ROBUST TECHNOLOGY REGULATION*

Andrew Koh
MIT

Sivakorn Sanguanmoo
MIT

Job Market Paper

November 6, 2025
latest version here

Abstract

We analyze how uncertain technologies should be robustly regulated and how regulation should evolve with new information. An *adaptive sandbox* comprising a zero marginal tax up to an evolving quantity limit is (i) *robust*: it delivers optimal payoff guarantees when the agent's learning process and/or preferences are chosen adversarially; (ii) *dominant*: it outperforms other robust and regular mechanisms across all agent learning processes and preferences; (iii) *time-consistent*: it is the only robust mechanism that can be implemented without commitment. Robustness is *important*: absent robust regulation, worst-case payoffs can be arbitrarily poor and are induced by weak but growing optimism that encourages excessive risk-taking. Our results offer optimality foundations for existing policy and speak directly to current debates around managing emerging technologies.

*Koh: MIT Department of Economics; ajkoh@mit.edu. Sanguanmoo: MIT Department of Economics; sanguanm@mit.edu.

We are grateful to Stephen Morris, Drew Fudenberg, and Daron Acemoglu for guidance, support, and many conversations. We also thank Ian Ball, Simon Board, Roberto Corrao, Théo Durandard, Piotr Dworczak, Matt Elliott, Andrea Galeotti, Joshua Gans, Alexis Ghersengorin, Gillian Hadfield, Anton Korinek, Jiangtao Li, Zihao Li, Jonathan Libgober, Daniel Luo, Parag Pathak, Jean Tirole, Mark Whitmeyer, Alex Wolitzky, Frank Yang, Weijie Zhong, as well as audiences at the 2025 SITE session on 'Dynamic Games', the 2025 AEA Meeting session on 'Policy Implications of Transformative AI', Johns Hopkins, and MIT Theory and Public Finance lunches for helpful comments. Andrew Koh acknowledges support from the Gordon B. Pye Dissertation Fellowship and the Centre for the Governance of AI (GovAI) where part of this work was done. First posted version: August 2024.

1 INTRODUCTION

The first robot from Greco-Roman antiquity was Talos, a giant bronze automaton who hurled boulders at passing ships. The Greeks could not build Talos, but they did try—Archimedes built catapults that rained 'immense masses of stones... with incredible din and speed' (Plutarch, *Marcellus* 14-19). Such technologies were seen, in antiquity, as something to be unleashed.[1] Restraint, by contrast, is a more modern preoccupation that accompanied the industrial revolution and the rise of the regulatory state. Recent advancements across robotics, biology, and artificial intelligence mean that we could soon build our own Talos. What are we to make of these new and perhaps transformative technologies? How, if at all, should we regulate them?

Motivated by these questions, we build a theory of technology regulation that takes three 'stylized facts' seriously:

1. **Uncertainty** about whether a new technology is beneficial or harmful;[2]
2. **Misalignment** between the firm developing the technology and wider society;
3. **Learning** as the technology is developed.

In our model, a real-valued technology level reflects the extent of irreversible R&D or deployment, more of which is beneficial if the technology is safe, but harmful if it is dangerous. The firm developing the technology is less risk-averse than the regulator whose preferences reflect those of wider society. This differential risk aversion might arise if the firm does not internalize potential harms, stands to gain disproportionately from the potential benefits, or is caught up in an 'R&D arms race'. Both parties learn as the technology is developed, but the firm learns more than the regulator and might do so in quite varied ways e.g., extra independent or correlated information, or simply observe the same information more quickly.

The firm's equilibrium technology choices are jointly shaped by *policy*—the regulator's chosen paths of taxes or subsidies that might evolve with new information,

[1] The description of Talos is given by Apollonius Rhodius in the Argonautica (4.1638-1688). The word 'automaton' (αὐτόματον) was first used in the Iliad (5.749) to describe the automatic wheeled tripos of Hephaestus. Aeschylus glorifies Prometheus for giving humans 'all greatness' and 'all arts' (τέχνη/techne).

[2] For instance, there is substantial amount of disagreement among experts and forecasters on the tail risks posed by artificial intelligence (Karger et al., 2023) where the 25th and 75th percentile disagree by a factor of about 100.

preferences—the firm's willingness to develop the technology given its beliefs, and *learning*—what the firm has learnt thus far, what it might yet learn as it develops the technology further, and how it expects future regulation to unfold. We develop broad principles for regulating new technologies, as well as guidance for how regulation should evolve with new information.

Economically, we argue for an *adaptive sandbox* that imposes a zero marginal tax on the technology up to an evolving quantity limit. This quantity limit adapts with the regulator's interim information, loosening and tightening as the regulator grows more or less optimistic. The adaptive sandbox is (i) *robust*: it delivers the best payoff guarantee across all agent learning processes and/or preferences; (ii) *undominated* and *dominates* all robust mechanisms that satisfy a basic regularity property; and (iii) *time-consistent*: it is the only robust mechanism that can be implemented without commitment. We further argue that robustness is *important*: sans robust regulation, the worst-case learning process is natural and progressively induces the firm to take on excessive risk—this can magnify even small wedges in preferences to yield unboundedly poor regulator payoffs. Taken together, our results offer a principled defense of quantity over price instruments for technology regulation, as well as sharp refinements among different kinds of quantity instruments.[3]

Conceptually, our framework offers a new perspective on how mechanisms should simultaneously *evolve* with new information (adaptivity) and *safeguard* against the worst-case (robustness). Our robustness notion is new, and requires optimal mechanisms to learn under ambiguity—but it also reduces to more familiar robustness notions when the regulator does not learn. Our notion of adaptive mechanisms lives in the rich middle ground between ex ante and ex post regulation (Laffont and Tirole, 1993)—we use it to shed light on how policy might avoid pitfalls associated with ex ante regulation which fails to exploit interim information, and ex post regulation (e.g., liability regimes) which relies excessively on verification and/or tort law.[4]

Practically, our results speak directly to existing debates around how new and potentially transformative technologies should be regulated. The adaptive sandbox in our model qualitatively resembles real-world policies ('regulatory sandboxes')

[3] As Weitzman (1974) notes, "the average economist in the Western marginalist tradition has at least a vague preference toward indirect control by prices, just as the typical non-economist leans toward the direct regulation of quantities."

[4] Our results also lend precision and guidance to a recent trend in environmental and AI policy advocating for 'anticipatory' or 'adaptive' regulation (Bennear and Wiener, 2019; Reuel and Undheim, 2024; Schrepel, 2025).

that have recently been introduced to regulate new financial technologies,[5] and are increasingly deployed to regulate artificial intelligence.[6] Despite their growing prevalence, surprisingly little is known—either in theory or practice—about these policies. We bridge this gap by making precise when and why these kinds of adaptive quantity limits work well and, in so doing, offer foundations for the oft-articulated goal of sandboxes as providing a "dynamic, evidence-based approach to regulation" (OECD, 2023).

Outline of results.

Robustness. We are interested in mechanisms that perform well even if we are prepared to assume little about what agents might learn, or what preferences they hold. *Learning-robust* mechanisms deliver optimal worst-case guarantees across all learning processes the agent might face. *Dually-robust* mechanisms deliver optimal worst-case guarantees across all learning processes and agent preferences. We show that an adaptive sandbox mechanism consisting of a zero marginal tax on the technology up to an evolving hard limit—beyond which further R&D or deployment is disallowed—is both learning- and dually-robust. This is driven by two key properties:

- *Aligned sensitivity to information*: before the sandbox limit, the zero marginal tax minimally distorts the shape of relative incentives between firm and regulator. To show this, we develop new comparative static results relating risk-aversion to the quantity of experimentation under arbitrary learning processes. This implies that under any learning process, the sandbox positively correlates stopping incentives: if the firm wishes to stop pushing the technology, then so would the regulator *if* they had access to the firm's private information.
- *Minimal sensitivity to information*: at the sandbox limit, firms are forced to ignore their interim information and stop pushing the technology. This safeguards the principal against learning processes we call *weak optimism* that anti-correlate continuation incentives: the agent grows progressively more optimistic and thus wishes to continue development or deployment, but not quickly enough—if the regulator held the same belief they would prefer to stop. Hard limits safeguard

[5] By 2020, regulators in over 50 jurisdictions have adopted some form of sandbox to regulate new financial technologies (Appaya et al., 2020). By 2023, approximately 100 sandbox initiatives have been implemented (OECD, 2023).

[6] Article 57 of the EU AI Act (European Parliament, 2024) stipulates that member states must establish AI regulatory sandboxes by August 2026.

against such learning processes.

Undominated and dominant. Why might we prefer sandboxes over other robust mechanisms (of which there are infinitely many)? We offer a simple but, we think, compelling reason: the adaptive sandbox is undominated, and dominates all other dually-robust mechanisms that satisfy a basic regularity property that rules out vacillating between taxing and rewarding pushing the technology further. That is, for any learning process and any agent preference—even those away from the worst case—the adaptive sandbox delivers higher regulator payoffs.[7] Dominance is driven by a third key property:

- *Maximal sensitivity to information*: across all mechanisms that induce aligned sensitivity (a necessary condition for robustness), our adaptive sandbox keeps the firm maximally sensitive to interim adverse information: under any learning process and any firm preference, sandboxes do better at positively correlating the events that the firm and regulator both wish to halt.

Taken together, robustness, dominance, and regularity are illustrated in Figure 1 (a) where each circle represents a distinct desideratum. The optimal adaptive sandbox is the unique adaptive mechanism that fulfills all three properties.

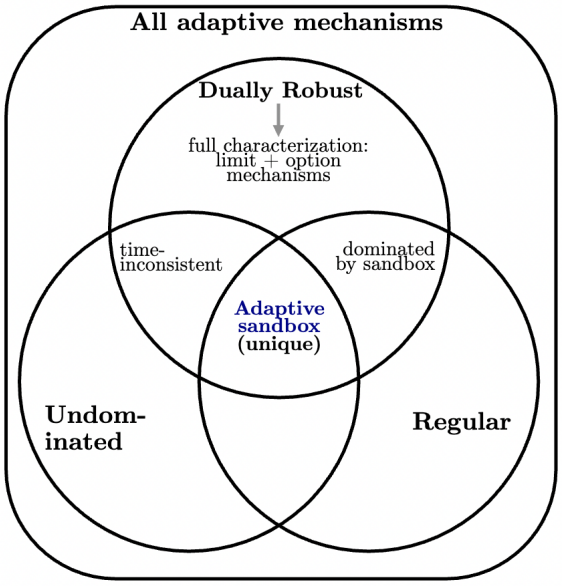
Time-consistency. Beyond evaluating mechanisms according to their ex ante performance (as does robustness and dominance), we might also be interested in whether our regulator has incentives to follow-through ex interim, at every possible history both on and off the equilibrium path of play.[8] The adaptive sandbox is time-consistent: whenever the agent halts or continues pushing the technology, whatever the regulator has learnt thus far, she has no incentive to revise the mechanism. Time-consistency is driven by a kind of positive selection induced by the sandbox's aligned sensitivity: if an agent stops, the regulator must infer that her own optimal technology level has already passed. This, in turn, prevents a Coase conjecture-like logic—in which the agent strategically halts to induce more lenient regulation—from taking hold. What is more, *all* other dually-robust mechanisms are time-inconsistent—each of them presents the regulator's future selves with the temptation to deviate to a different mechanism such as to screen the agent's interim belief.

The importance of robustness. We do not think robustness is always—or even often!—

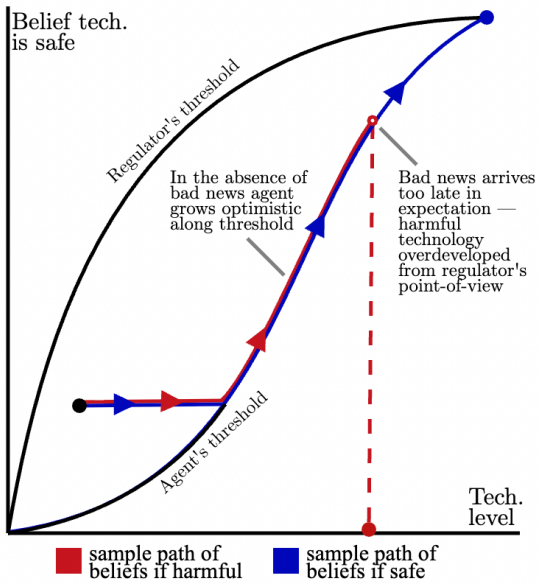
[7] Weakly higher for every learning process and agent preference, and strictly higher for some learning process and agent preference.

[8] Indeed, such problems of commitment loom large in monopoly or banking regulation (Laffont and Tirole, 1993).

Figure 1: Outline of results



(a) Desiderata for mechanisms



(b) Worst-case sans robustness

a good guide to policy. But we argue that it is especially *important* in the context of regulating risky technologies: sans robust regulation, the worst-case learning process is both natural and delivers poor regulator payoffs. To make this precise, we fully characterize the form and value of the worst-case learning process under non-robust mechanisms. They takes the form of an inhomogeneous 'bad news' Poisson process so that, in the absence of news that the technology is harmful, the agent becomes gradually more confident in its safety—but only at a rate that keeps her indifferent between continuing and halting the technology. We call such learning processes *weak optimism*, as illustrated in Figure 1 (b). Weak optimism hurts welfare by maximizing the distribution of false positives i.e., the technology is overdeveloped or deployed when it is harmful (red path), while inducing no false negatives i.e., the technology is always maximized when it is beneficial (blue path). In the absence of robust regulation, even small wedges in risk aversion can be substantially magnified by weak optimism to induce unboundedly poor payoffs for the regulator. Hard limits safeguard against exactly this.

Weak optimism arises naturally whenever the harms of new technologies are lumpy and extreme (e.g., pandemic induced by biotechnology; market crash induced by new financial instruments; AI catastrophe from misalignment) since the absence of disaster induces growing optimism that the technology is safe. Silicon Valley

assures us that the risks from new technologies are under control because *they* learn as the technology is developed. Thus writes Sam Altman, CEO of OpenAI: "the best way to make an AI system safe is by iteratively and gradually releasing it... learning from experience" (Altman, 2025).[9] But our results imply that in the absence of stringent regulation, the possibility of learning per se should not assure *us*—in fact, a little learning is dangerous, and can deliver worse outcomes than no learning at all. Indeed, a number of AI scientists pushing the technological frontier have raised concerns that powerful AI systems can be dangerous because of misalignment (Greenblatt et al., 2024), loss of human control (Kulveit et al., 2025), or wider socioeconomic disruptions, but nonetheless persist with R&D or deployment of these technologies—we think this is consistent with weak optimism. That the worst-case learning process is both qualitatively and quantitatively important suggest that our notions of robustness should be a first-order concern for policymakers.

Related literature. Our work relates directly to the burgeoning literature on the consequences of new technologies. Our focus is how policy should be designed when the private sector is misaligned[10] and thus relate to Acemoglu and Lensman (2024); Gans (2025); Guerreiro et al. (2025) who study regulation under a public and known learning process in which the agent does not have better information. Our work differs in several substantial ways. First, our agent has superior information which can be leveraged. Second, we seek optimal performance guarantees across all learning processes and/or preferences. Third, we analyze adaptive policies that evolve with the regulator's information. Our optimal policies thus differ substantially. Nonetheless, our characterization of the worst-case learning processes speaks directly to the policies proposed by these papers, and makes precise when and how price-instruments like Pigouvian taxes or coarse quantity-instruments like quotas perform poorly.

Our first desideratum is robustness and thus relates to the literature on robust mechanism design that seeks performance guarantees over information structures (Bergemann and Morris, 2005) or preferences (Armstrong, 1995). Work here has

[9] Likewise argues Dario Amodei: "We always monitor the models... so that we have a continuous process where we don't get taken by surprise." (Amodei, 2025)

[10] See Jones (2016, 2024, 2025); Aschenbrenner and Trammell (2024) for analysis of the welfare-maximizing solution without misaligned incentives. Acemoglu (2021) surveys potential socioeconomic harms posed by AI systems used in prediction; Hadfield and Koh (2025) survey risks distinctive to AI agents that can plan and execute complex tasks over long horizons; Feyen et al. (2021) surveys the promises and perils of new financial technologies.

primarily focused on static mechanisms.[11] But R&D is a fundamentally dynamic process intertwined with learning—our regulator thus seeks payoff guarantees over all adaptive information structures more informative than the principal's own. In the special case with no principal learning, this is related to Libgober and Mu (2021) who analyze informationally robust dynamic pricing.[12] Even here, an important difference is their monopolist only seeks to maximize profits; by contrast, our principal is concerned with shaping the *joint* distribution over equilibrium technology levels and the state. Thus, our optimal mechanisms and their underlying economic forces differ substantially. With principal learning, our robustness notion is conceptually new because optimal mechanisms must 'learn under ambiguity' (Epstein and Schneider, 2007), but retains the spirit of informational robustness.

Our second desideratum is dominance which is in the tradition of Wald (1950); Savage (1972). We show the optimal adaptive sandbox is undominated in the sense of Borgers, Li, and Wang (2024). Undominated does not imply robust, nor does robust imply undominated (Börgers, 2017). We show, however, that dominance and robustness can be combined to obtain sharp prescriptions: our adaptive sandbox is the unique mechanism that is robust, undominated, and regular. In particular, it dominates all other robust and regular mechanisms and, in this regard, is in similar spirit to work by Guo and Shmaya (2023); Dworczak and Pavan (2022) who rank robust policies away from the worst-case in context of static project choice and persuasion. Our dominance notion is demanding, and requires us to evaluate the performance of mechanisms under *any* learning process and agent preference. To do so, we establish new comparative static results relating risk-aversion to optimal experimentation that hold for general payoffs and learning processes—including those that are non-continuous and non-Markov—thereby nesting, unifying, and generalizing past work (Chancelier et al., 2009; Keller et al., 2019).

We also relate to the literature on instrument choice pioneered by Weitzman (1974) though our (i) policy is fully flexible and can adapt to the regulator's evolving information; and (ii) the agent's experimentation problem is dynamic. When the regulator

[11] In the context of pricing and auctions, see e.g. (Bergemann, Brooks, and Morris, 2015, 2017; Du, 2018; Brooks and Du, 2023). See Frankel (2014) for a more recent analysis of robustness to preferences. Also related is Guo and Shmaya (2019) who study robust monopoly regulation and Carroll (2017) who studies robust multidimensional screening.

[12] See also Li, Libgober, and Mu (2022) who study dynamic pricing under robustness concerns without commitment; in our setting, the regulatory sandbox mechanism is time-consistent. Hannan (1957); Vovk (1990) are early analyses on robustness in dynamic environments; Chassang (2013); Penta (2015) are important early work on dynamically robust mechanisms.

does not learn, our sandbox has a fixed boundary and thus resembles the static 'interval delegation' solutions that often arise in delegation problems (Holmström, 1977; Amador, Werning, and Angeletos, 2006) though the forces driving our results are different. In our setting, the zero marginal tax ensures that any deviation by nature to a different learning process must improve the principal's payoff, while the hard limit is necessitated by robustness concerns. Our results additionally offer a rationale for 'delegation-like' instruments to emerge endogenously under concerns for *both* robustness and dominance.[13]

Finally, our results relate to work on optimal experimentation. Much of this literature has assumed a particular exogeneous learning processes for tractability e.g., conclusive good news (Keller, Rady, and Cripps, 2005) or bad news (Keller and Rady, 2015); see, e.g., Guo (2016). By contrast, our agent's learning process is private and chosen adversarially and flexibly. Normatively, this allows us to encode non-Bayesian uncertainty about the learning process. Economically, this allows us to identify the form of worst-case learning processes that robust mechanisms safeguard against. To do so, we recast nature's problem as one of dynamic information design and have found it helpful to draw on techniques from some of our earlier work which analyzed these problems (Koh and Sanguanmoo, 2022; Koh, Sanguanmoo, and Zhong, 2024).[14] However, the present setting differs in meaningful ways because our fictitious designer (nature) has state-dependent preferences which is novel to the literature. Our outer problem of optimizing dynamic mechanisms is related to the elegant work of Kruse and Strack (2015, 2019) who study the 'inverse stopping problem' of designing a boundary to implement a target stopping level under a known diffusion (Markov with continuous sample paths). A key difference is that our regulator evaluates policies according to the worst-case over all learning processes, including those that are non-Markov and non-continuous.

2 MODEL

2.1 Technology and uncertainty. The *technology state* takes values in $\Theta = \{0, 1\}$ where $\theta = 1$ corresponds to the technology being beneficial and $\theta = 0$ corresponds to being

[13] If the principal wanted a robust *or* undominated policy, not all of them are 'delegation-like'.

[14] The literature on dynamic information design in stopping problems has focused on a designer with preferences over stopping levels (Ely and Szydlowski, 2020; Orlov, Skrzypacz, and Zryumov, 2020; Koh and Sanguanmoo, 2022; Saeedi, Shen, and Shourideh, 2024), or over both times and actions (Koh, Sanguanmoo, and Zhong, 2024).

harmful.[15] The *technology space* is a compact set $\mathcal{L} \subseteq \mathbb{R}$ that is either discrete or continuous and we write $L := \max \mathcal{L}$. Depending on the specific technology in question, the level $l \in \mathcal{L}$ might correspond to the degree of R&D, extent of deployment, and so on. Importantly, we assume that technology is irreversible.[16]

2.2 Learning. There is a *principal* representing a regulator, and an *agent* representing a firm developing the technology. They share a common prior $\mu_{0-} \in \Delta(\Theta) := [0,1]$, where we use the notation '0–' to denote the ex ante level such as to accommodate changes in beliefs exactly at level 0. Both principal and agent learn about θ as the technology is developed.

The principal's learning process is represented by a right-continuous filtration $\mathcal{G} := (\mathcal{G}_l)_{l\in\mathcal{L}}$ generated by some right-continuous signal process $(X_l)_{l\in\mathcal{L}}$ with measure $\mathbb{P}^\theta$.[17] We assume the following technical conditions that will simplify our exposition:

(i) $\mathbb{P}^0$ and $\mathbb{P}^1$ are mutually absolutely continuous; and

(ii) if $\mathbb{P}^1 \neq \mathbb{P}^0$ then for any $M, \epsilon > 0$, $\mathbb{P}\left(\log \frac{d\mathbb{P}^1}{d\mathbb{P}^0}\Big|_{\mathcal{G}_{l+\epsilon}} < -M \,\middle|\, \mathcal{G}_l\right) > 0$.

These conditions are fulfilled whenever $(X_l)_{l\in\mathcal{L}}$ is a jump diffusion (e.g., a Lèvy process with mutually absolutely continuous diffusion coefficients and jump measures) though the principal's learning might be more generally path-dependent, containing information both about state θ as well as about the informativeness of the future signal process. The agent learns (weakly) more: her learning process is represented by the filtration $(\mathcal{F}_l)_l$ that is finer than the principal's:

$$\mathcal{G}_l \subseteq \mathcal{F}_l \text{ for all } l \in \mathcal{L}. \tag{R}$$

The set of possible learning processes for the agent is given by

$$\mathbb{F} := \left\{(\mathcal{F}_l)_l : \mathcal{F} \text{ is a right-continuous filtration fulfilling (R)}\right\}$$

[15] All results in this paper extends to any set of ordered states with the exception of Theorem 4 on the form of the worst-case learning process that needs to be modified.

[16] See, e.g., Weitzman (1998) for an influential model of growth building on this idea that once we have learnt how to build something e.g., a bomb, synthetic pathogen, etc. we cannot unlearn it. Technology might also be locked-in because of networked adoption, behavioral biases, decentralization (as with the release of AI model weights), or even because the technology itself resists displacement (Russell, 2019).

[17] That is, $\mathbb{P}^\theta$ is the probability measure of the signal process over the Skorokhod space $\mathbb{D}(\mathcal{L};\mathbb{R}^d)$ when the technology state is θ. There is an appropriately filtered probability space in the background that we suppress in the main text to focus on the economics; details are in Appendix A.

and we use $\mu_l := \mathbb{E}[\theta \mid \mathcal{F}_l]$ to denote the agent's beliefs about the state. The agent's beliefs $(\mu_l)_l$ evolve (i) as a martingale by Bayes' rule; and (ii) right-continuously since $(\mathcal{F}_l)_l$ is right continuous.

We emphasize that the set of agent learning processes $\mathbb{F}$ is rich, capturing not simply how the agent's information about the state evolves, but also about the informativeness of the agent's and principal's future learning process. This allows us to capture various ways in which the private sector might be more informed, as the following examples illustrate:

1. **No principal learning.** $\mathcal{G}_l = \mathcal{G}_{0-}$ a.s. We have already noted that the agent's beliefs $\mu_l = \mathbb{E}[\theta \mid \mathcal{F}_l]$ evolves as a càdlàg martingale. Moreover, any càdlàg martingale—potentially non-Markov or non-continuous—can be induced by some learning process $\mathcal{F} \in \mathbb{F}$.

2. **Independent extra information.** The agent observes the principal's signal process $(X_l)_l$ and, in addition, observes an additional conditionally independent signal process $(Y_l)_l$.

3. **Correlated extra information.** The agent observes the principal's signal process $(X_l)_l$ and, in addition, observes an additional signal process $(Y_l)_l$ with law adapted to the natural filtration $\mathcal{G}$ induced by $(X_l)_l$.

4. **Learning with a lag.** The agent observes the signal process $(X_l)_l$ and the principal observes the process $(Y_l)_l$ which is simply the agent's information but with a lag of $\gamma \geq 0$: $Y_{l+\gamma \wedge L} = X_l$.

2.3 Preferences. If the agent develops the technology up to level l at state θ, the agent's utility is $u(\theta, l)$ and the principal's utility is $v(\theta, l)$.[18] It will be helpful to work with indirect utilities by defining

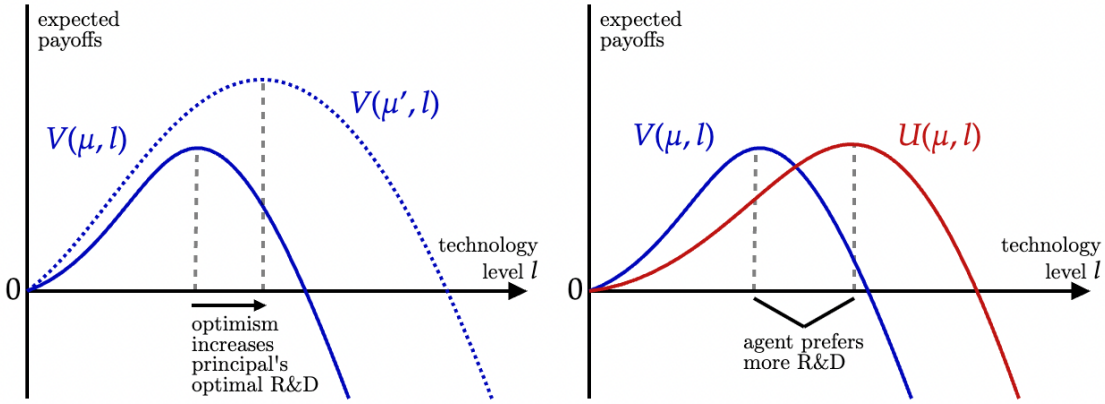
$$U(\mu, l) := \mathbb{E}_{\theta \sim \mu}\Big[u(\theta, l)\Big] \quad \text{and} \quad V(\mu, l) := \mathbb{E}_{\theta \sim \mu}\Big[v(\theta, l)\Big].$$

Increasing the technology level is beneficial if and only if the state is good i.e., $v(1, \cdot)$ is strictly increasing while $v(0, \cdot)$ is strictly decreasing. We assume $v(1, \cdot)$ and $v(0, \cdot)$ are continuously differentiable and normalize payoffs from zero technology to zero so $V(\mu, 0) = 0$.

[18] We extend the second argument of v from $\mathcal{G}$ to the positive reals, that is, $v : \Theta \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ which is without loss since $\mathcal{L} \subseteq \mathbb{R}_{\geq 0}$.

These assumptions imply that the principal prefers a higher technology level when she is more optimistic[19] as illustrated in Figure 2 (a) where the principal's indirect utility at belief $\mu \in (0,1)$ is given by the solid blue line, and the same for a more optimistic belief $\mu' > \mu$ is given by the dotted blue line.[20]

Figure 2: Illustration of assumptions on V and U



(a) Increasing in optimism

(b) Principal risk aversion

We maintain the following assumption on the agent's preference U:

Assumption 1 (Principal is more risk averse)**.** There exists an increasing convex function $g : \mathbb{R} \to \mathbb{R}$ such that $u(\theta, l) = g \circ v(\theta, l)$.

Let $\mathbb{U}$ denote the set of agent preferences which fulfill Assumption 1. An implication of Assumption 1 is that for each belief μ, the principal's static choice of technology level (sans further information) is weakly lower than that of the agent's. This is illustrated in Figure 2 (b) where, at the belief μ, the location of the peak of the agent's indirect utility is further out than that of the principal's.

Assumption 1 might capture intrinsic differences in risk preferences, but also arises naturally in the presence of externalities e.g., if the costs of developing or deploying the uncertain technology are only partially borne, if the benefits from the technology are disproportionately enjoyed by the firm, or if multiple firms are competing in an 'R&D arms race' so that the ensuing equilibrium is as if a representative firm is

[19] See Milgrom and Shannon (1994).

[20] In Figure 2 (a) $V(\mu, \cdot)$ is depicted as quasiconcave but this is not assumed.

more risk-seeking. We illustrate these via a range of concrete examples in Online Appendix III.[21]

2.4 Adaptive mechanisms. Let $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ denote the extended reals. An adaptive mechanism is an $\overline{\mathbb{R}}$-valued stochastic process $(\phi_l)_{l \in \mathcal{L}}$ adapted to the principal's filtration $\mathcal{G}$ such that, if the agent stops at technology level l, the regulator imposes a tax of ϕ_l. The set of adaptive mechanisms is thus

$$\Phi := \left\{ (\phi_l)_{l \in \mathcal{L}} : \begin{array}{c} \text{(i) } (\mathcal{G}_l)_l\text{-adapted;} \\ \text{(ii) pathwise lower semi-continuous;} \\ \text{(iii) class (D) on } \left\{|\phi_l| < +\infty\right\}. \text{[22]} \end{array} \right\}.$$

Condition (i) states that the mechanism cannot 'look into the future'—it can only condition on present information; conditions (ii) and (iii) are mild technical conditions that guarantee the existence of solutions to the agent's stopping problem. Stopping the technology at level l under belief μ yields the expected payoffs:

$$\underbrace{U^{\phi}(\mu, l) := U(\mu, l) - \phi_l}_{\text{Agent's payoff}} \quad \text{and} \quad \underbrace{V(\mu, l).}_{\text{Principal's payoff}}$$

Note that ϕ_l does not feature in the principal's payoff because we are interested in efficiency: if ϕ is a tax, this might be paired with a lump sum rebate; alternatively, ϕ might reflect non-monetary instruments (e.g., legal, administrative, or regulatory barriers).[23] By letting adaptive mechanisms take values in the extended reals, we capture both 'price instruments' ($\phi \in \mathbb{R}$), 'quantity' instruments' ($\phi \in \{-\infty, +\infty\}$), and mixtures thereof.[24] Say mechanisms ϕ, ϕ' are *equivalent* if $\phi = \phi' + \alpha$ for some $\mathcal{G}_{0-}$-measurable random variable α. Equivalent mechanisms induce identical incentives for the agent (conditional on participation).

[21] We develop an example of an economy in which consumption depends on the technology state $\theta \in \Theta$ and level $l \in \mathcal{L}$ and, in so doing, relate this to Jones (2024).

[22] Recall that a stochastic process $(X_l)_l$ is class (D) if the set $\{X_\ell : \ell \text{ is a } \mathcal{G}\text{-adapted stopping level}\}$ is uniformly integrable. Class (D) processes form the smallest class that ensures an optimal stopping level for the agent facing any mechanism $\phi \in \Phi$ exists.

[23] Our results can be modified to accommodate revenue motives for the regulator (see Proposition 6 in Online Appendix IV). Neither additional randomization nor elicitation of type reports can improve upon the principal's payoff gaurantee (see Proposition 4 in Online Appendix IV).

[24] Such 'mixed' instruments are in the spirit of the closing remarks of Weitzman (1974).

Our framework offers a unified umbrella for analyzing regulatory timing (Laffont and Tirole, 1993). Adaptive mechanisms live in the rich middle ground between ex ante regulation that fails to make use of interim information, and ex post regulation that is often infeasible.[25] Table 1 illustrates.

ex ante	ex interim ('adaptive')	ex post
Policy is $\mathcal{G}_{0-}$-adapted	Policy is $(\mathcal{G}_l)_l$-adapted	Policy is $\mathcal{G}_\infty$-adapted
Feasible	Feasible	Infeasible
Suboptimal	Second-best	First-best

Table 1: Comparison of regulatory timing

2.5 Agent's problem. The agent with preferences U, facing mechanism ϕ and learning process $\mathcal{F}$, first decides whether or not to participate in the mechanism. If she does not participate, the technology is not developed and payoffs are zero for principal and agent. If she does, the agent solves the following optimal stopping problem:

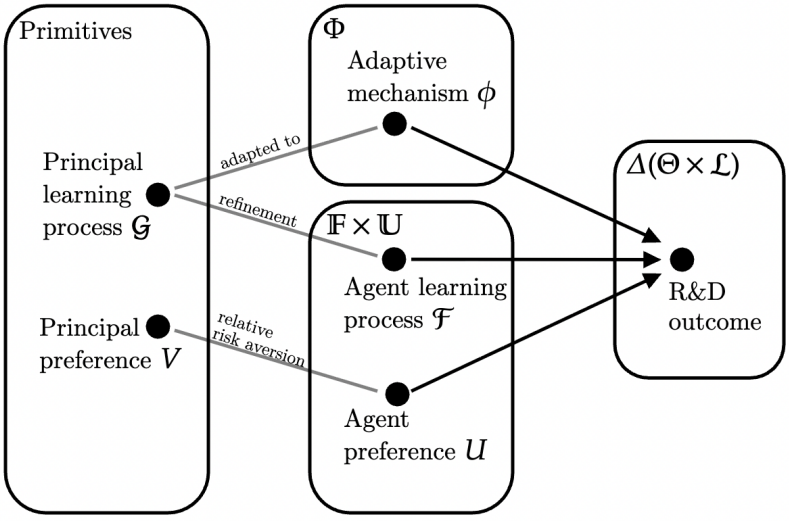
$$\sup_{\ell} \mathbb{E}\Big[U(\mu_\ell, \ell) - \phi_\ell\Big] \quad \text{s.t. } \ell \text{ is } \mathcal{F}\text{-adapted.} \tag{O}$$

where the agent's technology choices depend jointly on the adaptive mechanism ϕ, the learning process $\mathcal{F}$, and her preferences U. Let $\ell^*(\phi, \mathcal{F}, U)$ denote the largest stopping level that solves (O).[26] All our results hold for a more general selections of optimal stopping levels; we discuss this in Online Appendix VI. Figure 3 takes stock of how different elements of our model interact to jointly shape technology outcomes.

[25] ex post regulation achieves efficiency by conditioning on the true state e.g., via liability regimes (Shavell, 1984) which has been proposed for regulating AI (Gans, 2025). Among the standard objections to ex post regulation e.g., the weakness of tort law, limited liability, etc., we highlight one that is distinctive to technology regulation: state-contingent regulation is feasible only if it induces sufficient technology levels for the regulator to learn in every state. But this is self-defeating since the point of regulation is to shape the joint distribution over technology levels and states.

[26] Existence of solutions to (O) follows from (i) pathwise lower semi-continuity of ϕ; (ii) uniform integrability of the stopped process $(\phi_l)_l$; and (iii) compactness of the technology space $\mathcal{L}$. Among these optimizers, a largest one exists because $U - \phi$ is upper-semicontinuous in expectation along stopping times (Kobylanski and Quenez, 2012; El Karoui, 1979); see Online Appendix VI for a formal treatment.

Figure 3: How policy, learning, and preferences shape outcomes



2.6 Desiderata. We outline three desiderata for adaptive mechanisms.

Robustness. Suppose our principal evaluates mechanisms according to their payoff guarantee. If the agent's preference U is known but the principal is uncertain about what the agent might learn in the interim, she faces the following learning-robustness problem:

$$\sup_{\phi \in \Phi} \inf_{\mathcal{F} \in \mathbb{F}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U)\Big)\Big] \tag{L}$$

Call mechanisms that achieve (L) *learning-robust*.

If the principal is uncertain both about the agent's preferences (U) and learning process ($\mathcal{F}$), she faces the dual-robustness problem:

$$\sup_{\phi \in \Phi} \inf_{\substack{\mathcal{F} \in \mathbb{F} \\ U \in \mathbb{U}}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U)\Big)\Big] \tag{D}$$

Call mechanisms that achieve (D) *dually-robust*.

