infeasible *for any penalty. Social welfare is (weakly) lower than under perfect inquiry and (weakly) higher than under the No Inquiry Equilibrium (NIE).*

Intuition Under harm-based liability, the agent faces penalties whenever an illegal task is performed-*even if she inquired*. Inquiry therefore leaves her with *residual legal exposure* (because false positives occur). To support inquiry in equilibrium, the penalty must be *high enough* to make action without inquiry unattractive, yet *not so high* that the good principal can no longer profitably cover both the inquiry cost and the agent's residual exposure. These two requirements bound the penalty from opposite sides and can conflict when inquiry is noisy (low precision) or costly, making inquiry *privately infeasible* even though it would raise welfare. By contrast, the lawmaker can always *turn inquiry off*-simply raise the penalty until action without inquiry is deterred and funding inquiry is unprofitable. Because inquiry no longer perfectly screens (some illegal tasks slip through and some legal tasks are blocked), the best attainable welfare is (weakly) below the perfect-inquiry benchmark; because the lawmaker can at least deter action without inquiry, it is (weakly) above the No Inquiry Equilibrium.

Numerical Example (harm-based). Relative to perfect inquiry, welfare under semi-pooling is lower by exactly the probability mass of false signals, while pooling and inactive welfare are unchanged. Figure 3 illustrates: for parameters where semi-pooling exists, its welfare lies between perfect-inquiry screening and the No Inquiry Equilibrium; when feasibility collapses, the best attainable equilibrium reverts to pooling or inactive.

4.2.2 Compliance-based liability (false positives exempt)

Legal posture. Compliance-based penalty applies if the agent did not inquire, and exempt her from penalty for false positive. Here the agent who meets a stipulated inquiry standard (precision α) is *exempt* if an illegal task occurs because of a false positive. This is analogous to negligence-style safe harbors: due care (proper inquiry) defeats liability.

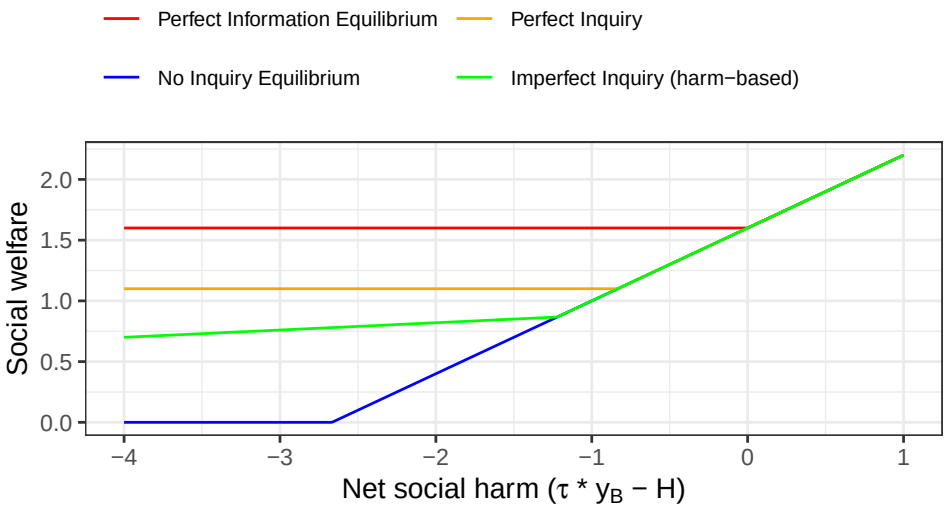


Figure 3: Social welfare under harm-based imperfect inquiry

Notes: Parameters are $\theta = 0.4$, $y_G = 4$, $k = 0.5$, $\alpha = 0.8$, $\tau = 1$, and $H = 8$. The x-axis plots net social harm, $\tau y_B - H$; the y-axis plots social welfare. Curves compare Perfect Information Equilibrium (PIE), No Inquiry Equilibrium (NIE), Perfect Inquiry (PIq), and Imperfect Inquiry (harm-based), which treats post-inquiry violations like willful ignorance. For $\tau y_B - H \geq 0$, pooling is efficient and curves coincide. As net harm turns negative, PIq can switch from pooling to screening at the inquiry threshold. The harm-based curve lies weakly below PIq because noisy inquiry both lets some illegal tasks slip through and blocks some legal ones, and also tightens the incentive bounds needed to sustain screening. Where those bounds conflict, screening is infeasible for any penalty and the harm-based curve drops to NIE (or zero) even when PIq sustains positive welfare.

Economic properties. Exemption removes the agent's residual penalty exposure once she inquires, *expanding* the feasibility of screening compared to harm-based rules. However, after uninformed action is ruled out, the penalty no longer affects post-inquiry incentives, so the lawmaker *loses selection leverage*: penalties cannot be used to turn screening off; screening can persist even when market shut-down is preferred.

Proposition 4.3 (Imperfect inquiry with compliance-based exemption)**.** *Under compliance-based liability (false positives are exempt), the lawmaker can induce the socially desirable behavior for all parameter values* except *in two non-empty ranges:*

(i) Inquiry is socially desirable but not as equilibrium behavior: even though the lawmaker prefers the agent to inquire, no penalty can induce inquiry as equilibrium behavior.

(ii) Inquiry is not socially desirable but as equilibrium behavior: the lawmaker prefers no market entry, yet any penalty still leaves inquiry-based participation in play and it cannot be ruled out.

Relative to perfect inquiry, the best attainable welfare is weakly lower. In range (ii), if inquiry-based activity remains among the outcomes, welfare can be weakly below the No Inquiry Equilibrium.

Intuition. Under compliance-based liability, an agent who inquires and then follows the signal ("act only when the signal indicates legality") bears *no* legal risk from false

positives. This makes inquiry privately attractive; the penalty T now bites only when she would act *without* inquiry. Two non-empty failures follow. First, if the good principal's expected legal surplus after inquiry is too small to cover the inquiry cost, inquiry can be *socially desirable* yet *privately infeasible* for any penalty-because T cannot raise the good principal's surplus. Second, if that surplus is large, inquiry remains feasible *for all sufficiently high* T; the lawmaker can no longer use penalties to switch inquiry off. Thus when full exclusion is preferred, inquiry may persist in equilibrium and deliver (weakly) lower welfare than the No Inquiry Equilibrium.

Numerical Example (compliance-based). When screening is desirable but infeasible under harm-based rules, exemption can restore feasibility and raise welfare above the No Inquiry Equilibrium. Conversely, when exclusion is preferred, exemption can make screening *harder to rule out*. Figure 4 illustrates a case in which screening persists alongside inactive and lowers welfare relative to the benchmark without screening.

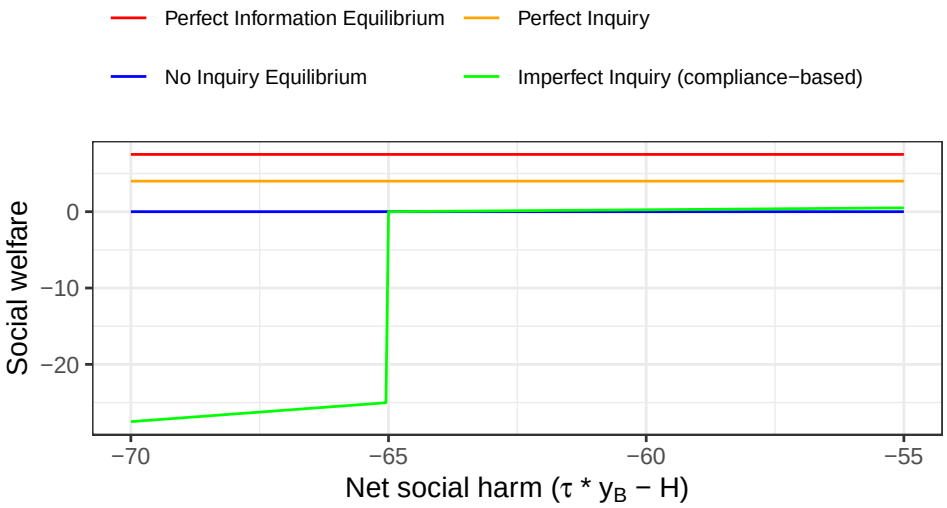


Figure 4: Social welfare under compliance-based imperfect inquiry

Notes: Parameters are $\theta = 0.5$, $y_G = 15$, $k = 3.5$, $\alpha = 0.9$, $\tau = 1$, and $H = 80$. The x-axis plots net social harm, $\tau y_B - H$; the y-axis plots social welfare. Curves compare Perfect Information Equilibrium (PIE), No Inquiry Equilibrium (NIE), Perfect Inquiry (PIq), and Imperfect Inquiry (compliance-based), where agents who meet the inquiry standard are exempt from penalties for false positives. Exemption makes screening easy to sustain (the curve tracks PIq where feasible), but removes the penalty lever that would otherwise switch screening off. As a result, screening can persist alongside inactivity even when full exclusion would yield higher welfare, and in such regions social welfare can fall weakly below the NIE curve. In this parameterization, misselection arises for $\tau y_B - H < -65$ (i.e., $y_B < 15$).

4.2.3 Dual penalties (willful ignorance vs. post-inquiry illegality)

Legal posture. A dual scheme assigns T_n to illegal performance without inquiry (willful ignorance) and T_i to illegal performance after inquiry (knowledge/mistake). It nests harm-based ($T_i = T_n$) and approximates compliance-based by taking $T_i \approx 0$. Jurisprudentially, it separates culpability based on mental state and evidentiary verifiability.

Economic properties. Dual penalties *decouple* the two binding margins under noise: use T_n to deter uninformed action (selection leverage) and set T_i to preserve inquiry

feasibility by limiting residual exposure after inquiry.

Proposition 4.4 (Imperfect inquiry with dual penalties)**.** *Consider imperfect inquiry and a dual-penalty rule differentiating willful ignorance and knowing crime. The lawmaker can implement the* socially desirable *continuation behavior in all cases, except when inquiry is desirable but the good type cannot finance the expected inquiry cost.*

Relative to single-penalty regimes, dual penalties weakly expand implementability and weakly raise the maximal attainable welfare; welfare remains (weakly) below the perfect-inquiry benchmark.

Intuition. With two levers the lawmaker separates the two margins. The no-inquiry penalty T_n prices uninformed action: raising T_n makes acceptance without inquiry expensive and can always shut down pooling; lowering T_n preserves it when desired. The after-inquiry penalty T_i prices the agent's *residual* legal exposure from false positives; lowering T_i reduces the expected transfer she needs to be willing to inquire, and raising T_i makes inquiry unattractive. Under a semi-pooling offer (both types post the same transfer and the agent inquires), her minimal per-performance payment equals

$$t^{\text{semi}} = \tfrac{k+(1-\theta)(1-\alpha)T_i}{\varphi}, \quad \text{where } \varphi := \theta\alpha + (1-\theta)(1-\alpha)$$

is the probability that inquiry leads to performance. The good principal pays t^{semi} only when his task is actually performed (probability α), so semi-pooling is privately feasible iff $y_G \geq t^{\text{semi}}$, i.e. $\varphi y_G \geq k + (1-\theta)(1-\alpha)T_i$. By setting T_i as low as needed (down to 0), the lawmaker makes feasibility easiest; thus inquiry is implementable iff $\varphi y_G > k$. Conversely, if inquiry is *not* desired, choosing T_i large pushes $t^{\text{semi}} > y_G$ so no one funds screening; T_n then selects between pooling and inactivity. Hence dual penalties can always "switch screening on or off" except in the single obstruction $\varphi y_G \leq k$. Because (T_n, T_i) nest harm-based ($T_i = T_n$) and compliance-based ($T_i = 0$) rules, the maximal welfare attainable with dual penalties weakly dominates what either single-penalty regime can achieve, and remains (weakly) below the perfect-inquiry benchmark when $\alpha < 1$.

Numerical Example (dual-penalty). Under dual penalties, the realized welfare at a given equilibrium is as under the other regimes (transfers and penalties cancel in welfare); the gain comes from *which* equilibrium is implementable and selected. The lawmaker calibrates penalties by setting T_n high enough to deter uninformed action while keeping T_i low enough to maintain inquiry feasibility.