Our robustness notions recover, as special cases, some of those studied in the literature. When the principal does not learn (i.e., $\mathcal{G}_l = \mathcal{G}_{0-}$ a.s.) and knows the agent's preference but not her learning process, learning-robustness coincides with informational robustness for static (Bergemann and Morris, 2005) and dynamic (Libgober and Mu, 2021) mechanisms. When the principal does not learn but is certain that the agent knows the state (i.e., θ is $\mathcal{F}_0$-measurable), robustness to the agent's preferences is similar to that studied by (Armstrong, 1995; Frankel, 2014). We think learning-

and dual-robustness are apt for new technologies rife with substantial disagreement, not just about the state ($\theta \in \Theta$), but also about what might be learnt over the course more R&D or deployment of the technology ($\mathcal{F} \in \mathbb{F}$)[27] and what preference the private sector has ($U \in \mathbb{U}$). Conceptually, our robustness notion is novel and requires mechanisms to learn under ambiguity, continually adapting to interim information while safeguarding against the worst-case learning-process and/or preference.

Dominance and undominated. We will be interested in assessing mechanisms not just according to their payoff guarantee, but also evaluated away from the worst case. For two different adaptive mechanisms $\phi, \phi' \in \Phi$, say ϕ *dominates* ϕ' if, for any learning process $\mathcal{F} \in \mathbb{F}$ and any agent preference $U \in \mathbb{U}$,

$$\mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U)\Big)\Big] \geq \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi', \mathcal{F}, U)\Big)\Big]$$

with strict inequality for some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$. Our dominance criteria allows the principal to entertain the possibility that the agent's learning process and/or preferences are not adversarially chosen. This in similar spirit to work by Wald (1950); Savage (1972). The mechanism ϕ is *undominated* if it is not dominated by another adaptive mechanism. This notion has been recently analyzed by Borgers, Li, and Wang (2024) in the context of static mechanisms.[28]

Time-consistency. A final desideratum is for the regulator to have incentives to follow-through with the mechanism, both on and off the equilibrium path. This matters for feasibility rather than optimality—time consistency determines whether a mechanism can be credibly run. Indeed, such problems looms large in financial regulation (Brunnermeier et al., 2009) so we think it is worth analyzing whether the same is true of technology regulation. We defer our definition and analysis of time-consistency to Section 3.3.

[27] For instance, computer scientists disagree about how much we can learn about the safety of AI from empirical testing (see, e.g., Cohen et al. (2024) and Anthropic (2025)).

[28] Moreover, if we require the strict inequality to be over a positive Lebesgue measure set, this has been recently defined as 'strong dominance' by Borgers, Li, and Wang (2024). While there is no analog of 'positive Lebesgue measure' on the space of learning processes, all examples of dominance in this paper are 'strong dominance' with respect to the Wiener measure on the space of càdlàg paths of the agent's 'first-order' belief $\mathbb{E}[\theta \mid \mathcal{F}_t]$.

3 ROBUST, DOMINANT, AND TIME-CONSISTENT REGULATION

We begin by defining a simple family of adaptive mechanisms.

Definition 1 (Adaptive sandboxes)**.** $\phi \in \Phi$ is an adaptive sandbox induced by the $\mathcal{G}$-adapted stopping level ℓ if:

$$\phi_s = \begin{cases} 0 & \text{if } s \leq \ell \\ +\infty & \text{if } s > \ell. \end{cases}$$

Adaptive sandboxes prescribe a zero marginal tax on the technology until the stopping level ℓ, beyond which the agent is not allowed to push the technology further. Sandboxes are adaptive mechanisms since the location of the limit ℓ is itself a $\mathcal{G}$-adapted stopping level.

Let ϕ^* be the adaptive sandbox induced by stopping level $\overline{\ell}$ solving

$$\sup_{\ell} \mathbb{E}\Big[V(\mu_\ell, \ell)\Big]$$
$$\text{s.t. } \ell \text{ is } \mathcal{G}\text{-adapted.}$$

ϕ^* is illustrated in Figure 4. To ease exposition, we will assume $\overline{\ell}$ is the unique solution to the above (which holds generically).

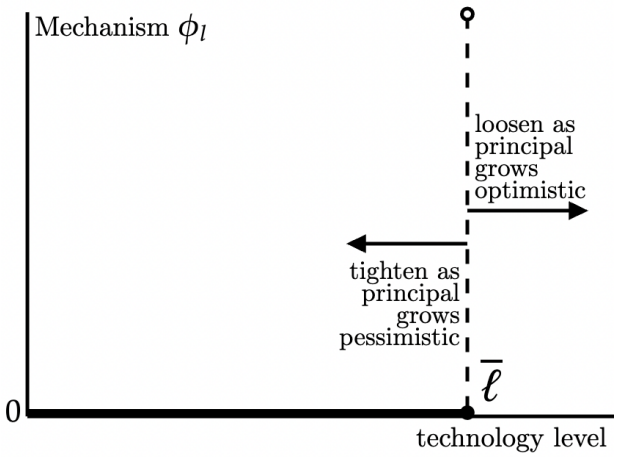


Figure 4: Adaptive sandbox

3.1 Robustness.

Theorem 1 (Robustness)**.** *The adaptive sandbox ϕ^* is learning-robust and dually-robust i.e., it solves* (L) *and* (D)*.*

Theorem 1 establishes the learning- and dual-robustness of ϕ^*. The proof is quite simple and instructive so we will sketch the main steps here.

Step 1: Robustness problem upper-bounded by direct control. A basic observation is that the value of learning-robust problem is upper-bounded by 'direct control' in

which the principal chooses the technology level directly:

$$\begin{aligned}\sup_{\phi\in\Phi}\inf_{\mathcal{F}\in\mathbb{F}}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{F},U)\Big)\Big] &\leq \sup_{\phi\in\Phi}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)\Big]\\ &\leq \sup_s \mathbb{E}\Big[V(\mu_s,s)\Big]\\ &\quad \text{s.t. } s \text{ is } \mathcal{G}\text{-adapted}\end{aligned}$$

where the first inequality is because $\mathcal{G}\in\mathbb{F}$ is *a* possible learning process, and the second is because the agent's technology choice ℓ^* is *a* $\mathcal{G}$-adapted stopping level so the principal might just as well control experimentation directly. It suffices to show that our adaptive sandbox ϕ^* achieves this upper-bound over all learning processes $\mathcal{F}\in\mathbb{F}$.

Step 2: ϕ^* achieves the upper-bound. First and most immediately, the adaptive sandbox ϕ^* delivers a *participation guarantee*: pushing the technology up to the sandbox limit—always a feasible strategy under any $\mathcal{F}\in\mathbb{F}$—is always weakly better than not participating. Next, we develop a simple but powerful comparative static result relating risk aversion to optimal experimentation:

Lemma 1 (Experimentation comparative static)**.** *Fix any learning and preference pair* $(\mathcal{F},U)\in\mathbb{F}\times\mathbb{U}$ *and any pair of* $\mathcal{F}$*-stopping levels* $\ell_1\leq\ell_2$*. The set of* $\mathcal{F}$*-stopping levels that solve* $\sup_{\ell\in[\ell_1,\ell_2]}\mathbb{E}[V(\mu_\ell,\ell)]$ *is dominated by the set of* $\mathcal{F}$*-stopping levels that solve* $\sup_{\ell\in[\ell_1,\ell_2]}\mathbb{E}[U(\mu_\ell,\ell)]$ *in the strong set order.*[29]

Lemma 1 nests and unifies previous work analyzing how different kinds of risk-aversion can either prolong or shorten experimentation (Keller et al., 2019; Chancelier et al., 2009). Lemma 1 reconciles these results, and holds path-by-path for any learning process, including those that induce non-Markov and/or non-continuous beliefs, without requiring explicit solutions to the optimal stopping problem.[30]

Now suppose that under the sandbox ϕ^*, the agent stops strictly before the sandbox limit i.e., $\ell^*<\overline{\ell}$. Since the adaptive sandbox imposes a zero-marginal tax, in the absence of any regulation the agent would continue to stop at ℓ^* facing the truncated technology space $[0,\overline{\ell}]$. By Lemma 1, this means that if the principal had access to

[29] Recall that for a lattice L and any $A,B\subseteq L$, A dominates B in the strong set order if for any $a\in A$ and $b\in B$, $a\vee b\in A$ and $a\wedge b\in B$.

[30] The connection between risk-aversion, patience, and stopping has a long history in economics (Keller et al., 2019; Chancelier et al., 2009; Quah and Strulovici, 2013). In Online Appendix II we discuss how Lemma 1 fully or partially generalizes these results, and allows the analyst to make predictions about the duration of experimentation when risk- and time-preferences are both changing.

the agent's information, she would also prefer to stop at $\mathcal{F}_{\ell^*}$.

This *aligned sensitivity* is depicted in Figure 5 where each of the red paths depict sample paths of the agent's stopped beliefs. By minimally distorting the shape of relative incentives, the adaptive sandbox generates positive correlation between the events that both principal and agent prefer to stop, thereby guaranteeing that any deviation by nature $\mathcal{G} \to \mathcal{F}$ must improve the principal's payoff.

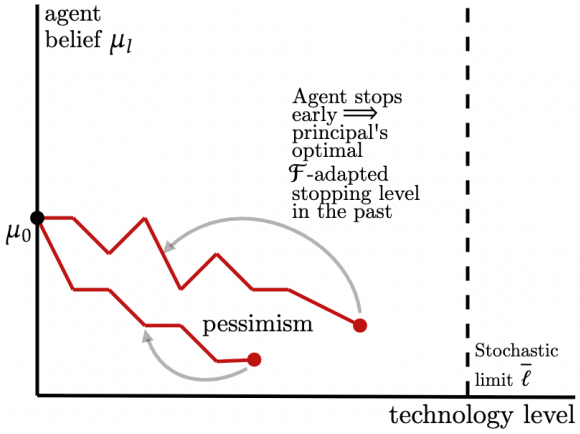


Figure 5: Aligned sensitivity before limit

Finally suppose that under the sandbox ϕ^*, the agent stops at the sandbox limit i.e., $\ell^* = \overline{\ell}$. But the principal's payoff from such paths coincide with that of 'direct experimentation'. Then partitioning up the set of paths into those in which the agent either stops early ($\ell^* < \overline{\ell}$) or pushes until the sandbox limit ($\ell^* = \overline{\ell}$) and taking total expectation implies that the principal's payoff under the adaptive sandbox ϕ^* and over all learning process $\mathcal{F} \in \mathbb{F}$ achieves the upper-bound constructed in Step 1, yielding learning-robustness. Dual-robustness follows by noticing that the adaptive sandbox ϕ^* does not depend on the agent's preference U. ◊

The hard limit induces *minimal sensitivity* to information—whatever the agent's interim beliefs, she is not allowed to push past it. Although this is stringent, there is a sense in which this is exactly what robustness demands. Consider, for instance, the realization of the blue path ('strong optimism') in Figure 6 in which the agent has privately become pretty optimistic and the principal, if she knew this, would like to slacken the sandbox limit.

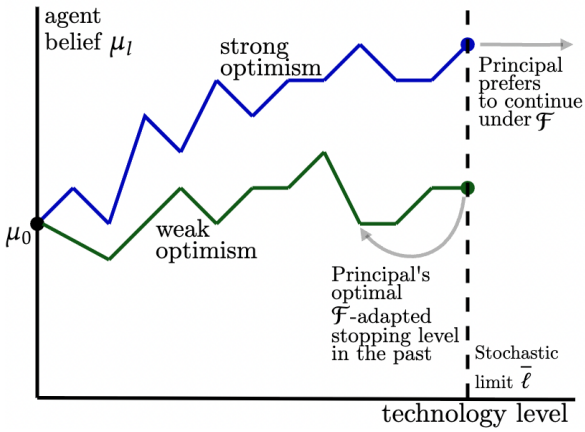


Figure 6: Minimal sensitivity at limit

Under the green path ('weak optimism'), however, the principal prefers stopping

although the agent is willing to take on more risk. It is precisely because the agent's learning process is adversarially chosen that adaptive mechanisms cannot distinguish between these kinds of optimism to guarantee positive correlation between the events that both agent and principal prefer a higher technology level. Indeed, we will see in Section 4 that in the absence of robust regulation, the worst-case learning process negatively correlates these events by inducing a specific kind of weak optimism that maximizes risk-taking—hard limits safeguard against exactly this.

The adaptive sandbox ϕ^* is relatively simple. This echoes a broader theme from the literature on robust mechanism design (Bergemann and Morris, 2005; Carroll, 2019) that robustness is particularly appealing when it prescribes mechanisms that are simple, intuitive, and mirror those used in practice. By contrast, the Bayesian problem is intractable and sensitive to the prior. An exception is when the principal does not learn and the agent's learning process $\mathcal{F}$ is (i) known; (ii) continuous; and (iii) Markov—then, results from Kruse and Strack (2015, 2019) can be adapted to solve for the optimal mechanism—this will take the form of a nonlinear (interior) tax whose form is quite sensitive to the agent's learning process. But the specificity of this exception suggests that the Bayesian problem is more generally hard and detail-dependent. On this view, Theorem 1 represents the best a boundedly rational regulator can do.[31]

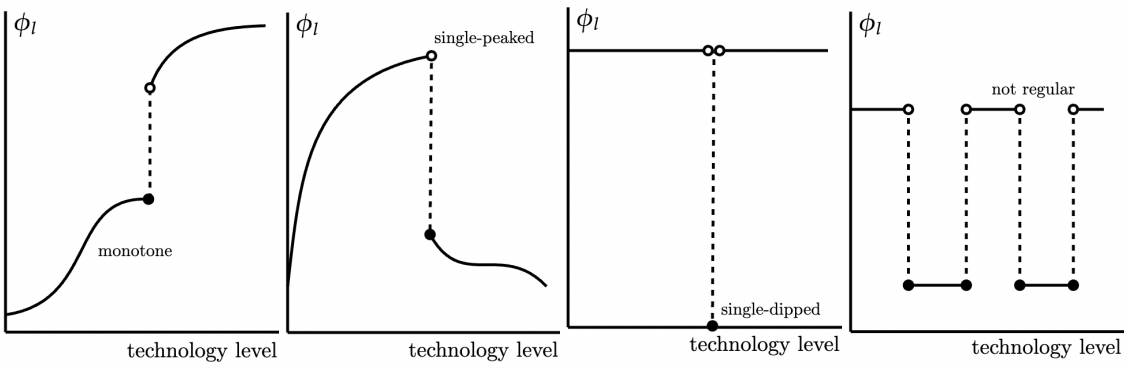
Indeed, qualitative features of our adaptive sandbox—zero marginal tax, quantity limits, and adaptivity—broadly resembles real-world policies that have been deployed by policymakers to manage uncertain technologies e.g., real-world 'regulatory sandboxes' for financial technology (Financial Conduct Authority, 2015) and AI (European Parliament, 2024), or phases of FDA drug trials (FDA, 2008). We detail these connections at the end of this section.

3.2 Undominated and dominance. Theorem 1 establishes that the adaptive sandbox ϕ^* is both learning- and dually-robust. This leaves open the question as to whether other mechanisms might also be robust and, if so, why might we prefer one mechanism over another? We will offer a simple but, we think, compelling reason: ϕ^* is dominant within a natural class of mechanisms we now define.

Definition 2 (Regularity)**.** An adaptive mechanism $\phi \in \Phi$ is *regular* if it is almost-surely single-peaked or single-dipped.

[31] This justification is analogous to that of Carroll (2017)'s treatment of multidimensional screening which is well-known to be hard.

Figure 7: Illustration of regularity



(a) Nonlinear tax Regular (b) Tax→subsidize Regular (c) Quota Regular (d) Oscillatory Not regular

Regularity requires mechanisms to not vacillate between penalizing and rewarding the agent for pushing the technology. This is, in part, motivated by practical considerations: regular mechanisms are "not too complex" (Diamond and Saez, 2011).[32] At the same time, regularity is rich enough to capture a wide suite of instruments such as (potentially stochastic) nonlinear taxes and/or subsidies, quotas, and mixtures thereof. Sample paths of regular mechanisms are illustrated in panels (a)-(c) of Figure 7. Panel (d) illustrates an irregular mechanism which oscillates between a low and a high tax.[33]

Theorem 2. *The adaptive sandbox ϕ^* is undominated. If the regulator chooses among regular mechanisms, all other dually-robust mechanisms are dominated by ϕ^*.*

The first part of Theorem 2 establishes that ϕ^* is undominated and can thereby be rationalized over any other adaptive mechanism—for any alternate mechanism ϕ, we can find some learning process $\mathcal{F}$ and preference U so that ϕ^* delivers strictly higher regulator payoffs than ϕ. The second part of Theorem 2 states that ϕ^* dominates all other mechanisms ϕ which are regular and dually-robust.[34] Thus, if our designer has lexicographic attitudes toward uncertainty about the agent's learning process and

[32] Diamond and Saez (2011) argue for this in the context of income taxation but we think this applies equally to taxes on R&D or deployment.

[33] Regularity is more permissive than the monotonicity condition typically imposed in security design (Nachman and Noe, 1994; DeMarzo and Duffie, 1999) since any tax schedule that penalizes R&D or deployment is increasing and thus regular.

[34] Our focus on dominance is in the spirit of Börgers (2017) who argue that simply focusing on robustness may be dominated; Theorem 2 makes precise that ϕ^* is the unique undominated, robust, and regular mechanism.

preferences and, perhaps for practical reasons, is restricted to regular mechanisms which do not vacillate between punishing and rewarding pushing the technology, then ϕ^* is uniquely optimal.

This is illustrated in Figure 8: holding fixed the agent's preference $U \in \mathbb{U}$, the figure illustrates how the principal's payoff varies with the learning process $\mathcal{F} \in \mathbb{F}$. The principal's payoff under the adaptive sandbox ϕ^* (blue) is minimized when the agent has no additional information ($\mathcal{F}_l = \mathcal{G}_l$ a.s.). But away from the worst-case, it performs better than other robust mechanisms e.g., an adaptive quota (red).

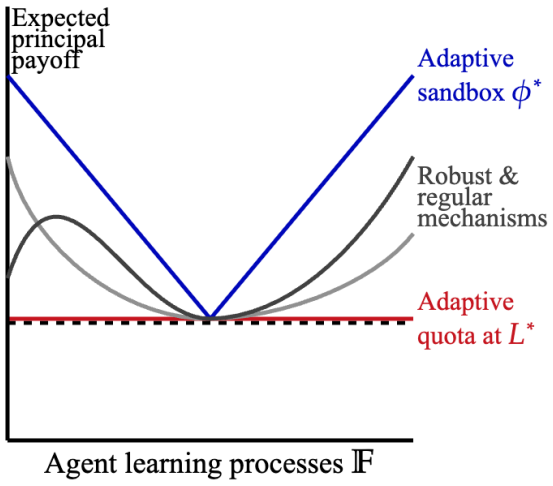


Figure 8: Illust. of dominance

What drives the dominance of ϕ^*? We have already highlighted how the zero marginal tax induces aligned sensitivity to new information by positively correlating the events that the agent and principal both prefer to stop. Among mechanisms with this property, the zero marginal tax within the sandbox induces *maximal sensitivity* to new adverse information—we will briefly sketch the key ideas behind the argument for dominance here. The complete proof of Theorem 2 is quite involved and is in Appendix A.

Step 1: Characterization of dual-robustness. We begin with a complete characterization of dually-robust mechanisms.

Limit with option mechanisms. ϕ is a *limit with option mechanism* induced by the $\mathcal{G}$-adapted level ℓ if:

(i) *Limit:* $\phi_s = +\infty$ a.s. for all $s > \ell$; and

(ii) *Option:* $\phi_s \geq \phi_\ell$ a.s. for all $\mathcal{G}$-adapted stopping levels $s \leq \ell$.

Such mechanisms impose a quantity limit on the technology at some $\mathcal{G}$-adapted level ℓ and, in this regard, resemble sandboxes. Differently, they only demand that for any $\mathcal{G}$-stopping level $s \leq \ell$, the agent has the option of pushing the technology until

ℓ to enjoy a *guaranteed subsidy* of $\phi_s - \phi_\ell \geq 0$ relative to stopping at s. Since a zero marginal tax is one such option, all sandboxes are limit with option mechanisms. Limit with option mechanisms completely characterize dual-robustness.

Proposition 1. *ϕ is dually-robust if and only if it is a limit with option mechanism induced by $\overline{\ell}$.*

The proof of Proposition 1 is fairly technical and deferred to Appendix A. We will sketch the broad intuition here. The limit at $\overline{\ell}$—which, recall, is the solution to the principal's direct experimentation problem $\sup_\ell \mathbb{E}[V(\mu_\ell, \ell)]$—is driven by concerns that the agent might grow (with high probability) just a little more optimistic, but also turn out to be quite risk-seeking and so take on excessive risk. Limits safeguard against the prospect that equilibrium technology levels are distorted upwards in this manner, and is why 'quantity-like' instruments must emerge under robustness concerns. By contrast, the option of receiving a guaranteed subsidy if the agent were to push the technology till ℓ (though she might be heavily taxed for stopping in the interim) safeguards against the prospect that regulation might stifle innovation by distorting the technology downward. Limit with option mechanisms simultaneously safeguard against both kinds of distortions which is exactly what dual-robustness demands.

Proposition 1 implies that if a dually-robust mechanism is also regular, it must be almost surely decreasing until the hard limit at $\overline{\ell}$. That is, at any interim history, the agent knows that she will *always* receive a subsidy for pushing the technology within the sandbox limit. This is illustrated by Figure 9 which depicts three possible sample paths continuing from some interim history.

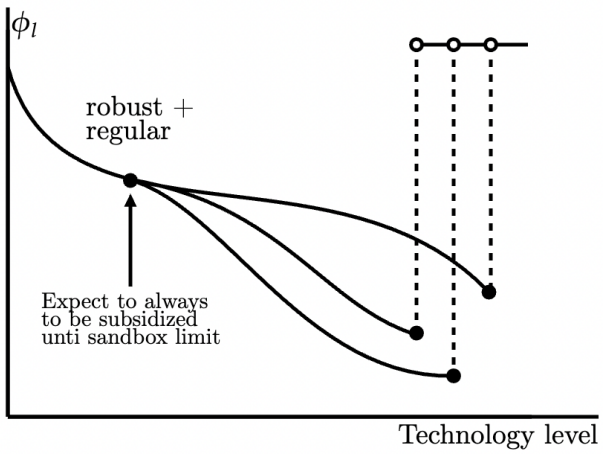


Figure 9: Subsidy within sandbox

Step 2: Coupling. Take any dually-robust and regular mechanism ϕ. We have already observed that ϕ must subsidize the technology until the hard limit at $\overline{\ell}$. Now for any learning process $\mathcal{F} \in \mathbb{F}$ and any agent preference $U \in \mathbb{U}$, we construct the simplest

coupling of the optimal stopping problems induced by ϕ and ϕ^* on an enlarged probability space such that (i) the learning processes is perfectly correlated across problems; and (ii) the marginal distribution of learning processes for each individual problem coincides with $\mathcal{G}$ and $\mathcal{F}$.

On this coupled problem, the agent always stops sooner facing the adaptive sandbox ϕ^* than under the alternative ϕ:

$$\ell^*(\phi^*, \mathcal{F}, U) \leq \ell^*(\phi, \mathcal{F}, U) \text{ a.s.}$$

precisely because ϕ imposes a marginal subsidy on the technology. This is depicted by Figure 10 which illustrates two sample paths of beliefs (red and dark red). On each of these paths, the agent stops sooner facing ϕ^* than under the alternate mechanism ϕ.

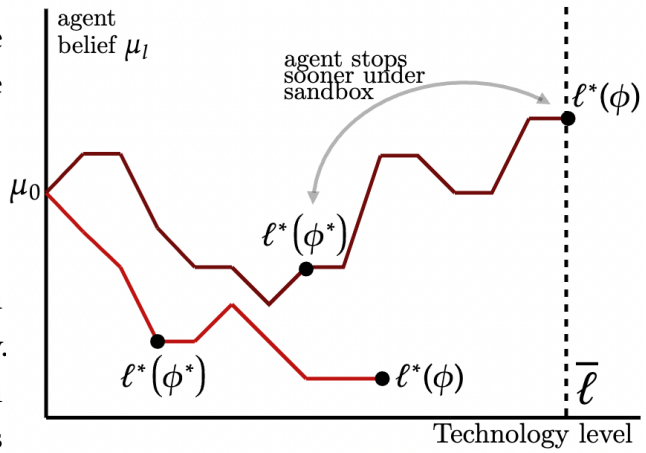


Figure 10: Pathwise comparison

This is the precise sense in which the regulatory sandbox induces *maximal sensitivity* to information.

Step 3: Maximal + aligned sensitivity $\implies$ dominance. To finish, recall that our argument from Theorem 1 and our comparative static Lemma 1 implied that the adaptive sandbox ϕ^* induced aligned sensitivity—whenever the agent prefers to stop, so does the principal as evaluated through the lens of the agent's private information:

$$\left\{\begin{array}{c}\text{Agent prefers to stop facing}\\ \text{mechanism } \phi^* \text{ and information } \mathcal{F}_{\ell^*}\end{array}\right\} \subseteq \left\{\begin{array}{c}\text{Principal prefers to stop facing}\\ \text{information } \mathcal{F}_{\ell^*}\end{array}\right\}.$$

But if this happens it is already too late—we are already past the principal's optimal technology level and the best thing to do is to stop the technology immediately. This implies that under this coupling, ϕ^* delivers a higher principal payoff path-by-path and so too in expectation. ♢

Theorem 2 speaks directly to current debates around how regulatory sandboxes should be implemented in practice. In a 10-year legal retrospective on regulatory

sandboxes, Allen (2025) notes the *"transformation of financial regulators into cheerleaders and sponsors for the innovations they've selected for their sandboxes"* and, as a result, facilitated the proliferation of harmful technologies like predatory lending. On this view, real-world sandboxes were implemented with a positive marginal subsidy within its boundaries which make tech firms less sensitive to interim information—they are not incentivized to halt even if they turn pessimistic, thereby leading to the over-proliferation of harmful technologies. Theorem 2 refines and unifies such different intuitions about why real-world sandboxes might not have worked well, as well as guidance for how they should be operationalized.

3.3 Time consistency. Robustness and dominance are ex ante criteria for assessing mechanisms before they are run. Might regulators be tempted to deviate to a different mechanism at interim histories? If so, this can render such mechanisms infeasible. Motivated by these concerns, we develop an analysis of time-consistency.

Time-consistency. ϕ is *time-consistent* if for each $\mathcal{G}$-stopping level ℓ:

(i) **Stopping.** If the agent stops at ℓ, ϕ is conditionally undominated i.e., there is no mechanism $\phi' \in \Phi$ such that

$$\mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U)\Big)\Big|\mathcal{G}_\ell, \ell^* = \ell\Big] \leq \mathbb{E}\Big[V\Big(\mu_{\ell'}, \ell'\Big)\Big|\mathcal{G}_\ell, \ell^* = \ell\Big]$$
$$\text{where } \ell' \text{ is the largest solution to } \sup_{s \geq \ell} \mathbb{E}[U^{\phi'}(\mu_s, s)]$$

with strict inequality for some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$.

(ii) **Continuing.** If the agent continues at ℓ, ϕ is undominated i.e., there is no mechanism $\phi' \in \Phi$ such that

$$\mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U)\Big)\Big|\mathcal{G}_\ell, \ell^* > \ell\Big] \leq \mathbb{E}\Big[V\Big(\mu_{\ell'}, \ell'\Big)\Big|\mathcal{G}_\ell, \ell^* > \ell\Big]$$
$$\text{where } \ell' \text{ is the largest solution to } \sup_{s \geq \ell} \mathbb{E}[U^{\phi'}(\mu_s, s)]$$

with strict inequality for some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$.

Time-consistency requires mechanisms to not be dominated following from at any history. There are three distinct ways it can fail. The first resembles the logic of the Coase conjecture: the agent might, knowing that the regulator lacks commitment, deviate from continuing to stopping, or from stopping to continuing such as to

influence future regulation.[35] The second is because the set of learning processes and preferences the regulator regards as possible shrinks in the interim so a mechanism that might have been previously rationalized for performing well at some learning-preference pair $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ might become dominated.[36] Finally, as the technology level increases irreversibly, regulators might simply lose incentives to follow-through at interim histories. Time-consistency rules out all these possibilities.

Our notion of time-consistency is quite weak, and does not require us to make strong positive assumptions on the objective the regulator running the mechanism 'actually has' e.g., she might have the robust objective, maximize expected utility, have a concern for misspecification, and so on.[37] Nonetheless, this minimal notion of time-consistency is enough to refine the set of robust mechanisms:

Theorem 3 (Time-consistency)**.** *The adaptive sandbox ϕ^* is time-consistent. If the technology space $\mathcal{L}$ is finite then ϕ^* is the only time-consistent and dually-robust mechanism.*

We first sketch why the sandbox is time-consistent. Then, we sketch why it is the only dually-robust mechanism that is time-consistent.

Why is the sandbox time-consistent? There are three kinds of histories:

1. Stopping before sandbox limit. If the agent stops strictly before the sandbox i.e., at technology level ℓ where $\ell = \ell^* < \overline{\ell}$, then by aligned sensitivity, the principal is certain that her own optimal stopping level (viewed through the agent's filtration $\mathcal{F}$) has already passed so the best thing continuing from that history is to let the agent stop. In fact, the sandbox ϕ^* is not only undominated, but it also *dominates* all other dually-robust mechanisms from this history.

2. Continuing before sandbox limit. If the agent chooses to continue pushing the technology before the sandbox i.e., at technology level $\ell < \overline{\ell}$, then the regulator can rule out some possibilities e.g., that the agent has conclusively learnt that the technology is harmful, but her ambiguity set over $\mathbb{F} \times \mathbb{U}$ remains large so maintaining the sandbox is conditionally undominated.[38]

[35] Here, the agent believes that the regulator will deviate to some dominant mechanism.

[36] The intuition here is similar to time-inconsistency generated by learning under ambiguity (Epstein and Schneider, 2003).

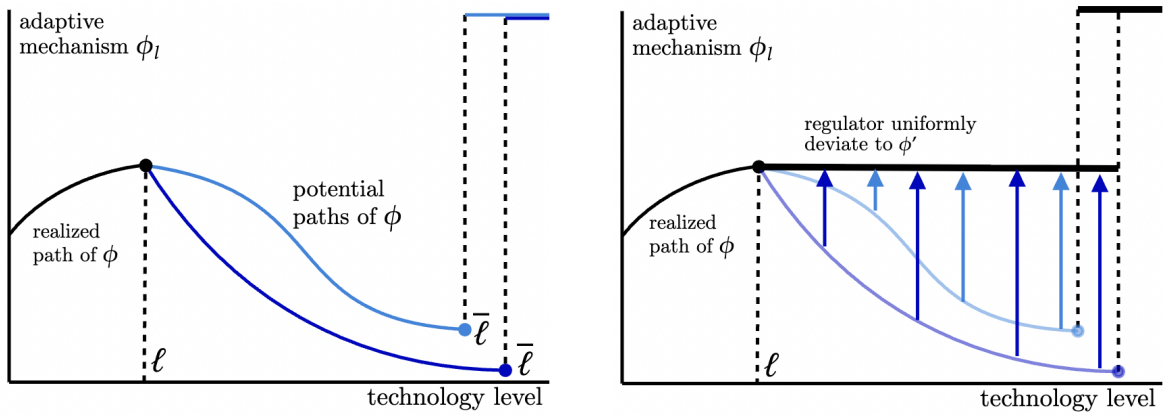
[37] We might, instead, demand that under ϕ the mechanism is 'conditionally robust' at every history so that if the regulator actually had a robust objective she has no incentives to deviate. This notion is related to learning under ambiguity (Epstein and Schneider, 2003) and a variant (with no principal learning) has been studied in the context of monopoly pricing (Li, Libgober, and Mu, 2022). Our adaptive sandbox ϕ^* is also time-consistent under this alternate notion.

[38] This can be thought of as a kind of (non-Bayesian) positive selection (Tirole, 2016).

3. Stopping at sandbox limit. If the agent pushes the technology until the sandbox limit i.e., $\ell^* = \overline{\ell}$, the regulator cannot rule out the possibility that the agent's information at $\overline{\ell}$ is such that if the regulator observed the agent's information she would prefer to stop. That is, the fear that the agent might only be weakly optimistic rationalizes keeping the hard limit in place.

Why are other robust mechanisms not time-consistent? Consider an alternate mechanism ϕ and recall dual-robustness which requires $\phi_\ell \geq \phi_{\overline{\ell}}$ a.s. for all $\ell \leq \overline{\ell}$ (via our characterization from Proposition 1 above). But suppose that with positive probability this inequality is strict at some $\mathcal{G}$-stopping level ℓ i.e., ϕ almost surely decreases until $\overline{\ell}$.

Figure 11: Why are other robust mechanisms not time-consistent?



(a) Original mechanism ϕ

(b) Dominant deviation at ℓ

This candidate mechanism ϕ is depicted in Figure 11 (a) where the blue paths reward pushing the technology between ℓ and $\overline{\ell}$. Now consider the uniform deviation to an alternate mechanism ϕ' continuing from ℓ which 'irons out' the mechanism along each path until the random level $\overline{\ell}$. This deviation by the **bold black paths** in Figure 11 (b), noticing that (i) this uniform deviation is feasible i.e., the resultant mechanism ϕ' remains $\mathcal{G}$-adapted; and (ii) ϕ' is simply an adaptive sandbox continuing from ℓ. If the agent stops pushing the technology following this deviation to ϕ', then since the sandbox is aligned sensitive, this means that the principal's optimal technology level has already passed so this deviation delivers a strict improvement to the regulator's payoff. It is precisely this temptation to screen out weakly optimistic agents that is in tension with time-consistency.

3.4 Connection to policy. We have thus far offered a rather abstract treatment of adaptive sandboxes. We now draw connections to real-world policy. Then, we will explain how our results might offer guidance for how policy can be improved.

Over the past century, regulators have gradually converged on various kinds of 'dynamic quantity limits' to regulate uncertain technologies. Drugs are the most paradigmatic example of this. In 1935, the Michigan state Supreme Court recognized that "the fact that if the general practice of medicine and surgery is to progress, there must be a certain amount of experimentation carried on." But it was not until the 1960s with the Kefauver-Harris Drug Amendments that strengthened FDA oversight over drug experimentation, requiring pharmaceutical companies to demonstrate that it was safe to conduct clinical trails on humans. Regulation continued evolving into the present phased system in which the firm is progressively allowed to experiment on a growing number of humans in each phase (Junod, 2008).

A similar progression has taken place in financial and AI regulation. In 2015, the UK's Finanical Conduct Authority (FCA) proposed 'regulatory sandboxes' to regulate new financial technologies in which participating firms could experiment with otherwise untested financial products e.g., insurance tech, peer-to-peer lending, distributed ledgers etc. for six months, upon which they would seek approval to be released into the wider market. As of 2020 over 50 jurisdictions have implemented versions of 'regulatory sandboxes' to regulate new financial technologies (Appaya et al., 2020). Such 'regulatory sandboxes' are increasingly enacted to regulate artificial intelligence across Europe (European Parliament, 2024) and the US (Cruz, 2025). Our mechanism ϕ^* reflects the important features common across these policies:

1. *Zero marginal tax as fostering innovation*. In the first use of regulatory sandboxes, UK regulators described it as a "'safe space' in which businesses can test innovative products... without immediately incurring all the normal regulatory consequences" (Financial Conduct Authority, 2015). This was implemented in 2016, and has been credited for making London a fintech hub.[39] Within our framework, the zero marginal tax before the limit corresponds to relaxing (monetary and non-monetary) regulatory burdens e.g., lengthy and costly approval processes. Theorem 1 showed that this zero marginal tax was sufficient to induce *aligned sensitivity*. Theorem 2 showed that subsidies, by contrast, are dominated

[39] Cornelli et al. (2024) estimate that the UK fintech sandbox led new financial technology firms to raise 15% more capital.

because it dulls the firm's sensitivity to interim adverse information. This makes precise the intution for why sandboxes have not always worked well (Allen, 2025; Sarro, 2025), and guidance for how they might be improved.

2. *Limits as safeguards*. Real-world implementation of sandboxes have imposed limits on different aspects of the technology. With fintech sandboxes, this has differed across jurisdictions e.g., (i) time limits for firms to roll-out new financial products (UK, Australia, Japan); (ii) scale limits which impose caps on the number of customers and/or deposit volumes (UK and Switzerland); or (iii) scope limits which fences off certain kinds of applications (virtually all jurisdictions).[40] More recently, the EU AI Act imposes a sandbox to "facilitate the development, training, testing and validation of innovative AI systems for a limited level" (European Parliament, 2024), and more broadly regulators in the EU and US have imposed thresholds on training compute (Heim and Koessler, 2024). A common theme across such policies is some form of 'quantity limits' on research, development, or deployment.