4.3 Calibration and comparative analysis under imperfect inquiry

Relative to the perfect-inquiry benchmark, the placement of penalties changes in two ways. First, the lower bound that makes inquiry attractive rises as accuracy falls: with noisier signals, the agent requires higher compensation to cover her expected exposure to penalties when she investigates. Formally, the indifference cutoff $k/[\theta(1-\theta)]$ is replaced by

$$L(\alpha) = \tfrac{k}{\theta(1-\theta)(2\alpha-1)}, \qquad \alpha \in (1/2, 1),$$

so $L'(\alpha) < 0$: higher accuracy reduces the penalty needed to trigger inquiry. Second, screening is now subject to a feasibility constraint,

$$U(\alpha) = \tfrac{\varphi(\alpha)y_G - k}{(1-\theta)(1-\alpha)}, \qquad \varphi(\alpha) = \theta\alpha + (1-\theta)(1-\alpha),$$

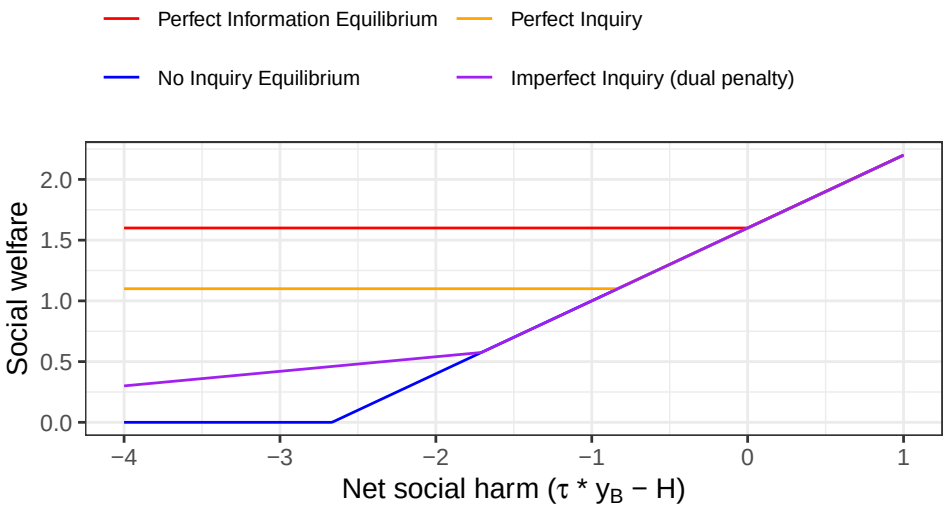


Figure 5: Social welfare under dual-penalty imperfect inquiry

Notes: Parameters are $\theta = 0.4$, $y_G = 4$, $k = 0.5$, $\alpha = 0.8$, and $\tau = 1$. The x-axis plots net social harm, $\tau y_B - H$; the y-axis plots social welfare. Curves compare Perfect Information Equilibrium (PIE), No Inquiry Equilibrium (NIE), Perfect Inquiry (PIq), and Imperfect Inquiry (dual-penalty), where the lawmaker can set separate penalties for action without inquiry and for post-inquiry violations. For $\tau y_B - H \geq 0$, pooling is efficient and all curves coincide. As net harm turns negative, PIq switches from pooling to screening. The dual-penalty curve also selects screening where feasible but lies weakly below PIq because noisy inquiry both lets some illegal tasks slip through and blocks some legal ones. When $\tau y_B - H$ is sufficiently negative, screening ceases to be profitable under noise and the dual-penalty curve falls to zero (inactive), while PIq can still sustain positive welfare by screening.

so that if accuracy is too low or inquiry costs too high, this constraint binds and the inquiry window collapses even when welfare would favor screening. By contrast, the cutoff for acceptance without inquiry, $T \leq y_G/(1-\theta)$, and the requirement that T be at least as large as the bad type's output y_B remain the same as under perfect inquiry.

These thresholds deliver transparent comparative statics. A higher α enlarges the inquiry window (lower L, higher U), making screening easier to implement. A higher k raises L and lowers U, squeezing the inquiry window and pushing the economy toward pooling or inactivity. Greater θ or higher y_G expand feasibility by relaxing the acceptance cap and raising U, while a larger y_B makes deterrence harder under harm-based rules and raises the cutoff for inactivity. Net harm $(\tau y_B - H)$ affects only welfare rankings, not feasibility.

Across regimes, the calibration logic is unified. With a single penalty, the lawmaker calibrates T directly into the range that sustains the socially desirable outcome: raising it to the indifference cutoff when inquiry must be induced, or lowering it to the feasibility bound when inquiry must be deterred. Under harm-based rules, inquiry can always be deterred by raising T but not always sustained when desirable because the feasibility cap may bind. Under compliance-based rules, both problems appear: if $\varphi(\alpha)y_G \leq k$, inquiry is desirable but infeasible; conversely, once uninformed action is deterred and $\varphi(\alpha)y_G > k$, inquiry persists for all higher T, so exclusion cannot be restored by penalties alone. Dual penalties separate the margins: T_n regulates inquiry versus no-inquiry, while T_i prices residual false-positive risk. This flexibility allows the lawmaker to sustain inquiry when-

ever $\varphi(\alpha)y_G > k$ and to switch it off when exclusion is optimal, though at the practical cost of specifying and enforcing two distinct penalties. All explicit cutoff expressions and comparative statics are collected in Appendix B.4.

5 Applications

The preceding analysis derived welfare comparisons and illustrated them with diagrams based on stylized parameter values. These figures already showed how different penalties map into pooling, screening, or exclusion, and how welfare rankings depend on θ, α, k, y_G, and y_B. What remains is to connect these predictions to real-world settings. This section illustrates the model with legal cases where variation in parameters such as harm H, economic contribution τ, or inquiry accuracy α played a decisive role.

Environmental enforcement: Reserve Mining (1975) vs. HF Sinclair Navajo (2025)

Reserve Mining (D. Minn. 1974; 8th Cir. 1975). Reserve Mining discharged taconite tailings into Lake Superior, releasing asbestos-like fibers into the drinking water of Duluth and nearby communities. The district court, after reviewing epidemiological evidence, concluded that the cancer and respiratory risks were intolerable even under uncertainty, and ordered an immediate shutdown, stressing that "human health must come first."[5] On appeal, the Eighth Circuit softened this stance, holding that "no harm to the public health has been shown to have occurred to this date and the danger to health is not imminent."[6] Instead of closure, the company was ordered to prepare on-land disposal facilities at an estimated cost of $243-300 million (around $1.3-1.6 billion in 2024 dollars).[7] At the time, toxicological studies placed ingestion risks only slightly above random inference ($\alpha \approx 0.55$-0.6), and monitoring costs k were high because long-term sampling campaigns required millions in expenditure.[8] Regional economic contribution was substantial: 3,050 direct jobs and 12,000 dependent jobs.[9]

This sequence illustrates willful ignorance: the company had internal warnings but avoided commissioning comprehensive studies, relying on scientific uncertainty to continue operations. The district court prioritized harm H as catastrophic and imminent, while the appellate court emphasized economic contribution τ and downgraded H to "not imminent." With α low and k high, the feasibility of semi-pooling (screening) was tenuous. In the model, this would place inquiry near the boundary of feasibility; in practice, the courts tolerated continued operation in the short run but coupled it with phased compliance obligations. This resembles present-period pooling combined with an order that compels transformation into compliance in subsequent periods.

5. *United States v. Reserve Mining Co.*, 380 F. Supp. 11, 26-28 (D. Minn. 1974).

6. *Reserve Mining Co. v. EPA*, 514 F.2d 492, 538 (8th Cir. 1975). EPA = Environmental Protection Agency.

7. *Reserve Mining Co. v. Herbst*, 262 N.W.2d 596, 605 (Minn. 1977) ("over $300 million" for on-land disposal); 514 F.2d at 505 ($243m estimate for Milepost 7 project).

8. See Gerald Markowitz & David Rosner, *Deceit and Denial: The Deadly Politics of Industrial Pollution* (2002), ch. 6.

9. *TIME*, Oct. 22, 1973, "Environment: Crisis in Silver Bay."

HF Sinclair Navajo (D.N.M. 2025). In 2025, HF Sinclair Navajo Refining entered a consent decree requiring a $35 million civil penalty and $137 million in injunctive compliance investments, including flare gas recovery, wastewater upgrades, and advanced monitoring through Continuous Emissions Monitoring Systems (CEMS) and Leak Detection and Repair (LDAR).[10] Management had received repeated violation notices but delayed installing available monitoring technology until forced by enforcement.

Here, inquiry was feasible but deliberately avoided. Modern monitoring precision was high ($\alpha \approx 0.9$), with EPA quality assurance protocols for CEMS targeting relative accuracy within 10-20%.[11] Inquiry costs k were moderate, with EPA manuals estimating CEMS operation and maintenance at $13,000-27,000 per unit-year and LDAR programs costing tens of thousands annually.[12] Sectoral compliance was high ($\theta \approx 0.8$-0.9), as most U.S. refineries had adopted CEMS/LDAR by 2020.[13] The harm H was well documented: benzene and formaldehyde exposure from refinery emissions carry significant cancer and respiratory risks according to EPA's National Air Toxics Assessment.[14]

At these values, harm-based screening was feasible, and compliance-based rules predict that once penalties eliminate the no-inquiry branch, inquiry persists. The scale of the decree easily cleared this threshold. In static terms, the outcome resembles semi-pooling; in practice, the consent decree translated this into a dynamic remedy, combining short-run tolerance with mandated transition into full compliance.

A final nuance concerns the difference between our one-shot model and real-world remedies. In the model, the bad type is either tolerated (pooling) or excluded (inactivity). By contrast, environmental enforcement often allows continued operation conditional on transformation—firms are ordered to invest in monitoring and technology that make future violations impossible. This can be understood as the dynamic counterpart of our framework: in the first period, the bad type is punished but tolerated; to remain in the market in subsequent periods, it must transform into a compliant good type. Consent decrees and mitigation orders therefore function as exclusion followed by conditional re-entry, an extension that underscores how our static analysis maps onto repeated regulatory practice.[15]

5.1 Empirical predictions

The case studies illustrate how judicial outcomes align with the model's logic, but the framework also yields testable comparative statics that go beyond single disputes. These

10. U.S. Department of Justice (DOJ), Press Release, Jan. 17, 2025; EPA Settlement Summary, Apr. 25, 2025, *United States v. HF Sinclair Navajo Refining LLC* (D.N.M. 2025). EPA = Environmental Protection Agency.

11. EPA, *Clean Air Markets: Continuous Emission Monitoring Systems (CEMS)*, technical guidance.

12. EPA, *Technical Support Document for CEMS Costs*, 2016 (O&M $13-27k per monitor-year); EPA, *Leak Detection and Repair: A Best Practices Guide*, 2007.

13. EPA, *Air Facility System Compliance Reports* (2022).

14. U.S. EPA, *National Air Toxics Assessment* (2018 cycle).

15. See, e.g., U.S. EPA, *Consent Decree: Cummins Inc.* (1998, amended 2006) (requiring phased engine monitoring and compliance upgrades over multiple years), available at https://19january2021snapshot.epa.gov/sites/static/files/2013-09/documents/cumminscd.pdf(Last visited: 21st Sep 2025); Joseph A. Hester, "Consent Decrees as Emergent Environmental Law," 85 *Mo. L. Rev.* 2020 (discussing how consent decrees often substitute for statutory precision by imposing phased compliance obligations); see also U.S. EPA, *Consent Decree: The Williams Companies Inc.* (2023) (requiring compliance certification and monitoring within 180 days), available at https://www.epa.gov/system/files/documents/2023-04/thewilliamscompaniesinc-cd.pdf. (Last visited: 21st Sep 2025)

predictions can be taken to data in various legal areas.

First, the model predicts that inquiry is more likely when compliant actors are common (θ high), legal surplus is large (y_G), monitoring is accurate (α high), and costs are low (k). Sectors with higher baseline compliance and cheaper monitoring technologies should therefore display higher rates of inquiry.

Second, liability design leaves distinct empirical footprints. Under harm-based rules, screening should collapse when monitoring costs rise or accuracy falls: for example, facilities may reduce inquiry intensity during periods of equipment failure or tightened QA protocols. By contrast, under compliance-based rules, once the safe-harbor threshold is crossed, inquiry persists even if exclusion would be more efficient. This predicts a divergence: screening rates should be more sensitive to cost shocks under harm-based regimes than under compliance-based ones.

Third, the model predicts that observed penalties should align with implementability thresholds. In compliance-based regimes, sanctions that induce inquiry should fall just above the threshold $T \geq k/[\varphi(1-\theta)]$, since higher penalties no longer affect post-inquiry incentives. In harm-based regimes, penalties that allow pooling to persist despite violations should remain at or below $T \leq y_G/(1-\theta)$, while penalties chosen to induce inquiry should exceed this cutoff. In dual-penalty regimes, front-end penalties T_n should be set just above $y_G/(1-\theta)$ to eliminate pooling, while back-end penalties T_i are tuned to govern inquiry quality. Empirically, one would therefore expect clustering of penalties near these cutoffs: regulators and courts tend to impose the minimum sufficient sanction to induce the desired equilibrium, or to hold penalties below the level that would destabilize an equilibrium they wish to tolerate.

Together, these predictions imply cross-sectional correlations between compliance rates and inquiry, differences in sensitivity to cost shocks across liability regimes, and penalty magnitudes that line up with theoretical thresholds. This provides a bridge from the theoretical model to enforcement data, enabling systematic tests beyond qualitative case studies.

6 Further Discussion

6.1 Endogenous inquiry precision.

Now assume the agent chooses the precision of inquiry, $\alpha \in (1/2, 1)$ as an additional strategy, facing an increasing cost $k(\alpha)$ ($k(1/2) = 0$, $k(1) = \infty$, $k' > 0, k'' \geq 0$). In semi-pooling, the agent's *residual* legal exposure after inquiry is $(1-\theta)(1-\alpha)$ multiplied by the penalty that applies *after* inquiry: under harm-based rules this is T, under compliance-based rules it is 0, and under dual penalties it is T_i. The agent's private choice of precision therefore solves a simple trade-off between the marginal saving in expected residual exposure and the marginal cost $k'(\alpha)$. Two implications follow. First, precision is *increasing* in the penalty that prices residual exposure (none under compliance, T under harm-based, T_i under dual): compliance yields the lowest privately chosen precision; dual penalties allow the lawmaker to raise precision by targeting T_i without simultaneously inflating the no-inquiry margin. Second, the *socially* optimal precision balances the social benefit of better screening (fewer false positives and false negatives) against $k'(\alpha)$; with dual penalties the lawmaker can implement this target (subject to the good type's profitability) by setting T_i to match that first-order condition, while using T_n only to regulate the no-inquiry branch. This arrangement preserves the selection lever

(via T_n) and aligns inquiry quality (via T_i). However, verifiability is strictly required: α (or a sufficient proxy) and due-diligence effort must be auditable so that transfers and liability can condition on the chosen precision. Formal statements and proofs appear in Appendix C.1. A close analogue arises in compliance law, where regulators adjust standards upward as inquiry can be made more accurate at similar cost. For example, U.S. environmental rules now mandate continuous emissions monitoring once the technology became feasible, replacing earlier self-reporting schemes.

6.2 Psychological inquiry costs and social preferences

Beyond tangible effort, inquiry often carries non-material burdens-embarrassment, fear of social disapproval, and reputational loss (Hellman 2009; Alexander and Ferzan 2009; Grossman and Van Der Weele 2017). Some agents also internalize others' welfare to varying degrees (Yaffe 2018).

Psychological (non-material) costs of inquiry. Let the agent's private inquiry cost remain k, but let the lawmaker place weight $\sigma \in [0, 1]$ on its non-material component in *welfare*. Private incentives to inquire are unchanged (they depend on k, not on σ); the social valuation of inquiry is attenuated: the semi-pooling welfare term is $\theta y_G - \sigma k$ rather than $\theta y_G - k$. Hence, as σ falls, the region where inquiry is *socially* preferred expands, yet implementability is unaffected-inducing inquiry still requires ruling out action without inquiry and ensuring the legal surplus covers k. Under imperfect inquiry, the same logic holds; with dual penalties one simply reads the no-inquiry side with T_n and the inquiry side with the agent's residual exposure governed by T_i.[16]

Social preference (moral) concern under pooling. Suppose an agent who performs an *illegal* task suffers an internal moral cost $m \geq 0$. Under pooling (no inquiry), acceptance depends on *expected* disutility, so the minimal transfer that induces acceptance is

$$X^* = (1 - \theta)\big(T + m\big)$$

(with a dual scheme, replace T by the no-inquiry penalty T_n). Thus m acts like an *additive expected penalty*: when the lawmaker aims to deter action without inquiry, a larger m permits a lower legal penalty to achieve the same deterrence; when the lawmaker instead prefers to *preserve* pooling (e.g., when the bad type's net contribution is non-negative or only mildly negative), a large m can make pooling infeasible by pushing X^* above a type's payoff. Under imperfect inquiry, these feasibility effects are unchanged because m bites only on the no-inquiry margin.

Interaction and robustness. Psychological inquiry costs and social preferences pull on different levers. Discounting non-material inquiry burdens in welfare ($\sigma < 1$) shifts *social desirability* toward inquiry but does not change the agent's private cutoff to inquire; moral concern ($m > 0$) tightens the *feasibility* of pooling by raising the transfer needed for no-inquiry acceptance. With imperfect inquiry ($\alpha < 1$), these directions are intact: noise lowers the welfare of inquiry-based screening and narrows its existence window, but σ and m affect the same margins as under perfect inquiry. The comparative statics

16. Formal thresholds with (σ, T) under a single penalty and (σ, T_n, T_i) under dual penalties are collected in Appendix C.4.

above do not depend on adopting any particular liability scheme; they describe how σ and m enter the lawmaker's trade-offs under each regime. Formal windows and welfare comparisons are in the appendix C.4.

6.3 Enforcement intensity vs. penalty severity.

Assume that conviction is not necessarily perfect, but with probability $p \in (0, 1]$. If p is exogenous and enforcement is costless, we can simply read T as the expected penalty pT throughout; under dual penalties read (T_n, T_i) as (pT_n, pT_i). Now let us relax the assumption of "costless" conviction and/or enforcement, and introduce cost into the penalty system.

Case A: p costless ($p = 1$), T costly. When enforcement is costless (thus $p = 1$) but penalty *severity* T absorbs real resources (e.g., incarceration), only equilibria that impose penalties on the path reduce welfare by expected resource cost. Under perfect inquiry, pooling bears $(1-\theta)C(T)$ while semi-pooling (screening) and inactive bear none. Under imperfect inquiry with harm-based liability, semi-pooling additionally bears $(1-\theta)(1-\alpha)C(T)$ (false positives); under compliance-based liability, inquiring agents are exempt and semi-pooling again bears no resource cost; under imperfect inquiry with dual-penalty liability, semi-pooling additionally bears $(1-\theta)(1-\alpha)C(T_i)$ (false positives). Hence the lawmaker should always choose the *minimal* T that implements the desired continuation behavior and avoid on-path penalties when $C(\cdot)$ is large.

Case B: T costless, p costly. If instead *intensity* p is resource-costly (with convex $K(p)$) and severity T is free, the conclusions are symmetric: equilibria that impose penalties on the path now bear expected $C_p(p)$; the welfare-maximizing choice is the *minimal* p that implements the desired equilibrium; and designs that avoid on-path penalties (semi under compliance-based rules; inactive) are strictly more attractive when $C_p(p)$ is large.

Case C: both p and T costly. When both instruments are costly, choose the cheapest mix of (p, T) (or (p, T_n, T_i) under dual penalties) that satisfies the relevant incentive/existence constraints. With convex costs $C_T(\cdot)$ and $C_p(p)$, an interior solution equalizes marginal resource cost per unit of expected penalty; corner solutions arise when one instrument is much cheaper. Under dual penalties, T_n prices the no-inquiry margin and T_i prices residual risk after inquiry, so it is efficient to set the unused component to zero whenever it does not bind.

6.4 Market structure and liability sharing

Our baseline assumes a single principal makes a take-it-or-leave-it offer to a single agent. The main comparative statics and implementability logic extend to alternative market structures and liability sharings with only cosmetic changes to the feasibility thresholds. We summarize four useful variants and refer to Appendix C.5 for formal statements and proofs.

Single principal with uncertainty (one side, one payoff). If neither party knows legality ex ante, set $y_G = y_B \equiv y$ and interpret type uncertainty as uncertainty about the legal state. The taxonomy and welfare ranking mirror the baseline. Under perfect inquiry, *pooling* (no inquiry) is feasible iff $y \geq (1-\theta)T_n$; *screening* (inquiry) is feasible iff $\theta y \geq k$. Under imperfect inquiry with dual penalties, replace the screening threshold by

$\varphi\, y \geq k + (1-\theta)(1-\alpha)T_i$, where $\varphi := \theta\alpha + (1-\theta)(1-\alpha)$. Hence the lawmaker's design problem and the penalty levers carry over verbatim.

No distinct principal (self-solicitation). In settings like soliciting contraband services, the "principal" and "agent" collapse into the same decision-maker; transfers vanish and only the decision to inquire (at cost k) versus act without inquiry remains. The screening condition becomes the agent's *private* surplus test (as above with y), and the pooling condition becomes her acceptance of expected legal risk. Qualitatively, this is the single-principal case with zero rents: inquiry occurs iff the legal-surplus term clears k, and willful ignorance arises iff expected penalties are (privately) coverable. The lawmaker's penalty levers and welfare comparison are unchanged.

Agent bargaining power (or competition among principals). If the agent can extract more than the minimal acceptance transfer (e.g., Nash bargaining, or many principals bidding), *pooling* feasibility shrinks (principals must cover higher transfers) while *screening* feasibility expands (the agent can be funded to cover k and any residual exposure). Implementability therefore tilts toward inquiry. When inquiry is socially desirable, bargaining power helps the lawmaker (lower penalty suffices to rule out no-inquiry action); when inclusion of both task types via pooling is desirable, strong agent power can make pooling infeasible. The lawmaker then trades off desirability against feasibility as in the baseline, but with the windows shifted toward screening.

Liability allocation as "just transfers." The same Coase-style logic behind bargaining power carries over to who bears legal penalties. If parties are risk-neutral, penalties are (fines) costless to society, and indemnity promises are enforceable, then shifting liability shares between the principal (he) and the agent (she)-or allowing the principal to warranty the agent's penalties-simply reassigns transfers without changing welfare. What *can* change is implementability: principal-side liability can restore a lever to deter no-inquiry behavior when compliance-based rules would otherwise make screening hard to switch off, whereas full indemnity to the agent effectively replicates an exemption and removes that lever. In short, the logic that "negotiation power translates into pure transfers" extends to shared-liability and warranty arrangements as well; when penalties or enforcement are costly, or indemnity is non-contractible, this equivalence breaks and design trade-offs reappear (outside our baseline).

6.5 Robustness to unknown θ.

If the agent does not know the true θ (share of good principals), let her act on a perceived value $\tilde{\theta}$ (posterior mean under Bayes; worst case $\underline{\theta}$ under ambiguity aversion). Then all *private* thresholds in our analysis hold with θ replaced by $\tilde{\theta}$ (e.g., the no-inquiry acceptance transfer becomes $(1-\tilde{\theta})T_n$, and screening is feasible under imperfect inquiry iff $\tilde{\varphi}\, y_G \geq k$ with $\tilde{\varphi} := \tilde{\theta}\alpha + (1-\tilde{\theta})(1-\alpha)$). Social desirability still evaluates welfare at the *true* θ. Ambiguity (lower $\tilde{\theta}$) simultaneously makes no-inquiry less attractive (higher expected penalty) and reduces the expected upside from inquiry (lower $\tilde{\varphi}$), so implementability can shift either way; our comparative statics otherwise carry through verbatim. Dual penalties remain the most robust instrument: set T_n high to deter no-inquiry across the relevant $\tilde{\theta}$ range, and keep T_i low to avoid overburdening inquiry when feasible.

This formulation also aligns with the doctrine's "sufficient suspicion" requirement: a court can impose a minimum suspicion threshold on $1 - \tilde{\theta}$ without altering our equilibrium taxonomy. It also clarifies how different *mens rea* categories map into the model. In our baseline, non-inquiry is a fully rational choice and thus corresponds to *deliberate ignorance*. By contrast, *negligence* could be represented by mistaken beliefs (systematically misperceiving θ), and *recklessness* by agents who recognize risk but underweight it relative to incentives. These cases lie outside the rational baseline but illustrate how the model can accommodate the doctrinal distinction between negligent, reckless, and deliberate ignorance.

6.6 Jurisprudential concerns.

Several doctrinal debates around willful ignorance can be mapped into our framework. First, the parity question asks whether deliberate ignorance should be punished like knowledge. In equilibrium, knowing crime is strictly dominated by non-inquiry once ignorance is punished at parity or above, so the equal-culpability debate has no bite except in the implausible case where ignorance is punished more severely than knowledge. Relatedly, some argue culpability should depend on whether knowledge would have changed behavior. Our model shows this criterion is moot: inquiry followed by knowing violation is never optimal, so all harmful conduct flows through ignorance.

Second, the legality principle (*nulla poena sine lege*) cautions against punishing under vague or shifting standards. In our framework, vagueness is captured by reduced accuracy α: inquiry may not yield a clear answer. Compliance-based or dual-penalty rules accommodate this by lowering residual risk once inquiry is documented in good faith.

Third, concerns about overbreadth and chilling effects correspond directly to our implementability frontier: excessively high penalties on non-inquiry can make screening infeasible and collapse valuable markets. Finally, doctrine distinguishes negligence, recklessness, and deliberate ignorance. Our baseline assumes rational agents, so non-inquiry is deliberate; negligence and recklessness could be modeled as systematic misperception of θ or underweighting of recognized risk. Some courts and commentators also place willful ignorance between recklessness and knowledge, on the view that the actor "in fact" does not know. We treat deliberate ignorance at parity with knowledge, reflecting its strategic nature. This choice has no effect on equilibrium outcomes, since knowing crime is strictly dominated and never arises.

7 Conclusion

This paper offers a tractable framework for analyzing willful ignorance under asymmetric information and legal penalties. The central message is an implementability one. With perfect inquiry, penalties can be used to screen out harmful behavior while preserving valuable activity, so the welfare-maximizing behavior can always be made to exist and be selected. With imperfect inquiry, noise simultaneously lowers the value of screening and decouples feasibility from desirability; for non-empty regions of primitives, inquiry is either unattainable when desirable or difficult to switch off when exclusion is preferred.

We show how liability design mediates that trade-off. Harm-based rules keep a strong selection lever but can choke off desirable screening; compliance-based rules relax feasibility but risk coexistence and misselection; dual-penalty rules separate these roles and

weakly dominate single-penalty rules in implementability, failing only when the legal surplus cannot cover expected inquiry cost. These insights translate into simple guidance: calibrate penalties directly to the cutoff conditions implied by accuracy and inquiry costs. When accuracy is imperfect, use separate penalties for uninformed action and for residual wrongdoing after inquiry. In this sense, ignorance is not a shield: the absence of knowing crime in equilibrium is an endogenous consequence of incentives rather than evidence of a higher moral standard, and liability rules must therefore target the strategic margin of non-inquiry.