3. *Adaptivity*. Our sandbox limit $\bar{\ell}$ loosens and tightens with the principal's interim belief. This reflects the practice of regulatory sandboxes in fintech—to push beyond the boundary e.g., to scale up the customer base of new producers—firms must convince the regulator to extend the sandbox by producing evidence that the technology is beneficial for consumers[41]—failing which the firm is forced to halt roll-out of the technology. Similar kinds of 'step-by-step' quantity limits predates regulatory sandboxes: the FDAs regulations on clinical research progressively rolls out a new drug in phases to a progressively larger number of human participants, halting experimentation (failing) if interim results make regulators pessimistic that the drug is safe and efficient.[42]

4. *Encouraging learning*. A common theme articulated across regulatory agencies is that real-world sandboxes are "intended to foster business learning, i.e., the development and testing of innovations in a real-world environment" (European

[40] For instance, the Financial Conduct Authority in the UK have excluded cryptocurrency-based derivative products from their sandbox. Similarly, the Monetary Authority in Singapore have excluded cryptocurrency exchanges and 'initial coin offerings' from their sandbox.

[41] At the sandbox limit, firms can 'apply to remove the restrictions on... permissions to become a fully authorised firm.' (Financial Conduct Authority, 2023). A similar policy is in place in Australia (Australian Securities and Investments Commission, 2024).

[42] For an outline of current regulation FDA clinical trials see https://www.fda.gov/patients/drug-development-process/step-3-clinical-research.

Parliament Briefing, 2024). Our results crystalize this ambition: by loosening the limit whenever the principal turns more optimistic, this carves out extra room for the agent to experiment. But if the agent then halts experimentation prematurely, aligned sensitivity to information ensures that the principal also prefers to halt.

We emphasize that our results should not be taken as an endorsement of quantity limits *tout court*. Policymakers can, of course, get things wrong—they might be too stringent with these limits (as has been argued in the case of drug trials), or too lax (as has been argued for AI). Instead, our results formalize the intuition for why certain *kinds* of regulatory instruments should be preferred: adaptive quantity limits are simple, transparent, and delivers optimal payoff guarantees (Theorem 1) in an otherwise complicated environment (what beliefs do you have over $\mathbb{F} \times \mathbb{U}$?) Moreover, we offered sharp refinements over the universe of quantity instruments via dominance and time-consistency (Theorems 2 and 3). Taken together, our results imply that *if* we took robustness seriously, the adaptive sandbox ϕ^* is, in effect, the uniquely optimal mechanism. Should we?

4 THE IMPORTANCE OF ROBUSTNESS

We will argue that in the context of technology regulation, robustness is *important* in two regards. First, the worst-case learning process under non-robust mechanisms is natural—that is, robust mechanisms safeguard against something real. Second, the worst-case payoff under non-robust mechanisms can be arbitrarily poor, even when the regulator and firms' risk preference differ slightly. Taken together, we think this offers a strong, if not indefeasible, reason to care about robust technology regulation.[43]

We start by characterizing the worst-case learning process under non-robust mechanisms. We will focus on the case in which the principal does not learn i.e., $\mathcal{G}_l = \mathcal{G}_0$ a.s. for all $l \in \mathcal{L}$. This approximates the 'short run' which we view as a conceptual shorthand for the period of analysis in which the principal's beliefs are fixed,[44] or for technologies with large 'regulator lags'. Let $\overline{\mathbb{F}}$ denote the set of all agent learning

[43]There are, of course, more 'standard' reasons for robustness: simplicity, difficulty of forming beliefs over the space of learning processes, axioms for ambiguity etc. that we will not belabor. For discussions of simplicity see Carroll (2019); for informational robustness see Bergemann and Morris (2005); for axiomatic approaches to learning under ambiguity see Epstein and Schneider (2003).

[44]Analogous to the Marshallian short run (Marshall, 1890).

processes and let $\underline{\Phi}$ denote the set of mechanisms under no principal learning. A few definitions are in order.

Definition 3 (Bad news process)**.** The belief process induced by $\mathcal{F} \in \overline{\mathbb{F}}$ is a *bad news process* with *continuation belief path* $\gamma : \mathcal{L} \to [0,1]$ if

$$\operatorname{supp} \mathbb{E}\left[\theta \middle| \mathcal{F}_l\right] \subseteq \left\{0, \gamma(l)\right\}$$

where γ is nondecreasing and right-continuous. Note that since $\mathbb{E}[\theta \mid \mathcal{F}_l]$ is an $\mathcal{F}$-martingale, γ fully specifies the law of the belief process.

Bad news processes are such that, at each technology level, bad news arrives at an inhomogeneous rate so the agent either conclusively learns that the technology is harmful or, in the absence of such news, becomes increasingly more optimistic and follow the path γ.[45] This is illustrated in Figure 12 where the gray lines show different belief paths and the red region denotes the support of the equilibrium joint distribution over technology levels and beliefs.

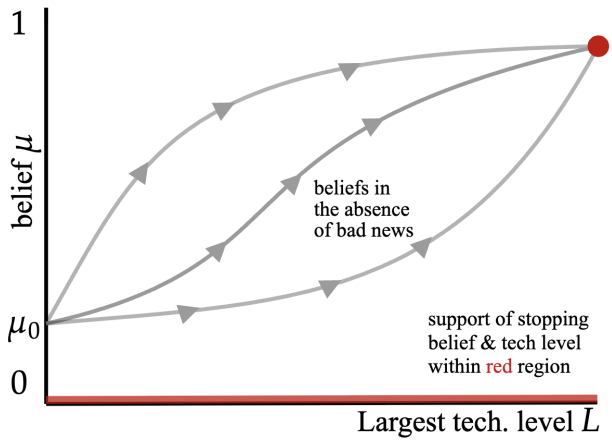


Figure 12: Bad news processes

The set of bad news processes is large. Let us construct a specific one. For the indirect utility $f : \Delta(\Theta) \times \mathcal{L} \to \mathbb{R}$, define:

$$\underbrace{l_f^*(\mu) := \max \operatorname*{argmax}_{l \in \mathcal{L}} f(\mu, l)}_{\text{Static optimal technology level at belief } \mu} \quad \text{and} \quad \underbrace{\mu_f^*(l) := \min\left\{\mu \in \Delta(\Theta) : l_f^*(\mu) \geq l\right\}}_{\text{Smallest belief rationalizing } l}.$$

$l_f^*(\mu)$ is the largest optimal 'one-shot' technology level under belief μ in the absence of further information; $\mu_f^*(l)$ is the belief threshold—the most pessimistic belief that can rationalize pushing the technology until level l.

[45] The connection between this arrival rate and beliefs is $\gamma(l) = \mu_0/(1 - F(l))$ where F is the (unnormalized) CDF of the arrival level of bad news.

Definition 4 (Weak optimism)**.** The bad news process $\mathcal{F} \in \overline{\mathbb{F}}$ induces *weak optimism* for indirect utility f if it has continuation belief path

$$\gamma(l) = \begin{cases} \mu_0 & \text{if } l < l_f^*(\mu_0); \\ \mu_f^*(l) & \text{if } l \geq l_f^*(\mu_0). \end{cases}$$

Bad news processes which induce weak optimism deliver conclusive evidence of bad news that the technology is harmful at a rate so that the agent's interim belief $\gamma(l)$ in the absence of bad news is exactly equal to the minimal belief $\mu_f^*(l)$ that can rationalize pushing the technology until level l in the absence of further information. The following result establishes that this is indeed the worst-case in the absence of robust regulation:

Theorem 4. *The form of the worst-case learning process and value of robustness is as follows:*

(i) ***Form.*** *If the agent's preference U is known, then for any mechanism ϕ such that $U^\phi = g \circ V$ for some increasing convex function g, the worst-case learning process is a bad news process that induces weak optimism for U^ϕ.*

(ii) ***Importance.*** *If the agent's preference U is also chosen adversarially, then for any mechanism ϕ which does not impose a hard limit, dual-robustness is arbitrarily important:*

$$\lim_{L\to\infty} \inf_{\substack{\mathcal{F}\in\overline{\mathbb{F}} \\ U\in\mathbb{U}}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] = -\infty \quad \textit{if} \quad \lim_{L\to+\infty} \left|\frac{V(0,L)}{V(1,L)}\right| = +\infty.$$

Theorem 4 pins down the learning processes which minimizes the principal's payoff under any mechanism ϕ that preserves the order of risk preferences.[46] The proof proceeds by recasting nature's problem as a dynamic information design problem to maximize $-V$, noting that here nature has state-dependent preferences which is novel to the literature[47]—this is fairly technical and we defer it to Appendix A for

[46] This is satisfied by a large class of mechanisms and payoff functions. A previous version of this paper established that if $l^*_{U^\phi}$ and l^*_V are continuous the worst-case learning process continues to be some some bad-news process, though not necessarily binding the agent's continuation incentive at every history.

[47] To handle this, we modestly extend techniques developed in some of our previous work (Koh, Sanguanmoo, and Zhong, 2024) that established a general toolkit for such problems.

we have quite a lot to say about the economics.

The form and functioning of weak optimism. Weak optimism hurts the regulator by (stochastically) maximizing the technology level when it is harmful, subject to the constraint that the agent does, in fact, want to push on with the technology—i.e., by *magnifying false positives*.

This is illustrated by the red curve in Figure 13 which shows the CDF of technology levels conditional on $\theta = 0$. This is chosen so that, in the absence of bad news, the agent grows progressively more emboldened to develop the technology. Conversely, the technology always fully developed when the technology is safe ($\theta = 1$) i.e., there are no false negatives as illustrated in blue.

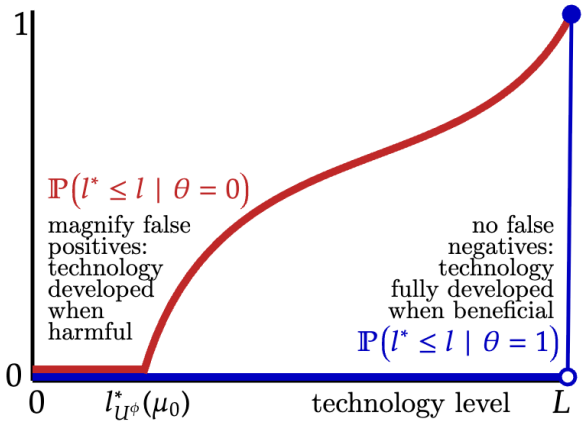


Figure 13: Worst-case joint dist.

Indeed, a distinctive feature of the worst-case learning process is that it progressively exploits *only* the static wedge between the principal's and agent's preferences. The agent's payoff from incrementally increasing the technology level can be heuristically decomposed as:

$$\underbrace{\text{Value of more technology at current belief}}_{\substack{\gtrless 0 \text{ under arbitrary learning process} \\ \text{Worst-case: } 0}} + \underbrace{\text{Value of future info from more technology}}_{\substack{\text{always} \geq 0 \text{ (free disposal)} \\ \text{Worst-case: } 0}}$$

where the first term gives the value of additional technology levels holding fixed the current belief, while the second gives the value of learning more about the technology which might inform future technology choices. For a typical learning process the agent has positive value for future information which incentivizes risk-taking e.g., the first term can be negative while the second positive. Weak optimism essentially holds both terms down to zero:[48] the agent's beliefs are dynamically steered such

[48]The first term is strictly positive before any information arrives i.e., up until level $l^*_{U^\phi}(\mu_0)$ and zero thereafter.

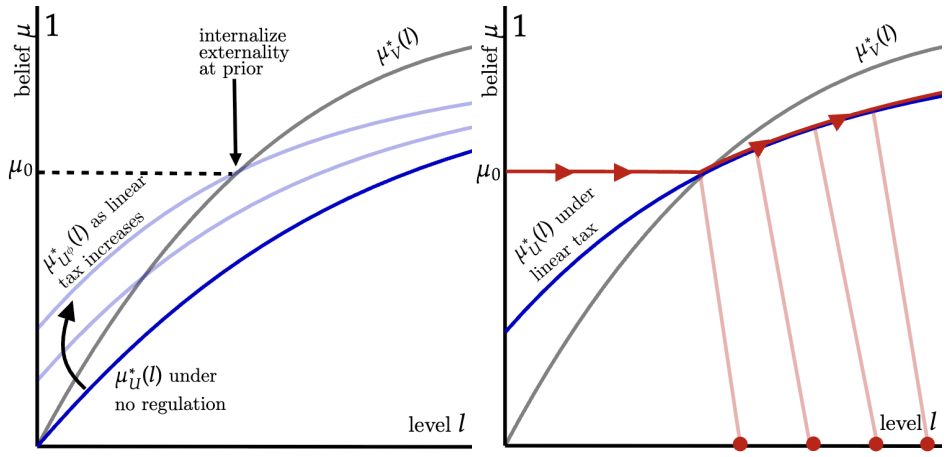
that she is continually kept indifferent between pushing the technology and stopping, while her value from future information is kept at zero at all histories.

The form of weak optimism carries two further implications. First, fixing the mechanism ϕ, and agent preference U, the worst-case learning process and attendant regulator payoff can be straightforwardly computed. We leverage this in Example 1 below to quantify the worst-case under CRRA utility. Second, the worst-case payoff remains possible whenever the agent employs any *undominated stopping rule*. This includes agents who are misspecified or ambiguity-averse over what they might learn in the future (Riedel, 2009), precisely because weak optimism exploits only the static wedge in risk-preferences. Indeed, computer scientists have raised worries that they might not learn much about the dangers of artificially intelligent agents as they continue to be developed.[49] Theorem 4 implies that tech firms' conservatism about future information per se is not enough to protect society writ large.

Policy implications of weak optimism. Theorem 4 warns against correcting externalities only 'locally'. Consider, for instance, the linear Pigouvian tax $\phi(l) = \alpha \cdot l$ with slope $\alpha > 0$ chosen to internalize externality under the prior. Panel (a) of Figure 14 illustrates the belief thresholds μ^*_V (gray) and $\mu^*_{U^\phi}$ (blue). As the Pigouvian tax increases, the belief threshold for the agent increases accordingly, and the slope can be chosen such that the externality is perfectly internalized under the prior: $l^*_{U^\phi}(\mu_0) = l^*_V(\mu_0) = \overline{\ell}$.Panel (b) of Figure 14 depicts the worst-case learning process under this tax: the agent learns nothing about the safety of the technology until $\overline{\ell}$, then bad news that the technology is unsafe arrives at a rate such that in its absence, the agent becomes progressively more confident in the technology (red arrow) which encourages—from the principal's point-of-view—excess risk-taking.

[49] This is formalized in Online Appendix IV.4. Here our agent correctly interprets the information that arrives and updates via Bayes' rule. Her ambiguity and/or misspecification is over the law of future information which shapes choices in the present. This notion is closely related to that of Riedel (2009) who studies a stopping problem where the agent perfectly observes realizations of the payoff process, but faces ambiguity over the future law.

Figure 14: Weak optimism under linear taxes

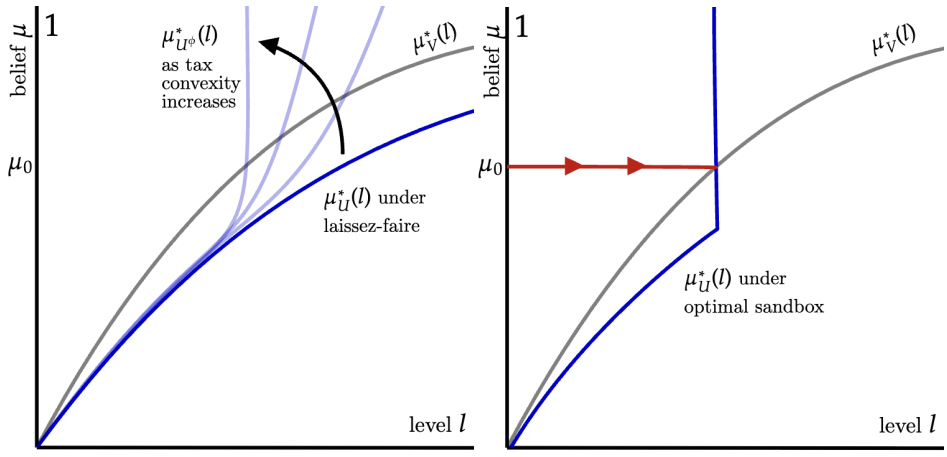


(a) Increasing the linear tax

(b) Worst-case learning process

Now consider panel (a) of Figure 15 which depicts the agent's belief thresholds under a sequence of mechanisms of 'increasing convexity'. As before, this can be chosen such as to internalize the externality locally at the prior. But differently, at more optimistic beliefs the agent now prefers a lower technology level than the principal, while at less optimistic beliefs the agent prefers more. Panel (b) of Figure 15 depicts a regulatory sandbox as the limit of such mechanisms. From Theorem 1 (i), aligned sensitivity ensures that any deviation by nature to a more informative learning process must benefit the principal so the worst-case learning process is, in fact, no learning.

Figure 15: How convexity safeguards against weak optimism



(a) Increasing tax convexity

(b) Worst-case learning process

The plausibility of weak optimism. Weak optimism identified by Theorem 4 (i) correspond to technologies that pose rare disaster risk: financial crises induced by the proliferation of new financial instruments, pandemic induced by gain-of-function research, or AI-induced catastrophes induced by misalignment. For instance, prominent computer scientists have noted that *"sufficiently capable LTPA, safety testing is likely to be either dangerous or uninformative"* (Cohen et al., 2024). This corresponds to a bad news process where the lack of a disaster makes the agent (financial engineer, biologist, AI scientist)—via Bayes' rule—progressively more optimistic that the technology is safe. But in the absence of stringent regulation, our result shows that learning per se is not enough to safeguard society—in fact, a little learning can be much worse than no learning at all. Indeed, with AI development there is substantial concerns among AI scientists at the technological frontier on dangers of powerful AI systems (Greenblatt et al., 2024)—but not enough to induce stopping. We think this is consistent with our notion of 'weak optimism'.

The quantitative importance of robustness. Theorem 4 (ii) establishes the quantitative importance of robustness: when both the agent's preference and learning process are chosen adversarially, then without a hard limit payoffs are arbitrarily poor if, in the large technology limit (with 'superhuman' AI systems, say) the loss to payoffs when the technology is dangerous ($\theta = 0$) dwarfs gains when the technology is beneficial ($\theta = 1$). We emphasize, however, that learning *per se* can magnify even small wedges in differential risk-aversion as the following example illustrates:

Example 1 (Worst-case under CRRA utility)**.** Suppose, following Jones (2024) that both agent and principal have CRRA preferences over consumption with risk-aversion parameter $\gamma_P > \gamma_A \geq 1$. Consumption is given by

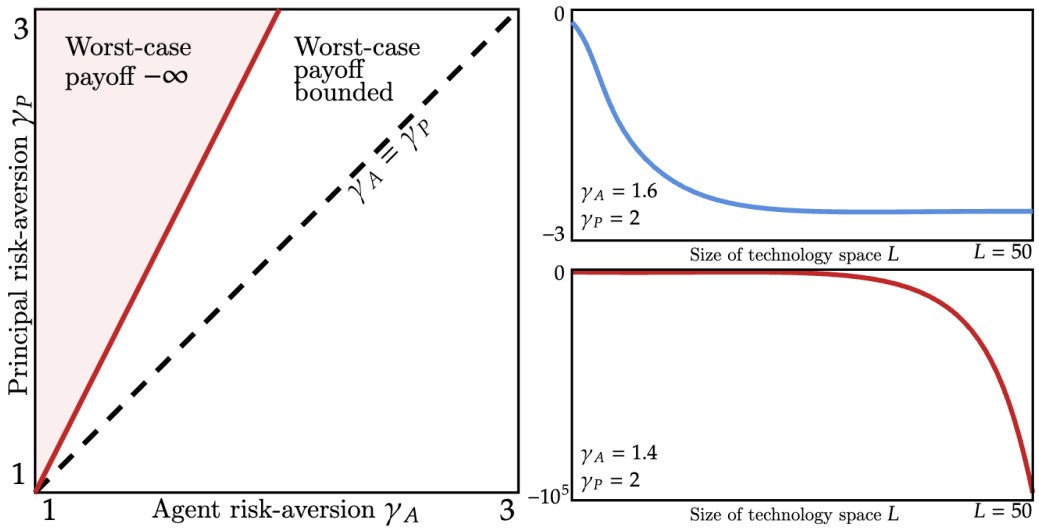
$$c(\theta, l) = \begin{cases} \exp\left(a \cdot l\right) & \text{if } \theta = 1 \\ \exp\left(-b \cdot l\right) & \text{if } \theta = 0 \end{cases}$$

with $a, b > 0$. Then for any linear Pigouvian tax $\phi_l = \alpha \cdot l$ that satisfies the condition in Theorem 4,

$$\lim_{L\to\infty} \inf_{\mathcal{F}\in\overline{\mathbb{F}}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] = -\infty \iff U, V \text{ s.t. } \frac{\gamma_P - 1}{\gamma_A - 1} > 1 + \frac{a}{b}.$$

Figure 16 illustrates worst-case payoffs as the principal and agent's risk-aversion coefficients vary. Panel (a) illustrates parameter regions under which the principal's payoffs under weak optimism is unboundedly poor. Panel (b) illustrates the principal's payoff under weak optimism when $\gamma_P = 2$ and $\gamma_A = 1.6$, noting that it asymptotes to approximately -3; by contrast, when the agent is just a little less risk-averse with $\gamma_A = 1.4$ the principal's payoff under weak-optimism diverges to $-\infty$ exponentially quickly with size of the technology space L.

Figure 16: Worst case (weak optimism) payoffs under CRRA utility
$a = b = 1$



We take this example to suggest that with *transformative* technologies—those that can drastically shape consumption-equivalents—even small wedges in risk preferences can translate into vastly different welfare outcomes under the worst-case learning process of weak optimism. We think this is reason for caution as reflected by our robustness notions. ◊

Indeed, that robustness is both quantitatively and qualitatively important allows us to partially bridge robustness and Bayesian optimality:

Proposition 2 (Informal)**.** *Suppose that the regulator's prior over learning processes and preferences $f \in \Delta(\mathbb{F} \times \mathbb{U})$ is 'diffuse'. Then as $L \to +\infty$, all mechanisms that attain*

$$\sup_{\phi \in \Phi} \int_{\substack{(U,\mathcal{F}) \\ \in \mathbb{U} \times \mathbb{F}}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U)\Big)\Big] df$$

imposes a hard limit at a finite level.

Proposition 2 states that when the principal's prior f is sufficiently 'diffuse' over a simple parametric class of Lèvy signal processes and CRRA agent preferences, Bayesian optimality requires hard limits. A precise statement is in Online Appendix V. The intuition is straightforward: without hard limits, the richness of the regulator's prior implies the existence of a positive probability set $\mathbb{S} \subseteq \mathbb{F} \times \mathbb{U}$ under which payoffs are unboundedly poor. Thus, our results imply that not only do hard limits perform well at *a* prior (Wald, 1950; LeCam, 1955), but that it is optimal under *flat* priors—sufficient uncertainty of the Bayesian kind is another reason for robust technology regulation.

Finally, Theorem 4 can be deployed to rank non-robust mechanisms according to their payoff-guarantees. This offers guidance for a regulator who might be concerned about robustness but faces political or other constraints and must choose the 'least bad' policy from a restricted set. In particular, if one mechanism induces more agent risk-aversion than another, it delivers better payoff guarantees when the agent's learning process is chosen adversarially:

Proposition 3. *Fix two mechanisms ϕ and ϕ' such that there exists an increasing convex function $g : \mathbb{R} \to \mathbb{R}$ such that $U^{\phi'}(\theta, l) = g \circ U^{\phi}(\theta, l)$ for every $\theta \in \Theta$ and $l \in \mathcal{L}$. Then worst-case payoffs are ordered as follows:*

$$\inf_{\mathcal{F} \in \overline{\mathbb{F}}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] \geq \inf_{\mathcal{F} \in \overline{\mathbb{F}}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi', \mathcal{F}, U))\Big].$$

5 POLICY IMPLICATIONS AND EXTENSIONS

We have argued that a regulator concerned about robustness should adapt policy to evolving information by relaxing and/or tightening a quantity limit on risky experimentation (Theorems 1, 2 and 3). We now distill several policy implications before briefly discussing how our framework and results can be extended in various directions.

A taxonomy of learnability for technologies. Our results offer simple guidance for how stringently different technologies should be regulated as a function of their learnability—how easily and predictably they can be learnt about—as well as the kinds of information regulators should seek out.

Some technologies are *transparently learnable*. These include clinical FDA drug trials, randomized lab or field interventions, and more broadly technologies for which the methods of modern science—double-blind randomization, placebo controls, and significance testing—can be brought to bear. Within our framework, the technology level l might correspond to the number of humans that have taken the drug so the regulator's learning process $\mathcal{G}$ comprises of iid signals resembling the Wald problem (Wald, 1947). Indeed, this is in essence what the FDA already does for clinical trials in which a drug of uncertain safety and efficacy is rolled out to a progressively larger number of participants, with the option for either party to halt experimentation: *within* each phase, the participating company might voluntarily choose to halt the trial; *at the boundary* of each phase, the FDA might stop the trial from proceeding. This practice mirrors our adaptive sandbox quite closely[50] and for such transparently learnable technologies, the challenge regulators face is to monitor the agent's experimentation such as to close the gap between $\mathcal{G}$ and $\mathcal{F}$ e.g., through enforcing the reporting of clinical trials, pre-analysis plans, and detecting fraud.[51]

Other technologies are *opaque*, beset by difficulties not just around learning about the technology, but also uncertainty about the learning process itself. We think transformative technologies like artificial intelligence often fall into this category—what, if anything, will the next generation of models teach us about whether powerful

[50] See also McClellan (2022) for a distinct analysis of 'approval mechanisms' that can be viewed as a once-and-for-all decision on whether a drug is approved—in which case it can be deployed limitlessly.

[51] There has been complaints that regulatory agencies like the FDA and FAA have been too conservative. We emphasize that our framework and results are consistent with this view since they are about the form of instrument choice—dynamic quantity limits—rather than the specifics around where these limits should be.

AI systems further out into the future can be aligned to human values? What kinds of biosecurity or labor market risks will such models pose? For such technologies, regulators might not learn *by default* as they are developed, but might nonetheless invest in *active learning* i.e., choosing $\mathcal{G}$ at some cost. Indeed, computer scientists have recently advocated for these kinds of evidence-seeking AI policy that "proactively helps to produce more information." (Casper, Krueger, and Hadfield-Menell, 2025). But this raises further questions: what kinds of information should regulators acquire, and how much resources should they expand to do so? Our framework is able to speak directly to such questions since our analysis can be viewed as the 'inner problem' that fixes the principal's learning process $\mathcal{G}$ and optimizes over $\mathcal{G}$-adapted mechanisms. This paves the way to analyze the 'outer problem' of how $\mathcal{G}$ should be chosen subject to constraints and/or costs—a rich toolkit from dynamic information acquisition (Moscarini and Smith, 2001; Zhong, 2022; Hébert and Woodford, 2023; Sannikov and Zhong, 2024) can be drawn upon here.

How much are we giving up for robustness? A natural worry is that robustness demands *too much*—that is, the payoff difference between robust and Bayes-optimal mechanisms can be large and, if so, our adaptive sandbox might be excessively stringent. We resist this claim for two distinct reasons.

If regulator learning is fast i.e., $\mathcal{G}$ is informative about the technology, then the difference between robustness and Bayes-optimal mechanisms must be small since the value of the latter is upper-bounded by the regulator directly observing the agent's learning $\mathcal{F}$ which, in turn, is upper-bounded by instantaneously learning the truth. In such cases, the adaptivity of our sandbox paired with fast regulator learning—e.g., for what we have called transparently learnable technologies—ensures that we are not giving up too much.

If regulator learning is slow i.e., $\mathcal{G}$ is not very informative about the technology, then as we have seen from Section 4, non-robust mechanisms are susceptible to worst-case learning processes that are both natural and generate unboundedly poor regulator payoffs. Thus the value of Bayes-optimal mechanisms[52] is now upper-bounded by the performance of mechanisms that impose hard limits on the technology (see Proposition 2) so the payoff gap between robustness and Bayes-optimal mechanisms once again cannot be too large.

Extensions and generalizations. We have deliberately worked within the simplest

[52] For a sufficiently 'rich' prior over $\mathbb{F} \times \mathbb{U}$; see Proposition 2.

model in which equilibrium technology choices are jointly shaped by learning, preferences, and policy. But there is quite a lot we can add to enrich the model along various dimensions—we outline some possibilities here and collect details in Online Appendix IV.

1. **Randomization and elicitation have zero value.** Our definition of mechanisms were those adapted to the principal's filtration which does not allow for additional randomization or type elicitation. Nonetheless, our regulator cannot improve upon her worst-case guarantee by doing either, precisely because under robust mechanisms, the worst-case learning process offers the agent no additional information beyond the principal's own. That our regulator cannot do better via randomization stands in contrast to work on robust pricing and contracting (Libgober and Mu, 2021); that our regulator cannot do better via elicitation of type reports is reminiscent of (Kruse and Strack, 2019).

2. **Revenue motives.** Suppose that the principal is interested in raising revenue i.e., we are in a (partially or fully) transferable utility environment e.g., the regulator's payoff from the technology until level l in state θ is: $v(\theta, l) + \lambda \cdot \phi_l$ for some $\lambda \geq 0$. An adaptive sandbox that solves the experimentation problem for a *weighted sum* of the principal's and agent's payoffs

$$\sup_{\ell} \mathbb{E}\Big[V(\mu_\ell, \ell) + \lambda \cdot U(\mu_\ell, \ell)\Big] \quad \text{s.t. } \ell \text{ is } \mathcal{G}\text{-adapted}$$

   augmented with a carefully chosen participation tax is learning-robust.

3. **Endogenous agent learning.** Suppose that the agent's learning process $\mathcal{F}$ is flexibly chosen at some cost which is increasing in the flow variation of the belief process (Moscarini and Smith, 2001; Zhong, 2022; Hébert and Woodford, 2023). When the principal is uncertain about how difficult it is to acquire information i.e., about the agent's cost function and/or variation measure, the adaptive sandbox ϕ^* remains robustly optimal.

4. **Behavioral robustness.** If the regulator does not learn, the sandbox ϕ^* delivers optimal worst-case guarantees whenever the agent employs any *undominated stopping rule*. Undominated stopping rules impose minimal assumptions on the agent's rationality by *only* ruling out stopping at a given pair of beliefs and technology levels (μ, l) when more technology is strictly better in the absence of

further information. These include stopping levels which are optimal, but also allows for agents to be misspecified or ambiguity averse about the future learning process, misoptimize, or be myopic.

5. **Technology shapes the state.** Suppose that the state $\theta_l \in \Theta$ evolves as an absorbing Markov process where it transitions from 0 to 1 at (inhomogenous) Poisson rate $\lambda_l \geq 0$, and state 1 is absorbing. This might correspond to a breakthrough in the technology that could make a technology safe. The sandbox mechanism ϕ^* remains robustly optimal. In particular, the regulator's beliefs about the state $\mathbb{E}[\theta \mid \mathcal{G}_l]$ exhibits a positive drift even in the absence of any information, and the sandbox limit $\overline{\ell}$ takes this into account by loosening accordingly.

6. **Multidimensional Technology.** Suppose that our technology space $\mathcal{L}^{\mathsf{M}} \subseteq \mathbb{R}^d$ is now a compact subset of the $d-$dimensional reals, and continue to assume that uncertainty Θ is totally ordered i.e., a higher-state corresponds to higher payoffs for each $\boldsymbol{l} \in \mathcal{L}^{\mathsf{M}}$. If the regulator does not learn, a *multidimensional sandbox* that imposes a zero tax on the technology, and a hard limit at a manifold $\mathcal{O}_* \subset \mathcal{L}^{\mathsf{M}}$ (see Figure 17) is robustly optimal, undominated, and time-consistent.

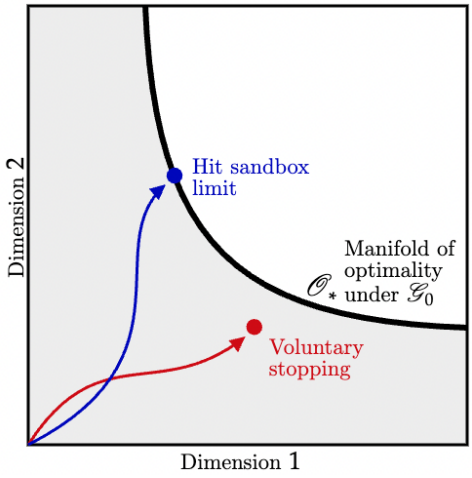


Figure 17: Multidim. sandbox

6 CONCLUSION

We have developed a parsimonious model in which learning, preferences, and policy jointly shape the equilibrium technology level. We showed that adaptive sandboxes mechanisms comprising of a zero marginal tax on the technology up to an evolving hard limit is robust to the agent's learning process and/or preferences (Theorem 1), undominated and dominant among robust and regular mechanisms (Theorem 2), and uniquely time-consistent among robust mechanisms (Theorem 3). Our results clarify when and why dynamic quantity limits on risky experimentation e.g., regulatory sandboxes and FDA clinical trials work well, as well as guidance for how they might be improved.

What do sandboxes safeguard against? We showed that they served as bulwarks against weak optimism: sans robust regulation, the agent's learning process is adversarially chosen to entice excessive risk-taking. This can magnify even small wedges in risk aversion to deliver unboundedly poor principal payoffs (Theorem 4). The worst-case learning process is natural and arises when new technologies pose tail risks so learning is lumpy e.g., financial crises induced by synthetic instruments, pandemics induced by gain-of-function research, or AI catastrophes. Robustness is not always (or even often!) a good guide to policy choice. But we have showed that it is especially important in the context of technology regulation—this, we think, are reasons to take it seriously.

Taken together, our results offer concrete guidance for policymakers who, in the coming years, will face difficult choices around how to regulate new and transformative technologies like artificial intelligence whose promises and perils loom large over human life. There are, of course, further questions around political feasibility, international coordination, and so on. These constraints might shrink the set of regulatory mechanisms, rendering our adaptive sandbox infeasible—what then? We think this is an important and exciting avenue for further work.

REFERENCES

Acemoglu, D. (2021): "Harms of AI," Tech. rep., National Bureau of Economic Research.

Acemoglu, D. and T. Lensman (2024): "Regulating transformative technologies," *American Economic Review: Insights*, 6, 359–376.

Allen, H. (2025): "Regulatory Sandboxes: One Decade On," *Working Paper*.

Altman, S. (2025): "Reflections," *Blog*.

Amador, M., I. Werning, and G.-M. Angeletos (2006): "Commitment vs. flexibility," *Econometrica*, 74, 365–396.

Amodei, D. (2025): "CEO Speaker Series With Dario Amodei of Anthropic," Council on Foreign Relations, conversation moderated by Michael Froman. Council on Foreign Relations, CEO Speaker Series.

Anthropic (2025): "Anthropic: Responsible Scaling Policy," .

Appaya, M. S., H. L. Gradstein, and M. Haji Kanz (2020): "Global Experiences from Regulatory Sandboxes," *World Bank Report*.

Armstrong, M. (1995): "Delegating decision-making to an agent with unknown preferences," *Unpublished Working Paper*.

Aschenbrenner, L. and P. Trammell (2024): "Existential Risk and Growth," *Global Priorities Institute WP*, 2024.

Auster, S., Y.-K. Che, and K. Mierendorff (2024): "Prolonged learning and hasty stopping: the wald problem with ambiguity," *American Economic Review*, 114, 426–461.

Australian Securities and Investments Commission (2024): "INFO 248 Enhanced regulatory sandbox," Accessed 2025.

Bennear, L. S. and J. B. Wiener (2019): "Adaptive regulation: instrument choice for policy learning over time," *Obtenido de Universidad de Harvard: https://www. hks. harvard. edu*.

Bergemann, D., B. Brooks, and S. Morris (2015): "The limits of price discrimination," *American Economic Review*, 105, 921–957.

——— (2017): "First-price auctions with general information structures: Implications for bidding and revenue," *Econometrica*, 85, 107–143.

Bergemann, D. and S. Morris (2005): "Robust mechanism design," *Econometrica*, 1771–1813.

Börgers, T. (2017): "(No) Foundations of dominant-strategy mechanisms: a comment on Chung and Ely (2007)," *Review of Economic Design*, 21, 73–82.

Borgers, T., J. Li, and K. Wang (2024): "Undominated mechanisms," .

Brooks, B. and S. Du (2023): "On the Structure of Informationally Robust Optimal Mechanisms," *Available at SSRN 3663721*.

Brunnermeier, M. K., A. Crockett, C. Goodhart, A. Persaud, and H. S. Shin (2009): *The fundamental principles of financial regulation*, Centre for Economic Policy Research (Great Britain).

Carroll, G. (2017): "Robustness and separation in multidimensional screening," *Econometrica*, 85, 453–488.

——— (2019): "Robustness in mechanism design and contracting," *Annual Review of Economics*, 11, 139–166.

Casper, S., D. Krueger, and D. Hadfield-Menell (2025): "Pitfalls of Evidence-Based AI Policy," *arXiv preprint arXiv:2502.09618*.

Chancelier, J.-P., M. De Lara, and A. de Palma (2009): "Risk aversion in expected intertemporal discounted utilities bandit problems," *Theory and decision*, 67, 433–440.

Chassang, S. (2013): "Calibrated incentive contracts," *Econometrica*, 81, 1935–1971.

Cohen, M. K., N. Kolt, Y. Bengio, G. K. Hadfield, and S. Russell (2024): "Regulating advanced artificial agents," *Science*, 384, 36–38.

Cornelli, G., S. Doerr, L. Gambacorta, and O. Merrouche (2024): "Regulatory sandboxes and fintech funding: evidence from the UK," *Review of Finance*, 28, 203–233.

Cruz, T. (2025): "Strengthening Artificial Intelligence Normalization and Diffusion By Oversight and eXperimentation Act (SANDBOX Act)," S. [Bill Number Pending], 119th Congress, 1st Session. U.S. Senate bill draft.

DeMarzo, P. and D. Duffie (1999): "A liquidity-based model of security design," *Econometrica*, 67, 65–99.

Diamond, P. and E. Saez (2011): "The case for a progressive tax: From basic research to policy recommendation," *Journal of Economic Perspectives*, 25, 165–190.

Du, S. (2018): "Robust mechanisms under common valuation," *Econometrica*, 86, 1569–1588.

Dworczak, P. and A. Pavan (2022): "Preparing for the worst but hoping for the best: Robust (bayesian) persuasion," *Econometrica*, 90, 2017–2051.

El Karoui, N. (1979): "Les aspects probabilistes du contrôle stochastique," in *École d'été de Probabilités de Saint-Flour IX-1979*.

Ely, J. C. and M. Szydlowski (2020): "Moving the goalposts," *Journal of Political Economy*, 128, 468–506.

Epstein, L. G. and M. Schneider (2003): "Recursive multiple-priors," *Journal of*

Economic Theory, 113, 1–31.
——— (2007): "Learning under ambiguity," *The Review of Economic Studies*, 74, 1275–1303.
European Parliament (2024): "REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," *Official Journal of the European Union*.
European Parliament Briefing (2024): "Artificial intelligence act and regulatory sandboxes," *European Parliament Briefing*.
FDA (2008): "Guidance for Industry CGMP for Phase 1 Investigational Drugs," .
Feyen, E., J. Frost, L. Gambacorta, H. Natarajan, and M. Saal (2021): "Fintech and the digital transformation of financial services: implications for market structure and public policy," *BIS papers*.
Financial Conduct Authority (2015): "Regulatory sandbox," *Financial Conduct Authority*.
——— (2023): "Apply to the Regulatory Sandbox," Accessed 2025.
Forges, F. (1986): "An approach to communication equilibria," *Econometrica: Journal of the Econometric Society*, 1375–1385.
Frankel, A. (2014): "Aligned delegation," *American Economic Review*, 104, 66–83.
Gans, J. (2025): "Regulating the Direction of Innovation," *Journal of Public Economics*, Forthcoming.
Greenblatt, R., C. Denison, B. Wright, F. Roger, M. MacDiarmid, S. Marks, J. Treutlein, T. Belonax, J. Chen, D. Duvenaud, et al. (2024): "Alignment faking in large language models," *arXiv preprint arXiv:2412.14093*.
Greenshtein, E. (1996): "Comparison of sequential experiments," *The Annals of Statistics*, 24, 436–448.
Guerreiro, J., S. Rebelo, and P. Teles (2025): "Regulating artificial intelligence," Tech. rep., National Bureau of Economic Research.
Guo, Y. (2016): "Dynamic delegation of experimentation," *American Economic Review*, 106, 1969–2008.
Guo, Y. and E. Shmaya (2019): "Robust monopoly regulation," *arXiv preprint arXiv:1910.04260*.
——— (2023): "Regret-Minimizing Project Choice," *Econometrica*, 91, 1567–1593.
Hadfield, G. and A. Koh (2025): "An Economy of AI Agents," *Working paper prepared for the NBER Handbook on the Economics of Transformative AI*.
Hannan, J. (1957): "Approximation to Bayes risk in repeated play," *Contributions to the Theory of Games*, 3, 97–139.
Hébert, B. and M. Woodford (2023): "Rational inattention when decisions take time," *Journal of Economic Theory*, 208, 105612.
Heim, L. and L. Koessler (2024): "Training compute thresholds: Features and functions in AI regulation," *arXiv preprint arXiv:2405.10799*.
Holmström, B. R. (1977): *On Incentives and Control in Organizations.*, Stanford University.
Jones, C. I. (2016): "Life and growth," *Journal of political Economy*, 124, 539–578.
——— (2024): "The ai dilemma: Growth versus existential risk," *American Economic Review: Insights*, 6, 575–590.
——— (2025): "How Much Should We Spend to Reduce AI's Existential Risk?" Tech. rep., National Bureau of Economic Research.
Junod, S. (2008): "FDA and clinical drug trials: a short history," *FDLI Update*, 55.
Karger, E., J. Rosenberg, Z. Jacobs, M. Hickman, R. Hadshar, K. Gamin, and P. Tetlock (2023): "Forecasting Existential Risks Evidence from a Long-Run

Forecasting Tournament," *Forecasting Research Institute*.
Keller, G., V. Novák, and T. Willems (2019): "A note on optimal experimentation under risk aversion," *Journal of Economic Theory*, 179, 476–487.
Keller, G. and S. Rady (2015): "Breakdowns," *Theoretical Economics*, 10, 175–202.
Keller, G., S. Rady, and M. Cripps (2005): "Strategic experimentation with exponential bandits," *Econometrica*, 73, 39–68.
Kobylanski, M. and M.-C. Quenez (2012): "Optimal stopping time problem in a general framework," *Electron. J. Probab*, 17, 1–28.
Koh, A. and S. Sanguanmoo (2022): "Attention capture," *arXiv preprint arXiv:2209.05570*.
Koh, A., S. Sanguanmoo, and W. Zhong (2024): "Persuasion and Optimal Stopping," *arXiv preprint arXiv:2406.12278*.
Kruse, T. and P. Strack (2015): "Optimal stopping with private information," *Journal of Economic Theory*, 159, 702–727.
——— (2019): "An inverse optimal stopping problem for diffusion processes," *Mathematics of Operations Research*, 44, 423–439.
Kulveit, J., R. Douglas, N. Ammann, D. Turan, D. Krueger, and D. Duvenaud (2025): "Gradual disempowerment: Systemic existential risks from incremental AI development," *arXiv preprint arXiv:2501.16946*.
Laffont, J.-J. and J. Tirole (1993): *A theory of incentives in procurement and regulation*, MIT press.
LeCam, L. (1955): "An extension of Wald's theory of statistical decision functions," *The Annals of Mathematical Statistics*, 26, 69–81.
Li, Z., J. Libgober, and X. Mu (2022): "Sequentially Optimal Pricing under Informational Robustness," *arXiv preprint arXiv:2202.04616*.
Libgober, J. and X. Mu (2021): "Informational robustness in intertemporal pricing," *The Review of Economic Studies*, 88, 1224–1252.
Lizzeri, A., E. Shmaya, and L. Yariv (2024): "Disentangling exploration from exploitation," Tech. rep., National Bureau of Economic Research.
Marshall, A. (1890): "Principles of Economics," *London: Mac-Millan*, 1–627.
McClellan, A. (2022): "Experimentation and approval mechanisms," *Econometrica*, 90, 2215–2247.
Milgrom, P. and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.
Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.
Moscarini, G. and L. Smith (2001): "The optimal level of experimentation," *Econometrica*, 69, 1629–1644.
Myerson, R. B. (1986): "Multistage games with communication," *Econometrica: Journal of the Econometric Society*, 323–358.
Nachman, D. C. and T. H. Noe (1994): "Optimal design of securities under asymmetric information," *The Review of Financial Studies*, 7, 1–44.
OECD (2023): "Regulatory sandboxes in artificial intelligence," *OECD Digital Economy Papers*.
Orlov, D., A. Skrzypacz, and P. Zryumov (2020): "Persuading the principal to wait," *Journal of Political Economy*, 128, 2542–2578.
Penta, A. (2015): "Robust dynamic implementation," *Journal of Economic Theory*, 160, 280–316.
Quah, J. K.-H. and B. Strulovici (2013): "Discounting, values, and decisions," *Journal of Political Economy*, 121, 896–939.

Reuel, A. and T. A. Undheim (2024): "Generative AI needs adaptive governance," *arXiv preprint arXiv:2406.04554*.
Riedel, F. (2009): "Optimal stopping with multiple priors," *Econometrica*, 77, 857–908.
Russell, S. (2019): *Human compatible: AI and the problem of control*, Penguin Uk.
Saeedi, M., Y. Shen, and A. Shourideh (2024): "Getting the Agent to Wait," *arXiv preprint arXiv:2407.19127*.
Sannikov, Y. and W. Zhong (2024): "Exploration and Stopping," .
Sarro, D. (2025): "Sandbox Fictions," *Osgoode Hall Law Journal [forthcoming]*.
Savage, L. J. (1972): *The foundations of statistics*, Courier Corporation.
Schrepel, T. (2025): "Adaptive Regulation," *Available at SSRN 5416454*.
Shavell, S. (1984): "Liability for harm versus regulation of safety," *The Journal of Legal Studies*, 13, 357–374.
Simon, L. K. and M. B. Stinchcombe (1989): "Extensive form games in continuous time: Pure strategies," *Econometrica: Journal of the Econometric Society*, 1171–1214.
Tirole, J. (2016): "From bottom of the barrel to cream of the crop: Sequential screening with positive selection," *Econometrica*, 84, 1291–1343.
Vovk, V. G. (1990): "Aggregating strategies," in *Proceedings of the third annual workshop on Computational learning theory*, 371–386.
Wald, A. (1947): "Foundations of a general theory of sequential decision functions," *Econometrica, Journal of the Econometric Society*, 279–313.
——— (1950): "Statistical Decision Functions," .
Weitzman, M. L. (1974): "Prices vs. Quantities," *Review of Economic Studies*, 41, 315–329.
——— (1998): "Recombinant growth," *The Quarterly Journal of Economics*, 113, 331–360.
Zhong, W. (2022): "Optimal dynamic information acquisition," *Econometrica*, 90, 1537–1582.

APPENDIX TO ROBUST TECHNOLOGY REGULATION

Organization. The appendix is organized as follows. Appendix A develops some mathematical preliminaries and proves results in Section 3 (Lemma 1 and Proposition 1, Theorems 1, 2 and 3). Appendix B proves results in Section 4 (Theorem 4, Propositions 2 and 3).

A PROOFS OF RESULTS IN SECTION 3

A.1 Mathematical preliminaries and notation. Filtrations can be delicate so it is worthwhile to flesh out some details. Our probability space is $\left(\Omega,(\mathcal{F},\mathcal{G}),\mathbb{P}\right)$ where $\mathcal{F}:=(\mathcal{F}_l)_{l\in\mathcal{L}}$ and $\mathcal{G}:=(\mathcal{G}_l)_{l\in\mathcal{L}}$ filter Ω. We will keep the space Ω abstract but emphasize that it is sufficiently rich to capture both randomness in what is learnt about the state ('first-order'), randomness in what is learnt about what will be learnt ('second-order'),

as well as higher-order learning. $\mathcal{G} = (\mathcal{G}_l)_l$ is the natural filtration generated by some right-continuous signal process $(X^\theta_l)_l$. $\mathcal{F} = (\mathcal{F}_l)_l$ is any right-continuous filtration such that $\mathcal{F}_l \supseteq \mathcal{G}_l$ for each $l \in \mathcal{L}$. As in the main text, we denote all such filtrations with $\mathbb{F}$.

A.2 Comparative statics for stopping. We start with the following lemma in which we compare the principal and agent's preferences over stopping levels (which need not be optimal). We will use it to show Lemma 1, as well as some of our subsequent proofs that will rank the agent's and principal's preferences over stopping levels.

Lemma 2. *For any two $\mathcal{F}$-stopping levels ℓ^-, ℓ^+ such that $\ell^- \leq \ell^+$ a.s.:*

$$\left\{ V(\mu_{\ell^-}, \ell^-) \underset{(<)}{\leq} \mathbb{E}\Big[V(\mu_{\ell^+}, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] \right\} \subseteq \left\{ U(\mu_{\ell^-}, \ell^-) \underset{(<)}{\leq} \mathbb{E}\Big[U(\mu_{\ell^+}, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] \right\}.$$

Proof of Lemma 2. We proceed in several steps.

Step 1: Lower-bound value of continuing to ℓ^+ under U and $\mathcal{F}_{\ell^-}$.

Consider that

$$\begin{aligned}
\mathbb{E}\Big[U(\mu_{\ell^+}, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] &= \mathbb{E}\Big[\mu_{\ell^+} \cdot U(1, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] + \mathbb{E}\Big[(1-\mu_{\ell^+}) \cdot U(0, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] \\
&= \mathbb{E}\Big[\mu_{\ell^+} \cdot g \circ V(1, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] + \mathbb{E}\Big[(1-\mu_{\ell^+}) \cdot g \circ V(0, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] \\
&\geq \mu_{\ell^-} \cdot g\Bigg(\underbrace{\mathbb{E}\left[\frac{\mu_{\ell^+}}{\mu_{\ell^-}} \cdot V(1, \ell^+)\middle|\mathcal{F}_{\ell^-}\right]}_{=:\overline{V}_1}\Bigg) \\
&\quad + (1-\mu_{\ell^-}) \cdot g\Bigg(\underbrace{\mathbb{E}\left[\frac{1-\mu_{\ell^+}}{1-\mu_{\ell^-}} \cdot V(0, \ell^+)\middle|\mathcal{F}_{\ell^-}\right]}_{=:\overline{V}_0}\Bigg), \qquad (1)
\end{aligned}$$

where the last inequality is from a version of Jensen's inequality for random measures e.g., for the inequality in the first term we have:

$$\begin{aligned}
&\mathbb{E}\Big[\mu_{\ell^+} \cdot g \circ V(1, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big] \\
&= \int_{\omega \in \Omega} \mu_{l^+} \cdot g\Big(V(1, \ell^+)\Big) d\mathbb{P}(\cdot|\mathcal{F}_{\ell^-})
\end{aligned}$$

$$= \int \mu_{\ell^+} d\mathbb{P}(\cdot|\mathcal{F}_{\ell^-}) \cdot \left(\int_{\omega\in\Omega} g\Big(V(1,\ell^+)\Big) \cdot \frac{\mu_{\ell^+}}{\int \mu_{\ell^+} d\mathbb{P}(\cdot|\mathcal{F}_{\ell^-})} d\mathbb{P}(\cdot|\mathcal{F}_{\ell^-}) \right)$$

(Random transform of measure)

$$\geq \mu_{\ell^-} g\left(\mathbb{E}\left[\frac{\mu_{\ell^+}}{\mu_{\ell^-}} V(1,\ell^+)\middle|\mathcal{F}_{\ell^+}\right]\right) \qquad \text{(Jensen + Optional Stopping + } \ell^+ > \ell^-\text{)}$$

and an analogous argument applies to the second term. Since $\ell^+ \geq \ell^-$ and $V(1,\cdot)$ is increasing, we have $V(1,\ell^+) \geq V(1,\ell^-)$ while $V(0,\cdot)$ is decreasing so $V(0,\ell^+) \leq V(0,\ell^-)$. Thus,

$$\begin{aligned}
\overline{V}_1 := \mathbb{E}\left[\frac{\mu_{\ell^+}}{\mu_{\ell^-}} V(1,\ell^+)\middle|\mathcal{F}_{\ell^-}\right] &\geq \mathbb{E}\left[\frac{\mu_{\ell^+}}{\mu_{\ell^-}} V(1,\ell^-)\middle|\mathcal{F}_{\ell^-}\right] && (V(1,\cdot) \text{ increasing}) \\
&= \frac{V(1,\ell^-)}{\mu_{\ell^-}} \mathbb{E}[\mu_{\ell^+}|\mathcal{F}_{\ell^-}] \\
&= V(1,\ell^-). && \text{(Optional Stopping \& } \{\ell^+ > \ell^-\}\text{)}
\end{aligned}$$

By a similar argument, $\overline{V}_0 \leq V(0,\ell^-)$, and since $V(1,l) \geq V(0,l)$ we have $\overline{V}_1 \geq \overline{V}_0$.

Step 2: Upper-bound value of stopping at ℓ^- under U.
Define the following $\mathcal{F}_{\ell^-}$-measurable event

$$A := \left\{\omega \in \Omega : V(\mu_{\ell^-}, \ell^-) \leq \mathbb{E}\Big[V(\mu_{\ell^+}, \ell^+)\Big|\mathcal{F}_{\ell^-}\Big]\right\}$$

which is the event that at $\mathcal{F}_{\ell^-}$ with preference V it is optimal to stop.

The inequality constraint within the event A can be equivalently written in terms of $\overline{V}_0$ and $\overline{V}_1$ as follows:

$$\mu_{\ell^-} \cdot \overline{V}_1 + (1-\mu_{\ell^-}) \cdot \overline{V}_0 \geq \mu_{\ell^-} \cdot V(1,\ell^-) + (1-\mu_{\ell^-}) \cdot V(0,\ell^-).$$

Since we have already argued from Step 1 that $\overline{V}_1 \geq V(1,\ell^-)$ and $\overline{V}_0 \leq V(0,\ell^-)$, the above inequality implies that under the event A

$$\Big(\mu_{\ell^-} \cdot \overline{V}_1, (1-\mu_{\ell^-}) \cdot \overline{V}_0\Big) \quad \text{majorizes} \quad \Big(\mu_{\ell^-} \cdot V(1,\ell^-), (1-\mu_{\ell^-}) \cdot V(0,\ell^-)\Big).$$

The weighted Karamata inequality then implies, under the event A

$$\begin{aligned}
\mu_{\ell^-} \cdot g(\overline{V}_1) + (1-\mu_{\ell^-}) \cdot g(\overline{V}_0) &\geq \mu_{\ell^-} \cdot g \circ V(1,\ell^-) + (1-\mu_{\ell^-}) \cdot g \circ V(0,\ell^-) \\
&= U(\mu_{\ell^-}, \ell^-). \qquad (2)
\end{aligned}$$

Step 3: Putting things together

Equations (1) and (2) imply that under the event A

$$\begin{aligned}\mathbb{E}\Big[U(\mu_{\ell^+},\ell^+)\Big|\mathcal{F}_{\ell^-}\Big] &\geq \mu_{\ell^-}\cdot g(\overline{V}_1)+(1-\mu_{\ell^-})\cdot g(\overline{V}_0) && \text{(Equation (1))}\\ &\geq U(\mu_{\ell^-},\ell^-) && \text{(Equation (2))}\end{aligned}$$

so at $\mathcal{F}_{\ell^-}$ with preference U it is also optimal to stop, as required. The proof for the strict inequality is exactly the same. □

We now show Lemma 1.

Proof of Lemma 1. Fix the learning process $\mathcal{F}\in\mathbb{F}$, $\mathcal{F}$-stopping levels $\ell_1\leq\ell_2$. For the indirect utilities $U,V:\Delta(\Theta)\times\mathcal{L}\to\mathbb{R}$ let ℓ_U and ℓ_V be any selection of optimal stopping levels that solve

$$\begin{array}{ll}\sup_{s\in[\ell_1,\ell_2]}\mathbb{E}\Big[U(\mu_s,s)|\mathcal{F}_{\ell_1}\Big] & \text{and} \quad \sup_{s\in[\ell_1,\ell_2]}\mathbb{E}\Big[V(\mu_s,s)|\mathcal{F}_{\ell_1}\Big]\\ \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted} & \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}\end{array}$$

respectively. We will show that $\ell^\vee:=\ell_U\vee\ell_V$ and $\ell^\wedge:=\ell_U\wedge\ell_V$ also solve the first and second problems respectively.

Define the event $\widetilde{\Omega}:=\{\omega\in\Omega:\ell_V>\ell_U\}$. If $\mathbb{P}(\widetilde{\Omega})=0$ there is nothing to show since $\ell_V\leq\ell_U$ a.s. so we suppose $\mathbb{P}(\widetilde{\Omega})>0$.

For any $\mathcal{F}_{\ell_U}$-measurable event $A\subset\widetilde{\Omega}$, consider the stopping time

$$\ell'_A=\begin{cases}\ell_U & \text{if } \omega\in A\\ \ell_V & \text{otherwise.}\end{cases}$$

This is indeed a stopping time because $\ell_U<\ell_V$ under the event $A\subset\widetilde{\Omega}$. The optimality of ℓ_V implies

$$\begin{aligned}&\mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)|\mathcal{F}_{\ell_1}\Big]\geq\mathbb{E}\Big[V(\mu_{\ell'_A},\ell'_A)|\mathcal{F}_{\ell_1}\Big]\\ \implies\quad &\mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)1\{\omega\in A\}\Big]\geq\mathbb{E}\Big[V(\mu_{\ell_U},\ell_U)1\{\omega\in A\}\Big]\\ \implies\quad &\mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)\Big|\mathcal{F}_{\ell_U}\Big]\geq V(\mu_{\ell_U},\ell_U)\quad\text{a.s. under the event }\widetilde{\Omega}\end{aligned}\tag{3}$$

From Lemma 2, since $U = g \circ V$ we have the same under preference U:

$$\mathbb{E}\Big[U(\mu_{\ell_V}, \ell_V)\Big|\mathcal{F}_{\ell_U}\Big] \geq U(\mu_{\ell_U}, \ell_U) \quad \text{a.s. under the event } \widetilde{\Omega}. \tag{4}$$

To finish, we will show that (i) the stopping time $\ell^\vee := \ell_U \vee \ell_V$ yields an identical expected value as ℓ_V for preference V under $\mathcal{F}_{\ell_1}$ which, recall, is the coarsest filtration over our space $[\ell_1, \ell_2]$; and (ii) the stopping time $\ell^\wedge := \ell_U \wedge \ell_V$ yields the same expected value as ℓ_U for preference U under $\mathcal{F}_{\ell_1}$.

Equation (4) implies

$$\begin{aligned}
\mathbb{E}\Big[U(\mu_{\ell_V}, \ell_V)1\{\omega \in \widetilde{\Omega}\}\Big|\mathcal{F}_{\ell_1}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[U(\mu_{\ell_V}, \ell_V)\cdot 1\{\omega \in \widetilde{\Omega}\}\,\Big|\,\mathcal{F}_{\ell_U}\Big]\Big|\mathcal{F}_{\ell_1}\Big] \\
&\qquad\qquad \text{(Iterated expectation)} \\
&= \mathbb{E}\Big[1\{\omega \in \widetilde{\Omega}\}\cdot \mathbb{E}\Big[U(\mu_{\ell_V}, \ell_V)\,\Big|\,\mathcal{F}_{\ell_U}\Big]\Big|\mathcal{F}_{\ell_1}\Big] \\
&\geq \mathbb{E}\Big[U(\mu_{\ell_U}, \ell_U)\cdot 1\{\omega \in \widetilde{\Omega}\}\,\Big|\,\mathcal{F}_{\ell_1}\Big]. \qquad \text{(Equation (4))}
\end{aligned}$$

Thus,

$$\begin{aligned}
0 &\leq \mathbb{E}\Big[U(\mu_{\ell_U}, \ell_U)\Big|\mathcal{F}_{\ell_1}\Big] - \mathbb{E}\Big[U(\mu_{\ell^\vee}, \ell^\vee)\Big|\mathcal{F}_{\ell_1}\Big] \\
&= \mathbb{E}\Big[U(\mu_{\ell_U}, \ell_U)1\{\omega \in \widetilde{\Omega}\}\Big|\mathcal{F}_{\ell_1}\Big] - \mathbb{E}\Big[U(\mu_{\ell_V}, \ell_V)1\{\omega \in \widetilde{\Omega}\}\Big|\mathcal{F}_{\ell_1}\Big] \\
&\leq 0,
\end{aligned}$$

which implies all the above inequalities, including Equation (4), must be equalities. Thus,

$$\mathbb{E}\Big[U(\mu_{\ell_U}, \ell_U)\Big|\mathcal{F}_{\ell_1}\Big] = \mathbb{E}\Big[U(\mu_{\ell^\vee}, \ell^\vee)\Big|\mathcal{F}_{\ell_1}\Big],$$

implying $\ell^\vee$ solves the optimal stopping problem $\sup_{s\in[\ell_1,\ell_2]}\mathbb{E}[V(\mu_s, s)|\mathcal{F}_{\ell_1}]$.

It remains to show $\ell^\wedge$ yields the same expected value as ℓ_U for preference U under $\mathcal{F}_{\ell_1}$. Since Equation (4) must be an equality:

$$\mathbb{E}\Big[U(\mu_{\ell_V}, \ell_V)\Big|\mathcal{F}_{\ell_U}\Big] = U(\mu_{\ell_U}, \ell_U) \quad \text{a.s. under the event } \widetilde{\Omega}.$$

Now observe we have

$$\mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)\Big|\mathcal{F}_{\ell_U}\Big] \underset{\text{Lemma 2}}{\overset{\text{Equation (3)}}{\gtrless}} V(\mu_{\ell_U},\ell_U) \quad \text{a.s. under the event } \widetilde{\Omega}.$$

so this must be an equality. Hence we have

$$\mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)\Big|\mathcal{F}_{\ell_U}\Big] = V(\mu_{\ell_U},\ell_U) \quad \text{a.s. under the event } \widetilde{\Omega}.$$

This implies (once again using the fact that $\mathcal{F}_{\ell_1} \subseteq \mathcal{F}_{\ell_U}$ and the law of iterated expectation):

$$\begin{aligned}
\mathbb{E}\Big[V(\mu_{\ell_U},\ell_U)1\{\omega\in\widetilde{\Omega}\}\Big|\mathcal{F}_{\ell_1}\Big] &= \mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)1\{\omega\in\widetilde{\Omega}\}\Big|\mathcal{F}_{\ell_1}\Big] \\
\implies \mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)\Big|\mathcal{F}_{\ell_1}\Big] &= \mathbb{E}\Big[V(\mu_{\ell_V},\ell_V)1(\omega\in\widetilde{\Omega}) + V(\mu_{\ell_V},\ell_V)1(\omega\notin\widetilde{\Omega})\Big|\mathcal{F}_{\ell_1}\Big] \\
&= \mathbb{E}\Big[V(\mu_{\ell_U},\ell_U)1(\omega\in\widetilde{\Omega}) + V(\mu_{\ell_V},\ell_V)1(\underbrace{\omega\notin\widetilde{\Omega}}_{\ell_V<\ell_U})\Big|\mathcal{F}_{\ell_1}\Big] \\
&= \mathbb{E}\Big[V(\mu_{\ell^\wedge},\ell^\wedge)\Big|\mathcal{F}_{\ell_1}\Big],
\end{aligned}$$

implying $\ell^\wedge$ solves the optimal stopping problem $\sup_{s\in[\ell_1,\ell_2]}\mathbb{E}[V(\mu_s,s)|\mathcal{F}_{\ell_1}]$, as required. □

Inspecting the proof yields the same result for any set of states on which payoffs can be totally ordered by replacing the binary state with a vector and invoking the same Schur convexity argument. In Online Appendix II we show that this nests and generalizes the main results of Keller et al. (2019); Chancelier et al. (2009).

A.3 Proof of Theorem 1.

Proof. Recall that in the main text we defined the adaptive sandbox mechanism as a $\overline{\mathbb{R}}$-valued stochastic process ϕ^* which is adapted to the principal's filtration $\mathcal{G}$.

Step 1. ϕ^* achieves the payoff guarantee under the learning process $\mathcal{F}=\mathcal{G}$.
Observe

$$\begin{aligned}
(\mathrm{L}) &:= \sup_{\phi\in\Phi}\inf_{\mathcal{F}\in\mathbb{F}}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)\Big] \\
&\leq \sup_{\phi}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)\Big]
\end{aligned}$$

$$\leq \sup_{s} \mathbb{E}\Big[V\big(\mu_s, s\big)\Big] \quad \text{s.t. } s \text{ is } \mathcal{G}\text{-adapted} \qquad (\ell^* \text{ is } \mathcal{G}\text{-adapted})$$
$$= \mathbb{E}\Big[V\big(\mu_{\overline{\ell}}, \overline{\ell}\big)\Big] \qquad (\text{defn. of } \overline{\ell})$$

Step 2. No additional info is the worst-case under ϕ^*. Take any $(\mathcal{F}, U) \in \mathbb{F}\times\mathbb{U}$. Notice that at filtration $\mathcal{F}_{\ell^*}$ with $\ell^* < \overline{\ell}$ we have

$$\mathbb{E}\Big[U(\mu_{\ell^*}, \ell^*)\Big|\mathcal{F}_{\ell^*}\Big] \geq \sup_{\substack{s\geq \ell^* \\ \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}}} \mathbb{E}\Big[U(\mu_s, s) - \phi^*_s\Big|\mathcal{F}_{\ell^*}\Big] \qquad (l^* \text{ solves (O)})$$
$$= \sup_{\substack{\overline{\ell}\geq s\geq \ell^* \\ \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}}} \mathbb{E}\Big[U(\mu_s, s)\Big|\mathcal{F}_{\ell^*}\Big] \qquad (\text{defn. of adaptive sandbox})$$

and from Lemma 1 this implies that at $\mathcal{F}_{\ell^*}$

$$\mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*)\Big|\mathcal{F}_{\ell^*}\Big] = \sup_{\substack{\overline{\ell}\geq s\geq \ell^* \\ \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}}} \mathbb{E}\Big[V(\mu_s, s)\Big|\mathcal{F}_{\ell^*}\Big]$$

because, for any stopping time s that solves the optimal stopping problem in the RHS of the above equality, $\ell^* = \ell^* \wedge s$ must also solve the same stopping problem.

Now denoting $\gamma \in \Delta(\mathcal{L})$ as the probability distribution i.e., some Borel measure (of $\ell^*(\phi^*, \mathcal{F}, U)$), we thus have:

$$\mathbb{E}\Big[V\big(\mu_{\ell^*}, \ell^*(\phi^*, \mathcal{F}, U)\big)\Big|\big\{\ell^* < \overline{\ell}\big\}\Big] = \int_{\ell^*<\overline{\ell}} \sup_{\substack{\overline{\ell}\geq s>\ell^* \\ \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}}} \mathbb{E}\Big[V(\mu_s, s)\Big|\mathcal{F}_{\ell^*}\Big]\, d\gamma(\ell^*)$$
$$\geq \int_{\ell^*<\overline{\ell}} \sup_{\substack{\overline{\ell}\geq s>\ell^* \\ \text{s.t. } s \text{ is } \mathcal{G}\text{-adapted}}} \mathbb{E}\Big[V(\mu_\ell, \ell)\Big|\mathcal{F}_{\ell^*}\Big]\, d\gamma(\ell^*)$$
$$\geq \mathbb{E}\Big[V(\mu_{\overline{\ell}}, \overline{\ell})\Big|\big\{\ell^* < \overline{\ell}\big\}\Big].$$

From the law of total expectation,

$$\mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] = \mathbb{P}\Big(\big\{\ell^* = \overline{\ell}\big\}\Big)\cdot\mathbb{E}\Big[V\big(\mu_{l^*}, \ell^*(\phi^*, \mathcal{F}, U)\big)\Big|\big\{\ell^* = \overline{\ell}\big\}\Big]$$
$$+ \mathbb{P}\Big(\big\{\ell^* < \overline{\ell}\big\}\Big)\cdot\mathbb{E}\Big[V\big(\mu_{\ell^*}, \ell^*(\phi^*, \mathcal{F}, U)\big)\Big|\big\{\ell^* < \overline{\ell}\big\}\Big]$$

hence

$$\mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] \geq \mathbb{E}\Big[V\big(\mu_{\overline{\ell}}, \overline{\ell}\big)\Big].$$

To conclude, observe

$$\inf_{\mathcal{F}\in\mathbb{F}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi^*, \mathcal{F}, U)\Big)\Big] \geq \mathbb{E}\Big[V(\mu_{\overline{\ell}}, \overline{\ell})\Big] \qquad \text{(Step 2)}$$

$$\geq (\mathrm{L}) \qquad \text{(Step 1)}$$

i.e., ϕ^* solves (L). Since ϕ^* does not depend on U, it also solves (D). □

Outline of proof of Proposition 1 and Theorem 2. We proceed in several steps. Lemma 3 argues that 'floor mechanisms' (those that impose an infinite subsidy) are either not dually-robust, or are equivalent to a quota. Next, Lemma 4 shows a dually robust mechanism must have a hard limit at $\overline{\ell}$. Lemma 5 shows a dually robust mechanism must have zero marginal transfer until the hard limit. These two parts together imply the forward direction of Proposition 1. Finally, Lemma 7 shows the converse of Proposition 1. Then, Proposition 1 is used to show *phi** dominates robust and regular mechanisms; that ϕ^* is undominated is showed separately. Figure 18 illustrates.

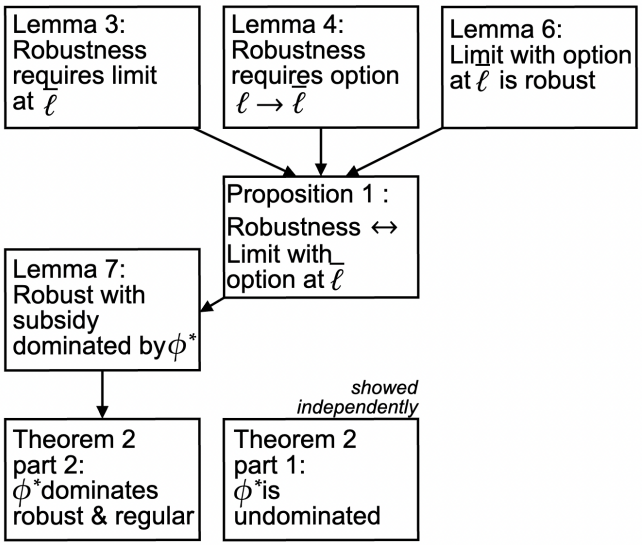


Figure 18: Outline of proof steps for Theorem 2

A.4 Proof of Proposition 1. We prove the complete characterization of dually-robust mechanisms given in the main text. Denote the set of dually-robust adaptive mechanisms with

$$\Phi^* := \Big\{\phi \in \Phi : \phi \text{ solves (D)}\Big\}.$$

We first show that it if a dually-robust mechanism delivers an infinite subsidy with

positive probability, it must do so only at $\bar{\ell}$:

Lemma 3. *If $\phi \in \Phi^*$ and there exists a $\mathcal{G}$–stopping level s such that such that $\phi(s) = -\infty$ with positive probability, then $s = \overline{\ell}$ almost surely.*

Proof of Lemma 3. Consider the agent's filtration $(\mathcal{G}_l)_l = (\mathcal{F}_l)_l$ and the agent's utility function $U = V$. Under the mechanism ϕ, the latest stopping level s_0 is an optimal stopping level for the agent because it yields the positively infinite utility. However, since the principal's optimal stopping is unique and s_0 is adapted to the filtration $(\mathcal{G}_l)_l$, we must have $s_0 = \overline{\ell}$ almost surely.

Thus, under $\phi \in \Phi^*$, $\overline{\ell}$ is always the agent's optimal stopping no matter what his filtration is. This implies the principal's value under ϕ always equals $\mathbb{E}[V(\theta, \overline{\ell})]$ regardless of the agent's filtration.

To show that ϕ is dominated by the sandbox mechanism, it suffices to find the agent's filtration that yields strictly better the principal's outcome than that under no further information and the sandbox mechanism ϕ^*. Consider the agent's filtration $(\mathcal{G}_l)_l$ such that $\mathcal{G}_l = \mathcal{F}_l \vee \sigma(\theta)$, i.e., the agent fully learn the state, and the agent's utility function $U = V$. If $\theta = 0$, the agent always stop at level 0 under the sandbox mechanism. On the other hand, if $\theta = 1$, the agent always stops as late as possible, which is $\overline{\ell}$, under the sandbox mechanism. Thus, the principal's expected utility is

$$\underbrace{(1-\mu_0)\mathbb{E}\Big[V(0,0)\Big]}_{\theta=0} + \mathbb{E}\Big[V(1,\overline{\ell})1\{\theta=1\}\Big] > \mathbb{E}\Big[V(0,\overline{\ell})1\{\theta=0\}\Big] + \mathbb{E}\Big[V(1,\overline{\ell})1\{\theta=1\}\Big]$$
$$= \mathbb{E}\Big[V(\theta,\overline{\ell})\Big]$$

which is the robustly optimal payoff for the principal, as desired. □

Lemma 3 implies that the only floor mechanism that is dually-robust is one that is equivalent to the following quota mechanism $\phi' \in \Phi^*$ such that $\phi'(s) = 0$ a.s. for all $s \neq \bar{\ell}$, and $\phi(\bar{\ell}) = -\infty$ a.s. Hence, we will focus on mechanisms for which $\phi(s) > -\infty$ a.s. for every $\mathcal{G}$-stopping level s.

We now show every robustly optimal mechanism must have hard limit at $\overline{\ell}$. We begin with a useful lemma.