The framework's strengths are its clarity about what penalties can and cannot accomplish and its portability: it provides a template for studying inquiry technologies, evidentiary standards, and enforcement frictions. Future work could extend the framework to dynamic and heterogeneous environments-e.g., repeated interactions with reputation, heterogeneity in inquiry costs, subsidies for inquiry and information disclosure regulation. Together, these can deepen the welfare foundations for regulating willful ignorance in practice.

References

Alexander, Larry, and Kimberly Kessler Ferzan. 2009. *Crime and culpability: A theory of criminal law.* Cambridge University Press.

Alter, Adam L, Julia Kernochan, and John M Darley. 2007. "Morality influences how people apply the ignorance of the law defense." *Law & Society Review* 41 (4): 819–864.

Charlow, Robin. 1991. "Wilful ignorance and criminal culpability." *Tex. L. Rev.* 70:1351.

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." *Economic Theory,* 67–80.

Grossman, Zachary, and Joel J Van Der Weele. 2017. "Self-image and willful ignorance in social decisions." *Journal of the European Economic Association* 15 (1): 173–217.

Hellman, Deborah. 2009. "Willfully blind for good reason." *Criminal Law and Philosophy* 3:301–316.

Kartik, Navin, Marco Ottaviani, and Francesco Squintani. 2007. "Credulity, lies, and costly talk." *Journal of Economic theory* 134 (1): 93–116.

Kirfel, Lara, and Ivar R Hannikainen. 2023. "Why blame the ostrich? Understanding culpability for willful ignorance." *K., Prochownik, S. Magen,(Eds.), Advances in experimental philosophy of law,* 75–98.

Luban, David. 1998. "Contrived ignorance." *Geo. LJ* 87:957.

Polinsky, A Mitchell, and Steven Shavell. 2000. "The economic theory of public enforcement of law." *Journal of economic literature* 38 (1): 45–76.

Sarch, Alexander. 2018. "Willful ignorance in law and morality." *Philosophy Compass* 13 (5): e12490.

Sarch, Alexander F. 2014. "Willful ignorance, culpability, and the criminal law." . *John's L. Rev.* 88:1023.

Simons, Kenneth W. 2021. "The willful blindness doctrine: Justifiable in principle, problematic in practice." *Ariz. St. LJ* 53:655.

Yaffe, Gideon. 2018. "The point of mens rea: The case of willful ignorance." *Criminal Law and Philosophy* 12 (1): 19–44.

Appendix A Tables and Figures

In this appendix, we show the extensive games from the baseline model and from imperfect inquiry.

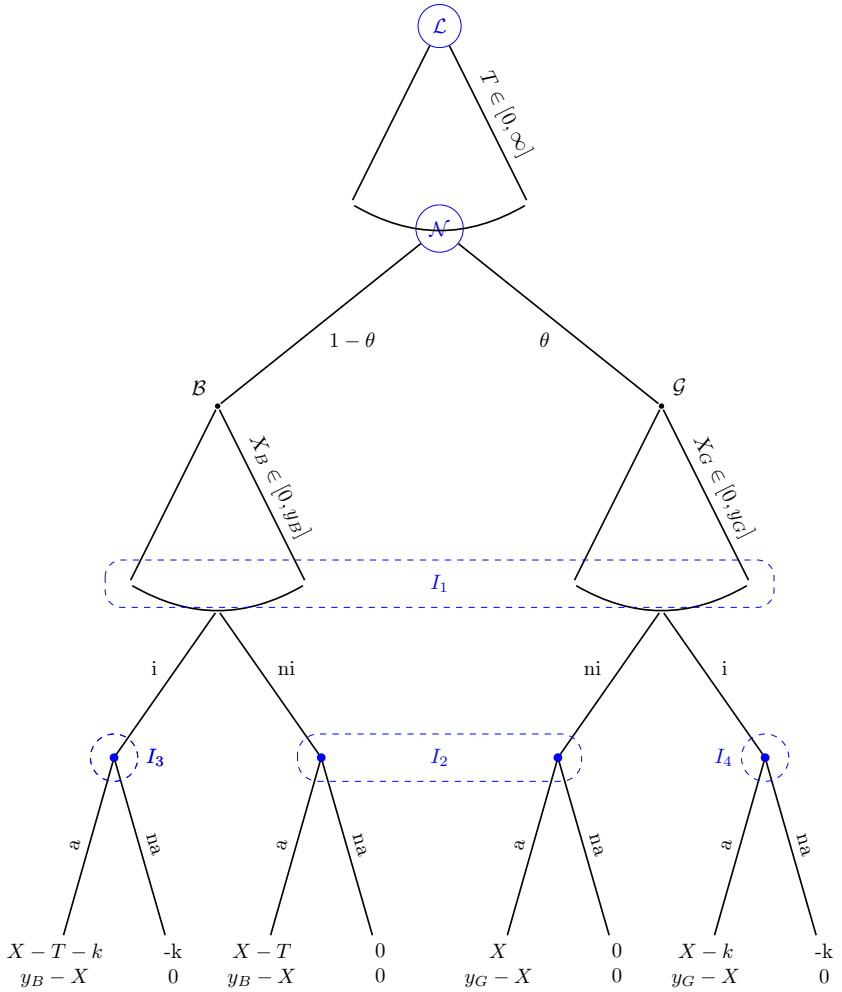


Figure A.1: Baseline $\mathcal{LPA}$ Game (perfect inquiry)

Notes: The figure illustrates the lawmaker–principal–agent (LPA) game under perfect inquiry. Nature selects the principal's type; the principal offers a transfer; the agent decides whether to inquire and whether to accept the task; payoffs depend on legality and penalties. Transfers and penalties are shown next to branches.

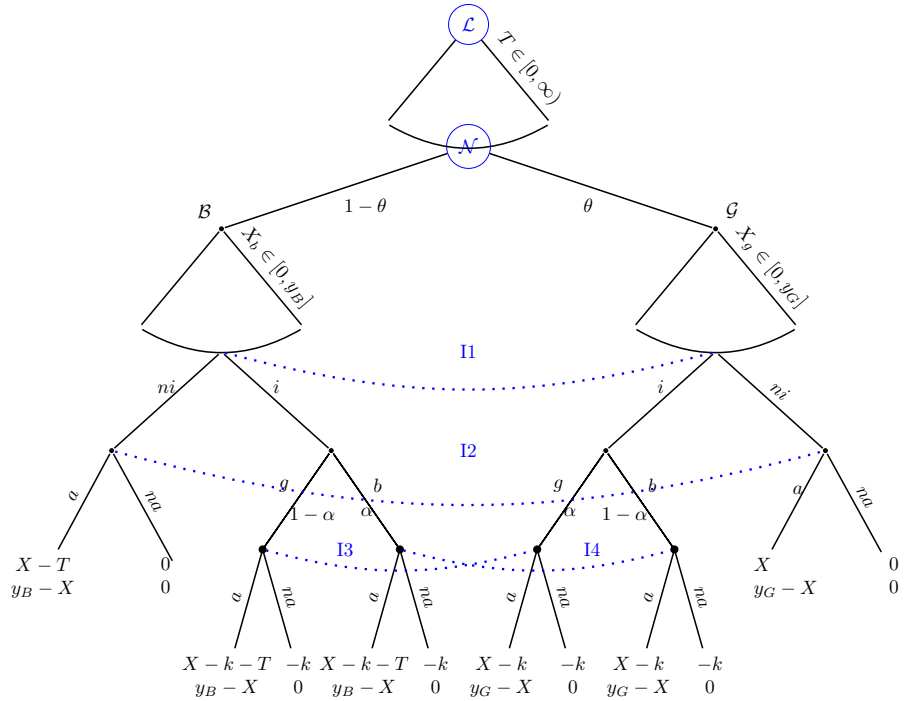


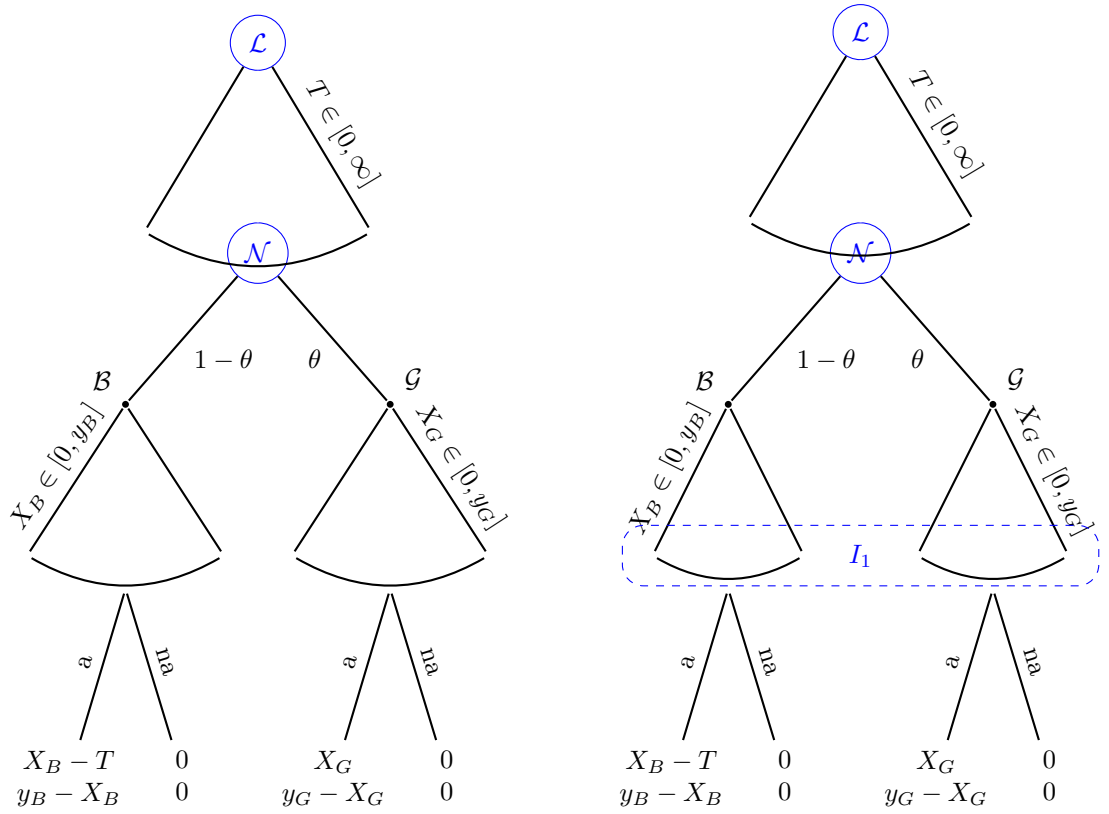
Figure A.2: $\mathcal{LPA}$ Game (Imperfect Inquiry)

Notes: The figure illustrates the lawmaker–principal–agent (LPA) game when inquiry is imperfect. A false positive incurs a penalty T under harm-based liability. For compliance-based and dual-penalty rules, the penalty for false positives changes accordingly.

Appendix B General Proof

B.1 Proof of Lemma 3.1

The games played in Perfect Information Equilibrium (PIE) and No Inquiry Equilibrium (NIE) are depicted in Figure B.3.



(a) $\mathcal{LPA}$ game (PIE) (b) $\mathcal{LPA}$ game (NIE)

Figure B.3: Strategic Decision Making of the Principal and the Agent ((Im)perfect Information)

Notes: Panel (a) shows the LPA game under Perfect Information Equilibrium (PIE), where the principal's type is publicly known and strategies are chosen accordingly. Panel (b) shows the game under No Inquiry Equilibrium (NIE), where types are not observed and no inquiry occurs.

Proof. **(i) PIE: strategies and welfare.** Type ω is publicly known. For G, the task is legal, so the agent accepts at $X_G = 0$, and any $X_G > 0$ strictly lowers the principal's payoff; thus $X_G^* = 0$ and G is performed. For B, the agent accepts iff $X_B \geq T$. The minimal acceptance offer is $X_B^* = T$ when feasible (i.e., $T \leq y_B$, so $y_B - T \geq 0$). The lawmaker chooses T to implement the welfare-maximizing admission decision for B:

$$\tau y_B - H \geq 0 \;\Rightarrow\; T \leq y_B \;\; \text{(perform } B\text{)}, \qquad \tau y_B - H < 0 \;\Rightarrow\; T > y_B \;\; \text{(deter } B\text{)}.$$

Hence

$$\mathrm{SW}_{\mathrm{PIE}} = \begin{cases} \theta y_G + (1-\theta)(\tau y_B - H), & \text{if } \tau y_B - H \geq 0, \\ \theta y_G, & \text{if } \tau y_B - H < 0. \end{cases}$$

(ii) NIE: equilibrium and welfare. Types are not observed and, by definition, there is no inquiry. Without credible signaling, any acceptance decision must apply to both types. In pooling, let both types offer X. The agent's expected utility is $X-(1-\theta)T$, so at the minimal acceptance offer

$$X^* = (1-\theta)T.$$

Type G can profitably offer X^* iff $y_G - (1-\theta)T \geq 0$, i.e.

$$(1-\theta)T \leq y_G. \qquad (*)$$

Given $y_B > y_G$, condition $(*)$ also ensures B's profitability.