Lemma 4 (Necessity of limit)**.** *Fix a dually-robust mechanism $\phi \in \Phi^*$. Let s be a stopping level with respect to the filtration $\mathcal{G}$. Then, $\mathbb{E}[\phi_s | \mathcal{G}_{\overline{\ell}}] = +\infty$ under the event $\{\omega : s > \overline{\ell}\}$ almost surely.*

Proof. Suppose towards a contradiction that ϕ is dually-robust but there exists some $\mathcal{G}$-stopping level s such that

$$\mathbb{P}\left(\left\{\mathbb{E}[\phi_s \mid \mathcal{G}_{\overline{\ell}}] < +\infty\right\} \cap \left\{s > \overline{\ell}\right\}\right) > 0.$$

Since $s > \overline{\ell}$, we must have

$$\mathbb{E}\Big[V(0,\overline{\ell}) - V(0,s)\Big|\mathcal{G}_{\overline{\ell}}\Big] > 0$$

almost surely under the event $\{\omega : s > \overline{\ell}\}$. Since $\phi(\overline{\ell}) > -\infty$ and $\mathbb{P}(\theta = 1|\mathcal{G}_{\overline{\ell}}) > 0$ almost surely, there exist $\overline{M}, \underline{M} > 0$ and $\underline{\mu}, \delta > 0$ such that

$$\mathbb{P}\left(\underbrace{\left\{\mathbb{E}[\phi_s \mid \mathcal{G}_{\overline{\ell}}] < \overline{M}\right\} \cap \left\{\phi(\overline{\ell}) > -\underline{M}\right\} \cap \left\{\mu_{\overline{\ell}} \geq \underline{\mu}\right\} \cap \left\{\mathbb{E}[V(0,\overline{\ell}) - V(0,s)|\mathcal{G}_{\overline{\ell}}] > \delta\right\} \cap \left\{s > \overline{\ell}\right\}}_{=:\Omega^*}\right) > 0.$$

Note that Ω^* is measurable with respect to $\mathcal{G}_{\ell^*}$.

We will construct the agent's preferences U such that

$$\mathbb{E}\Big[U(\mu_s, s) - \phi_s\Big|\mathcal{G}_{\overline{\ell}}\Big] > \mathbb{E}\Big[U(\mu_{\overline{\ell}}, \overline{\ell}) - \phi_{\overline{\ell}}\Big|\mathcal{G}_{\overline{\ell}}\Big]$$

under the event Ω^*. Choose

$$U(1,l) = \alpha\beta \cdot V(1,l) \quad \text{and} \quad U(0,l) = \beta \cdot V(0,l)$$

where α and β are constants to be determined later and $\alpha > 1$.

Now observe we can lower-bound the difference (sans mechanism) in the agent's payoff between stopping levels s and $\overline{\ell}$:

$$\begin{aligned}
&\mathbb{E}\Big[U(\mu_s, s) - U(\mu_{\overline{\ell}}, \overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] \\
&\quad = \mathbb{E}\Big[U(\mu_{\overline{\ell}}, s) - U(\mu_{\overline{\ell}}, \overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] && \text{(Optional stopping theorem)} \\
&\quad = \mu_{\overline{\ell}} \cdot \mathbb{E}\Big[U(1,s) - U(1,\overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] + (1 - \mu_{\overline{\ell}}) \cdot \mathbb{E}\Big[U(0,s) - U(0,\overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] \\
&\quad \geq \underline{\mu} \cdot \alpha \cdot \beta \cdot \mathbb{E}\Big[V(1,s) - V(1,\overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] + (1 - \underline{\mu}) \cdot \beta \cdot \mathbb{E}\Big[V(0,s) - V(0,\overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] \\
&\quad = \beta \cdot \mathbb{E}\Big[\underline{\mu}\alpha(V(1,s) - V(1,\overline{\ell})) + (1 - \underline{\mu}) \cdot \Big(V(0,s) - V(0,\overline{\ell})\Big)\Big|\mathcal{G}_{\overline{\ell}}\Big]
\end{aligned}$$

Since V is continuously differentiable, we have, for every $l_0 \neq l_1 \in \mathcal{L}$,

$$\left|\frac{V(0,l_0)-V(0,l_1)}{V(1,l_0)-V(1,l_1)}\right| \leq \sup_l \frac{\partial V(0,l)/\partial l}{\partial V(1,l)/\partial l} =: M.$$

Note that M is also well-defined when technology space is discrete because there are finite values of $l_0 \neq l_1$. Now choose

$$\alpha = \max\left\{\frac{2M(1-\underline{\mu})}{\underline{\mu}}, 1\right\} \quad \text{and} \quad \beta = \frac{\overline{M}+\underline{M}}{(1-\underline{\mu})\delta}.$$

Then, our above lower-bound simplifies further to:

$$\begin{aligned}\mathbb{E}\Big[U(\mu_s,s)-U(\mu_{\overline{\ell}},\overline{\ell})\Big|\mathcal{G}_{\overline{\ell}}\Big] &\geq \beta(1-\underline{\mu})\cdot \underbrace{\mathbb{E}\Big[V(0,\overline{\ell})-V(0,s)\Big|\mathcal{G}_{\overline{\ell}}\Big]}_{\geq\delta} \\ &\geq \mathbb{E}\Big[\phi_s\Big|\mathcal{G}_{\overline{\ell}}\Big]-\phi_{\overline{\ell}},\end{aligned}$$

as desired.

Now consider the agent's learning process $\mathcal{F}=\mathcal{G}$ (no extra information) and the preference U we constructed above. Under the adaptive sandbox ϕ^*, the agent's optimal stopping level is $\overline{\ell}$ from our comparative static Lemma 1. Under the mechanism ϕ, however, $\overline{\ell}$ is no longer an agent's optimal stopping because stopping at level $s>\overline{\ell}$ is strictly better than $\overline{\ell}$ under the event Ω^*.[53] But since $\overline{\ell}$ was defined as the latest optimal stopping level that solved the principal's direct experimentation problem

$$\sup_{\ell}\mathbb{E}[V(\mu_\ell,\ell)] \quad \text{s.t. } \ell \text{ is } \mathcal{G}\text{-adapted}$$

and $\mathbb{P}(\Omega^*)>0$, by total probability the principal's payoff is strictly lower than its payoff guarantee so ϕ cannot be dually-robust. □

Lemma 5 shows that the any dually-robust mechanism ϕ must be such that for any $\mathcal{G}$-adapted stopping level ℓ for which $\ell>\bar{\ell}$ with positive probability, $\mathbb{P}(\phi_\ell=+\infty)>0$. We now argue that this implies the same for any $\mathcal{F}$-adapted level. To see this, choose

$$\hat{\ell} := \inf\Big\{l>\bar{\ell}: \phi_l<+\infty\Big\}\wedge L$$

[53] Choosing to stop at time s under the event Ω^* and stop at $\overline{\ell}$ otherwise is still a valid stopping level because Ω^* is $\mathcal{G}_{\overline{\ell}}$-measurable.

and notice that this is $\mathcal{G}$-adapted. Now for any $\mathcal{F}$-adapted stopping level ℓ, we claim

$$\mathbb{P}(\phi_\ell = +\infty) \geq \mathbb{P}(\phi_{\hat{\ell}} = +\infty)$$

because for each path of $\phi(\omega)$, if there exists some level $s > \bar{\ell}$ such that $\phi(\omega)_s < +\infty$ then $\phi(\omega)_{\hat{\ell}} < +\infty$ and if not, then $\phi(\omega)$ must be equal to infinity everywhere so $\phi(\omega)_\ell = \phi(\omega)_{\hat{\ell}} = +\infty$. But since $\hat{\ell}$ is $\mathcal{G}$-adapted, $\mathbb{P}(\phi_{\hat{\ell}} = +\infty) > 0$ hence $\mathbb{P}(\phi_\ell = +\infty) > 0$. Hence, for any learning process $\mathcal{F}$, the agent never stops after $\bar{\ell}$ as formalized by the following corollary:

Corollary 1. *For every $\phi \in \Phi^*$ and the agent's preference and filtration $(U, \mathcal{F})$, we must have $\ell^*(\phi, \mathcal{F}, U) \leq \bar{\ell}$ almost surely.*

Now we prove the transfer before the hard limit must be weakly higher that that at the hard limit for every dually robust mechanism.

Lemma 5 (Necessity of option)**.** *For every $\phi \in \Phi^*$, we must have $\phi_\ell \geq \phi_{\bar{\ell}}$ for any $\mathcal{G}$-stopping level $\ell \leq \bar{\ell}$ almost surely.*

Proof of Lemma 5. We proceed in four steps.

Step 1: Construct agent's learning process $\mathcal{F}^\epsilon$. Suppose towards a contradiction there exists a $\mathcal{G}$-stopping level ℓ_0 such that the event

$$A^* := \{\omega \in \Omega : \phi_{\ell_0} < \phi_{\bar{\ell}}\}$$

in which the agent faces a strictly positive subsidy between ℓ_0 and $\bar{\ell}$ happens with positive probability. Define $A_1^* := A \cap \{\omega : \theta = 1\}$. This event must happen with strictly positive probability since $\mathbb{P}(\theta = 1 | \mathcal{G}_L) \in (0, 1)$ almost surely.

Let ν be Bernoulli random variable such that $\mathbb{P}(\nu = 1) = \epsilon$ which is is independent of the principal's learning $\mathcal{G}_L$. Define the random variable

$$Z^\epsilon := 1\{\omega \in A_1^*, \nu = 1\}.$$

Now construct the agent's filtration $(\mathcal{F}_l^\epsilon)_l$ as follows:

$$\mathcal{F}_l^\epsilon := \sigma\left(\underbrace{\left\{E \cap \{\omega : l < \ell_0\} : E \in \mathcal{G}_l\right\}}_{\text{Same info before } \ell_0} \bigcup \underbrace{\left\{E \cap \{\omega : l \geq \ell_0, Z^\epsilon = z\} : E \in \mathcal{G}_l, z \in \{0, 1\}\right\}}_{\text{Observe } Z^\epsilon \text{ after } \ell_0}\right).$$

In words, our construction does the following:

- Before level ℓ_0, the agent learns the same information as the principal.
- At and after level ℓ_0, the agent observes the value of Z^ϵ:
    1. Under event $(A_1^*)^c$, the agent still receives the same information as the principal does.
    2. Under event A_1^*, the agent learns the current transfer is strictly less than the transfer at the limit $\overline{\ell}$ but the technology state is good ($\theta = 1$) with probability ϵ.

In particular, for every stopping level ℓ, if $\ell < \ell_0$, then $\mathcal{F}_\ell^\epsilon = \mathcal{G}_\ell$. On the other hand, if $\ell \geq \ell_0$, then $\mathcal{F}_\ell^{\ell^*} = \mathcal{G}_\ell \vee \sigma(Z^\epsilon)$. The intuition why this filtration makes the principal worse off is the agent is disincentivized to continue pushing the technology under event A_1^* even though the technology state is good because she knows that she will be taxed more in the future.

Step 2: Under $\mathcal{F}^\epsilon$, the agent stops after ℓ_0. Let $s_U^\epsilon := \ell^*(\phi, \mathcal{F}_l^\epsilon, U)$ be the largest stopping level that solves the agent's optimal stopping problem under preference U and filtration $\mathcal{F}^\epsilon$. We want to show that $s_U^\epsilon \geq l_0$ almost surely. To do so, we will construct a few auxillary stopping levels and 'thread' s_U^ϵ through them.

Take *any* stopping level $\tilde{s}$ solves the agent's optimal stopping problem under U and filtration $\mathcal{F}_l^\epsilon$. Define

$$s' := \begin{cases} \overline{\ell} & \text{if } \tilde{s} \geq \ell_0, \\ \tilde{s} & \text{if } \tilde{s} < \ell_0, \end{cases} \qquad s_1' := \begin{cases} \tilde{s} & \text{if } \tilde{s} \geq \ell_0, \\ \overline{\ell} & \text{if } \tilde{s} < \ell_0. \end{cases}$$

Note that s_1' is adapted $\mathcal{F}^\epsilon$, and s' is adapted to $\mathcal{G}$. Consider that

$$0 \leq \mathbb{E}[U^\phi(\theta, \tilde{s})] - \mathbb{E}[U^\phi(\theta, s_1')] = \mathbb{E}[U^\phi(\theta, s')] - \mathbb{E}[U^\phi(\theta, \overline{\ell})],$$

where the first inequality follows from the optimality of $\tilde{s}$. This implies $\mathbb{E}[U^\phi(\theta, \overline{\ell})] \leq \mathbb{E}[U^\phi(\theta, s')]$.

Next we will once again make the familiar observation that if $\phi \in \Phi^*$ then the agent's latest optimal stopping stopping level $\ell^*(\phi, \mathcal{G}, U)$ under no extra information ($\mathcal{G} = \mathcal{F}$) must be equal to $\overline{\ell}$ almost surely. To see this, notice that dual robustness of ϕ demands $\mathbb{E}[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{G}, U))] \geq \mathbb{E}[V(\mu_{\overline{\ell}}, \overline{\ell})]$. However, but since $\ell^*(\phi, \mathcal{G}, U)$ is $\mathcal{G}$-adapted so we must have $\ell^*(\phi, \mathcal{G}, U) = \overline{\ell}$ almost surely. That is, the agent must optimally stop at the

principal's optimal stopping level under every robustly optimal mechanism with no extra information.

Since s' is also adapted to $\mathcal{G}$, we must have $\mathbb{E}[U^\phi(\theta,\overline{\ell})] = \mathbb{E}[U^\phi(\theta,s')]$, implying the above inequality must be an equality and so

$$\mathbb{E}[U^\phi(\theta,\tilde{s})] = \mathbb{E}[U^\phi(\theta,s_1')]$$

i.e., s_1' is also an optimal stopping level under U and filtration $\mathcal{F}_l^\epsilon$. This implies $s_U^\epsilon \geq s_1'$ almost surely since s_U^ϵ was defined as the largest optimal stopping level. But since $s_1' \geq \ell_0$, we must have $s_U^\epsilon \geq \ell_0$ almost surely, as desired.

Step 3: Decompose s_U^ϵ. Since we have showed that $s_U^\epsilon \geq \ell_0$ a.s., it must solve the following optimal stopping problem.

$$\sup_{s\geq\ell_0} \mathbb{E}[U(\theta,s) - \phi_s] \quad \text{s.t. } s \text{ is adapted to the filtration } \mathcal{F}_l^\epsilon.$$

We can decompose s_U^ϵ as follows:

$$s_U^\epsilon := \begin{cases} s^1 & \text{if } Z^\epsilon = 1 \\ s^\epsilon & \text{if } Z^\epsilon = 0 \end{cases}$$

almost surely, where s^1 is a $\mathcal{F}^\epsilon$-stopping level that solves

$$\sup_{s\geq\ell_0} \mathbb{E}\Big[U^\phi(\theta,s)\cdot 1\{\omega\in A_1^*\}\Big] \quad \text{s.t. } s \text{ is adapted to } \mathcal{G}_l \tag{OSP-1}$$

and s^ϵ is a $\mathcal{F}^\epsilon$-stopping level that solves

$$\sup_{s\geq\ell_0} \mathbb{E}\Big[U^\phi(\theta,s)\cdot 1\{Z^\epsilon = 0\}\Big] \quad \text{s.t. } s \text{ is adapted to } \mathcal{G}_l. \tag{OSP-ϵ}$$

To see this, observe that

$$\begin{aligned}\mathbb{E}[U^\phi(\theta,s)] &= \mathbb{E}[U^\phi(\theta,s)|Z^\epsilon=1]\mathbb{P}(Z^\epsilon=1) + \mathbb{E}[U^\phi(\theta,s)1\{Z^\epsilon=0\}] \\ &= \mathbb{E}[U^\phi(\theta,s)\cdot 1\{\omega\in A_1^*\}]\epsilon + \mathbb{E}[U^\phi(\theta,s)\cdot 1\{Z^\epsilon=0\}].\end{aligned}$$

This implies, $s^\epsilon|Z^\epsilon=1$ and $s^\epsilon|Z^\epsilon=0$ must solve (OSP-1) and (OSP-ϵ), respectively.

Step 4: The principal's payoff is strictly below dual-robust payoff guarantee.

Notice that the principal's payoff when the agent's preference is U and her filtration

is $\mathcal{F}^\epsilon$ can be written as:

$$\begin{aligned}\mathbb{E}[V(\theta, s_U^\epsilon)] &= \mathbb{E}[V(\theta,s)1\{\omega \in A_1^*\}]\epsilon + \mathbb{E}[V(\theta,s)1\{Z^\epsilon = 0\}] \\ &\leq \underbrace{\mathbb{E}[V(\theta,s^1)1\{\omega \in A_1^*\}]\epsilon + \Phi(\epsilon)}_{=:\Psi(\epsilon)},\end{aligned}$$

where

$$\Phi(\epsilon) := \sup_{s \geq \ell_0} \mathbb{E}\Big[V(\theta,s)\cdot 1\{Z^\epsilon = 0\}\Big] \quad \text{s.t. } s \text{ is } \mathcal{G}\text{-adapted.}$$

Our goal is to show that $\Psi'(0+) < 0$ for some choice of the agent's utility function. We start by investigating the term $\mathbb{E}[V(\theta,s^1)1\{\omega \in A_1^*\}]$. We choose the agent's utility function as $U = \epsilon V$ such that

$$\epsilon < \frac{\mathbb{E}\Big[(\phi_{\overline{\ell}} - \phi_{\ell_0})1\{\omega \in A_1^*\}\Big]}{\mathbb{E}\Big[(V(1,\overline{\ell}) - V(1,\ell_0))1\{\omega \in A_1^*\}\Big]}.$$

This is well-defined because $\mathbb{P}(A_1^*) > 0$ so that both of the numerator and denominator are strictly positive. Consider that

$$\begin{aligned}\mathbb{E}[(U(1,\ell_0) - \phi(\ell_0))1\{\omega \in A_1^*\}] &= \mathbb{E}[(\epsilon V(1,\ell_0) - \phi_{\ell_0})1\{\omega \in A_1^*\}] \\ &> \mathbb{E}[(\epsilon V(1,\overline{\ell}) - \phi_{\overline{\ell}})1\{\omega \in A_1^*\}].\end{aligned}$$

This implies choosing $\overline{\ell}$ is suboptimal under the event A_1^*, so $s^1 \neq \overline{\ell}$ positive probability. Lemma 4 implies $s^1 \leq \overline{\ell}$, where the inequality is strict with positive probability. Since $V(1,\cdot)$ is strictly increasing, we must have

$$\begin{aligned}\mathbb{E}[V(\theta,s^1)1\{\omega \in A_1^*\}] &= \mathbb{E}[V(1,s^1)1\{\omega \in A_1^*\}] \\ &< \mathbb{E}[V(1,\overline{\ell})1\{\omega \in A_1^*\}] \\ &= \mathbb{E}[V(\theta,\overline{\ell})1\{\omega \in A_1^*\}].\end{aligned}$$

Next, we compute $\Phi'(0+)$. Consider that $\mathbb{E}[V(\theta,s)1\{Z^\epsilon = 0\}] = \mathbb{E}[X_s^\epsilon]$, where $X_l^\epsilon := \mathbb{E}[V(\theta,t)1\{Z_0^\epsilon = 0\}|\mathcal{G}_l]$. This means X_s^ϵ is $\mathcal{G}$-adapted. Thus, the Snell envelope is well-defined, and the maximum is attainable. For every stopping level s adapted to $\mathcal{G}_l$ and a nonnegative real number $\epsilon > 0$, define $f(\epsilon,s) := \mathbb{E}[V(\theta,s)1\{Z^\epsilon = 0\}]$. Thus, $\Phi(\epsilon) = \sup_{s\geq\ell_0}\mathbb{E}[f(\epsilon,s)]$ subject to s is adapted to $\mathcal{G}_l$.

We know that $\Phi(0)$ is uniquely attained by $s = \overline{\ell}$ because $\mathbb{P}(Z^0 = 0) = 0$. Moreover, we have the following lemma whose proof is technical and relegated to Online Appendix I.

Lemma 6. *For any decreasing sequence $(\epsilon_n)_n$ such that $\epsilon_n \downarrow 0$, define s^{ϵ_n} as the latest stopping time that solves (OSP-ϵ_n) for every n. Then, $s^{\epsilon_n} \uparrow \overline{\ell}$ as $n \to \infty$.*

Now consider that

$$\begin{aligned} f(\epsilon, s) &= \mathbb{E}[V(\theta, s)] - \mathbb{E}[V(\theta, s) \cdot Z^\epsilon] \\ &= \mathbb{E}[V(\theta, s)] - \epsilon V[(\theta, s) 1\{\omega \in A_1^*\}]. \end{aligned}$$

Thus, $\frac{\partial f}{\partial \epsilon} = -\mathbb{E}[V(\theta, s) 1\{\omega \in A_1^*\}]$, which is constant in ϵ. From Theorem 3 of Milgrom and Segal (2002) since Φ is right-differentiable at 0 and

$$\begin{aligned} \Phi'(0+) &= \lim_{\epsilon \to 0^+} \frac{\partial f}{\partial \epsilon}(s^\epsilon, 0) \\ &= -\lim_{n \to \infty} \mathbb{E}\Big[V(\theta, s^{\epsilon_n}) \cdot 1\{\omega \in A_1^*\} \Big]. \end{aligned}$$

Since $s^{\epsilon_n} \uparrow \overline{\ell}$ as $n \to \infty$ we have

$$\Phi'(0+) = -\lim_{n \to \infty} \mathbb{E}[V(\theta, s^{\epsilon_n}) 1\{\omega \in A_1^*\}] = -\mathbb{E}[V(\theta, \overline{\ell}) 1\{\omega \in A_1^*\}],$$

by the dominated convergence theorem. This implies

$$\Psi'(0+) = \mathbb{E}[V(\theta, s^1) 1\{\omega \in A_1^*\}] - \mathbb{E}[V(\theta, \overline{\ell}) 1\{\omega \in A_1^*\}] < 0.$$

Thus, there exists $\epsilon > 0$ such that $\Psi(\epsilon) < \Psi(0)$, which means ϕ is not robustly optimal, as desired. □

Lemma 4 and Lemma 5 imply the forward direction of Proposition 1 i.e., necessacity. The next lemma establishes sufficiency.

Lemma 7. *If ϕ satisfies two conditions stated in Proposition 1, then ϕ is dually-robust.*

Proof of Lemma 7. Given any agent's filtration $(\mathcal{F}_l)_l$ and agent's utility function U, let ℓ_U^* be the agent's corresponding optimal stopping level facing ϕ. This implies $\ell_U^* \leq \overline{\ell}$. Moreover, the optimality of ℓ_U^* implies

$$U^\phi(\mu_{\ell_U^*}, \ell_U^*) \geq \mathbb{E}\Big[U^\phi(\mu_{\overline{\ell}}, \overline{\ell}) \Big| \mathcal{F}_{\ell_U^*} \Big] = \mathbb{E}\Big[U^\phi(\mu_{\ell_U^*}, \overline{\ell}) \Big| \mathcal{F}_{\ell_U^*} \Big] \text{ a.s.}$$

$$\implies U(\mu_{\ell_U^*}, \ell_U^*) - \mathbb{E}\Big[U(\mu_{\ell_U^*}, \bar{\ell})\Big|\mathcal{F}_{\ell_U^*}\Big] \geq \phi_{\ell_U^*} - \mathbb{E}\Big[\phi_{\bar{\ell}}\Big|\mathcal{F}_{\ell_U^*}\Big] \geq 0 \text{ a.s.},$$

where the last inequality is because ϕ offers the option of continuing under $\bar{\ell}$ and enjoying a weakly positive subsidy.

This implies $U(\mu_{\ell_U^*}, \ell_U^*) \geq \mathbb{E}\Big[U(\mu_{\ell_U^*}, \bar{\ell})\Big|\mathcal{F}_{\ell_U^*}\Big]$ almost surely. Since $\bar{\ell} \geq \ell_U^*$ almost surely and $U = g \circ V$ for some convex function g, Lemma 2 implies[54]

$$V(\mu_{\ell_U^*}, \ell_U^*) \geq \mathbb{E}\Big[V(\mu_{\ell_U^*}, \bar{\ell})\Big|\mathcal{F}_{\ell_U^*}\Big] \text{a.s.}$$

Thus,

$$\begin{aligned}
\mathbb{E}[V(\mu_{\ell_U^*}, \ell_U^*)] &\geq \mathbb{E}\Big[\mathbb{E}[V(\mu_{\ell_U^*}, \bar{\ell})|\mathcal{F}_{\ell_U^*}]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[V(\mu_{\bar{\ell}}, \bar{\ell})|\mathcal{F}_{\ell_U^*}]\Big] \\
&= \mathbb{E}[V(\mu_{\bar{\ell}}, \bar{\ell})] \\
&= \mathbb{E}[V(\mu_0, \bar{\ell})]
\end{aligned}$$

where the first and last equalities are implied by the optional stopping theorem, as desired. □

A.5 Proof of Theorem 2. We start with an important lemma stating that any mechanism whose transfer is decreasing until the hard limit is dominated by the sandbox mechanism.

Lemma 8. *Every mechanism $\phi \in \Phi^*$ such that $\phi(\ell_1) \geq \phi(\ell_2)$ for every $\mathcal{G}$-stopping level $\ell_1 \leq \ell_2 \leq \bar{\ell}$ almost surely is dominated by ϕ^*.*[55]

Proof. Fix any $U \in \mathcal{U}$ and any agent's filtration $\mathcal{F}$. We will construct the simplest coupling of the agent's stopping behavior under ϕ^* and under ϕ: let the realized learning process be the same (that is, the processes are indistinguishable). Suppose the agent's optimal stopping level under ϕ^* is $\ell_{*,U}$. Then, $\ell_{*,U}$ must solves the restricted stopping problem:

$$\sup_{s \leq \bar{\ell}} \mathbb{E}[U(\theta, s)] \quad \text{where } s \text{ is } (\mathcal{F}_l)_l\text{-adapted stopping level.}$$

[54] Note here that ℓ_U^* need not be an optimal stopping level for the agent with preference U facing the empty mechanism so Lemma 1 does not apply. But Lemma 2 does since it compares preferences over two stopping times.

[55] This is also true under an adversarial selection of optimal stopping level. See Online Appendix VI.2

We will show that the agent's latest optimal stopping level ℓ^*_ϕ under $(\phi, \mathcal{F})$ must satisfy $\ell^*_\phi \geq \ell_{*,U}$ almost surely. Define $\ell_1 = \ell^*_\phi \vee \ell_{*,U}$ and $\ell_0 = \ell^*_\phi \wedge \ell_{*,U}$. The optimality of ℓ^*_ϕ implies

$$\begin{aligned}
0 &\leq \mathbb{E}[U^\phi(\theta, \ell^*_\phi)] - \mathbb{E}[U^\phi(\theta, \ell_1)] \\
&= \mathbb{E}[(\phi_{\ell_{*,U}} - \phi_{\ell^*_\phi}) 1\{\ell_{*,U} \geq \ell^*_\phi\}] + \mathbb{E}[(U(\theta, \ell^*_\phi) - U(\theta, \ell_{*,U})) 1\{\ell_{*,U} \geq \ell_\phi\}] \\
&\leq \mathbb{E}[(U(\theta, \ell^*_\phi) - U(\theta, \ell_{*,U}) 1\{\ell_{*,U} \geq \ell^*_\phi\}] \qquad (\phi \text{ is decreasing up until } \overline{\ell})
\end{aligned}$$

The optimality of $\ell_{*,U}$ implies

$$0 \leq \mathbb{E}[U(\theta, \ell_{*,U})] - \mathbb{E}[U(\theta, \ell_0)] = \mathbb{E}[(U(\theta, \ell_{*,U}) - U(\theta, \ell^*_\phi)) 1\{\ell_{*,U} \geq \ell^*_\phi\}]$$

Thus, two above inequalities must be equalities, so ℓ_1 must be the agent's optimal stopping level under $(\phi, \mathcal{F})$ as well. The definition of ℓ^*_ϕ implies $\ell^*_\phi \geq \ell_1$, meaning $\ell^*_\phi \geq \ell_{*,U}$, as desired.

Now applying Lemma 1 to the restricted (random) domain $[0, \overline{\ell}]$ we have the solution to the fictional problem in which the principal is in control of the stopping level from time $\ell_{*,U}$ onward

$$\sup_{\ell_{*,U} \leq s \leq \overline{\ell}} \mathbb{E}\Big[V(\mu_s, s) \Big| \mathcal{F}_{\ell_{*,U}} \Big] \quad \text{s.t. } s \text{ is } (\mathcal{F}_l)_l\text{-adapted}$$

is simply $s = \ell_{*,U}$ i.e., to stop immediately.

Our argument thus far has been path-wise. Now letting $F_{\phi^*} \in \Delta([0, L])$ be the distribution of the agent's stopping level under ϕ^*, we have:

$$\begin{aligned}
\mathbb{E}\Big[V(\boldsymbol{\mu}, \phi^*, U) \Big] &= \int \mathbb{E}\Big[V(\mu_s, s) \Big| \ell_{*,U} = s \Big] dF_{\phi^*} && \text{(Total expectation)} \\
&= \int \left\{ \sup_{\ell_{*,U} \leq \ell \leq \overline{\ell}} \mathbb{E}\Big[V(\mu_\ell, \ell) \Big| \ell_{*,U} = s \Big] \right\} dF_{\phi^*} \\
& && \text{(Stopping at } \ell_{*,U} \text{ is optimal for principal)} \\
&\geq \int_{s \leq \overline{\ell}} \mathbb{E}\Big[V(\mu_{\ell^*_\phi}, \ell^*_\phi) \Big| \ell_{*,U} = s \Big] dF_{\phi^*} && (\ell^*_\phi \geq \ell_{*,U} \text{ almost surely}) \\
&= \mathbb{E}\Big[V(\boldsymbol{\mu}, \phi, U) \Big]
\end{aligned}$$

where the first equality is from the law of total probability, the first inequality is

because we showed the solution to the principal's restricted stopping problem is to stop at $\ell_{*,U}$ and furthermore conditioning on just $\{\ell_{*,U} = s\}$ is a coarser event, and the last inequality is because $\ell^*_\phi \geq \ell_{*,U}$ almost surely. This establishes that ϕ^* performs weakly better under any learning process and any agent preference. □

We are now ready to prove Theorem 2.

Proof of Theorem 2. We show each part of Theorem 2 in turn.

Step 1: ϕ^* dominates other dually-robust and regular mechanisms.

Here we focus on the case that the principal's learning process $\mathcal{G}$ and preferences V (both primitives of our environment) are such that $\overline{\ell} < L$ almost surely. The other case is more technical and is handled in Online Appendix I.

Consider any regular and robustly optimal mechanism ϕ. For every $\mathcal{G}$-stopping level $\ell_1 \leq \ell_2 \leq \overline{\ell}$, we will show that $\phi_{\ell_1} > \phi_{\ell_2}$. We know that $\phi_{\ell_1}, \phi_{\ell_2} \geq \phi_{\overline{\ell}}$ almost surely by Proposition 1. Since $\overline{\ell} < L$, we must have $\phi_L = \infty$. This implies $\phi_{\ell_1}, \phi_{\ell_2} \geq \phi_{\overline{\ell}} < \phi_L$. Since ϕ is regular, $\phi_{\ell_1} \geq \phi_{\ell_2}$ a.s. i.e., ϕ is decreasing almost surely up until $\overline{\ell}$. Then, Lemma 8 implies ϕ is dominated by ϕ^*. This delivers the second half of Theorem 2.

Step 2: ϕ^* is undominated. Suppose towards a contradiction that a mechanism ϕ dominates the sandbox mechanism. Then, the principal's worst case under ϕ must be weakly greater than that under ϕ^*. However, since $\phi^* \in \Phi^*$, we must have $\phi \in \Phi^*$. We will show that $\phi_{\ell^*} = \phi_{\overline{\ell}}$ almost surely for every $\mathcal{G}$-adapted stopping level ℓ^* such that $\ell^* \leq \overline{\ell}$.

Define the following random variable:

$$\varphi^{\ell^*} := \begin{cases} 1 & \text{if } \phi_{\ell^*} > \phi_{\overline{\ell}} \\ 0 & \text{if } \phi_{\ell^*} = \phi_{\overline{\ell}} \\ -1 & \text{if } \phi_{\ell^*} < \phi_{\overline{\ell}} \end{cases}$$

which encodes information about whether the tax between ℓ^* and $\bar{\ell}$ is positive or negative.

We choose the agent's filtration $\mathcal{F}_l^{\ell^*}$ as follows:

$$\mathcal{F}_l^{\ell^*} := \sigma\Big(\Big\{A \cap \{\omega : l < \ell^*\} : A \in \mathcal{G}_l\Big\} \\ \cup \Big\{A \cap \{\omega : l \geq \ell^*, \theta = \theta_0, \varphi^{\ell^*} = \varphi\} : A \in \mathcal{G}_l, \theta_0 \in \Theta, \varphi \in \{-1, 0, 1\}\Big\}\Big).$$

In words, we are constructing a filtration under which, at the fixed stopping level $\ell^* \leq \overline{\ell}$, the agent (i) learns the true state θ; and (ii) whether the transfer at the hard limit is higher or lower than the current transfer i.e., she learns about what the principal will learn in the future. In particular, for every stopping level ℓ, if $\ell < \ell^*$, then $\mathcal{F}_\ell^{\ell^*} = \mathcal{G}_\ell$. On the other hand, if $\ell \geq \ell^*$, then $\mathcal{F}_\ell^{\ell^*} = \mathcal{G}_\ell \vee \sigma(\theta, \varphi^{\ell^*})$.

We now construct the agent's preferences by setting $U = \epsilon \cdot V$ for some small $\epsilon > 0$ that will be chosen later. Under our adaptive sandbox mechanism, since the agent's utility is a linear transformation of the principal's, the principal's expected utility under mechanism ϕ is bounded above by the value of the following optimal stopping problem:

$$\sup_{s \leq \overline{\ell}} \mathbb{E}\Big[V(\theta, s)\Big] \quad \text{s.t. } s \text{ is adapted to } (\mathcal{F}_l^{\ell^*})_l$$

Now suppose that s^* is optimal stopping under the sandbox mechanism and the agent's filtration $(\mathcal{F}_l^{\ell^*})_l$. We know that $s^* \leq \overline{\ell}$ since continuing past the limit $\bar{\ell}$ is always dominated by stopping. This implies s^* must solve the above optimal stopping problem. We claim

$$s^* = \begin{cases} \ell^* & \text{if } \theta = 0 \\ \overline{\ell} & \text{otherwise} \end{cases} \quad \text{almost surely.}$$

To see this, we will show it is always suboptimal to stop before ℓ^* with positive probability. Define

$$s' := \begin{cases} \overline{\ell} & \text{if } s^* < \ell^* \\ s^* & \text{if } s^* \geq \ell^* \end{cases} \qquad s_1' := \begin{cases} s^* & \text{if } s^* < \ell^* \\ \overline{\ell} & \text{if } s^* \geq \ell^* \end{cases}$$

Note that both s' and s_1' are adapted to the agent's filtration $\mathcal{F}^{\ell^*}$, and that s_1' is in addition adapted to $\mathcal{G}$. Then,

$$\begin{aligned} \mathbb{E}[V(\theta, s')] - \mathbb{E}[V(\theta, s^*)] &= \mathbb{E}[(V(\theta, \overline{\ell}) - V(\theta, s^*))1\{s^* < \ell^*\}] \\ &= \mathbb{E}[V(\theta, \overline{\ell})] - \mathbb{E}[V(\theta, s_1')] \\ &> 0, \end{aligned}$$

where the last inequality follows from the fact that $\overline{\ell}$ is the unique optimal stopping of the principal under the filtration $\mathcal{G}$, and s_1' is adapted to the filtration $\mathcal{G}$. This

contradicts the optimality of s', as desired. Thus, $s^* \geq \ell^*$. If $\theta = 0$, the principal wants to stop immediately at ℓ^*. If $\theta = 1$, the principal wishes to continue until $\overline{\ell}$. Thus, s^* does, in fact, solve the principal's stopping problem and the principal's payoff under the adaptive sandbox ϕ^* against the agent's learning process $(\mathcal{F}_l^{\ell^*})_l$ and preference $U = \epsilon \cdot V$ hits the upper-bound.

Next, we show that s^* cannot be the agent's optimal stopping level under ϕ and the filtration $\mathcal{F}^{\ell^*}$. Consider the following alternative stopping level

$$s = \begin{cases} \ell^* & \text{if } \theta = 1 \text{ and } \varphi^{\ell^*} = -1 \\ \overline{\ell} & \text{if } \theta = 0 \text{ and } \varphi^{\ell^*} = 1 \\ s^* & \text{otherwise.} \end{cases}$$

In words, the agent stops at ℓ^* if the state is good and the agent has learnt that $\phi_{\ell^*} > \phi_{\bar{\ell}}$ i.e., she will be taxed if she continues until $\bar{\ell}$, and continues until $\bar{\ell}$ if the state is bad and the agent has learnt that she will be subsidized. This is clearly a $\mathcal{F}^{\ell^*}$-stopping level. We show it is optimal for the agent.

The difference of the agent's utility under stopping levels s and s^* is

$$\begin{aligned} &\mathbb{E}[U^\phi(\theta, s)] - \mathbb{E}[U^\phi(\theta, s^*)] \\ &\qquad = \underbrace{\mathbb{E}\Big[(U^\phi(1,\ell^*) - U^\phi(1,\overline{\ell}))1\{\theta = 1, \varphi^{\ell^*} = -1\}\Big]}_{=:A} \\ &\qquad + \underbrace{\mathbb{E}\Big[(U^\phi(0,\overline{\ell}) - U^\phi(0,\ell^*))1\{\theta = 0, \varphi^{\ell^*} = 1\}\Big]}_{=:B}. \end{aligned}$$

Now recalling that $U = \epsilon \cdot V$, we can rewrite A and B as:

$$\begin{aligned} A &= \epsilon \cdot \underbrace{\mathbb{E}[(V(1,\ell^*) - V(1,\overline{\ell}))1\{\theta = 1, \varphi^{\ell^*} = -1\}]}_{\leq 0} + \underbrace{\mathbb{E}[(\phi_{\overline{\ell}} - \phi_{\ell^*})1\{\theta = 1, \varphi^{\ell^*} = -1\}]}_{\geq 0} \\ B &= \epsilon \underbrace{\mathbb{E}[(V(0,\overline{\ell}) - V(1,\ell^*))1\{\theta = 0, \varphi^{\ell^*} = 1\}]}_{\leq 0} + \underbrace{\mathbb{E}[(\phi_{\ell^*} - \phi_{\overline{\ell}})1\{\theta = 0, \varphi^{\ell^*} = 1\}]}_{\geq 0} \end{aligned}$$

We can choose sufficiently small but positive $\epsilon > 0$ so that both A and B are weakly positive and strictly positive whenever $\mathbb{P}(\theta = 1, \varphi^{\ell^*} = -1) > 0$ and $\mathbb{P}(\theta = 0, \varphi^{\ell^*} = 1) > 0$, respectively. Thus, if either $\mathbb{P}(\theta = 1, \varphi^{\ell^*} = -1) > 0$ or $\mathbb{P}(\theta = 0, \varphi^{\ell^*} = 1) > 0$, then the agent strictly prefers stopping level s to s^*, which contradicts the optimality of s^*.

Under ϕ, we must thus have

$$\mathbb{P}\Big(\Big\{\theta = 1, \varphi^{\ell^*} = -1\Big\} \cup \Big\{\theta = 0, \varphi^{\ell^*} = 1\Big\}\Big) = 0 \implies \mathbb{P}\Big(\varphi^{\ell^*} \in \{-1, 1\}\Big) = 0.$$

Hence, $\phi_{\ell^*} = \phi_{\overline{\ell}}$ almost surely, as desired. But since the stopping level $\ell^* \leq \bar{\ell}$ was arbitrary, this implies $\phi_\ell = \phi_{\overline{\ell}} = \phi_0$ for any $\ell \leq \bar{\ell}$. Since ϕ_0 is $\mathcal{G}_0$-measurable, this implies that ϕ is equivalent to the adaptive sandbox mechanism ϕ^*, as desired. □

A.6 Proof of Theorem 3 (Time-consistency). Consider any $\mathcal{G}$-stopping level $\ell_0 \in (0, L)$. We first show the sufficiency direction that ϕ^* is time-consistent starting following from ℓ_0. Then, we show the necessity direction that among dually-robust mechanisms, only ϕ^* is time-consistent.

ϕ^* is time-consistent upon stopping.

We wish to show that ϕ^* is conditionally undominated upon the agent stopping at ℓ_0. Suppose, towards a contradiction, that there exists some other other mechanism ϕ' that conditionally dominates ϕ^*. This implies for every $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ and $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0},

$$\mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A_{\ell_0}, \ell^* = \ell_0\}\Big] \leq \mathbb{E}\Big[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) \cdot 1\{\omega \in A_{\ell_0}, \ell^* = \ell_0\}\Big],$$

with strict inequality for some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ and some $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0}.

For each $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0}, we can partition it as follows:

$$A_{\ell_0} = \underbrace{\Big\{A_{\ell_0} \cap \{\omega : \ell_0 < \overline{\ell}\}\Big\}}_{A^<_{\ell_0}} \cup \underbrace{\Big\{A_{\ell_0} \cap \{\omega : \ell_0 \geq \overline{\ell}\}\Big\}}_{A^\geq_{\ell_0}}.$$

Note that the event $\{\ell_0 = \ell^*\} \cap A^<_{\ell_0}$ is the event that the agent stops strictly before the sandbox limit while $\{\ell_0 = \ell^*\} \cap A^\geq_{\ell_0}$ is the event that the agent stops right at the sandbox limit. We proceed in two cases.

Case A: Under $A^\geq_{\ell_0}$ the ϕ^* is undominated. Consider any filtration $(\mathcal{F}_l)_l$ such that the agent and the principal learn identically before level ℓ_0, i.e., $\mathcal{F}_{\ell_0} = \mathcal{G}_{\ell_0}$, and pick any agent preference $U \in \mathbb{U}$. Now let ℓ^*_V be the principal's optimal stopping level *as if* they observe the agent's information $(\mathcal{F}_l)_l$. An identical argument to Step 2 of the proof of Theorem 2 implies that if $\ell_0 \leq \overline{\ell}$, then $\ell_0 \leq \ell^*_V$, i.e., it is suboptimal for the principal to stop before ℓ_0 under the event $\{\ell_0 \leq \overline{\ell}\}$ because $\mathcal{F}_{\ell_0} = \mathcal{G}_{\ell_0}$. By Lemma 1, we must also have, if $\ell_0 \leq \overline{\ell}$, then $\ell_0 \leq \ell^*_V \leq \ell^*$. Thus, $\{\ell^* = \ell_0\} = \{\overline{\ell} = \ell_0\}$ because $\ell^* \leq \overline{\ell}$.

In words, if the agent stops at ℓ_0, then ℓ_0 must coincide with the limit.

This implies

$$A^{\geq}_{\ell_0} \cap \{\ell_0 = \ell^*\} = A_{\ell_0} \cap \{\ell_0 = \overline{\ell}\} =: A^{=}_{\ell_0},$$

which is $\mathcal{G}_{\ell_0}$-measurable, i.e., the principal does not learn any additional information from the observation that the agent has stoped at ℓ_0. Since ϕ' conditionally dominates ϕ^* upon stopping at ℓ_0, for every $\mathcal{G}_{\ell_0}$-measurable set $A'_{\ell_0} \subset A^{=}_{\ell_0}$,

$$\begin{aligned}
&\mathbb{E}[V(\mu_{\ell^*}, \ell^*) 1\{\omega \in A'_{\ell_0}, \ell^* = \ell_0\}] \leq \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) 1\{\omega \in A'_{\ell_0}, \ell^* = \ell_0\}] \\
\Longrightarrow &\mathbb{E}[V(\mu_{\ell^*}, \ell^*) 1\{\omega \in A'_{\ell_0}\}] \leq \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) 1\{\omega \in A'_{\ell_0}\}] \\
\Longrightarrow &\mathbb{E}[V(\mu_{\ell^*}, \ell^*) | \mathcal{G}_{\ell_0}] \leq \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) | \mathcal{G}_{\ell_0}] \quad \text{a.s. under the event } A^{=}_{\ell_0}
\end{aligned}$$

for every $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ such that $\mathcal{F}_{\ell_0} = \mathcal{G}_{\ell_0}$. In other words, ϕ' dominates ϕ^* in terms of expected payoff under the event $A^{=}_{\ell_0}$ with no restriction of learning process and agent's utility function beyond level ℓ_0. Since ϕ^* remains an adaptive sandbox after level ℓ_0, we can apply the same construction from the proof that ϕ^* is undominated from Step 2 of Theorem 2 otherwise by the same argument from Step 2 of Theorem 2, we can construct some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ such that $\mathcal{F}_{\ell_0} = \mathcal{G}_{\ell_0}$ under which ϕ^* yields strictly higher payoff than ϕ', a contradiction. Hence, choosing the initial level ℓ_0 instead of 0 to show that ϕ' must be equivalent to ϕ^* under the event $A^{=}_{\ell_0}$. Thus,

$$\mathbb{E}[V(\mu_{\ell^*}, \ell^*) \underbrace{1\{\omega \in A^{\geq}_{\ell_0}, \ell^* = \ell_0\}}_{= A^{=}_{\ell_0}}] = \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) 1\{\omega \in A^{\geq}_{\ell_0}, \ell^* = \ell_0\}]. \tag{5}$$

Case B: Under $A^{<}_{\ell_0}$ ϕ^* is undominated. Next, we analyze ϕ' under the event $A^{<}_{\ell_0}$. The following lemma implies ϕ must also have a hard limit at $\overline{\ell}$.

Lemma 9. *For every $\mathcal{G}$-stopping level $\ell_1 > \bar{\ell}$, we must have $\phi'_{\ell_1} = \infty$ almost surely under the event $\{\phi'_{\ell_0} > -\infty\} \cap A^{<}_{\ell_0}$. Moreover, $\phi'_{\ell_1} > -\infty$ almost surely under the event $\{\phi'_{\ell_0} = -\infty\} \cap A^{<}_{\ell_0}$.*

We relegate the proof to Online Appendix I. The proof is similar to Lemma 4, but we need choose $(U, \mathcal{F})$ more carefully so that the event that the agent stops at ℓ_0 happens with strictly positive probability. This implies stopping after $\overline{\ell}$ under the event $A^{<}_{l_0} \cap \{\phi'_{\ell_0} > -\infty\}$ with positive probability is always suboptimal. Moreover, under the event $A^{<}_{\ell_0} \cap \{\phi'_{\ell_0} = -\infty\}$, stopping after $\overline{\ell}$ is suboptimal as well because it gives finite utility, whereas stopping at ℓ_0 gives the positive infinite utility. Then, we

yield the following corollary.

Corollary 2. *$\ell_{\phi'} \le \overline{\ell}$ almost surely under the event $A^{<}_{\ell_0}$.*

Since ϕ' conditionally dominates ϕ^*, under any $(\mathcal{F}, U)$ we obtain:

$$\begin{aligned} &\mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \le \mathbb{E}\Big[V(\mu_{\ell_\phi}, \ell_\phi) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \\ \overset{\text{Lemma 2}}{\Longrightarrow} &\mathbb{E}\Big[U(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \le \mathbb{E}\Big[U(\mu_{\ell_{\phi'}}, \ell_{\phi'}) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \end{aligned} \tag{6}$$

since $A^{<}_{\ell_0}$ is measurable with respect to $\mathcal{G}_{\ell_0}$ and $\ell_{\phi'} \ge \ell^*$, and the last inequality becomes strict when $\mathbb{P}(\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0 < \ell_{\phi'}) > 0$.

However, the agent optimally stops at ℓ^* under the adaptive sandbox mechanism instead of stopping at $\ell_{\phi'}$ under the event $A^{<}_{\ell_0} \cap \{\ell^* = \ell_0\}$, which is $\mathcal{F}_{\ell_0}$-measurable. We must have

$$\begin{aligned} &\mathbb{E}\Big[U^{\phi^*}(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \ge \mathbb{E}\Big[U^{\phi^*}(\mu_{\ell_{\phi'}}, \ell_{\phi'}) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \\ \Longrightarrow &\mathbb{E}\Big[U(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \ge \mathbb{E}\Big[U(\mu_{\ell_{\phi'}}, \ell_{\phi'}) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big] \end{aligned}$$

because $\ell_{\phi'} \le \overline{\ell}$. These together imply

$$\mathbb{E}\Big[U(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}] = \mathbb{E}\Big[U(\mu_{\ell_{\phi'}}, \ell_{\phi'}) \cdot 1\{\omega \in A^{<}_{\ell_0}, \ell^* = \ell_0\}\Big].$$

Then, Equation (6) must become an equality. Hence,

$$\mathbb{P}\Big(\Big\{\omega \in A^{<}_{\ell_0}\Big\} \cap \Big\{\ell^* = \ell_0 < \ell_{\phi'}\Big\}\Big) = 0.$$

Together with Equation (5), we obtain

$$\mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*) \cdot 1\{\omega \in A_{\ell_0}, \ell^* = \ell_0\}\Big] = \mathbb{E}\Big[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) \cdot 1\{\omega \in A_{\ell_0}, \ell^* = \ell_0\}\Big],$$

for every $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0} and every $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$. From Cases A and B, we have that ϕ' and ϕ yield identical payoffs for every $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0} and every $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$, contradicting the hypothesis that ϕ' conditionally dominates ϕ^*, as desired. This yields the first part of Theorem 3.

ϕ^* is time-consistent upon continuing. We wish to show that ϕ^* is conditionally undominated on this history A_{ℓ_0}. Suppose, towards a contradiction, that there exists some other other mechanism ϕ that conditionally dominates ϕ^*. This implies, for

every $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ and $\mathcal{G}_{\ell_0}$-measurable set,

$$\mathbb{E}[V(\mu_{\ell^*}, \ell^*) 1\{\omega \in A_{\ell_0}, \ell^* > \ell_0\}] \leq \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) 1\{\omega \in A_{\ell_0}, \ell^* > \ell_0\}], \tag{7}$$

and the inequality is strict for some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$ and a $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0}. We proceed similarly as we consider ϕ^* under the event $A^{\geq}_{\ell_0}$.

We consider any filtration $(\mathcal{F}_l)_l$ that the agent and the principal learn equally before level ℓ_0, and any preference $U \in \mathbb{U}$. We showed earlier that, if $\ell_0 \leq \bar{\ell}$, then $\ell_0 \leq \ell^*$. Because of that sandbox limit, $\ell^* = \bar{\ell}$ under the event $\{\bar{\ell} \leq \ell_0\}$. Thus, $\{\ell^* > \ell_0\} = \{\bar{\ell} > \ell_0\}$ and the latter is $\mathcal{G}_{\ell_0}$-measurable. Hence, the principal does not learn any extra information by observing that the agent continues beyond ℓ^*. From Equation (7), we obtain

$$\mathbb{E}[V(\mu_{\ell^*}, \ell^*) | \mathcal{G}_{\ell_0}] \leq \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) | \mathcal{G}_{\ell_0}] \quad \text{under the event } \{\ell^* > \ell_0\}.$$

Since ϕ^* remains an adaptive sandbox after level ℓ_0, we can apply the same argument from the proof that ϕ^* is undominated from Theorem 2 to show that ϕ' must coincide with ϕ^* when $\ell \geq \ell_0$ under the event $\{\ell^* > \ell_0\}$. Thus,

$$\mathbb{E}[V(\mu_{\ell^*}, \ell^*) 1\{\omega \in A_{\ell_0}, \ell^* > \ell_0\}] = \mathbb{E}[V(\mu_{\ell_{\phi'}}, \ell_{\phi'}) 1\{\omega \in A_{\ell_0}, \ell^* > \ell_0\}]$$

for every $\mathcal{G}_{\ell_0}$-measurable set A_{ℓ_0}, but this contradicts that the inequality must be strict somewhere, as desired.

All other dually-robust mechanisms are not time-consistent. We now show the converse of Theorem 3. Consider any time-consistent mechanism ϕ under the finite technology space $\mathcal{L} := \{l_0, l_1, \ldots, l_n\}$. We will show via induction backwards in levels that for every $l \in \mathcal{L}$, $\phi_{l \wedge \bar{\ell}}$ must be a constant.

Base step: $\phi_{l_{n-1} \wedge \bar{\ell}} = \phi_{l_n \wedge \bar{\ell}}$ almost surely. Consider the event the agent continues at l_{n-1}. We pick the agent's filtration so that $\mathcal{F}_l = \mathcal{G}_l$ for every $l < n-1$. Since ϕ is a dually-robust mechanism, by Proposition 1 we must have $\phi_{l \wedge \bar{\ell}} \geq \phi_{\bar{\ell}}$. Under the event $\{\bar{\ell} > l_{n-1}\}$, we must have $\phi_{l_{n-1}} \geq \phi_{l_n}$. Thus, condition on the agent continuing until l_{n-1} under the event $\{\bar{\ell} > l_{n-1}\}$, the transfer is weakly decreasing starting from l_{n-1} until $\bar{\ell}$ almost surely i.e., $\phi_{\bar{\ell}} \leq \phi_{l_{n-1}}$.

If it is strictly decreasing, then from Lemma 8, ϕ is dominated by the sandbox mechanism ϕ' constructed such that $\phi'_{\bar{\ell}} = \phi'_{l_{n-1}}$ under the event $\{\bar{\ell} > l_{n-1}\}$ as long as $\phi_{l_{n-1} \wedge \bar{\ell}} > \phi_{l_n \wedge \bar{\ell}}$ with positive probability. Thus, for ϕ to be conditionally undominated at l_{n-1}, $\phi_{l_{n-1} \wedge \bar{\ell}} = \phi_{l_n \wedge \bar{\ell}}$ almost surely.

Inductive step: $\phi_{l_k \wedge \bar{\ell}} = \phi_{l_{k+1} \wedge \bar{\ell}}$ almost surely if $\phi_{l_{k+1} \wedge \bar{\ell}} = \phi_{l_{k+2} \wedge \bar{\ell}} = \cdots = \phi_{l_n \wedge \bar{\ell}}$. The proof is similar to the base step. We pick the agent's filtration $\mathcal{F}_l = \mathcal{G}_l$ for every $l < l_k$. Since ϕ is a dually-robust mechanism, we must have $\phi_{l \wedge \bar{\ell}} \geq \phi_{\bar{\ell}}$. Under the event $\{\bar{\ell} > l_k\}$, we must have $\phi_{l_k} \geq \phi_{\bar{\ell}} = \phi_{l_{k'} \wedge \bar{\ell}}$ for every $k' > k$ by the induction hypothesis. Thus, condition on the agent continuing until l_k under the event $\{\bar{\ell} > l_k\}$, transfer is weakly decreasing starting from l_k until $\bar{\ell}$ almost surely. However, from Lemma 8, ϕ is dominated by the sandbox mechanism ϕ' constructed such that $\phi'_\ell = \phi_{l_k}$ for all $l_k \leq \ell \leq \bar{\ell}$ under the event $\{\bar{\ell} > l_k\}$ as long as $\phi_{l_k \wedge \bar{\ell}} > \phi_{l_{k+1} \wedge \bar{\ell}}$ with positive probability. Thus, for ϕ to be conditionally undominated at l_k, $\phi_{l_k \wedge \bar{\ell}} = \phi_{l_{k+1} \wedge \bar{\ell}}$ almost surely.

The argument yields that $\phi_{l_k \wedge \bar{\ell}} = \phi_{l_0}$ for every $k \in \{0, \ldots, n\}$, i.e., ϕ_l must be almost surely constant up until $\bar{\ell}$. But then ϕ must be equivalent to our adaptive sandbox ϕ^* as desired.

B PROOFS OF RESULTS IN SECTION 4

B.1 Proof of Theorem 4. The proof of Theorem 4 is quite involved. We will prove the result via the following steps:

1. There exists a 'direct and extremal' worst-case learning process
2. A worst-case learning process is an obedient bad news process
3. The obedient bad news process has arrival of bad news supported on $[l^*_{U^\phi}(\mu_0), L)$ such that continuation beliefs are $\gamma(l) = \mu^*_{U^\phi}(l)$ for each level in the support.
4. Argue that as $L \to \infty$ worst-case payoffs when both $\mathcal{F}$ and U are chosen adversarially are arbitarily poor.

Steps 1-3 are outlined in Figure 19.

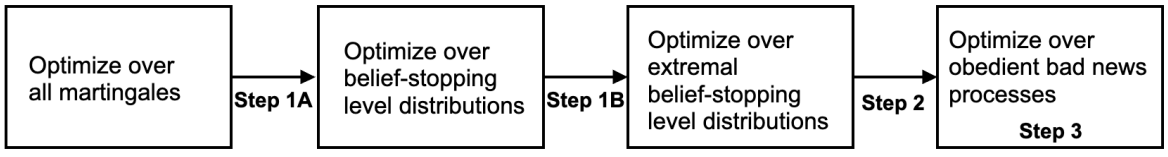


Figure 19: Outline of proof of Theorem 4

Proof of Theorem 4. We proceed along the steps outlined above.
Step 1. There is a direct and extremal worst-case process

Nature's problem is

$$\inf_{\mathcal{F}\in\mathbb{F}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \boldsymbol{\mu}, U)\Big] \tag{N}$$
$$\text{s.t. } \ell^*(\phi, \boldsymbol{\mu}, U) \text{ is a } (\mathcal{F}_l)_l\text{-adapted optimal stopping level for the agent}$$

From the revelation principle (Myerson, 1986; Forges, 1986) developed for dynamic information design (Koh and Sanguanmoo, 2022; Koh, Sanguanmoo, and Zhong, 2024), it is without loss to directly consider the joint distribution over beliefs and stopping levels $f \in \Delta(\Delta(\Theta)\times\mathcal{L})$ i.e., it is without loss to rule out learning processes with continuous sample paths i.e., those driven by Brownian noise such as the drift-diffusion model. In particular, if f fulfills a series of conditions laid out below, we can construct a jump process corresponding to the learning process which induces f via agent optimal stopping; conversely, if it does not fulfill these conditions, then we cannot find any learning process which induces f.

Thus, Nature's problem can be recast as that of choosing a joint distribution over agent's beliefs and stopping levels as follows:

$$\inf_{f\in\Delta(\Delta(\Theta)\times\mathcal{L})} \int V(\mu, l) f(d\mu, dl) \tag{N*}$$
$$\text{s.t.} \quad \int \mu f(d\mu, dl) = \mu_0 \tag{Martingale}$$
$$\int_{l>s} U^\phi(\mu, l) f(d\mu, dl) \geq \int_{l>s} U^\phi(\mu, s) f(d\mu, dl) \quad \forall s \in \mathcal{L} \tag{OC-C}$$
$$\mu \leq \mu^*_{U^\phi}(l) \quad \forall (\mu, l) \in \operatorname{supp} f \tag{OC-S}$$

We claim (N*) = (N) which follows from applying Lemma 1 of Koh, Sanguanmoo, and Zhong (2024). We briefly sketch what each of these conditions correspond to, and the intuition for why these conditions are tight. [Martingale] is a standard requirement on beliefs which any learning process $\boldsymbol{\mu}$, projected onto its induced stopping level, must continue to fulfill. [OC-C] is the requirement that the agent must find it optimal to continue pushing the technology given the joint distribution f.[56] [OC-S] is the

[56] Since U is linear in belief, this is equivalent to the condition on expected payoff from pushing the technology until the level prescribed by f vs stopping at s under the current belief:

$$\frac{\int_{l>s} U^\phi(\mu, l) f(d\mu, dl)}{\int_{l>s} f(d\mu, dl)} \geq U^\phi\left(\frac{\int_{l>s} \mu f(d\mu, dl)}{\int_{l>s} f(d\mu, dl)}, s\right) \quad \forall s \in \mathcal{L}.$$

requirement that the agent finds it optimal to stop. This is necessary because at belief and level pair $(\mu, l) \in \text{supp} f$, the agent cannot find it optimal to continue even in the absence of further information i.e., we cannot have $\mu > \mu^*_{U^\phi}(l)$. On the other hand, it is also sufficient since nature can choose the learning process so that agent indeed learns nothing following (μ, l).

We can further reduce the set of candidate joint distributions over beliefs and stopping levels. Define the set of direct and extremal belief-level pairs as follows:

$$\mathcal{E} := \left\{ f \in \Delta(\Delta(\Theta) \times \mathcal{L}) : \text{supp} f \subseteq \underbrace{\left\{\{0, \mu^*_{U^\phi}(l)\} \times [0, L)\right\}}_{\text{extremal beliefs before } L} \bigcup \underbrace{\left\{\{0,1\} \times \{L\}\right\}}_{\text{extremal beliefs at } L} \right\}.$$

Lemma 10. *There exists $f \in \mathcal{E}$ such that f solves* (N*).

This follows from noticing that nature's objective is a linear functional, V, U are linear in belief, [OC-C], [Martingale], and [OC-S] are linear constraints. Then applying Bauer's maximum principle, is without loss to extremize each belief μ at time l to extreme beliefs on the boundaries which, for level l, are exactly $\{0, \mu^*_{U^\phi}(l)\}$.

Step 2. Every worst-case learning process is an obedient bad news process
Say that $f \in \Delta(\Delta(\Theta) \times \mathcal{L})$ is an *obedient bad news process* if

$$\text{supp} f \subseteq \left\{\{0\} \times [0, L]\right\} \bigcup \left\{\{1\} \times L\right\}.$$

since f must fulfil [OC-C], this means that conditional on bad news not arriving, the agent pushes the technology.

Lemma 11 (Worst-case is obedient bad news)**.** *If ϕ is such that $U^\phi = g \circ V$ for some increasing convex function g then there exists an obedient bad news process f that solves* (N*).

Proof. Suppose that f solves (N*). We will modify it so that it is an obedient bad news process. From Lemma 10, it is without loss to suppose that $f \in \mathcal{E}$.

We will define a family of joint distributions parametrized by $l_0 \in \text{int}\,\mathcal{L}$. Write $\widehat{f_{l_0}} \in \Delta(\Delta(\Theta)\times[0,L])$ such that (i) the support of stopping beliefs are on 0 over the interval $[l_0, L)$, and on 0 and 1 at L i.e.,

$$\text{supp}\widehat{f_{l_0}} = \Big\{\{0\}\times[l_0,L]\Big\}\bigcup\Big\{\{1\}\times L\Big\}$$

which is depicted in Figure 20.

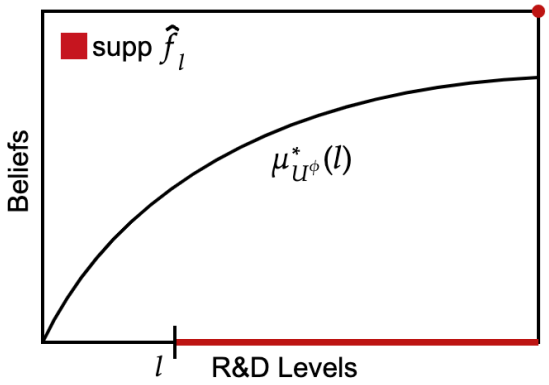


Figure 20: Support of $\hat{f}_{l_0}$

Moreover, the distribution $\widehat{f_{l_0}}$ is constructed such that (ii) the agent is indifferent between continuing and stopping for every $s \in [l_0, L]$:

$$\widehat{f_{l_0}}\Big(\{(1,L)\}\Big) = \mu^*_{U^\phi}(l_0) \qquad \widehat{f_{l_0}}\Big(\Big\{\{0\}\times[s,L]\Big\}\Big) = \frac{\mu^*_{U^\phi}(l_0)}{\mu^*_{U^\phi}(s)} - \mu^*_{U^\phi}(l_0).$$

and we use the convention that $\mu^*_{U^\phi}(L) = 1$ i.e., there is an atom at the belief-level pair $(0, L)$. This is well-defined because $\mu^*_{U^\phi}(s)$ is weakly increasing in s.

Lemma 12. *Under $\widehat{f_{l_0}}$, [OC-C] is tight for all times $s \geq l_0$.*

Proof. Direct calculation yields:

$$\begin{aligned}
\int_{l>s} U^\phi(\mu,l)\widehat{f_{l_0}}(d\mu,dl) &= U^\phi(1,L)\cdot\mu^*_{U^\phi}(l_0) - \Big[U^\phi(0,l)\cdot\Big(\frac{\mu^*_{U^\phi}(l_0)}{\mu^*_{U^\phi}(l)} - \mu^*_{U^\phi}(l_0)\Big)\Big]_{l=s}^{l=L} \\
&\quad + \int_{l>s}\frac{\partial U^\phi(0,l)}{\partial l}\Big(\frac{\mu^*_{U^\phi}(l_0)}{\mu^*_{U^\phi}(l)} - \mu^*_{U^\phi}(l_0)\Big)dl \\
&= U^\phi(1,L)\cdot\mu^*_{U^\phi}(l_0) + U(0,s)\cdot\Big(\frac{\mu^*_{U^\phi}(l_0)}{\mu^*_{U^\phi}(s)} - \mu^*_{U^\phi}(l_0)\Big) \\
&\quad + \mu^*_{U^\phi}(l_0)\int_{l>s}\underbrace{\frac{\partial U^\phi(0,l)}{\partial l}\Big(\frac{1}{\mu^*_{U^\phi}(l)} - 1\Big)}_{=-\partial U^\phi(1,l)/\partial l\,(\diamond)}dl \\
&= U^\phi(1,s)\mu^*_{U^\phi}(l_0) + U^\phi(0,s)\underbrace{\Big(\frac{\mu^*_{U^\phi}(l_0)}{\mu^*_{U^\phi}(s)} - \mu^*_{U^\phi}(l_0)\Big)}_{=\widehat{f_{l_0}}(\{\{0\}\times[s,L]\})}
\end{aligned}$$

$$= \int_{l>s} U^{\phi}(\mu, s) f(d\mu, dl)$$

where the first equality is via integration by parts and ($\diamond$) is because $\mu^*_{U^\phi}$ must almost everywhere (a.e.) satisfy,

$$\frac{\partial U^{\phi}\left(\mu^*_{U^\phi}(l), l\right)}{\partial l} = 0 \implies (1 - \mu^*_{U^\phi}(l))\frac{\partial U^{\phi}(0,l)}{\partial l} + \mu^*_{U^\phi}(l)\frac{\partial U^{\phi}(1,l)}{\partial l} = 0.$$

noting that U^ϕ is differentiable in its second argument a.e. because V is the weighted sum of monotone continuous functions which imply that they are differentable a.e., and U^ϕ inherits this because by assumption $U^\phi = g \circ V$ where g is convex. □

Now define $f^* \in \Delta(\Delta(\Theta) \times [0, L])$ as the 'weighted projection' of our original distribution f onto the beliefs $\{0, 1\}$ so that $\text{supp} f^* \subseteq \{(1, L)\} \cup \{(0, l) : l \in [0, L]\}$.

$$f^*\Big(\{(1,L)\}\Big) = f\Big(\{(1,L)\}\Big) + \int_{l\in[0,L)} \mu^*_{U^\phi}(l) 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl)$$

and for each $s \in \mathcal{L}$,

$$\begin{aligned} f^*\Big(\Big\{\{0\}\times[s,L]\Big\}\Big) =& f\Big(\Big\{\{0\}\times[s,L]\Big\}\Big) \\ &+ \int_{l\in[0,L)} \widehat{f_l}\Big(\Big\{\{0\}\times[s,L]\Big\}\Big) 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl). \end{aligned}$$

where we recall that $f \in \mathcal{E}$ is the distribution which is hypothesized to solve (N*). We first verify that f^* satisfies the constraints in (N*); this is a little tedious but the calculations are straightforward:

1. $[f^* \in \Delta(\Delta(\Theta) \times \mathcal{L})]$ Notice that f^* is a valid probability distribution because:

$$\begin{aligned} \int df^* &= f^*\Big(\{(1,L)\}\Big) + f^*\Big(\Big\{\{0\}\times[0,L]\Big\}\Big) \\ &= f\Big(\{(1,L)\}\Big) + f\Big(\Big\{\{0\}\times[0,L]\Big\}\Big) + \int_{l\in[0,L)} 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl) \\ &= \int df = 1 \end{aligned}$$

   where the second-last equality is because $f \in \mathcal{E}$ and the last is because f is by assumption a valid probability distribution.

2. [Martingale]: Directly integrating:

$$\begin{aligned}\int \mu f^*(d\mu, dl) &= \int \mu f(d\mu, dl) - \int \mu 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl) \\ &\quad + \int_{l\in[0,L)} \mu^*_{U^\phi}(l) 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl) \\ &= \int \mu f(d\mu, dl) \\ &= \mu_0.\end{aligned}$$

3. [OC-C]: for every $s \in [0, L)$, we can write the difference in [OC-C] as :

$$\begin{aligned}&\int_{l>s} U(\mu, l) f^*(d\mu, dl) - \int_{l>s} U(\mu, s) f^*(d\mu, dl) \\ &= \int_{l>s} \Big(U(\mu, l) - U(\mu, s)\Big) f^*(d\mu, dl) \\ &= \int_{l>s} \Big(U(\mu, l) - U(\mu, s)\Big) f(d\mu, dl) \qquad \text{(Definition of } f^*\text{)} \\ &\quad - \int_{l>s} \Big[U(\mu^*_{U^\phi}(l), l) - U(\mu^*_{U^\phi}(l), s)\Big] \cdot 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl) \\ &\quad + \int_l \left[\mu^*_{U^\phi}(l)\Big(U(1, L) - U(1, s)\Big) + \int_{l'\in[s,L)} (U(0, l') - U(0, s)) \widehat{f_l}(0, dl')\right] \\ &\qquad \cdot 1\Big\{\mu = \mu^*_{U^\phi}(l)\Big\} f(d\mu, dl)\end{aligned}$$

Note that from Lemma 12, since the agent is indifferent between continuing and stopping at $s \vee l$ under $\widehat{f_l}$ we have

$$\begin{aligned}0 &= \int_{l'>s\vee l} \Big(U(\mu, l) - U(\mu, s)\Big) \widehat{f_l}(d\mu, dl') \\ &= \mu^*_{U^\phi}(l)\Big(U(1, L) - U(1, s \vee l)\Big) + \int_{l'\in[s\vee l, L)} (U(0, l') - U(0, s \vee l)) \widehat{f_l}(0, dl').\end{aligned}$$

Hence

- If $s \geq l$, we have

$$\mu^*_{U^\phi}(l)\Big(U(1, L) - U(1, s)\Big) + \int_{l'\in[s,L)} (U(0, l') - U(0, s)) \widehat{f_l}(0, dl') = 0.$$

- If $s < l$, we have

$$\begin{aligned}
&\mu^*_{U^\phi}(l)\Big(U(1,L) - U(1,s)\Big) + \int_{l' \in [s,L)} (U(0,l') - U(0,s))\widehat{f_l}(0,dl') \\
&= \mu^*_{U^\phi}(l)(U(1,l) - U(1,s)) + (U(0,l) - U(0,s))\widehat{f_l}\Big(\{(0,l) : l \in (l,L]\}\Big) \\
&= U(\mu^*_{U^\phi}(l), l) - U(\mu^*_{U^\phi}(l), s).
\end{aligned}$$

  and these together imply

$$\int_{l>s} U(\mu,l) f^*(d\mu, dl) - U\left(\int_{l>s} \mu f^*(d\mu, dl), s\right) = \int_{l>s} \Big(U(\mu,l) - U(\mu,s)\Big) f(d\mu, dl) \geq 0,$$

  as desired.

4. [OC-S]: fulfilled by construction.

We have verified that f^* satisfies the conditions in (N*). We now show that f^* achieves a lower value than f:

$$\begin{aligned}
&\int V(\mu,l) f^*(d\mu, dl) \\
&= \int V(\mu,l) f(d\mu, dl) - \int V(\mu^*_{U^\phi}(l), l) 1\{\mu = \mu^*_{U^\phi}(l)\} f(d\mu, dl) \\
&\quad + \int \left(\mu^*_{U^\phi}(l) V(1,L) + \int V(0,l')\widehat{f_l}(0,dl')\right) 1\{\mu = \mu^*_{U^\phi}(l)\} f(d\mu, dl).
\end{aligned}$$

From Lemma 12, $\widehat{f_l}$, the agent is indifferent between continuing and stopping at l. From our comparative static Lemma 1, the principal must weakly prefer stopping at l under $\widehat{f_l}$ i.e.,

$$V(\mu^*_{U^\phi}(l), l) \geq \mu^*_{U^\phi}(l) V(1,L) + \int V(0,l')\widehat{f_l}(0,dl'),$$

for every l. Therefore,

$$\int V(\mu,l) f^*(d\mu, dl) \leq \int V(\mu,l) f(d\mu, dl),$$

as desired. □

Step 3. The arrival of bad news is supported on $[l^*_{U^\phi}(\mu_0), L)$

The previous step established that a worst-case learning process is obedient bad news; we now establish its specific form is that claimed in Theorem 4. An advantage of any obedient bad news process f is that the joint distribution over stopping beliefs and stopping levels can be summarized by a single cumulative distribution function $F : \mathcal{L} \to [0,1)$ where $F(s)$ gives the unconditional probability that bad news arrives before level $s \in \mathcal{L}$:

$$F(s) := f\Big(\{0\} \times [0,s]\Big)$$

and we require $F(L) = 1 - \mu_0$. This choice of F induces the principal's payoff

$$\mu_0 V(1,L) + \int_0^L V(0,s)dF(s),$$

subject to the agent's continuation constraint [OC-C] at each level s which can be rewritten in terms of F as:

$$\begin{aligned} &\mu_0 U^\phi(1,L) + \int_{l>s} U^\phi(0,l)dF(l) \geq \mu_0 U^\phi(1,s) + \int_{l>s} U^\phi(0,s)dF(l) \\ \iff &\mu_0\Big(U^\phi(1,L) - U^\phi(1,s)\Big) \geq \int_{l>s}\Big(U^\phi(0,s) - U^\phi(0,l)\Big)dF(l). \end{aligned}$$

We can thus restate problem (N*) in terms of optimizing over F:

$$\begin{aligned} &\inf_{F:[0,L]\to[0,1)} \int_0^L V(0,l)dF(l) \qquad (\text{N}^{**}) \\ &\text{s.t. } F \text{ is nondecreasing, } F(L) = 1-\mu_0, \text{ and} \\ &\underbrace{\mu_0\Big(U^\phi(1,L) - U^\phi(1,s)\Big)}_{=:H(s)} \geq \int_{l>s}\Big(U^\phi(0,s) - U^\phi(0,l)\Big)dF(l) \quad \text{for all } s \in \mathcal{L}. \end{aligned}$$

where the first constraint on the nondecreasingness of F and the value of $G(L)$ is equivalent to [Martingale] and the condition that the joint distribution f is a probability distribution; the second constraint is equivalent to [OC-C]; [OC-S] is automatically fulfilled. From the previous step, we know it is without loss to look for an obedient bad news process to solve (N*) and since there is a bijection between obedient bad news processes f and cumulative distribution functions F, (N*) = (N**).