If the lawmaker picks a low T satisfying $(*)$, the unique pooling equilibrium has both types offer X^*, the agent accepts, and welfare is $\theta y_G + (1-\theta)(\tau y_B - H)$. If T violates $(*)$, no type can profitably cover acceptance and the agent rejects all offers; welfare is 0. Optimizing over T yields the stated piecewise formula for $\mathrm{SW}_{\mathrm{NIE}}$.

(iii) Comparison. If $\tau y_B - H \geq 0$, PIE and NIE both perform G and B, so they coincide at $\theta y_G + (1-\theta)(\tau y_B - H)$. If $\tau y_B - H < 0$, then $\theta y_G + (1-\theta)(\tau y_B - H) \leq \theta y_G$, so NIE yields either 0 or $\theta y_G + (1-\theta)(\tau y_B - H)$, both weakly below $\theta y_G = \mathrm{SW}_{\mathrm{PIE}}$, with strict inequality whenever $(1-\theta)(\tau y_B - H) < 0$. □

B.2 Proof of Equilibrium Preliminaries

Proof. **(i) Nonexistence of separating equilibrium.** Suppose, per contra, that there exists a equilibrium $(\sigma_P^*, \sigma_A^*, \mu^*)$ with $X_B^* \neq X_G^*$. Since the offer X uniquely identifies the type on the equilibrium path, Bayes' rule yields $\mu^*(G \mid X_G^*) = 1$ and $\mu^*(G \mid X_B^*) = 0$. Hence information is already revealed by X, so any inquiry would be strictly dominated (by saving cost $k > 0$) and thus $\sigma_A^{i,*}(X_G^*) = \sigma_A^{i,*}(X_B^*) = ni$. Given $\mu^*(G \mid X_G^*) = 1$, sequential rationality implies the agent accepts any legal task at nonnegative transfer, so $\sigma_A^{a,*}(X_G^*, ni) = a$. Given $\mu^*(G \mid X_B^*) = 0$, the agent accepts the illegal task iff $X_B^* \geq T$: $\sigma_A^{a,*}(X_B^*, ni) = a \iff X_B^* \geq T$. Without loss of generality, $X_G^* = 0$: if $X_G^* > 0$, type G can deviate to $X_G' \in [0, X_G^*)$ and still be accepted, which strictly raises G's payoff. Consider the bad type's payoff under two exhaustive cases:

Case 1: $X_B^* < T$. Then $\sigma_A^{a,*}(X_B^*, ni) = na$ and type B's equilibrium payoff is 0. If B deviates to $X_B' = X_G^* = 0$, the observed offer coincides with the on-path G-offer. By Bayes' rule at this on-path history, the agent believes $\mu^*(G \mid X_G^*) = 1$ and accepts. The deviating payoff is $y_B > 0$, a strict gain.

Case 2: $X_B^* \geq T$. Then $\sigma_A^{a,*}(X_B^*, ni) = a$ and B's payoff is $y_B - X_B^* \leq y_B - T < y_B$. Deviating to $X_B' = X_G^* = 0$ again induces acceptance at the on-path G-offer and yields y_B, a strict gain. In both cases type B has a profitable deviation, contradicting equilibrium. Hence no separating equilibrium exists.

(ii) Illegal performance requires non-inquiry. Suppose, per contra, that along the equilibrium path the agent inquires ($\sigma_A^{i,*} = i$) and nevertheless performs the illegal task with positive probability. Under perfect inquiry, after observing the illegal signal the agent knows with certainty that the task is illegal. If she still performs upon learning illegality, then (by monotonicity) she also performs upon learning legality. Therefore, conditional on the offer, his *post-inquiry* action plan coincides with the *no-inquiry* plan (perform in both states). Ex ante, inquiry then strictly reduces the agent's payoff by $k > 0$ without changing behavior. Deviating to $\sigma_A^i = ni$ is strictly profitable, contradicting sequential rationality. Thus, whenever an illegal task is performed in equilibrium, the agent must not have inquired: the illegal performance is via willful ignorance. □

Remark B.1 (Imperfect inquiry). The argument in part (i) does not rely on $\alpha = 1$; separation implies no inquiry and the same mimicking deviation to the on-path G-offer rules out separating equilibria when $\alpha \in (1/2, 1)$. Part (ii) is specific to perfect inquiry:

with imperfect inquiry, illegal performance can occur after a *good* (erroneous) signal, which is not willful ignorance.

Going back to the computation of social welfare function, it is then clear that the social welfare takes only three values:

- *pooling*: $X_B = X_G$, $\sigma_A^i(\cdot) = ni, \sigma_A^a(\cdot) = a$, $SW = \theta y_G + (1-\theta)(\tau y_B - H)$
- *semi-pooling*: $X_B = X_G$, $\sigma_A^i(\cdot) = i, \sigma_A^a(X_B, i) = na$, $\sigma_A^a(X_G, i) = a$, $SW = \theta y_G$
- *inactive*: $X_B, X_G \in [0, T)$, $\sigma_A^i(\cdot) = ni, \sigma_A^a(\cdot) = na$, $SW = 0$.

B.3 Proof of Proposition 4.1

Lemma B.1 (Threshold geometry of existence sets under perfect inquiry)**.** *Under $\alpha = 1$ and, the continuation equilibrium existence conditions are:*

$$\begin{aligned}
&\text{Pool 1 (insurance-style pooling) } \textit{exists} && \iff \theta y_G > k \ \textit{ and } \ \tfrac{k}{\theta(1-\theta)} \le T \le y_G + \tfrac{k}{1-\theta},\\
&\text{Pool 2 (minimal-acceptance pooling) } \textit{exists} && \iff \theta y_G \le k \ \textit{ and } \ T \le \tfrac{y_G}{1-\theta},\\
&\text{Semi } \textit{exists} && \iff \theta y_G > k \ \textit{ and } \ T > \max\Big\{y_B, \ \tfrac{k}{\theta(1-\theta)}\Big\},\\
&\text{Inactive } \textit{exists} && \iff T > y_B.
\end{aligned}$$

Consequently, multiplicity arises only in:

$$\begin{cases}
\theta y_G > k, \ \ \max\{y_B, \frac{k}{\theta(1-\theta)}\} < T \le y_G + \frac{k}{1-\theta} & \Rightarrow \text{ Pool 1 } \textit{and} \text{ Semi } \textit{coexist},\\
\theta y_G \le k, \ \ y_B < T \le \frac{y_G}{1-\theta} & \Rightarrow \text{ Pool 2 } \textit{and} \text{ Inactive } \textit{coexist}.
\end{cases}$$

Proof. Transfers/penalties are redistributive; only feasibility and best responses matter. In pooling, let both types offer X. The agent's payoffs are $U_{ni}(X;T) = X - (1-\theta)T$ and $U_i(X) = \theta X - k$.

(2) Pool 1 (insurance-style pooling). *Sufficiency.* If $\theta y_G > k$ and $\frac{k}{\theta(1-\theta)} \le T \le y_G + \frac{k}{1-\theta}$, set $X_G = X_B = T - \frac{k}{1-\theta}$ and σ_A: ni and accept. Then $U_{ni} = X - (1-\theta)T = \theta T - \frac{k}{1-\theta} \ge 0$ and $U_i = \theta X - k = U_{ni}$, so ni by tie-break. Feasibility for G holds by $X \le y_G$. For B, since and $T \le y_G + \frac{k}{1-\theta}$, we also have $T \le y_B + \frac{k}{1-\theta}$, hence $y_B - (T - \frac{k}{1-\theta}) \ge 0$. Deviations $X' < T$ are rejected; any $X' \ge T$ lowers the deviator's payoff.

Necessity. Pool 1 requires $U_{ni} = U_i$ at $X = T - \frac{k}{1-\theta}$ (so $T \ge \frac{k}{\theta(1-\theta)}$) and feasibility for G ($X \le y_G$, i.e. $T \le y_G + \frac{k}{1-\theta}$). These bounds are compatible only if $k \le \theta y_G$, i.e. $\theta y_G > k$.

(1) Pool 2 (minimal-acceptance pooling). *Sufficiency.* If $\theta y_G \le k$ and $T \le \frac{y_G}{1-\theta}$, set $X_G = X_B = (1-\theta)T$ and σ_A: ni and accept. Then $U_{ni} = 0$ and $U_i = \theta X - k \le \theta y_G - k \le 0$, so ni is optimal. Payoffs: $y_G - (1-\theta)T \ge 0$ and, since, also $y_B - (1-\theta)T \ge 0$. Any $X' < T$ is rejected by convention; any $X' \ge T$ is accepted but yields $y_\omega - X' \le y_\omega - T < y_\omega - (1-\theta)T$, hence no profitable deviation.

Necessity. In any pooling equilibrium with ni, minimality implies $X = (1-\theta)T$. Feasibility for G requires $(1-\theta)T \le y_G$ (i.e. $T \le \frac{y_G}{1-\theta}$). Optimality of ni demands $U_i = \theta X - k \le 0$; since $X \le y_G$, this gives $\theta y_G \le k$.

(3) Semi (inquiry; perform iff legal). *Sufficiency.* If $\theta y_G > k$ and $T > \max\{y_B, \frac{k}{\theta(1-\theta)}\}$, take $X_G = X_B = \frac{k}{\theta}$ and σ_A: i; perform iff legal. Then $U_i = \theta X - k = 0$ while $U_{ni} = \frac{k}{\theta} - (1-\theta)T < 0$, so inquiry is strictly optimal. G's on-path payoff is $y_G - \frac{k}{\theta} > 0$; B is never performed on path. Any $X' < T$ is rejected; any $X' \geq T$ gives $y_B - X' \leq y_B - T < 0$. Also $T > \frac{k}{\theta(1-\theta)} \geq \frac{k}{\theta}$, so G cannot improve by deviating to $X' \geq T$ (since $y_G - X' \leq y_G - T < y_G - \frac{k}{\theta}$).

Necessity. Semi requires strict preference for inquiry over ni: $\theta X - k > X - (1-\theta)T$. At the minimal X with $\theta X - k \geq 0$ (namely $X = \frac{k}{\theta}$), this becomes $0 > \frac{k}{\theta} - (1-\theta)T$, i.e. $T > \frac{k}{\theta(1-\theta)}$. Feasibility for G at $X = \frac{k}{\theta}$ requires $\theta y_G > k$. Deterring B's profitable pooling deviations needs $T > y_B$.

(4) Inactive (no task performed). *Sufficiency.* If $T > y_B$ (hence $T > y_G$), let both types offer $X < T$. By convention the agent does not inquire and rejects any $X' < T$. Any $X' \geq T$ is accepted; G would earn $y_G - X' \leq y_G - T < 0$ and B would earn $y_B - X' < 0$.

Necessity. If $T \leq y_B$, type B can deviate to $X' = T$ and earn $y_B - T \geq 0$. Thus $T > y_B$ is necessary (and, since, also sufficient).

(5) Multiplicity. The intervals above yield exactly the two overlaps stated; no triple overlap occurs. □

Lemma B.2 (Welfare is T-invariant; existence is T-dependent)**.** *Under $\alpha = 1$,*

$$\mathrm{SW}_{pool} = \theta y_G + (1-\theta)(\tau y_B - H), \qquad \mathrm{SW}_{semi} = \theta y_G - k, \qquad \mathrm{SW}_{inact} = 0,$$

and none of these depends on T. By contrast, the existence *of each continuation equilibrium type depends on T through the thresholds in Lemma B.1.*

Proof. Transfers/penalties are redistributive, so T affects only which types are performed (existence), not the value of each type once selected. Threshold dependence follows from Lemma B.1. □

Proof of Proposition 4.1 :

Proof. For this proof, let $e^{\mathrm{SW}} \in \{\text{pooling}, \text{semi}, \text{inactive}\}$ denote a welfare-maximizing candidate under perfect inquiry (ignoring implementability). By Lemmas B.1 and B.2, welfare rankings are T-invariant within types, so it suffices to choose T so that $e^{\mathrm{SW}} \in \mathcal{E}(T)$. With our implementation convention, the implemented outcome then coincides with the socially desirable $e^\star$ from the main text.

Case 1: $e^\star = semi$. Then $\theta y_G > k$. Choose any $T > \max\{y_B, \frac{k}{\theta(1-\theta)}\}$; Semi exists. If desired, take $T > y_G + \frac{k}{1-\theta}$ to eliminate Pool 1. (Coexistence with Inactive may remain but is immaterial to implementability.)

Case 2: $e^\star \neq semi$. (a) Pooling desirable, $\theta y_G \leq k$: pick $T \in (0, \frac{y_G}{1-\theta}]$; then Pool 2 exists and Semi does not. *(b) Pooling desirable, $\theta y_G > k$:* set $T^\dagger := \frac{k}{\theta(1-\theta)}$; Pool 1 exists at $X = T^\dagger - \frac{k}{1-\theta}$ and Semi does not (strict threshold). *(c) Inactive desirable:* pick $T > y_B$; then Inactive exists (and, if uniqueness is desired, take $T > y_G + \frac{k}{1-\theta}$ to exclude both Pool types). In all cases there exists a T under which the equilibrium set contains an equilibrium of the socially desirable type $e^\star$, proving implementability. Since the No Inquiry Equilibrium can realize only pooling or inactivity (never Semi), perfect inquiry weakly dominates it, with strict dominance whenever $\mathrm{SW}_{\mathrm{semi}} > \max\{\mathrm{SW}_{\mathrm{pool}}, \mathrm{SW}_{\mathrm{inact}}\}$. □

B.4 Proof of Proposition 4.2

Lemma B.3 (Threshold geometry under harm-based liability with imperfect inquiry)**.** *Let*

$$\varphi := \theta\alpha + (1-\theta)(1-\alpha), \qquad p_3 := \tfrac{(1-\theta)\alpha}{(1-\theta)\alpha+\theta(1-\alpha)}, \qquad p_4 := \tfrac{(1-\theta)(1-\alpha)}{\varphi},$$

and define the bad-type incentive-compatibility threshold

$$T_B^{\mathrm{IC}} := \tfrac{\varphi\,\alpha\,y_B+(1-\alpha)k}{\alpha(1-\alpha+\alpha\theta)}.$$

Then under harm-based liability with $\alpha \in (\frac{1}{2}, 1)$, the continuation equilibrium existence conditions are:

Pool 1 (insurance-style pooling) *exists for* $\frac{k}{\theta(1-\theta)(2\alpha-1)} \;\le\; T \;\le\; \frac{y_G}{p_3} + \frac{k}{(1-\theta)\alpha}$,

Pool 2 (minimal-acceptance pooling) *exists for* $T \;\le\; \min\Big\{\frac{y_G}{1-\theta},\; \frac{k}{\theta(1-\theta)(2\alpha-1)}\Big\}$,

Semi (inquiry; act only on g) *exists for* $T > \max\Big\{\frac{k}{\theta(1-\theta)(2\alpha-1)},\; T_B^{\mathrm{IC}},\; \frac{y_G}{1-\theta}\Big\}$ *and* $T \le \frac{\varphi\, y_G - k}{(1-\theta)(1-\alpha)}$,

Inactive *exists for* $T \;>\; \max\Big\{y_B,\; \frac{y_G}{1-\theta}\Big\}$ *and* $(1-\theta)(1-\alpha)T + k \;>\; \varphi\, y_G$.

Consequently, multiplicity arises only in:

$\max\Big\{y_B,\; \frac{y_G}{1-\theta}\Big\} < T \le \frac{y_G}{p_3} + \frac{k}{(1-\theta)\alpha}$, $(1-\theta)(1-\alpha)T + k > \varphi y_G \Rightarrow$ Pool 1 *and* Inactive *coexist.*

In particular, whenever Semi is feasible (i.e. $k < \theta(2\alpha-1)y_G$), its lower bound exceeds Pool 1's upper bound, so Semi and Pool 1 cannot coexist*.*

Proof of Proposition 4.2 (implementation under harm-based rules, imperfect inquiry).

Proof. **Definitions.** Minimal transfers:

$$X_{\min}^{\mathrm{pool}} = \max\Big\{(1-\theta)T,\; p_3 T - \tfrac{k}{1-\varphi}\Big\}, \quad X^{\mathrm{semi}} = \tfrac{(1-\theta)(1-\alpha)T+k}{\varphi} = p_4 T + \tfrac{k}{\varphi}.$$

As in the main text, transfers/penalties are redistributive; welfare is T-invariant *within* a branch:

$$SW_{\mathrm{pool}} = \theta y_G + (1-\theta)(\tau y_B - H), \quad SW_{\mathrm{semi}} = \theta\alpha\, y_G + (1-\theta)(1-\alpha)(\tau y_B - H) - k, \quad SW_{\mathrm{inact}} = 0.$$

Let $e^{\mathrm{SW}} \in \{\mathrm{Pool}, \mathrm{Semi}, \mathrm{Inactive}\}$ denote a welfare-maximizing *type* (ignoring implementability).

Case 1: $e^\star = \mathrm{Pool}$ (all trades desirable).

(a) *Insurance-style pooling (Pool 1).* Choose T *strictly inside* the region:

$$T \in \Big[\tfrac{k}{\theta(1-\theta)(2\alpha-1)},\;\; \tfrac{y_G}{p_3} + \tfrac{k}{(1-\theta)\alpha}\Big].$$

We do *not* need to rule out Inactive: under our implementation convention, Inactive is weakly Pareto-dominated by Pool 1 and therefore not credited when both

coexist.[17]

(b) *Minimal-acceptance pooling (Pool 2).* Pick any

$$T \le \min\{\tfrac{y_G}{1-\theta},\ \tfrac{k}{\theta(1-\theta)(2\alpha-1)}\},$$

Semi fails by $T \le \frac{k}{\theta(1-\theta)(2\alpha-1)}$ and we do *not* need to rule out Inactive.

Case 2: $e^\star =$ Inactive (no trade desirable). Choose T strictly inside the inactivity region and outside both pooling regions, e.g.

$$T \ > \ M \ := \ \max\left\{ y_B,\ \tfrac{y_G}{1-\theta},\ \tfrac{\varphi\, y_G - k}{(1-\theta)(1-\alpha)},\ \tfrac{y_G}{p_3} + \tfrac{k}{(1-\theta)\alpha} \right\}.$$

Then Pool 2 fails by $T > \frac{y_G}{1-\theta}$, Pool 1 fails by $T > \frac{y_G}{p_3} + \frac{k}{(1-\theta)\alpha}$, and Semi fails by $(1-\theta)(1-\alpha)T + k > \varphi\, y_G$. Hence Inactive is *unique*.

Case 3: $e^\star =$ Semi (inquiry desirable). Semi must be feasible, i.e.

$$\max\left\{ \tfrac{k}{\theta(1-\theta)(2\alpha-1)},\ T_B^{\mathrm{IC}},\ \tfrac{y_G}{1-\theta} \right\} \ < \ \tfrac{\varphi\, y_G - k}{(1-\theta)(1-\alpha)}.$$

Since Semi's region is strictly above $\frac{y_G}{1-\theta}$ while Pool 1's upper bound is below $\frac{y_G}{1-\theta}$ whenever Semi is feasible, *Semi and Pool 1 cannot coexist.* Also Semi and Inactive cannot coexist because their feasibility inequalities for $(1-\theta)(1-\alpha)T + k$ are mutually exclusive. Therefore, selecting any

$$T \in \left(\max\left\{ \tfrac{k}{\theta(1-\theta)(2\alpha-1)},\ T_B^{\mathrm{IC}},\ \tfrac{y_G}{1-\theta} \right\},\ \ \tfrac{\varphi\, y_G - k}{(1-\theta)(1-\alpha)} \right)$$

puts T in the interior of Semi's region and outside all competitors. Hence Semi is *unique* if feasible. □

Welfare comparison and conclusion. Within each equilibrium type, SW is independent of T; hence the lawmaker's choice of T only selects between types. Pooling and Inactive equilibria deliver exactly the same welfare as in the No-Inquiry benchmark, while Semi yields

$$SW_{\text{semi}} = \theta\alpha\, y_G + (1-\theta)(1-\alpha)(\tau y_B - H) - k,$$

which is strictly below its perfect-inquiry counterpart $\theta y_G - k$ but strictly above the No-Inquiry outcome whenever Semi is feasible. Thus the lawmaker can always implement pooling or inactivity to deter inquiry when it is not socially desirable. However, there exist non-empty parameter ranges (lower α or higher k) in which Semi is welfare-maximizing yet infeasible, so inquiry cannot be induced by any penalty. In all cases, maximal welfare under imperfect inquiry is weakly below that under perfect inquiry and weakly above that under No Inquiry.

B.5 Proof of Proposition 4.3

Proof. **Definitions.** Let

$$\varphi := \theta\alpha + (1-\theta)(1-\alpha),$$

17. Non-emptiness: the interval $\left[\frac{k}{\theta(1-\theta)(2\alpha-1)},\ \frac{y_G}{p_3} + \frac{k}{(1-\theta)\alpha}\right]$ is non-empty iff $k \ge \theta(2\alpha-1)y_G$.

Minimal transfers:

$$X^{\text{semi}} = \tfrac{k}{\varphi}, \qquad X^{\text{pool}} = (1-\theta)T.$$

Agent utilities:

$$U_i = \varphi X - k, \qquad U_{ni} = X - (1-\theta)T.$$

Good-type feasibility for semi: $\varphi y_G \geq k$.

Welfare is T-invariant within branches:

$$SW_{\text{semi}} = \theta\alpha y_G + (1-\theta)(1-\alpha)(\tau y_B - H) - k, \quad SW_{\text{pool}} = \theta y_G + (1-\theta)(\tau y_B - H), \quad SW_{\text{inact}} = 0.$$

Case 1: $e^\star$ = Semi (inquiry desirable). Semi requires the agent to strictly prefer inquiry. At $X = k/\varphi$,

$$U_i = 0, \quad U_{ni} = \tfrac{k}{\varphi} - (1-\theta)T.$$

Thus inquiry is optimal whenever

$$T \;\geq\; \tfrac{k}{\varphi(1-\theta)}. \tag{C1}$$

Good-type feasibility requires

$$\varphi y_G \;>\; k. \tag{C2}$$

Together, (C1)–(C2) define the region in which inquiry is implementable. In this region, Semi is unique: Pool fails by (C1), and Inactive fails by good-type deviations. Hence Semi is implemented whenever it is feasible.

Case 2: $e^\star$ = Inactive (no trade desirable). Suppose welfare would prefer no action. If $\varphi y_G \leq k$, Semi is infeasible by (C2), and the lawmaker can set $T > y_B$ to rule out Pool, implementing Inactive uniquely. If $\varphi y_G > k$, however, Semi remains feasible at $X = k/\varphi$ for any $T \geq k/[\varphi(1-\theta)]$. Because compliance-based liability exempts inquiry from penalties on false positives, no choice of T can remove the inquiry equilibrium. Thus when exclusion is desirable but (C2) holds, inquiry persists in the equilibrium set.

Case 3: $e^\star$ = Pool (no inquiry desirable). Without inquiry, the agent accepts if $U_{ni} \geq 0$, i.e. $X \geq (1-\theta)T$. Minimal transfer is $X = (1-\theta)T$. The lawmaker can always set T small enough that $X \leq y_G, y_B$; hence Pool can be implemented whenever pooling is welfare-maximizing.

Welfare comparison and conclusion. Pooling and Inactive equilibria yield exactly the same welfare as the No-Inquiry benchmark. Semi yields

$$SW_{\text{semi}} = \theta\alpha y_G + (1-\theta)(1-\alpha)(\tau y_B - H) - k,$$

which converges to the perfect-inquiry benchmark $\theta y_G - k$ as $\alpha \to 1$, and is strictly below it for $\alpha < 1$. Thus, under compliance-based liability, inquiry can always be sustained when it is feasible but cannot always be deterred when socially undesirable. Maximal welfare is therefore weakly below perfect inquiry and may in case (ii) fall strictly below the No-Inquiry benchmark. □

B.6 Proof of Proposition 4.4

Proof. **Primitives and notation.** Let

$$\varphi := \theta\alpha + (1-\theta)(1-\alpha), \qquad 1-\varphi = \theta(1-\alpha) + (1-\theta)\alpha,$$

the probability (under pooling types) that inquiry yields a "good" signal and the task is performed. Under *dual penalties* we denote by T_n the penalty if the agent did *not* inquire and the task is illegal, and by T_i the penalty if the agent *did* inquire yet an illegal task is performed (false positive).

Step 1 (Minimal transfers). If both types offer the same *per-performance* transfer t and the agent inquires and performs only on a good signal, her expected utility is

$$U_i = \varphi\, t - k - (1-\theta)(1-\alpha)\, T_i.$$

By tie-breaking, the minimal transfer that induces inquiry is

$$t^{\text{semi}} = \tfrac{k+(1-\theta)(1-\alpha)\,T_i}{\varphi}. \tag{B.1}$$

If instead she accepts without inquiry, the minimal lump-sum transfer that induces acceptance equals her expected penalty:

$$X^{\text{pool}} = (1-\theta)\, T_n. \tag{B.2}$$

Step 2 (Turning inquiry *off* when it is not desirable). Make semi infeasible by choosing T_i large so that the good type cannot fund t^{semi}:

$$t^{\text{semi}} > y_G \iff T_i > \tfrac{\varphi\, y_G - k}{(1-\theta)(1-\alpha)}.$$

Since $\alpha > 1/2$, the denominator is positive, so such T_i exists whenever desired. With inquiry infeasible, use T_n to select the no-inquiry branch: implement pooling by keeping $X^{\text{pool}} \le \min\{y_G, y_B\}$ (the agent then weakly prefers no inquiry at the minimal transfer), or implement inactivity by taking $T_n > \max\{y_G, y_B\}/(1-\theta)$.