We now solve (N**) which is a fairly standard infinite-dimensional program: we proceed via weak duality by constructing an appropriate set of multipliers. Define

$\Lambda^* : \mathcal{L} \to \mathbb{R}_{\geq 0}$ as follows:

$$\Lambda^*(s) := \begin{cases} \dfrac{\partial V/\partial l(0,s)}{\partial U^\phi/\partial l(0,s)} & \text{if } s > l^*_{U^\phi}(\mu_0), \\ \dfrac{V\Big(0, l^*_{U^\phi}(\mu_0)\Big)}{U^\phi\Big(0, l^*_{U^\phi}(\mu_0)\Big)} & \text{if } 0 < s \leq l^*_{U^\phi}(\mu_0), \\ 0 & \text{if } s = 0, \end{cases}$$

Note that Λ is increasing, so $\frac{d\Lambda^*}{ds} \geq 0$ wherever it is differentiable. Define the Lagrangian

$$\Psi(F,\Lambda) := \int_0^L V(0,l)dF(l) + \int_0^L \Bigg(\int_{l>s} \Big(U^\phi(0,s) - U^\phi(0,l) \Big) dF(l) - H(s) \Bigg) d\Lambda(s)$$

which, from standard manipulation, can be rewritten

$$\begin{aligned} \Psi(F,\Lambda) &= \int_0^L V(0,l)dF(l) + \int_0^L \int_{l>s} \Big(U^\phi(0,s) - U^\phi(0,l) \Big) dF(l) d\Lambda(s) - \int_0^L H(s) d\Lambda(s) \\ &= \int_0^L \Bigg(V(0,l) + \int_0^s \Big(U^\phi(0,s) - U^\phi(0,l) \Big) d\Lambda(s) \Bigg) dF(l) - \int_0^L H(s) d\Lambda(s) \\ &= \int_0^L \Bigg(\int_0^s \frac{\partial V(0,l)}{\partial l} - \frac{\partial U^\phi(0,l)}{\partial l} \Lambda(l) dl \Bigg) dF(s) - \int_0^L H(s) d\Lambda(s) \end{aligned}$$

Now from our construction of Λ^*, we have

$$\frac{\partial V(0,s)}{\partial s} - \frac{\partial U^\phi(0,s)}{\partial s} \cdot \Lambda^*(s) = 0 \quad \text{for every } s > l^*_{U^\phi}(\mu_0).$$

For $0 \leq s \leq l^*_{U^\phi}(\mu_0)$, notice that

$$\begin{aligned} &\int_0^s \left(\frac{\partial V(0,l)}{\partial l} - \frac{\partial U^\phi(0,l)}{\partial l} \cdot \Lambda^*(l) \right) dl \\ &\qquad = V(0,s) - U^\phi(0,s) \cdot \frac{V(0, l^*_{U^\phi}(\mu_0))}{U^\phi(0, l^*_{U^\phi}(\mu_0))} > 0 \end{aligned}$$

because $\frac{\partial V/\partial l(0,s)}{\partial U^\phi/\partial l(0,s)}$ is increasing in s since $U^\phi = g \circ V$ for some increasing convex

function g. Therefore, we have the lower-bound (changing the order of integration)

$$\inf_F \Psi(F,\Lambda^*) = \inf_F \Bigg[\underbrace{\int_0^L \int_l^L \left(\frac{\partial V(0,s)}{\partial s} - \frac{\partial U^\phi(0,s)}{\partial s} \cdot \Lambda^*(s) \right) ds dF(l)}_{\geq 0} - \int_0^L g(l) d\Lambda^*(l) \Bigg]$$
$$\geq - \int_0^L g(l) d\Lambda^*(l).$$

But this can be attained by choosing F^* corresponding to $\widehat{f}_{l^*_{U^\phi}(\mu_0)}$ as follows:

$$F^*(s) = \begin{cases} 0 & s \leq l^*_{U^\phi}(\mu_0) \\ \widehat{f}_{l^*_{U^\phi}(\mu_0)}\Big(\{0\} \times [0,s] \Big) & s \in (l^*_{U^\phi}(\mu_0), L) \\ 1 - \mu_0 & s = L. \end{cases}$$

and observe that (i) by construction F^* is nondecreasing; (ii) $F^*(l^*_{U^\phi}(\mu_0)) = 0$; (iii) from Lemma 12, [OC-C] binds from for each $s \in (l^*_{U^\phi}, L)$. We have found a feasible F^* which attains the lower bound. But this also solves the primal since

$$\Psi(F^*,\Lambda^*) = \inf_F \Psi(F,\Lambda^*) \leq \sup_\Lambda \inf_F \Psi(F,\Lambda) \leq \inf_F \sup_\Lambda \Psi(F,\Lambda)$$

where the first equality is because F^* attains $\inf_F \Psi(F,\Lambda^*)$, and the last inequality is from weak duality.

It remains to verify that continuation beliefs take the form claimed in Theorem 4:

$$\gamma^{F^*}(l) = \begin{cases} \mu_0 & \text{if } l \leq l^*_{U^\phi}(\mu_0) \\ \mu^*_{U^\phi}(l) & \text{if } l > l^*_{U^\phi}(\mu_0). \end{cases}$$

The first case is immediate; for the second case $l > l^*_{U^\phi}(\mu_0)$, continuation beliefs can be rewritten directly as:

$$\gamma^{F^*}(l) = \frac{\mu_0}{\mu_0 + \Big(1 - F^*(l)\Big)} = \frac{\mu_0}{\mu_0 + \frac{\mu^*_{U^\phi}(l^*_{U^\phi}(\mu_0))}{\mu^*_{U^\phi}(l)} - \mu^*_{U^\phi}(l^*_{U^\phi}(\mu_0))} = \mu^*_{U^\phi}(l)$$

where the last equality is because $\mu^*_{U^\phi}(l^*_{U^\phi}(\mu_0)) = \mu_0$ by definition.

Step 4: Worst-case payoff arbitarily poor.

Consider that

$$\lim_{L\to\infty} V(\mu_0, L) = \lim_{L\to\infty} V(0,L)\left(\mu_0 \cdot \frac{V(1,L)}{V(0,L)} + (1-\mu_0)\right) = \lim_{L\to\infty} V(0,L)(1-\mu_0).$$

Since $V(0,L) < 0$ and

$$\lim_{L\to\infty} |V(0,L)| = \underbrace{\left|\frac{V(0,L)}{V(1,L)}\right|}_{\to\infty} \cdot \underbrace{|V(1,L)|}_{>V(1,1)>0} = \infty.$$

Thus, $\lim_{L\to\infty} V(\mu_0, L) = -\infty$. Fix some small $\epsilon > 0$. For each sufficiently large L, there exists an agent's utility function $U_L \in \mathbb{U}$ such that $l^*_{U_L^\phi}(\mu_0) \in [L-\epsilon, L]$. Under the no information learning process, the agent with the utility function U_L stops within $[L-\epsilon, L]$ with probability 1. Thus,

$$\inf_{\substack{\mathcal{F}\in\overline{\mathbb{F}}\\ U\in\mathbb{U}}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] \le \lim_{L\to\infty} V(\mu_0, L-\epsilon) - \mathbb{E}[V(\mu_0, \bar{\ell})] \to -\infty$$

which implies $\lim_{L\to+\infty} \inf_{\substack{\mathcal{F}\in\overline{\mathbb{F}}\\ U\in\mathbb{U}}} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*(\phi, \mathcal{F}, U))\Big] = -\infty$ as desired. □

B.2 Proof of Proposition 3.

Proof. We know from Theorem 4 that the principal's worst information structure induces a distribution of stopping belief and time as follows:

$$\widehat{f_{l^*_{U^\phi}(\mu_0)}}\Big(\{(1,L)\}\Big) = \mu_0 \qquad \widehat{f_{l^*_{U^\phi}(\mu_0)}}\Big(\{(0,l) : l \ge l^*_{U^\phi}(\mu_0)\}\Big) = \frac{\mu_0}{\mu^*_{U^\phi}(l)} - \mu_0.$$

Thus, the principal's worst-case payoff under mechanism ϕ can be rewritten as:

$$\mu_0 V(1,L) + \int_{l^*_{U^\phi}(\mu_0)}^{L} V(0,l)\widehat{f_{l^*_{U^\phi}}}(0, dl).$$

Since $U^{\phi'}(\theta, l) = g \circ U^\phi(\theta, l)$, Lemma 1 implies $l^*_{U^\phi}(\mu_0) \le l^*_{U^{\phi'}}(\mu_0)$ and $\mu^*_{U^\phi}(l) \ge \mu^*_{U^{\phi'}}(l)$. Thus, the distribution of $\widehat{f_{l^*_{U^\phi}}}(0,\cdot)$ is strictly dominated by that of $\widehat{f_{l^*_{U^{\phi'}}}}(0,\cdot)$ in the

first-order stochastic dominance sense. Since $V(0,l)$ is decreasing in l, we must have

$$\int_{l^*_{U^\phi}(\mu_0)}^{L} V(0,l)\widehat{f_{l^*_{U^\phi}}}(0,dl) \geq \int_{l^*_{U^{\phi'}}(\mu_0)}^{L} V(0,l)\widehat{f_{l^*_{U^{\phi'}}}}(0,dl),$$

implying the principal's payoff under mechanism ϕ is better than that under mechanism ϕ', as desired. □

ONLINE APPENDIX TO
'ROBUST TECHNOLOGY REGULATION'

ANDREW KOH SIVAKORN SANGUANMOO
ajkoh@mit.edu sanguanm@mit.edu

I REMAINING TECHNICAL PROOFS

We now show the remaining technical proofs that were omitted from the main text.

I.1 Proof of Lemma 6. We begin with the following lemma stating that optimal stopping levels that solve (OSP-ϵ) in the proof of Lemma 5 must form a kind of 'lattice'.

Lemma 13. *Suppose $\epsilon_1 > \epsilon_2$ be positive real numbers. Let η^{ϵ_1} and η^{ϵ_2} solve (OSP-ϵ_1) and (OSP-ϵ_2), respectively. Then, $\eta_* := \eta^{\epsilon_1} \wedge \eta^{\epsilon_2}$ and $\eta^* := \eta^{\epsilon_1} \vee \eta^{\epsilon_2}$ also solve (OSP-ϵ_1) and (OSP-ϵ_2), respectively.*

Proof of Lemma 13. Consider that

$$\begin{aligned}
&\mathbb{E}[V(\theta,\eta^{\epsilon_1})1\{Z^{\epsilon_1}=0\}] = \Phi(\epsilon_1) \geq \mathbb{E}[V(\theta,\eta_*)1\{Z^{\epsilon_1}=0\}] \\
&\implies \mathbb{E}[(V(\theta,\eta^{\epsilon_1}) - V(\theta,\eta^{\epsilon_2}))1\{\eta^{\epsilon_1} > \eta^{\epsilon_2}\}] \\
&\quad \geq \epsilon_1 \underbrace{\mathbb{E}[V(1,\eta^{\epsilon_1}) - V(1,\eta^{\epsilon_2}))1\{\eta^{\epsilon_1} > \eta^{\epsilon_2}, \omega \in A_1^*\}]}_{\geq 0}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}[V(\theta,\eta^{\epsilon_2})1\{Z^{\epsilon_2}=0\}] = \Phi(\epsilon_2) \geq \mathbb{E}[V(\theta,\eta^*)1\{Z^{\epsilon_2}=0\}] \\
&\implies \mathbb{E}[(V(\theta,\eta^{\epsilon_1}) - V(\theta,\eta^{\epsilon_2}))1\{\eta^{\epsilon_1} > \eta^{\epsilon_2}\}] \\
&\quad \leq \epsilon_2 \underbrace{\mathbb{E}[V(1,\eta^{\epsilon_1}) - V(1,\eta^{\epsilon_2}))1\{\eta^{\epsilon_1} > \eta^{\epsilon_2}, \omega \in A_1^*\}]}_{\geq 0}.
\end{aligned}$$

Since $\epsilon_1 > \epsilon_2$, both inequalities must be equalities so

$$\Phi(\epsilon_1) = \mathbb{E}[V(\theta,\eta_*)1\{Z^{\epsilon_1}=0\}] \quad \text{and} \quad \Phi(\epsilon_2) = \mathbb{E}[V(\theta,\eta^*)1\{Z^{\epsilon_2}=0\}]$$

as desired. □

Proof of Lemma 6. Suppose s^{ϵ_n} is the latest optimal stopping level that solves (OSP-ϵ_n). First, we show that s^{ϵ_n} is increasing in n almost surely. From Lemma 13, $s^{\epsilon_n} \vee s^{\epsilon_{n+1}}$

must solve (OSP-ϵ_{n+1}). The definition of $s^{\epsilon_{n+1}}$ implies $s^{\epsilon_n} \vee s^{\epsilon_{n+1}} \leq s^{\epsilon_{n+1}}$ almost surely. Thus, $s^{\epsilon_n} \leq s^{\epsilon_{n+1}}$ almost surely, as desired. Thus, $s^{\epsilon_n} \uparrow s'$ for some s' almost surely and s' is still a stopping level because $(\mathcal{G}_l)_l$ is a right-continuous filtration.

Next, we verify Φ must be a continuous function. Consider that, for every positive real numbers ϵ_1 and ϵ_2,

$$\begin{aligned}
\Phi(\epsilon_2) - \Phi(\epsilon_1) &= \mathbb{E}[V(\theta, s^{\epsilon_2}) 1\{Z^{\epsilon_2} = 0\}] - \mathbb{E}[V(\theta, s^{\epsilon_1}) 1\{Z^{\epsilon_1} = 0\}] \\
&\leq \mathbb{E}[V(\theta, s^{\epsilon_2}) 1\{Z^{\epsilon_2} = 0\}] - \mathbb{E}[V(\theta, s^{\epsilon_2}) 1\{Z^{\epsilon_1} = 0\}] \\
&= \mathbb{E}[V(\theta, s^{\epsilon_2}) 1\{Z^{\epsilon_1} = 1\}] - \mathbb{E}[V(\theta, s^{\epsilon_2}) 1\{Z^{\epsilon_2} = 1\}] \\
&= (\epsilon_1 - \epsilon_2) \mathbb{E}[V(\theta, s^{\epsilon_2}) 1\{\omega \in A_1^*\}] \\
&\leq M |\epsilon_1 - \epsilon_2|,
\end{aligned}$$

where $M := \sup_{\theta, l} |V(\theta, l)|$, which is well-define because the level space is compact. Similarly, $\Phi(\epsilon_1) - \Phi(\epsilon_2) \leq M|\epsilon_1 - \epsilon_2|$. This implies $|\Phi(\epsilon_1) - \Phi(\epsilon_2)| \leq M|\epsilon_1 - \epsilon_2|$, so Φ is continuous.

Finally, we show that $s' = \overline{\ell}$ almost surely. Consider that

$$\begin{aligned}
\mathbb{E}[V(\theta, s')] &= \lim_{n \to \infty} \mathbb{E}[V(\theta, s^{\epsilon_n})] && \text{(Dominated convergence theorem)} \\
&= \lim_{n \to \infty} \mathbb{E}[V(\theta, s^{\epsilon_n}) 1\{Z^{\epsilon_n} = 0\}] + \lim_{n \to \infty} \mathbb{E}[V(\theta, s^{\epsilon_n}) 1\{Z^{\epsilon_n} = 1\}] \\
&= \lim_{n \to \infty} \mathbb{E}[V(\theta, s^{\epsilon_n}) 1\{Z^{\epsilon_n} = 0\}] && (\mathbb{P}(Z^{\epsilon_n} = 1) \to 0 \text{ as } n \to \infty) \\
&= \lim_{n \to \infty} \Phi(\epsilon_n) \\
&= \Phi(0) && (\Phi \text{ is continuous}) \\
&= \mathbb{E}[V(\theta, \overline{\ell})].
\end{aligned}$$

Since $\overline{\ell}$ is the principal's unique optimal stopping level, we must have $s' = \overline{\ell}$ almost surely. Thus, $s^{\epsilon_n} \uparrow \overline{\ell}$, as desired. □

I.2 Proof of Theorem 2 when $\bar{\ell} < L$ w.p.p.. We assumed in the main text that $\overline{\ell} < L$ almost surely. We relax this assumption and show Theorem 2. We start with the following lemma stating that the hard limit is strictly before L with positive probability given any interim history.

Lemma 14. *For every stopping level $\ell < L$, we must have $\mathbb{P}(\overline{\ell} < L | \mathcal{G}_\ell) > 0$ almost surely.*

Proof of Lemma 14. Suppose towards a contradiction there exists a stopping level

$\ell < L$ such that the event

$$\mathbb{P}\left(\left\{\omega \in \Omega : \mathbb{P}(\overline{\ell} = L|\mathcal{G}_\ell) = 0\right\}\right) > 0.$$

Our assumption on the law of the principal's signal process $\mathbb{P}^\theta$ implies the event

$$A = \left\{\omega \in \Omega : \mathbb{P}(\overline{\ell} = L|\mathcal{G}_\ell) = 0, \mu^P_{\ell+\epsilon} < \delta\right\},$$

where $\mu^P_l := \mathbb{P}(\theta = 1|\mathcal{G}_\ell)$ also happens with positive probability, where we choose $\delta > 0$ so that $V(\delta, L) < V(\delta, l)$ for every $l < L$.[1] Note that the event A is $\mathcal{G}_{\ell+\epsilon}$-measurable. We define a stopping level ℓ' as follows

$$\ell' := \begin{cases} \overline{\ell} & \text{if } \omega \notin A \\ \ell + \epsilon & \text{if } \omega \in A \end{cases}$$

It is easy to see that ℓ' is a valid stopping level because A is $\mathcal{G}_{\ell+\epsilon}$-measurable. Under event A, we must also have $\overline{\ell} = L$ almost surely. Thus,

$$\begin{aligned} \mathbb{E}[V(\mu_{\overline{\ell}}, \overline{\ell}) - V(\mu_{\ell'}, \ell')] &= \mathbb{E}[(V(\mu_{\overline{\ell}}, \overline{\ell}) - V(\mu_{\ell'}, \ell'))1\{\omega \in A\}] \\ &= \mathbb{E}[(V(\mu_{\ell^*}, L) - V(\mu_{\ell+\epsilon}, \ell + \epsilon))1\{\omega \in A\}] \\ &= \mathbb{E}[(V(\mu_{\ell+\epsilon}, L) - V(\mu_{\ell+\epsilon}, \ell + \epsilon))1\{\omega \in A\}] \\ &\qquad\qquad \text{(optional stopping theorem)} \\ &< 0, \end{aligned}$$

where the inequality follows from $\mu_{\ell+\epsilon} < \delta$ and $V(\delta, L) < V(\delta, l)$ for every $l < L$, and $\mathbb{P}(A) > 0$. This contradicts the optimality of $\overline{\ell}$, as desired. □

Now we are ready to show the proof of Theorem 2 for the case that $\overline{\ell} = L$ with positive probability.

Remaining proof of Theorem 2. Consider any stopping times $\ell_1 \le \ell_2 \le \overline{\ell}$. We will show that $\phi_{\ell_1} \ge \phi_{\ell_2}$ almost surely. Suppose toward a contradiction that the event $A := \{\omega : \phi_{\ell_1} < \phi_{\ell_2}\}$ happens with positive probability. From Lemma 5, we have, under event A, $\phi_{\overline{\ell}} \le \phi_{\ell_1} < \phi_{\ell_2}$. Since A is $\mathcal{G}_{\ell_2}$-measurable, $\mathbb{P}(A \cap \{\overline{\ell} < L\}) > 0$. Under event $A \cap \{\overline{\ell} < L\}$, we have $\phi_{\ell_1} < \phi_{\ell_2} \ge \phi_{\overline{\ell}} < \infty = \phi_L$, which contradicts single-peakedness,

[1] We simply choose small δ so that $\frac{\delta}{1-\delta} < \sup_{l \in [0,L]} |\frac{\partial V(1,l)/\partial l}{\partial V(0,l)/\partial l}|$.

as desired. Thus, ϕ is decreasing until the hard limit, and then the remaining proof follows the same as the proof in the main text. □

I.3 Proof of Lemma 9.

Proof of Lemma 9. Suppose, towards a contradiction, there exists a stopping level $\ell_1 > \overline{\ell}$ such that the event

$$A_0 = \left\{\omega : \left\{\ell_1 - \epsilon > s \geq \overline{\ell} > \ell_0\right\} \bigcap \left\{\phi_{\ell_1} - \phi_{\ell_0} < M \text{ or } \phi_{\ell_1} = -\infty\right\}\right\} \cap A^{\leq}_{\ell_0}$$

happens with positive probability for some $\epsilon > 0, s \in \mathcal{L}$, and some $M < +\infty$. Note that the event is $\mathcal{G}_L$-measurable.

We construct an auxiliary random variable. Let S be a Bernoulli random variable (taking values in $\{0,1\}$) such that

$$\mathbb{P}\Big(S = 0 \Big| \mathcal{G}_L, \theta = 0\Big) = 1 \quad \text{and} \quad \mathbb{P}\Big(S = 0 \Big| \mathcal{G}_L, \theta = 1\Big) = \max\left\{\frac{\delta(1-\mu_L)}{\mu_L(1-\delta)}, 1\right\},$$

where $\mu_L = \mathbb{P}(\theta = 1|\mathcal{G}_L)$ and δ is sufficiently small so that $V(\delta, l_0) > V(\delta, l_1)$ for every $l_1 > l_0$. This condition is equivalent to

$$\frac{\delta}{1-\delta} < \inf_{l_1 > l_0} \frac{V(1,l_1) - V(1,l_0)}{V(0,l_0) - V(0,l_1)},$$

where the RHS is weakly greater than $\inf_{l \in [0,L]} \left|\frac{\partial V(1,l)/\partial l}{\partial V(0,l)/\partial l}\right|$, which is strictly positive. The belief about the state given by all principal's information and S is

$$\mathbb{P}(\theta = 1|\mathcal{G}_L, S = 1) = 1$$
$$\mathbb{P}(\theta = 1|\mathcal{G}_L, S = 0) = \frac{\mathbb{P}(S = 0|\mathcal{G}_L, \theta = 1)\mu_L}{\mathbb{P}(S = 0|\mathcal{G}_L, \theta = 1)\mu_L + \mathbb{P}(S = 0|\mathcal{G}_L, \theta = 0)(1-\mu_L)} = \min\{\delta, \mu_L\}.$$

Since A_0 is $\mathcal{G}_L$-measurable, the event $A^* := A_0 \cap \{\omega : S = 0, \mu_L > \underline{\delta}\}$ happens with positive probability for some $\underline{\delta} > 0$.

Consider the following agent's filtration:

$$\mathcal{F}^{\epsilon}_l := \sigma\left(\left\{X \cap \{\omega : l < \ell_0\} : X \in \mathcal{G}_l\right\} \cup \left\{X \cap \{\omega : l \geq \ell_0, S = s\} : X \in \mathcal{G}_L, s \in \{0,1\}\right\}\right).$$

In words, at ℓ_0, the agent learns everything the principal will learn in the future as

well as information from the extra signal S. In particular, for every stopping level ℓ, if $\ell < \ell_0$, then $\mathcal{F}_\ell = \mathcal{G}_\ell$. On the other hand, if $\ell \geq \ell_0$, then $\mathcal{F}_\ell = \mathcal{G}_L \vee \sigma(S)$.

This implies, for every l, under the event that $\{\ell_0 \leq l\}$,

$$\mathbb{P}(\theta = 1|\mathcal{F}_l) = \mathbb{P}(\theta = 1|\mathcal{G}_L, S) = \begin{cases} 1 & \text{if } S = 1 \\ \min\{\delta, \mu_L\} & \text{if } S = 0. \end{cases} =: \mu^U_{l_0}$$

Suppose $\ell^* := \ell^*(\phi^*, \mathcal{F}, U)$ is the agent's stopping time under the adaptive sandbox mechanism ϕ^*, facing the learning process $\mathcal{F}$, with preference U. We can use the same argument from Step 2 in the proof of Theorem 2 to show that $\ell^* \geq \overline{\ell} \wedge \ell_0$ almost surely. Under the event $\{S = 1, \overline{\ell} > \ell_0\}$, the agent learns the state is perfectly good at level ℓ_0, so he will continue developing until the hard limit $\overline{\ell}$ under the sandbox mechanism. Therefore, $\mathbb{P}(S = 0|\mathcal{G}_{\ell_0}, \ell^* = \ell_0) = 1$ under the event that $\{\overline{\ell} > \ell_0\}$, i.e., the principal believes that $S = 0$ with probability one after observing that the agent stopped at ℓ_0 strictly before the sandbox limit. Thus, under the event $\{\ell^* = \ell_0\}$, we must have $\mu^U_{\ell_0} \leq \delta$.

We now construct the agent's preference $U \in \mathbb{U}$ as follows. We set

$$U(\theta, l) = \begin{cases} V(1,s) + \beta(V(1,l) - V(1,s)), & \text{if } l \geq s \text{ and } \theta = 1 \\ V(\theta, l), & \text{otherwise,} \end{cases}$$

for sufficiently large $\beta > 1$ which will be chosen later. In words, the agent and the principal have aligned preferences until the level s, and then the agent becomes highly optimistic about the technology state. It is easy to see that $U \in \mathbb{U}$.[2]

First, we will show that $A^* \subset \{\omega : \ell^* = \ell_0\}$, i.e., under the event A^*, the agent must stop at ℓ_0 under the adaptive sandbox mechanism. Due to the hard limit, the optimal stopping ℓ^* must be before the hard limit $\overline{\ell}$. Since $\overline{\ell} < s$ under the event A^*, we must have $\ell^* < s$. Under event A^*, we must have

$$\begin{aligned} &\mathbb{E}[U(\mu^U_{\ell^*}, \ell^*)] \geq U(\mu^U_{\ell_0}, \ell_0) \\ \Longrightarrow &\mu^U_{\ell_0}(V(1,\ell^*) - V(1,\ell_0)) + (1 - \mu^U_{\ell_0})(V(0,\ell^*) - V(0,\ell_0)) \geq 0 \quad (U = V \text{ whenever } l < s) \end{aligned}$$

[2]This is because $U = g \circ V$, where $g : \mathbb{R} \to \mathbb{R}$ is a convex function such that

$$g(x) = \begin{cases} x & \text{if } x < V(1,s), \\ V(1,s) + \beta(x - V(1,s)) & \text{if } x \geq V(1,s), \end{cases}$$

$$\implies \delta(V(1,\ell^*) - V(1,\ell_0)) + (1-\delta)(V(0,\ell^*) - V(0,\ell_0)) \geq 0 \qquad \text{(Since } S=0,\ \mu^U_{\ell_0} \leq \delta\text{)}$$
$$\implies V(\delta,\ell^*) \geq V(\delta,\ell_0).$$

By the construction of δ, we must have $\ell^* = \ell_0$, as desired.

Suppose $\ell_\phi := \ell^*(\phi, \mathcal{F}, U)$ is the agent's stopping time under the mechanism ϕ, facing the learning process $\mathcal{F}$, with preference U (after the agent has been developing until the level ℓ_0). Thus, ℓ_ϕ must solves the following stopping problem

$$\sup_{\ell \geq \ell_0} \mathbb{E}[U^\phi(\theta,\ell)] \quad \text{s.t. } \ell \text{ is adapted to } \mathcal{F}_l$$

This implies $\ell_\phi \in \arg\max_{\ell \geq \ell_0} U^\phi(\mu_{\ell_0},\ell)$ almost surely because

$$\mathbb{E}[U^\phi(\theta,\ell)] = \mathbb{E}[\mathbb{E}[U^\phi(\theta,\ell)|\mathcal{G}_L,S]] \leq \mathbb{E}[\max_{\ell\geq\ell_0} \mathbb{E}[U^\phi(\theta,l)|\mathcal{G}_L,S]] = \mathbb{E}^\phi[\max_{\ell\geq l_0} U^\phi(\mu_{\ell_0},\ell)],$$

where the last inequality follows from the fact that $\mathbb{E}[\theta|\mathcal{G}_L,S] = \mu_{\ell_0}$ and U^ϕ is linear in belief. Consider that ℓ_ϕ must be strictly greater than ℓ_0 under the event A^* because we have

$$\begin{aligned}
U(\mu_{\ell_0},\ell_1) - U(\mu_{\ell_0},\ell_0) &\geq U(\underline{\delta},\ell_1) - U(\underline{\delta},\ell_0) \\
&= (\underline{\delta}\beta V(1,\ell_1) + (1-\underline{\delta})V(0,\ell_1)) - (\beta-1)\underline{\delta}V(1,s) - V(\underline{\delta},\ell_0) \\
&\qquad\qquad (\ell_1 > s) \\
&\geq (\underline{\delta}\beta V(1,s+\epsilon) + (1-\underline{\delta})V(0,s+\epsilon)) - (\beta-1)\underline{\delta}V(1,s) - V(\underline{\delta},\ell_0) \\
&\qquad\qquad (\ell_1 > s+\epsilon) \\
&\geq \underline{\delta}\beta \underbrace{\Big(V(1,s+\epsilon) - V(1,s)\Big)}_{>0} \\
&\qquad + \Big(\underline{\delta}V(1,s) + (1-\underline{\delta})V(0,s+\epsilon) - V(\underline{\delta},0)\Big) \\
&\qquad\qquad (V(\underline{\delta},0) > V(\underline{\delta},\ell_0)) \\
&\geq M+1,
\end{aligned}$$

by setting β large enough so that the second and third inequality hold.[3] Under the

[3] For the second inequality, a sufficient condition is

$$\beta > \frac{1-\underline{\delta}}{\underline{\delta}} \sup_{l\in[0,L]} \left|\frac{\partial V(0,l)/\partial l}{\partial V(1,l)/\partial l}\right| \implies \frac{\partial}{\partial l}\Big(\underline{\delta}\beta V(1,l) + (1-\underline{\delta})V(0,l)\Big) > 0 \quad \forall l$$

For the third inequality, the multiplier of β is strictly positive, so the inequality must hold when β is

event A^*, we know either $\phi_{\ell_1} - \phi_{\ell_0} < M$ or $\phi_{\ell_1} = -\infty$. If $\phi_{\ell_1} - \phi_{\ell_0} < M$ we must have

$$U^{\phi}(\mu_{\ell_0}, \ell_1) - U^{\phi}(\mu_{\ell_0}, \ell_0) \geq (M+1) - M = 1,$$

so stopping at ℓ_0 is not optimal under the event A^*, implying $\ell_\phi > \ell_0$ under the event A^*, as desired. On the other hand, if $\phi_{\ell_1} = -\infty$, then ℓ_0 is not the latest optimal stopping under the event A^*, implying $\ell_\phi > \ell_0$ under the event A^*, as desired.

Now we compare the principal's payoff under ϕ^* and ϕ under the event $\{\omega : \ell^* = \ell_0\} \cap A^<_{\ell_0},$. We showed earlier that $\mu_{\ell_0} \leq \delta$ under the event $\{\ell^* = \ell_0\}$. Because $V(\delta, \ell_0) > V(\delta, \ell)$ for every $\ell > \ell_0$, we must have $V(\mu_{\ell_0}, \ell_0) \geq V(\mu_{\ell_0}, \ell_\phi)$ under the event $\{\ell^* = \ell_0\}$, and the inequality is strict under the event A^* because $\ell_\phi > \ell_0$. Since the event A^* happens with positive probability, we must have

$$\begin{aligned}
&\mathbb{E}[V(\theta, \ell_\phi) 1\{\ell^* = \ell_0, \omega \in A^<_{\ell_0}\}] \\
&= \mathbb{E}[V(\mu_{\ell_0}, \ell_\phi)) 1\{\ell^* = \ell_0, \omega \in A^<_{\ell_0}\}] \\
&= \mathbb{E}[V(\mu_{\ell_0}, \ell_\phi) 1\{\omega \in A^*\}] + \mathbb{E}[V(\mu_{\ell_0}, \ell_\phi) 1\{\omega \notin A^*, \ell^* = \ell_0, \omega \in A^<_{\ell_0}\}] \quad (A^* \subset \{\omega : \ell^* = \ell_0\}) \\
&< \mathbb{E}[V(\mu_{\ell_0}, \ell_0) 1\{\omega \in A^*\}] + \mathbb{E}[V(\mu_{\ell_0}, \ell_0) 1\{\omega \notin A^*, \ell^* = \ell_0, \omega \in A^<_{\ell_0}\}] \\
&= \mathbb{E}[V(\mu_{\ell_0}, \ell_0) 1\{\ell^* = \ell_0, \omega \in A^<_{\ell_0}\}].
\end{aligned}$$

This implies ϕ performs strictly worse than ϕ^* under the event $A^<_{\ell_0}$ given the principal knows the agent stops ℓ_0, which contradicts ϕ conditionally dominates ϕ^* upon stopping at ℓ_0, as desired.

□

II EXPERIMENTATION UNDER RISK-AVERSION

We relate our comparative static on optimal experimentation to extant work on optimal stopping and risk aversion.

II.1 Generalizing learning processes. We show that Lemma 1 generalizes the main result of Keller et al. (2019) (henceforth KNW). We start by mapping their exponential bandit setting into our framework. Interpret l as the level (in KNW, time) the agent switches from the risky to the safe arm. Let $\theta = 1$ correspond to the risky arm paying out a transfer of $h > 0$ at rate $\lambda > 0$, and $\theta = 0$ correspond to never paying out. The safe arm pays out a transfer of $s > 0$ at rate 1.

sufficiently large.

Then, we have that if the arm is good:

$$u(1,l) = \mathbb{E}\Big[\int_0^l e^{-rs} \cdot 1\{\text{arrive at } l\} \cdot u(h)ds\Big] = \frac{\lambda \cdot u(h) - u(s)}{r}\Big(1 - e^{-rl}\Big) \qquad \text{(Bandit-1)}$$

where the second equality makes use of the fact (sometimes called 'Poissonization') that conditional K arrives over the interval $[0,l]$ the arrival locations are uniformly distributed over $[0,l]$.

But if the arm is bad:

$$u(0,l) = -\mathbb{E}\Big[\int_0^l e^{-rs} \cdot 1\{\text{arrive at } l\} \cdot u(s)ds\Big] = -\frac{u(s)}{r}\Big(1 - e^{-rl}\Big) \qquad \text{(Bandit-0)}$$

Suppose that for two increasing concave utility functions u_1, u_2 normalized to $u_1(0) = u_2(0) = 0$, we have that $u_2 = f \circ u_1$ for some increasing concave function f. We call u_2 more *flow risk averse* than u_1 where the qualifier 'flow' is to emphasize that this risk aversion is imposed on flow payoffs rather than on the timing of rewards.

Corollary 3 (Generalizing Keller et al. (2019))**.** *Consider the model with payoffs given by* (Bandit-1) *and* (Bandit-0)*.*

(i) If $h > s$ then under any learning process, the more flow risk averse agent stops sooner.

(ii) If $h < s$ then under under any learning process, the more flow risk averse agent stops later.

Proposition 1 of Keller et al. (2019) is recovered as the special case with learning process is given by the arrival of perfect good news with rate $\lambda > 0$.

By decoupling the learning process from the reward process, this is also in the spirit of (Lizzeri, Shmaya, and Yariv, 2024). To see Corollary 3, define the following function:

$$g(v) = \begin{cases} \dfrac{\lambda u_2(h) - u_2(s)}{\lambda u_1(h) - u_1(s)} \cdot v & \text{if } x \geq 0 \\ \dfrac{u_2(s)}{u_1(s)} \cdot v & \text{if } x < 0 \end{cases}$$

so that $v_2 = g \circ v_1$. Now this function is convex if and only if

$$\frac{\lambda u_2(h) - u_2(s)}{u_2(s)} \geq \frac{\lambda u_1(h) - u_1(s)}{u_1(s)}$$

$$\iff \frac{u_2(h)}{u_2(s)} \geq \frac{u_1(h)}{u_1(s)}$$

which requires $h < s$. By Lemma 1, an agent with preference v_2 stops later than the agent with preference v_1. But recalling that v_2 corresponds to the flow payoff of u_2 (more flow risk-averse), so this delivers the corollary. The learning process in KNW is a special case of 'perfect good news' arriving at rate $\lambda > 0$ so they can characterize the stopping boundary exactly.