Step 3 (Turning inquiry *on* when it is desirable). Minimize the semi transfer by lowering T_i (down to 0 if needed), so t^{semi} becomes as small as possible. Sustaining semi requires three conditions:

(i) Good-type feasibility.

$$t^{\text{semi}} \le y_G \quad\iff\quad (1-\theta)(1-\alpha)T_i + k \;\le\; \varphi\, y_G. \tag{F}$$

(ii) Inquiry optimal for the agent at t^{semi}. At t^{semi} we have $U_i = 0$ by construction, while $U_{ni} = t^{\text{semi}} - (1-\theta)T_n$. Hence $U_i \ge U_{ni}$ iff

$$(1-\theta)T_n \;\ge\; t^{\text{semi}} \quad\iff\quad T_n \;\ge\; \tfrac{k+(1-\theta)(1-\alpha)T_i}{\varphi(1-\theta)}. \tag{AO}$$

(iii) Bad-type IC against the deviation $X' = T_n$. Off path the agent accepts any $X' \ge T_n$ without inquiry, yielding $y_B - T_n$ to type B. On-path under semi, B earns $(1-\alpha)(y_B - t^{\text{semi}})$. Deterring the deviation requires

$$y_B - T_n \;\le\; (1-\alpha)\,(y_B - t^{\text{semi}}) \quad\iff\quad T_n \;\ge\; \alpha\, y_B + (1-\alpha)\, t^{\text{semi}}. \tag{BIC}$$

If $\varphi y_G > k$, take $T_i = 0$ to minimize $t^{\text{semi}} = k/\varphi$, and then choose

$$T_n \;\geq\; \max\Big\{\tfrac{k}{\varphi(1-\theta)},\; \alpha y_B + (1-\alpha)\tfrac{k}{\varphi}\Big\},$$

which satisfies (AO) and (BIC) with slack; (F) holds since $\varphi y_G > k$. Pooling is then irrelevant because the agent strictly prefers inquiry. If $\varphi y_G \leq k$, even $T_i = 0$ gives $t^{\text{semi}} \geq y_G$, so semi is privately infeasible.

Step 4 (Welfare and dominance over single-penalty regimes). The dual-penalty pair (T_n, T_i) nests harm-based ($T_n = T_i$) and compliance-based ($T_i = 0$) rules. Therefore the maximal welfare attainable under dual penalties weakly dominates the maximal welfare attainable under either single-penalty design. Since $\alpha < 1$ implies false positives/negatives under inquiry, the maximal welfare under dual penalties is (weakly) below the perfect-inquiry benchmark. The implementability frontier stated in the proposition follows from Step 2 (deterring inquiry) and Step 3 (sustaining inquiry) via the linear bounds (F), (AO), and (BIC). □

Imperfect inquiry: calibration and comparative statics

This appendix collects the full cutoff expressions and comparative statics that underlie the brief discussion in Section 4. They make precise how penalty placement depends on signal accuracy α, inquiry cost k, the prior θ, and the type-dependent outputs y_G, y_B.

Assume imperfect inquiry with precision $\alpha \in (1/2, 1)$ and define

$$\varphi(\alpha) := \theta\alpha + (1-\theta)(1-\alpha).$$

Lemma B.4 (Calibration under imperfect inquiry by liability rule)**.** *Let $e^\star \in \{\text{pooling}, \text{semi}, \text{inactive}\}$ be the socially desirable behavior.*

(a) Harm-based single penalty T.

(i) Semi-pooling. *Implementable if*

$$\max\Big\{y_B,\; \tfrac{k}{\theta(1-\theta)(2\alpha-1)}\Big\} < T < \tfrac{\varphi(\alpha)\, y_G - k}{(1-\theta)(1-\alpha)} \qquad \textit{and} \qquad \varphi(\alpha) y_G > k.$$

(ii) Pooling. *Implementable if*

$$T \leq \min\Big\{y_B,\; \tfrac{y_G}{1-\theta}\Big\} \quad \textit{when } \theta y_G \leq k, \qquad T = \tfrac{k}{\theta(1-\theta)(2\alpha-1)},\; T \leq y_B \quad \textit{when } \theta y_G > k.$$

(iii) Inactive. *Implementable if*

$$T > \max\Big\{y_B,\; \tfrac{y_G}{1-\theta},\; \tfrac{\varphi(\alpha) y_G - k}{(1-\theta)(1-\alpha)}\Big\}.$$

(b) Compliance-based single penalty T (safe harbor after inquiry).

(i) Semi-pooling. *Implementable if*

$$T \geq \tfrac{k}{\varphi(\alpha)(1-\theta)} \qquad \textit{and} \qquad \varphi(\alpha) y_G > k.$$

(ii) Pooling. *Implementable if*

$$T \le \min\Big\{y_B,\ \tfrac{y_G}{1-\theta}\Big\} \quad \textit{when } \theta y_G \le k, \qquad T < \tfrac{y_G}{1-\theta} \quad \textit{when } \theta y_G > k.$$

(iii) Inactive. *Implementable if*

$$\varphi(\alpha) y_G \le k \qquad \textit{and} \qquad T > y_B.$$

(c) Dual penalties (T_n, T_i)**.**

(i) Semi-pooling. *Implementable if*

$$T_n > \tfrac{y_G}{1-\theta} \qquad \textit{and} \qquad (1-\theta)(1-\alpha)\, T_i + k \le \varphi(\alpha)\, y_G.$$

(ii) Pooling. *Implementable if*

$$T_n \le \min\Big\{y_B,\ \tfrac{y_G}{1-\theta}\Big\},$$

or, if $\theta y_G > k$,

$$T_n = \tfrac{k}{\theta(1-\theta)(2\alpha-1)}, \quad T_n \le y_B, \qquad \textit{and} \qquad (1-\theta)(1-\alpha)\, T_i + k > \varphi(\alpha)\, y_G.$$

(iii) Inactive. *Implementable if*

$$T_n > y_B \qquad \textit{and} \qquad (1-\theta)(1-\alpha)\, T_i + k > \varphi(\alpha)\, y_G.$$

Proposition B.1 (Comparative statics of imperfect-inquiry cutoffs)**.** *For the harm-based rule (and the* T_i*-arm of the dual-penalty rule), define*

$$L(\alpha) := \tfrac{k}{\theta(1-\theta)(2\alpha-1)}, \qquad U(\alpha) := \tfrac{\varphi(\alpha)\, y_G - k}{(1-\theta)(1-\alpha)}.$$

1. $L'(\alpha) < 0$*: higher accuracy lowers the penalty required to induce inquiry.*
2. $U'(\alpha) > 0$ *if* $\theta y_G > k$*;* $U'(\alpha) = 0$ *if* $\theta y_G = k$*;* $U'(\alpha) < 0$ *if* $\theta y_G < k$*.*
3. $\partial L/\partial k > 0$*,* $\partial U/\partial k < 0$*;* $\partial U/\partial y_G > 0$ *while* L *does not depend on* y_G*.*
4. $L(\theta)$ *is minimized at* $\theta = 1/2$*;* $U(\alpha)$ *rises in* θ *whenever* $\alpha y_G > k$*.*
5. *Limits: as* $\alpha \to 1$*,* $L(\alpha) \to k/[\theta(1-\theta)]$ *and, if* $\theta y_G > k$*,* $U(\alpha) \to +\infty$*; as* $\alpha \downarrow 1/2$*, the interval* $(L(\alpha), U(\alpha))$ *shrinks and may vanish unless* y_G *is large relative to* k*.*
6. *Under compliance-based rules, once uninformed action is deterred, screening is feasible iff* $\varphi(\alpha) y_G > k$*. Pooling and inactivity cutoffs in* $y_G/(1-\theta)$ *and* y_B *remain unchanged.*

Appendix C Extensions-Proofs and Technical Notes

C.1 Endogenous Inquiry Precision

Setup and assumptions. In semi-pooling, after agreeing to inquire the agent chooses precision $\alpha \in (1/2, 1)$ at cost $k(\alpha)$. Assume:

A1. $\alpha \in (1/2, 1)$ and $k : (1/2, 1) \to \mathbb{R}+$ is C^1, strictly increasing and strictly convex, with $k(1/2) = 0$ and $\lim_{\alpha\uparrow 1} k(\alpha) = \infty$.

A2. Transfers and liability can condition on (auditable) inquiry and its recorded precision (or a verifiable proxy).

A3. Tie-breaking: indifferent agents accept; indifferent principals offer the smallest transfer that induces acceptance.

Let λ denote the penalty that prices the agent's *residual exposure after inquiry*:

$$\lambda \equiv \begin{cases} T & \text{(harm-based single penalty)}, \\ 0 & \text{(compliance-based exemption)}, \\ T_i & \text{(dual-penalty: after-inquiry arm)}. \end{cases}$$

In semi-pooling, residual exposure occurs with probability $(1-\theta)(1-\alpha)$. With verifiability, the good principal (he) pays the minimal transfer that covers the agent's private inquiry bill $k(\alpha)$ plus expected residual exposure $(1-\theta)(1-\alpha)\lambda$.

Lemma C.1 (Agent's privately optimal precision)**.** *Fix a semi-pooling outcome and a residual-exposure price $\lambda \geq 0$. The agent's privately optimal precision $\alpha^*(\lambda)$ is the unique solution to*

$$k'(\alpha) = (1-\theta)\lambda,$$

and satisfies $\alpha^*(\lambda) \in (1/2, 1)$, $\alpha^{*\prime}(\lambda) > 0$, $\lim_{\lambda\downarrow 0} \alpha^*(\lambda) = 1/2$, and $\lim_{\lambda\uparrow\infty} \alpha^*(\lambda) = 1$.

Proof. In semi-pooling the agent minimizes $k(\alpha) + (1-\theta)(1-\alpha)\lambda$ on $(1/2, 1)$. Strict convexity of k gives a unique minimizer; the FOC is $k'(\alpha) = (1-\theta)\lambda$. Monotonicity and limits follow from A1. □

Welfare with semi-pooling. If inquiry is undertaken with precision α, the welfare contribution (transfers cancel) is

$$\mathrm{SW}_{\mathrm{semi}}(\alpha) = \theta\,\alpha\,y_G \;+\; (1-\theta)(1-\alpha)\,(\tau y_B - H) \;-\; k(\alpha). \tag{C.3}$$

The derivative is

$$\tfrac{d}{d\alpha}\mathrm{SW}_{\mathrm{semi}}(\alpha) = \theta\,y_G \;-\; (1-\theta)(\tau y_B - H) \;-\; k'(\alpha),$$

hence the socially optimal precision $\widehat{\alpha}$ (when semi-pooling is the target behavior) solves

$$k'(\widehat{\alpha}) = \theta\,y_G - (1-\theta)(\tau y_B - H). \tag{C.4}$$

By A1, $\widehat{\alpha} \in (1/2, 1)$ is unique whenever the right-hand side is positive.[18]

Proposition C.1 (Instrument power across liability regimes)**.** *Let $\alpha^*(\lambda)$ be as in Lemma C.1, and $\widehat{\alpha}$ satisfy* (C.4)*. Then:*

18. If $\theta\,y_G \leq (1-\theta)(\tau y_B - H)$ (net benefit of accuracy non-positive), the lawmaker prefers the lowest precision in the admissible set; under A1 this is 1/2. The interesting case for screening has $\tau y_B - H < 0$, making the RHS strictly larger than θy_G.

(a) ***compliance-based:*** *$\lambda = 0$ implies $\alpha^* = 1/2$. The lawmaker cannot raise precision via penalties; inquiry accuracy is privately minimized.*

(b) ***harm-based:*** *$\lambda = T$ so α^* is increasing in T. Raising T also tightens the* no-inquiry *margin, potentially eliminating pooling even when inclusion is desirable.*

(c) ***Dual:*** *$\lambda = T_i$ so α^* is increasing in T_i while the no-inquiry margin is governed by T_n. Thus T_i targets inquiry quality without collateral effects on pooling, and T_n controls selection on the no-inquiry branch.*

Proof. Immediate from Lemma C.1 and the mapping of λ to the penalty by regime. □

Proposition C.2 (Implementing the socially optimal precision under dual penalties)**.** *Suppose semi-pooling is (socially) the target behavior and is* feasible *at $\widehat{\alpha}$, i.e.,*

$$\theta\,\widehat{\alpha}\,y_G \;\geq\; k(\widehat{\alpha}) + (1-\theta)\big(1-\widehat{\alpha}\big)\,T_i. \tag{C.5}$$

Under dual penalties, setting

$T_i^\star \;=\; \frac{k'(\widehat{\alpha})}{1-\theta}$ *and choosing any T_n that preserves the desired selection on the no-inquiry branch*

induces $\alpha^* = T_i^\star \mapsto \widehat{\alpha}$ and implements the socially optimal precision.

Proof. By Lemma C.1, α^* solves $k'(\alpha) = (1-\theta)T_i$. With $T_i^\star$ as above the agent's private FOC coincides with (C.4), hence $\alpha^* = \widehat{\alpha}$. Feasibility is guaranteed by (C.5). T_n can be tuned (independently) to preserve or eliminate pooling as desired. □

Corollary C.1 (Limits under single-penalty regimes)**.** *Under compliance-based rules, $\alpha^* = 1/2$ and the lawmaker cannot implement $\widehat{\alpha} > 1/2$ via penalties. Under harm-based rules, the lawmaker can raise α^* by increasing T, but doing so simultaneously affects the no-inquiry branch, creating selection trade-offs that dual penalties avoid.*

Existence and welfare. When semi-pooling is socially desirable, feasibility with endogenous α requires that the good type's expected legal surplus at $\alpha^*(\lambda)$ covers the privately chosen inquiry bill:

$$\theta\,\alpha^*(\lambda)\,y_G \;\geq\; k\big(\alpha^*(\lambda)\big) + (1-\theta)\big(1-\alpha^*(\lambda)\big)\lambda,$$

with λ mapped to T, 0, or T_i by regime. Welfare at the induced precision follows from (C.3). Dual penalties allow the lawmaker to (i) select behavior on the no-inquiry branch via T_n and (ii) align inquiry quality with $\widehat{\alpha}$ via T_i, subject only to the feasibility condition above.

C.2 Off-path beliefs and implementability

Let $\hat{\mu} \in [0,1]$ denote the agent's belief that the task is *legal* if she accepts an unexpected (off-path) offer without inquiry. Under a dual-penalty rule, let T_n be the penalty if she *did not* inquire, and T_i the penalty if she *did* inquire but still performs an illegal task (e.g., a false positive). Under single-penalty rule, both T_n and T_i merge into T. Let $\alpha \in (1/2, 1)$ denote inquiry precision, $\theta \in (0,1)$ the prior probability of a good principal (he), and $k > 0$ the inquiry cost.

Lemma C.2 (Off-path thresholds)**.** *(i) If the agent accepts an off-path offer X' without inquiry, her expected utility is $X' - (1-\hat{\mu})T_n$, so (by tie-breaking) she accepts the smallest such offer $X'_{\min} = (1-\hat{\mu})T_n$. (ii) A deviating bad principal earns $y_B - (1-\hat{\mu})T_n$ at $X'_{\min}$; he is deterred iff*

$$T_n \;>\; \tfrac{y_B}{1-\hat{\mu}}.$$

(iii) Suppose the on-path outcome is screening (semi-pooling): the good principal funds inquiry and the agent acts only on a "legal" signal. The minimal on-path transfer that induces inquiry equals

$$X^{\text{semi}} \;=\; k + (1-\theta)(1-\alpha)\,T_i,$$

i.e., the inquiry cost plus the agent's residual expected exposure under imperfect inquiry. To deter a good-type off-path deviation to "no inquiry" acceptance, it suffices that

$$(1-\hat{\mu})T_n \;\geq\; X^{\text{semi}} \quad\Longleftrightarrow\quad T_n \;\geq\; \tfrac{k+(1-\theta)(1-\alpha)T_i}{1-\hat{\mu}}.$$

Under perfect inquiry ($\alpha = 1$), this reduces to $T_n \geq k/(1-\hat{\mu})$.

Proof. Part (i) follows from the agent's off-path acceptance condition with no inquiry: $X' - (1-\hat{\mu})T_n \geq 0$, minimized at equality. Part (ii) plugs X'min into the deviating bad type's payoff $y_B - X'$min and requires it < 0. Part (iii) equates the off-path acceptance threshold to the on-path payment that exactly covers private inquiry cost k plus residual expected penalty $(1-\theta)(1-\alpha)T_i$. The perfect-inquiry simplification is immediate when $(1-\alpha) = 0$. □

The deterrence bound for the bad type, $y_B/(1-\hat{\mu})$, is (weakly) increasing in $\hat{\mu}$; the no-deviation bound for the good type, $\big[k+(1-\theta)(1-\alpha)T_i\big]/(1-\hat{\mu})$, is also (weakly) increasing in $\hat{\mu}$. Hence moving from the pessimistic benchmark $\hat{\mu} = 0$ to more optimistic beliefs ($\hat{\mu} > 0$) *weakly shrinks* the penalty windows that support either pooling or screening. This is the sense in which the pessimistic convention maximizes implementability.

C.3 Mixed strategies by the principal

Fix any liabilities (single-penalty or dual-penalty), $\alpha \in (1/2, 1]$, and $k > 0$. Consider a candidate profile where the bad type randomizes over two transfers, $X_\ell < X_h$, while the good type plays a single transfer. Let $\hat{\mu}(X)$ denote the agent's posterior that the task is legal upon observing X (Bayes-consistent on-path; arbitrary off-path subject to the convention above).

Proposition C.3 (Extreme-point best reply for the bad type)**.** *For any fixed off-path beliefs and the agent's best response, the bad type's expected payoff as a function of his own transfer X is piecewise linear with at most one kink (the point at which the agent switches from "inquire/reject when illegal" to "accept without inquiry"). Hence his best reply is attained at an extreme point: either the lowest X that still induces acceptance without inquiry, or the highest feasible X that yields acceptance. Generically, the maximizer is unique and pure.*

Proof. For each observed X, the agent's best response is threshold in X and $\hat{\mu}(X)$: below a cutoff, she inquires (and rejects when illegal); above it, she accepts without inquiry. Therefore the acceptance probability as a function of X is a step function with

a single jump. The bad type's expected payoff is $y_B - X$ when acceptance occurs and 0 otherwise, so it is piecewise linear in X with at most one kink at the jump. A piecewise-linear function with one kink attains its maximum at an endpoint unless parameters are knife-edge; thus the bad type's best reply is pure except on a measure-zero set. □

Implication. Apart from the trivial inactive case (rejection regardless of offers), equilibrium behavior is exhausted by the pure types analyzed in the main text (pooling or semi-pooling). Allowing the bad type to "mix" between legal and illegal tasks with some probability q simply rescales posteriors $\hat{\mu}(X)$ and leaves the threshold logic unchanged; it does not generate new equilibrium types.

C.4 Psychological Inquiry Costs and Social Preferences

Let the agent's *private* inquiry cost be k, but the lawmaker place weight $\sigma \in [0, 1]$ on its non-material component in *welfare*. Let $m \geq 0$ be an internal moral cost the agent suffers when she performs an illegal task without inquiry. Under dual penalties, T_n applies to no-inquiry acts and T_i to post-inquiry illegal acts (false positives).

Lemma C.3 (Welfare and feasibility with (σ, m))**.** *(i) The* private *inquiry condition is unchanged by σ and m; thresholds shift only with k, T_n, T_i, and α. (ii) The semi-pooling welfare term becomes $\theta y_G - \sigma k$ (perfect inquiry) and $\theta\alpha y_G + (1-\theta)(1-\alpha)(\tau y_B - H) - \sigma k$ (imperfect inquiry). (iii) Under pooling (no inquiry), the minimal acceptable transfer is $X^* = (1-\theta)(T_n + m)$ (replace T_n by T with a single penalty). Feasibility of pooling requires $y_G \geq X^*$ and $y_B \geq X^*$.*

Proof. (i) σ is a lawmaker's weight and does not enter the agent's private problem. m is borne only when the agent performs an illegal task under no inquiry; it does not affect the inquiry branch. (ii) Transfers cancel in welfare; the only change from σ is the scaled k. Under imperfect inquiry, the standard welfare term for semi-pooling is reduced by noise; σ scales k identically. (iii) Under pooling, the agent's expected disutility is $(1-\theta)(T_n + m)$; tie-breaking gives the threshold X^* and the feasibility conditions stated. □

Implications. Lower σ expands the region where inquiry is *socially* preferred but leaves implementability unchanged. Higher m substitutes for T_n when deterring no-inquiry action is the objective, but it also makes sustaining pooling harder when inclusion of both types is desirable. Dual penalties keep the levers distinct: tune T_n (with m) for the no-inquiry margin and T_i for residual exposure after inquiry.

C.5 Additional notes on market structure

Throughout, "screening" refers to the semi-pooling equilibrium in which the agent inquires and performs only when the signal (or belief) indicates legality. Penalty notation under imperfect inquiry: T_n applies when the agent acts *without* inquiry; T_i applies to post-inquiry illegal performance (false positives). Precision $\alpha \in (1/2, 1)$ and $\varphi := \theta\alpha + (1 - \theta)(1 - \alpha)$.

A. Single principal with uncertainty about legality

Lemma C.4 (Thresholds and welfare equivalence)**.** *Let $y_G = y_B \equiv y$ so that neither party can identify legality without inquiry.*

1. *Under perfect inquiry,* pooling *(no inquiry) is feasible iff $y \geq (1-\theta)T_n$;* screening *(inquiry) is feasible iff $\theta\, y \geq k$. If both are infeasible, the unique equilibrium is inactive. Welfare expressions match the baseline with y_G, y_B replaced by y.*

2. *Under imperfect inquiry with dual penalties,* pooling *feasibility remains $y \geq (1-\theta)T_n$.* Screening *is feasible iff*

$$\varphi\, y \;\geq\; k + (1-\theta)(1-\alpha)\, T_i, \qquad \varphi := \theta\alpha + (1-\theta)(1-\alpha).$$

Welfare under each equilibrium equals the baseline formulae with the same replacement and the (standard) reduction of screening welfare by noise.

Proof. Pooling: If the agent accepts without inquiry, her minimal acceptance transfer is $X^{\text{pool}}\text{min} = (1-\theta)T_n$; with $y_G = y_B = y$, feasibility is $y \geq X^{\text{pool}}\text{min}$.

Screening (perfect inquiry): With $\alpha = 1$, legality is revealed. The agent needs $X^{\text{scr}}_{\text{min}} = k$ to cover the inquiry cost. The (good) principal obtains y with probability θ, so screening is fundable iff $\theta y \geq k$.

Screening (imperfect inquiry): With $\alpha \in (1/2, 1)$, the agent faces residual exposure $(1-\theta)(1-\alpha)T_i$ from false positives; thus $X^{\text{scr}}\text{min} = k + (1-\theta)(1-\alpha)T_i$. The pooled offer leads to a "legal" signal with probability φ, hence fundability is $\varphi\, y \geq X^{\text{scr}}\text{min}$. Welfare follows by cancelling transfers and adding the standard noise loss on the screening branch. □

B. No distinct principal (self-solicitation)

Lemma C.5 (Equivalence to the single-principal benchmark)**.** *If the principal and agent are the same decision-maker (no transfers), the implementability thresholds coincide with Lemma C.4. In particular, under perfect inquiry the agent chooses screening iff $\theta\, y \geq k$; under imperfect inquiry with dual penalties, screening iff $\varphi\, y \geq k + (1-\theta)(1-\alpha)T_i$; pooling is feasible iff $y \geq (1-\theta)T_n$.*

Proof. Without transfers, decisions compare private surplus directly. *Pooling:* accept if $y - (1-\theta)T_n \geq 0 \iff y \geq (1-\theta)T_n$. *Screening:* with perfect inquiry, accept iff $\theta y - k \geq 0$; with noise, accept iff $\varphi y - \big(k + (1-\theta)(1-\alpha)T_i\big) \geq 0$. These are the thresholds in Lemma C.4. □

C. Agent bargaining power (or competition among principals) We allow the agent to extract rents above the minimal acceptance transfer. Let $r_{\text{pool}} \geq 0$ denote the additional rent (unconditional, paid upon acceptance) in pooling; let $r_{\text{scr}} \geq 0$ denote the additional rent in screening. We consider two common modalities: (i) *unconditional rent* (paid upon acceptance); (ii) *success-contingent rent* (paid only when the task is performed).

Lemma C.6 (Pooling shrinks; screening depends on rent modality)**.** *Under either perfect or imperfect inquiry:*

1. Pooling: *feasibility becomes* $\min\{y_G, y_B\} \geq (1-\theta)T_n + r_{pool}$. *Hence agent bargaining power (higher r_{pool})* strictly shrinks *the pooling region.*

2. Screening, unconditional rent: *feasibility becomes (perfect inquiry)* $\theta\, y_G \geq k + r_{scr}$*; (imperfect inquiry)* $\varphi\, y_G \geq k + (1-\theta)(1-\alpha)T_i + r_{scr}$*. Thus unconditional rents* shrink *the screening region.*

3. Screening, success-contingent rent: *if the extra rent* r_{scr} *is paid only when the task is actually performed, the expected rent cost is scaled by the success probability: (perfect inquiry)* $\theta\, y_G \geq k + \theta\, r_{scr}$*; (imperfect inquiry)* $\varphi\, y_G \geq k + (1-\theta)(1-\alpha)T_i + \varphi\, r_{scr}$*. Relative to unconditional rents of the same size, success-contingency* weakly expands *the screening region.*

Proof. Pooling: The minimal acceptance transfer rises from $(1-\theta)T_n$ to $(1-\theta)T_n + r_{\text{pool}}$, and both types must cover it, giving the stated constraint.

Screening: The agent's minimal transfer on the screening branch rises by r_{scr}. If r_{scr} is unconditional, it is a level shift; if success-contingent, it is multiplied by the success probability (θ with perfect inquiry; φ with noise). The principal's fundability condition equates expected revenue (success probability times y_G) to this augmented transfer. The stated inequalities follow. □

Remark C.1 (Comparative implementability). Because r_{pool} burdens *both* types while r_{scr} burdens only the good type and can be success-contingent, competition that shifts surplus to the agent typically shrinks pooling more than screening. Thus, relative to the baseline, market power on the agent side tilts implementability toward screening *if* extra rents are structured success-contingently; with unconditional rents, both regions shrink.