II.2 General comparative static over time and risk. Lemma 1 is useful for settings beyond exponential discounting. For instance, both Chancelier, De Lara, and de Palma (2009) (henceforth CDP) and Quah and Strulovici (2013) (henceforth QS) deliver conditions under which changing the preferences (flow risk aversion in CDP, impatience in QS) can lead an agent to experiment less. Lemma 1 can be viewed as partially generalizing these insights by incorporating risk-aversion both within and across time/levels.

CDP works within an (indexable) multi-armed bandit setting and argues that higher *flow risk aversion* ensures that the Gittins index for each risky arm is accordingly lower than that for an agent with lower flow risk aversion. QS apply their results on a *patience order* over discount rates to experimentation (see Section II.F of QS), showing that a *more impatient* agent explores less in multi-arm bandit environments.

We now develop the simplest model of experimentation that allows for both time and risk preferences. This partially nests CDP who fix time preference and vary risk preferences, and in the spirit of QS who fix risk preference and vary time preferences. There is a single risky arms with unknown flow payoff given by the independent random variable $X_l \sim F^\theta \in \Delta(\mathbb{R}_+)$ where $\theta \in \Theta$ parameterizes the distribution over risky arns and $\theta \geq \theta' \implies F^\theta \geq_{FOSD} F^{\theta'}$. The safe arms pays out a flow payoff of 1 per unit time and to make the problem non-trivial we assume there exists $\underline{\theta} \in \Theta$ such that $\mathbb{E}[X|\underline{\theta}] < 1$. Payoffs are aggregated across levels (in QS, time) via exponential discounting so we can define:

$$v(\theta, l) := \mathbb{E}\Big[\int_0^l \alpha(s) \cdot \Big[u(X_s) - u(1)\Big] ds \Big| \theta\Big]$$

for discount factor $\alpha : \mathcal{L} \to \mathbb{R}_+$ with bounded variation
and flow utility $u : \mathbb{R} \to \mathbb{R}$.

The *patience order* of QS requires that for two discount factors α, α' their ratio $\alpha(l)/\alpha(l')$ is increasing in l. The main result of QS shows that the patience order

is quite generally a sufficient condition to order the duration of experimentation (fixing other aspects of the decision problem). Within the above bandit setting, the condition on patience implies the existence of increasing convex functions $(g^\theta)_\theta$ such that $u(\theta,\cdot) = g^\theta \circ v(\theta,\cdot)$ for each $\theta \in \Theta$. In fact, we can pick a single $g : \mathbb{R} \to \mathbb{R}$ since just taking partials at $l \in \mathcal{L}$:

$$\left(\frac{\partial u(\theta,\cdot)}{\partial l}\middle/\frac{\partial v(\theta,\cdot)}{\partial l}\right) = \frac{\alpha(l)\mathbb{E}[u(X_l)|\theta]}{\alpha'(l)\mathbb{E}[u(X_l)|\theta]} = \frac{\alpha(l)}{\alpha'(l)}$$

which via straightforward manipulation means that $g' = \alpha(l)/\alpha'(l)$ is increasing and hence g is convex.

The *flow risk aversion* from the previous section ranks u, v in the usual way: there exists some increasing convex function f such that $u = f \circ v$. Supposing that the discount rate is identical and suppose, as is assumed in Theorem 1 of CDP that we have that $u(1) \leq v(1)$ while for all $\theta \in \Theta$ we have $\mathbb{E}[u(X)|\theta] \geq \mathbb{E}[v(X)|\theta]$ then

$$g^\theta(l) = \frac{\mathbb{E}[u(X)|\theta] - u(1)}{\mathbb{E}[v(X)|\theta] - v(1)} \cdot l$$

with slope either $[0,1)$ when the state is such that the safe arm is better, and $[1,+\infty)$ if the state is such that the risky arm is better for the agent with flow utility $u = f \circ v$ and with slight modification of the proof of Lemma 1 we can show that this implies that the agent who is more flow risk-averse stops experimentation sooner.

Of course, there there are experimentation problems in which both impatience and flow payoffs cannot be ranked along the patience order and flow risk-aversion. But some of them can still be ranked according to our notion and, in those situations, Lemma 1 applies.

III EXAMPLES OF RISK-AVERSION

The following example illustrates different channels through which differential risk appetites (Assumption 1) might arise:

Example 2 (Constant relative risk aversion)**.** We consider the case of AI risk and follow Jones (2024) by considering CRRA utility over consumption; differently, we will suppose that the damage from a higher technological level in the dangerous state is continuous which allows us to capture a wide set of potential harms.[4] Consumption

[4] Jones (2024) focuses on existential risk which he models as the event that consumption is zero for

is a random variable determined by the state and technology level:

$$\underbrace{c(\theta=1,l)=c_0\cdot\exp\Big(a\cdot l\Big)}_{\text{Beneficial} \implies \text{grows at rate } a>0} \quad \text{and} \quad \underbrace{c(\theta=0,l)=c_0\cdot\exp\Big(-b\cdot l\Big)}_{\text{Harmful} \implies \text{shrinks at rate } b>0}$$

for some constant $c_0 > 0$. The principal's preferences over consumption (which might represent the preferences of society) is CRRA with risk-aversion parameter $\gamma_P \neq 1$.[5] The following examples all fulfill Assumption 1:[6]

1. **Intrinsic differences in risk aversion.** The agent's preferences over consumption is also CRRA, but the Principal has a higher risk-aversion coefficient than the agent i.e., $\gamma_P > \gamma_A$ by choosing

$$u = g(v) = \frac{\Big(v(1-\gamma_P)+1\Big)^{\frac{1-\gamma_A}{1-\gamma_P}}-1}{1-\gamma_A}$$

2. **Negative externalities.** The principal and agent's preferences are identical if the technology is safe but if the technology is harmful, the agent does not fully internalize those harms e.g., because of limited liability. This might be represented by $U(0,l) = \alpha \cdot V(0,l)$ for $\alpha < 1$ so

$$g(v) = \begin{cases} v & \text{for } v \geq 0 \\ \alpha \cdot v & \text{for } v < 0 \end{cases}$$

3. **Disproportionate winners.** The principal and agent's preferences are identical if the technology is unsafe but if the technology is beneficial, the agent benefits disproportionately relative to the rest of society. This might be represented by $U(1,l) = \beta \cdot V(1,l)$ for $\beta > 1$ so

$$g(v) = \begin{cases} \beta v & \text{for } v \geq 0 \\ v & \text{for } v < 0 \end{cases}$$

4. **R&D arms race.** There are n identical firms playing a continuous-time stopping

all future periods. It is also straightforward to do an analogous exercise for other kinds of technologies (maps from (l,θ) to consumption) and other preferences (maps from consumption to utility) e.g., CARA.

[5] That is, $v(c(\theta,l)) = \bar{u} + \frac{c^{1-\gamma_P}}{1-\gamma_P}$ for $\bar{u} > 0$ chosen to normalize zero technology to zero utility.

[6] See Online Appendix III for details.

game where learning is public among firms. Let $\boldsymbol{l} = (l_i)_{i=1}^n$ be the profile of firms' R&D. Firm i's payoff is $U(0, \boldsymbol{l}) = V(0, \max \boldsymbol{l})$ if the technology is dangerous (common disaster), and $U(1, \boldsymbol{l}) = \beta_i \cdot V(1, l_i)$ if it is safe (private benefit) where $\beta_1 > \beta_2 > \ldots \beta_n > 0$ represent heterogeneous payoffs across firms. Then in every SPE of this game, all firms choose the same stopping level that a representative firm with $U(0, l) = V(0, l)$ and $U(1, l) = \beta_1 \cdot V(1, l)$. ◊

III.1 Microfoundation for R&D arms race. We now develop a simple model of R&D arms races and show the claim made in Example 2 about equilibrium play. Let $\mathcal{F} \in \mathbb{F}$ be any learning process, and let $\widetilde{\mathcal{F}}$ be the augmented filtration that contains information about other players' stopping levels i.e.,

$$\widetilde{\mathcal{F}}_\ell = \mathcal{F}_\ell \vee \underbrace{\sigma(\ell_1, \ell_2, \ldots \ell_n)}_{\substack{\text{Natural filtration} \\ \text{generated} \\ \text{by } (\ell_1, \ell_2, \ldots \ell_n)}}$$

Strategies are stopping levels that are $\widetilde{\mathcal{F}}$–adapted—letting players condition on the stopping behavior of others players 'at the same time', therbey allowing for *instantaneous* responses. This simplifies our exposition, is intuitive, and well-defined (see Simon and Stinchcombe (1989)).

Lemma 15. *For any learning process $\mathcal{F} \in \mathbb{F}$ and any mechanism $\phi \in \Phi$, let $\mathsf{SPE}(\mathcal{F}, \phi)$ be the set of subgame-perfect equilibrium:*

(i) For any $\boldsymbol{\ell}^* \in \mathsf{SPE}(\mathcal{F}, \phi)$, $\ell_1 = \ell_2 = \ldots \ell_n =: \ell^*$ almost surely.

(ii) ℓ^* solves

$$\sup_{\ell} \mathbb{E}\Big[U_1(\mu_\ell, \ell) - \phi_\ell\Big] \quad \textit{s.t. } \ell \textit{ is } \mathcal{F}\textit{-adapted}$$

where $U_1(0, l) = V(0, l)$ and $U_1(1, l) = \beta_1 \cdot V(1, l)$.

Proof. Let ℓ^i be any selection of firm i's optimal stopping level that solves $\mathbb{E}[U_1(\mu_\ell, \ell) - \phi(\ell)]$.

To show (i), take any firm i that stops at level ℓ_i strictly before all agents have stopped i.e., $\ell_i < \max \boldsymbol{\ell} =: \ell^{max}$. Then observe on this event:

$$\sup_{\ell > \ell_i} \mathbb{E}\Big[U_i(\mu_\ell, \ell) \Big| \widetilde{\mathcal{F}}_{\ell_i}\Big] \quad \text{s.t. } \ell \text{ is } \widetilde{\mathcal{F}}\text{-adapted}$$

$$
\begin{aligned}
&= \sup_{\ell > \ell_i} \mathbb{E}\Big[\mu_\ell \cdot \beta_i u(1,\ell) + (1-\mu_\ell)\cdot u(0,\ell^{max})\Big|\widetilde{\mathcal{F}}_{\ell_i}\Big] \\
&\quad \text{s.t. } \ell \text{ is } \widetilde{\mathcal{F}} - measurable \\
&\geq \mathbb{E}\Big[\mu_{\ell^{max}} \cdot \beta_i u(1,\ell^{max}) + (1-\mu_{\ell^{max}})\cdot u(0,\ell^{max})\Big|\widetilde{\mathcal{F}}_{\ell_i}\Big] \\
&\qquad\qquad (\ell^{max} \text{ is } \widetilde{\mathcal{F}}\text{-adapted}) \\
&= \mathbb{E}\Big[\mathbb{E}\Big[\mu_{\ell^{max}} \cdot \beta_i u(1,\ell^{max}) + (1-\mu_{\ell^{max}})\cdot u(0,\ell^{max})\Big|\ell^{max}=s,\widetilde{\mathcal{F}}_{\ell_i}\Big]\Big|\widetilde{\mathcal{F}}_{\ell_i}\Big] \\
&\qquad\qquad \text{(iterated expectations)} \\
&= \mathbb{E}\Big[\mu_{\ell_i} \cdot \beta_i u(1,\ell^{max}) + (1-\mu_{\ell_i})\cdot u(0,\ell^{max})\Big|\widetilde{\mathcal{F}}_{\ell_i}\Big] \qquad \text{(martingale)} \\
&> \mathbb{E}\Big[U_i(\mu_{\ell_i},\ell)\Big|\widetilde{\mathcal{F}}_{\ell_i}\Big] \qquad (u(0,\cdot) \text{ is strictly decreasing})
\end{aligned}
$$

a contradiction.

We now show (ii). For any $\mathcal{F}$-stopping level ℓ, if at filtration $\mathcal{F}_\ell$ we have

$$
\begin{aligned}
&\sup_{s>\ell} \mathbb{E}\Big[U_1(\mu_s,s) - \phi_s\Big|\mathcal{F}_\ell\Big] \; < U_1(\mu_\ell,\ell) - \phi_\ell \\
&\text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}
\end{aligned}
$$

then this must also be true of all other firms. Since the number of firms is finite and firms can respond immediately, backward induction implies that all firms must stop at such histories so $\ell^* = \ell$ a.s. Now suppose, instead, that

$$
\begin{aligned}
&\sup_{s>\ell} \mathbb{E}\Big[U_1(\mu_s,s) - \phi_s\Big|\mathcal{F}_\ell\Big] \; > U_1(\mu_\ell,\ell) - \phi_\ell \\
&\text{s.t. } s \text{ is } \mathcal{F}\text{-adapted}
\end{aligned}
$$

then firm 1 has an individual deviation to do strictly more R&D since

$$
\begin{aligned}
\sup_{s>\ell} \mathbb{E}\Big[U_1(\mu_s,s) - \phi_s\Big|\widetilde{\mathcal{F}}_\ell\Big] \;&=\; \sup_{s>\ell} \mathbb{E}\Big[\mu_s \cdot U_1(1,s) + (1-\mu_s)\cdot U_1(0,\ell^{max}) - \phi_s\Big|\widetilde{\mathcal{F}}_\ell\Big] \\
\text{s.t. } s \text{ is } \widetilde{\mathcal{F}}\text{-adapted} \qquad &\quad \text{s.t. } s \text{ is } \widetilde{\mathcal{F}}\text{-adapted} \\
&=\; \sup_{s>\ell} \mathbb{E}\Big[\mu_s \cdot U_1(1,s) + (1-\mu_s)\cdot U_1(0,s) - \phi_s\Big|\widetilde{\mathcal{F}}_\ell\Big] \\
&\quad \text{s.t. } s \text{ is } \widetilde{\mathcal{F}}\text{-adapted} \\
&\qquad\qquad \text{(Backward induction from stopping at } s\text{)}
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{s>\ell} \mathbb{E}\Big[U_1(\mu_s, s) - \phi_s \Big| \mathcal{F}_\ell\Big] \\
&\quad \text{s.t. } s \text{ is } \mathcal{F}\text{-adapted} \\
&\quad \text{(Information about others' stopping levels is trivial)} \\
&> U_1(\mu_\ell, \ell) - \phi_\ell \qquad \text{(by hypothesis)}
\end{aligned}
$$

and the last term on the RHS is, in turn, an upper-bound on firm 1's payoff from stopping at ℓ so $\ell^* > \ell$ a.s. Hence ℓ^* solves the single-agent stopping problem in the statement of Lemma 15. □

III.2 Quantitative results on worst-case payoffs. (Example 1) Suppose that consumption is given by

$$
c(\theta, l) = \begin{cases} \exp(a \cdot l) & \text{if } \theta = 1 \\ \exp(-b \cdot l) & \text{if } \theta = 0 \end{cases} \quad \text{for } a, b > 0
$$

and that the agent and principal's utility of consumption is CRRA and given by:

$$
u(\theta, l) = \frac{c(\theta, l)^{1-\gamma_A}}{1-\gamma_A} \quad \text{and} \quad v(\theta, l) = \frac{c(\theta, l)^{1-\gamma_P}}{1-\gamma_P}
$$

where $1 \leq \gamma_A \leq \gamma_P$. Under the empty laissez-faire mechanism i.e., $\phi = 0$ a.s., we have that the lowest belief that can rationalize pushing the technology up to l is:

$$
\mu_U^*(l) = \frac{1}{1 + \frac{a}{b} \cdot \exp\Big(l \cdot (1-\gamma_A) \cdot (a+b)\Big)}
$$

so under weak optimism defined in the main text, we have that

$$
\mathbb{P}\Big(l^* \leq l + l_U^*(\mu_0)\Big) = 1 - \frac{\mu_0}{\mu_U^*(l)}
$$

and let $F : \mathcal{L} \to [0, 1-\mu_0]$ be the (unnormalized) CDF of the technology level l^*. Since whenever $l^* < L$ the agent conclusively learns the state is bad, this pins down the equilibrium joint distribution between beliefs and technology levels. Plugging this into the regulator's utility and standard manipulation yields:

$$
\mathbb{E}\Big[V(l, \mu)\Big] = \mu_0 \cdot V(1, L) + (1-\mu_0) \cdot \mathbb{E}\Big[V(\mu, l) \Big| \theta = 0\Big]
$$

$$= \mu_0 \cdot V(1,L) + \int_0^L V(0,l)dF(l)$$
$$= \mu_0 \cdot V(1,L) + \frac{\mu_0 \cdot a \cdot (\gamma_A - 1) \cdot (a+b)}{b \cdot (1-\gamma_P) \cdot \lambda} \cdot \Big(\exp(\lambda L) - 1 \Big)$$

where $\lambda := b(\gamma_P - 1) - (\gamma_A - 1)(a+b)$. Direct calculation yields:

$$\lim_{L \to +\infty} \mathbb{E}\Big[V(l,\mu) \Big] = -\infty \iff \frac{\gamma_P - 1}{\gamma_A - 1} > 1 + \frac{a}{b}$$

and it is immediate to replicate this analysis for linear Pigouvian taxes.

IV EXTENSIONS AND GENERALIZATIONS

In the Section 5 of the main text we outline directions in which our framework and results could be extended and generalized. Here we flesh out some details.

IV.1 No value of elicitation and randomization. We start by augmenting our space of adaptive mechanisms to allow for additional independent randomization and type reports:

- **Introducing additional randomization.** Recall we started on the space $(\Omega, \mathcal{G}, \mathbb{P})$ and since our definition of adaptive mechanisms were $\mathcal{G}$-adapted processes, this did not accommodate additional (independent) randomization. We now enlarge our filtered probability space with an independent Brownian motion. Define an auxillary space $(\Omega', \mathcal{G}', \mathbb{P}')$ that supports an independent Brownian motion $(B_l)_{l \in \mathcal{L}}$. Our augmented space is then

$$\Big(\Omega \times \Omega', \mathcal{G} \times \mathcal{G}', \mathbb{P} \times \mathbb{P}' \Big)$$

  and define $(\tilde{\mathcal{G}}_l)_l$ as the product filtration of $\mathcal{G}$ and $\mathcal{G}'$.[7]

- **Introducing additional elicitation.** Consider a richer space of adaptive mechanisms in which the regulator attempts to elicit reports and condition future transfers on past reports. Let $\boldsymbol{r} := (r_s)_{s \leq l}$ denote a path of reports up to level l, each taking values in the rich message space $\mathbb{R}$. Let $\mathcal{R}_l$ denote the set of all possible

[7] That is, $\tilde{\mathcal{G}}_l = \sigma\Big(\mathcal{G}_s \otimes \{\emptyset, \Omega'\} \cup \{\emptyset, \Omega'\} \otimes \mathcal{G}'_s : s \leq l \Big)$ which is the standard way of defining the product filtration via cylinders.

paths of reports up to level l and $\mathcal{R} := \bigcup_{l \in \mathcal{L}} \mathcal{R}_l$. Define

$$\widehat{\mathcal{G}}_l := \tilde{\mathcal{G}}_l \vee \sigma(R_l)$$

as the filtration augmented with the path of reports.

An *augmented adaptive mechanism* is a $(\widehat{\mathcal{G}}_l)_l$-adapted stochastic process that satisfies the standard regularity conditions in the main text i.e., uniform integrability where it is finite, and lower-semicontinuity. Let $\widehat{\Phi}$ denote this space of augmented adaptive mechanisms. Given $\widehat{\phi} \in \widehat{\Phi}$ the agent solves a joint stopping and control problem of choosing both an (i) $\mathcal{F}$-adapted stopping time; and (ii) a dynamic reporting strategy which is a progressively measurable $\mathbb{R}$-valued process adapted to $\mathcal{F}$. Clearly the agent does not require additional randomization since (i) she is 'best-responding' against $\widehat{\phi}$; and (ii) is maximizing expected payoffs. We let $\ell^*(\widehat{\phi}, \mathcal{F}, U)$ denote the largest stopping time that is part of a solution to the agent's joint problem.

Proposition 4. *Randomization and elicitation have no value:*

$$\sup_{\widehat{\phi} \in \widehat{\Phi}} \inf_{\mathcal{F} \in \mathbb{F}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\widehat{\phi}, \mathcal{F}, U)\Big)\Big] = \sup_{\phi \in \Phi} \inf_{\mathcal{F} \in \mathbb{F}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*(\phi, \boldsymbol{\mu}, U)\Big)\Big] =: (\mathrm{L}).$$

Proof sketch. We sketch the argument since it follows that of Theorem 1 with a few modifications. First notice that Step 1 of the proof of Theorem 1 continues to apply: no additional information ($\mathcal{F} = \mathcal{G}$) remains an admissible choice by nature. Facing this mechanism, the adaptive sandbox ϕ^* (that does not require randomization) is optimal—loosely speaking, we can find an adaptive mechanism $\widehat{\phi} \in \widehat{\Phi}$ that is 'driven only by variation' in the projection of $\widehat{\mathcal{G}}$ onto $\mathcal{G}$. The point is that at this saddle point, there is no value to (i) elicitation since the agent has no extra information; and (ii) randomization since extra randomization cannot improve solutions to the principal's optimal experimentation problem. Step 2 then proceeds similarly to that of Theorem 1—under any alternate $\mathcal{F} \in \mathbb{F}$ the principal is made weakly better-off, and putting these steps together yields the result. □

The fact that the regulator cannot do better with elicitation is reminiscent of Kruse and Strack (2015, 2019); the fact that the regulator cannot do better with randomization is in contrast to Libgober and Mu (2021).

IV.2 Endogenous agent learning. Suppose that instead of exogeneously learning about the state as the technology develops, the learning process is optimally chosen by the agent at some cost which is increasing in the flow variation of the belief

process. This framework was proposed in the important papers of Zhong (2022); Hébert and Woodford (2023).[8] We augment it to accommodate outside learning (via $\mathcal{G}$)). Some notation:

Let $H : \Delta(\Theta) \to \mathbb{R}$ be a concave and continuous measure of uncertainty and $C : \mathbb{R} \to \mathbb{R}$ is a weakly increasing cost function with $C(0) = 0$. The *flow information* at level l is given by the random variable

$$I_l := \mathbb{E}\left[\lim_{l' \downarrow l} \frac{H\Big(\mathbb{E}[\theta|\mathcal{F}_{l'}]\Big) - H\Big(\mathbb{E}[\theta|\mathcal{F}_l \vee \mathcal{G}_{l'}]\Big)}{l' - l}\middle| \mathcal{F}_l\right]$$

Notice here that our variation measure (i) is *infinitesimal* i.e., it depends on how much beliefs move between l and l' (resembling the infinitesimal generator for Feller semigroups); (ii) depends only on movement in *first-order beliefs* about the state, thereby following a long tradition in information economics; and (iii) measures the *excess movement* since change is measured to the agent's 'counterfactual' level-l' belief *as if* she did not acquire any excess information interim information, but continued to observe the outside learning process so her first-order belief is given by $\mathbb{E}[\theta|\mathcal{F}_l \vee \mathcal{G}_{l'}]$.

Facing the mechanism ϕ and outside learning process $\mathcal{G}$, the agent thus solves the following stochastic control problem:

$$\sup_{\mathcal{F} \in \mathbb{F}, \ell} \mathbb{E}\left[U(\mu, \ell) - \phi(\ell) - \int_0^\ell C(I_s) ds\right]$$

where ℓ is a stopping level adapted to $(\mathcal{F}_l)_l$. Note that the choice of $\mathcal{F} \in \mathbb{F}$ already specifies a complete and contingent plan for how (first-order) beliefs from μ_l evolve. Let $\ell^*(\phi, (H, C), U)$ denote the agent's optimal solution to the above problem.

We suppose, instead, that the principal is unsure how hard it is for the agent to acquire additional information about the technology. To reflect this uncertainty, we let $\mathcal{E}$ denote the set of cost functions and uncertainty measure (C, H) pairs that fulfill the above assumptions. The adaptive sandbox ϕ^* remains robustly optimal when the ambiguity set is $\mathcal{E}$.

[8] Our environment is nearby, but non-nested within theirs: our agent's action space is singleton and the payoffs from stopping at different levels varies with the state. By contrast, Zhong (2022); Hébert and Woodford (2023) assumes that the agent always dislikes delay.

Proposition 5 (Robust regulation with endogenous agent information)**.** ϕ^* *solves*

$$\sup_{\phi\in\Phi}\inf_{(C,H)\in\mathcal{E}}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,(H,C),U)\Big)\Big]$$
$$\textit{and}\quad \sup_{\phi\in\Phi}\inf_{\substack{(C,H)\in\mathcal{E}\\U\in\mathbb{U}}}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,(H,C),U)\Big)\Big]$$

Proof sketch. Step 1 of the proof of Theorem 1 is unchanged, choosing the saddle point in which it is infinitely costly / impossible to acquire excess information over $\mathcal{G}$ so $\mathcal{F}=\mathcal{G}$. The challenge now is to show that away from the 'infinite cost' case, the all solutions to the agent's endogenous information acquisition problem must weakly improve the principal's payoff.

Given $\phi\in\Phi$ and $(H,C)\in\mathcal{E}$, given the agent's solution of endogenous information acquisition and stopping $\mathcal{F},\ell^*$ we will construct an *exogeneous* filtration $\mathcal{F}^{\mathsf{x}}$ that induces an identical joint distribution over stopping levels and first-order beliefs as follows:

$$\mathcal{F}^{\mathsf{x}}_l := \begin{cases}\mathcal{F}_l & \text{if } l\le\ell^*\\ \mathcal{F}_{\ell^*}\vee\mathcal{G}_l & \text{otherwise.}\end{cases}$$

Now let ℓ^{x} denote the agent's solution facing mechanism ϕ and *exogenous* information $\mathcal{F}^{\mathsf{x}}$. We proceed via cases.

Case 1: $\ell^{\mathsf{x}}<\ell^*$. Since the agent found it strictly better to continue at ℓ^{x} (and acquire information at some cost) we have

$$\begin{aligned}
0 &< \sup_{\ell>\ell^{\mathsf{x}}}\mathbb{E}\Big[U(\mu_\ell,\ell)-\int_{\ell^{\mathsf{x}}}^{\ell^*}C(I_s)ds-U\Big(\mathbb{E}[\theta|\mathcal{F}_{\ell^{\mathsf{x}}}],\ell^{\mathsf{x}}\Big)\Big|\mathcal{F}_{\ell^*}\Big] && (\ell^*>\ell^{\mathsf{x}})\\
&\le \sup_{\ell>\ell^{\mathsf{x}}}\mathbb{E}\Big[U(\mu_\ell,\ell)-U\Big(\mathbb{E}[\theta|\mathcal{F}_{\ell^{\mathsf{x}}}],\ell^{\mathsf{x}}\Big)\Big|\mathcal{F}_{\ell^*}\Big] && (C\ge 0)\\
&= \sup_{\ell>\ell^{\mathsf{x}}}\mathbb{E}\Big[U(\mu_\ell,\ell)-U\Big(\mathbb{E}[\theta|\mathcal{F}_{\ell^{\mathsf{x}}}],\ell^{\mathsf{x}}\Big)\Big|\mathcal{F}^{\mathsf{x}}_\ell\Big] && (\text{Construction of }\mathcal{F}^{\mathsf{x}})\\
&\le 0
\end{aligned}$$

a contradiction.

Case 2: $\ell^{\mathsf{x}}>\ell^*$. If the agent finds it strictly better to continue under the exogenous learning process past ℓ^*, then consider instead the deviation by the agent under the endogenous process under which at level ℓ^*, the agent does not acquire any further

information but instead replicated ℓ^{x}. At ℓ^* we have

$$\begin{aligned}
0 &< \sup_{\ell>\ell^*} \mathbb{E}\Big[U(\mu_\ell,\ell) - U\Big(\mathbb{E}[\theta|\mathcal{F}_{\ell^{\mathsf{x}}}],\ell^{\mathsf{x}}\Big)\Big|\mathcal{F}^{\mathsf{x}}_{\ell^*}\Big] && (\ell^* < \ell^{\mathsf{x}}) \\
&= \mathbb{E}\Big[U(\mu_\ell,\ell) - \underbrace{\int_{\ell^*}^{\ell^{\mathsf{x}}} C(I_s)ds}_{=0} - U\Big(\mathbb{E}[\theta|\mathcal{F}_{\ell^*}],\ell^*\Big)\Big|\mathcal{F}_{\ell^*}\Big] && \text{(Construction of deviation)} \\
&\leq \sup_{\substack{\mathcal{F}'\in\mathbb{F}|_{\mathcal{F}_\ell} \\ \ell>\ell^*}} \mathbb{E}\Big[U(\mu_\ell,\ell) - \int_{\ell^*}^{\ell^{\mathsf{x}}} C(I_s)ds - U\Big(\mathbb{E}[\theta|\mathcal{F}_{\ell^*}],\ell^*\Big)\Big|\mathcal{F}'_{\ell^*}\Big] && \text{(Replicating } \ell^{\mathsf{x}} \text{ is feasible)} \\
&\leq 0
\end{aligned}$$

a contradiction, where $\mathbb{F}_{\mathcal{F}_l}$ is the set of filtrations consistent with $\mathcal{F}_l$. Hence $\ell^{\mathsf{x}} = \ell^*$ a.s. Moreover, by construction of $\mathcal{F}^{\mathsf{x}}_\ell$ we have $\mathbb{E}[\theta|\mathcal{F}^{\mathsf{x}}_{\ell^*}] = \mathbb{E}[\theta|\mathcal{F}_{\ell^*}]$ so the induced joint distribution over stopping levels and beliefs are identical. We have showed that under any $\phi \in \Phi$ and any $(H,C) \in \mathcal{E}$, we can find some $\mathcal{F} \in \mathbb{F}$ so that the regulator's payoffs are identical. Then proceeding with Step 2 of the proof of Theorem 1 completes the result. □

IV.3 Revenue motives. In the main text we supposed that our principal has no revenue motives. Now consider the case in which the principal's payoff is

$$V(\mu_\ell,\ell) + \lambda \cdot \phi_\ell$$

for $\lambda > 0$. Suppose $U \in \mathbb{U}$ is known (so we focus on learning-robustness) and construct the sandbox ϕ^{R} as follows:

$$\phi^{\mathsf{R}}_l = \begin{cases} \mathsf{R}(U,\mathcal{G}) & \text{if } l \leq \ell_{\mathsf{R}} \\ +\infty & \text{otherwise.} \end{cases}$$

where ℓ_{R} solves

$$\sup_{\ell} \Big(V(\mu_\ell,\ell) + \lambda \cdot U(\mu_\ell,\ell)\Big)$$
$$\text{s.t. } \ell \text{ is } \mathcal{G}\text{-adapted.}$$

and

$$\mathsf{R}(U,\mathcal{G}) := \mathbb{E}\Big[U(\mu_{\ell_{\mathsf{R}}},\ell_{\mathsf{R}}) - U(\mu_0,0)\Big].$$

Proposition 6 (Robust regulation with revenue motives)**.** ϕ^{R} *solves*

$$\sup_{\phi\in\Phi}\inf_{\mathcal{F}\in\mathbb{F}}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{F},U)\Big)+\phi_{\ell^*}\Big].$$

Proof sketch. Only Step 1 of Theorem 1 needs to be modified: it is easy to see that given mechanism ϕ^{R} and under no additional information ($\mathcal{G}=\mathcal{F}$), the agent's participation constraint is tight since

$$\sup_{\ell\le\ell_{\mathsf{R}}}\mathbb{E}\Big[U(\mu_\ell,\ell)-\phi_\ell\Big]\ge\mathbb{E}\Big[U(\mu_{\ell_{\mathsf{R}}},\ell_{\mathsf{R}})-\phi_{\ell_{\mathsf{R}}}\Big]=U(\mu_0,0)$$

because continuing until the sandbox boundary is a feasible stopping time for the agent. Now for the principal, notice that ϕ^{R} is the best mechanism against $\mathcal{F}=\mathcal{G}$ facing this participation constraint since

$$\begin{aligned}
&\sup_{\phi\in\Phi}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)+\lambda\cdot\phi_{\ell^*}\Big]\\
&\quad=\sup_{\phi\in\Phi}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)\\
&\qquad\qquad+\underbrace{\lambda\cdot\phi_{\ell^*}-\lambda\cdot U\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)}_{\le -U(\mu_0,0)\text{ from participation}}+\lambda\cdot U\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)\Big]\\
&\quad\le\sup_{\phi\in\Phi}\mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)+\lambda\cdot U\Big(\mu_{\ell^*},\ell^*(\phi,\mathcal{G},U)\Big)\Big]-\lambda\cdot U(\mu_0,0).
\end{aligned}$$

and it is easy to verify that ϕ_{R} implements this upper-bound since by Lemma 1 under ϕ_{R} and $\mathcal{G}=\mathcal{F}$, $\ell^*=\ell_{\mathsf{R}}$ and ℓ_{R} is by definition the $\mathcal{G}$-stopping level that maximizes the term inside the brackets. Now Step 2 proceeds similarly to that of Theorem 1, noting that for any alternate learning process $\mathcal{F}\in\mathbb{F}$, the agent's participation constraint is still fulfilled (from Greenshtein (1996)) and a similar argument to that of Theorem 1 yields that any deviation (weakly) increases the principal's payoff to weakly increase since revenue, being extracted upon participation, is unaffected. □

IV.4 Undominated Stopping Rules. In the main text we assumed that the agent knew the learning process $\mathcal{F}\in\mathbb{F}$ and, facing the mechanism ϕ, solved her optimal stopping problem (O). This served as a helpful Bayesian benchmark in which the agent has (i) correct beliefs about the learning process she faces; and (ii) solves the optimal stopping problem accordingly. We now relax both of these assumptions

which will allow us to capture a wider class of agent behavior e.g., misspecifcation, ambiguity aversion, myopia, etc.

We start by defining undominated stopping rules. The set of *dominated* belief-level pairs under preference f is:

$$\mathsf{DOM}[f] := \left\{(\mu, l) \in \Delta(\Theta) \times \mathcal{L} : l \notin \operatorname{argmax}_{s \geq l} f(\mu, s)\right\}$$

That is, at belief μ, even in the absence of further information, the agent stops too early—she can do strictly better by pushing the technology further.

Definition 5. Fix the learning process $\mathcal{F} \in \mathbb{F}$ and preference $U \in \mathbb{U}$. The $\mathcal{F}$-adapted stopping level ℓ is *undominated* if

$$\mathbb{P}\Big((\mu_\ell, \ell) \in \mathsf{DOM}[U]\Big) = 0.$$

and let $\mathsf{USR}[U, \mathcal{F}]$ be the set of undominated stopping rules.

Undominated stopping rules include:

1. **Bayes optimal** stopping levels that solve (O) i.e., the agent has correct beliefs about the learning process. This was studied in the main text.

2. **Ambiguity** where the agent faces non-Bayesian uncertainty about the law of future information and solves

$$\sup_{\ell} \inf_{\mathcal{F}' \in \mathbb{F}} \mathbb{E}^{\mathcal{F}'}\Big[U^{\phi}(\mu_\ell, \ell)\Big] \quad \text{s.t. } l \text{ is } (\mathcal{F}_l)_l\text{-adapted}$$

   where we use $\mathbb{E}^{\mathcal{F}}$ to denote the expectation under the learning process $\mathcal{F}$.[9] This is closely related to the formulation in Riedel (2009) who studies optimal stopping under ambiguity.[10] Note that here, ambiguity is about the law of the learning process i.e., how informative future experiments will be as opposed to having ambiguous priors (scenario 2 in Epstein and Schneider (2007)[11]) or ambiguous signals (scenario 3).

3. **Misspecification** in which the agent has misspecified beliefs about the law of information i.e., the true law is not in the support of her prior and she solves her stopping problem under this erroneous belief.

[9] Here we suitably extend the probability space.
[10] The difference is that in Riedel (2009), the process directly specifies payoffs.
[11] See also Auster, Che, and Mierendorff (2024)

4. **Myopia** in which the agent does not internalize the learning value of pushing the technology, and simply chooses the best technology level under her present beliefs.

We first establish that ϕ^* remains learning- and dually-robust whenever the agent's stopping level is undominated. We will impose an additional regularity condition that $U(\mu,\cdot)$ is quasiconcave.

Proposition 7 (Robustness against undominated stopping rules)**.** *Suppose that $U(\mu,\cdot)$ is quasiconcave for each $\mu \in \Delta(\Theta)$ and that the principal does not learn. Then the sandbox ϕ^* with hard limit $\overline{\ell}$ is both learning- and dually-robust under any undominated stopping rule i.e., ϕ^* solves*

$$\sup_{\phi\in\Phi} \inf_{\substack{\mathcal{F}\in\mathbb{F}\\ \ell\in \mathsf{USR}[U^\phi,\mathcal{F}]}} \mathbb{E}\Big[V(\mu_\ell,\ell)\Big] \quad \text{and} \quad \sup_{\phi\in\Phi} \inf_{\substack{\mathcal{F}\in\mathbb{F}\\ U\in\mathbb{U}\\ \ell\in \mathsf{USR}[U^\phi,\mathcal{F}]}} \mathbb{E}\Big[V(\mu_\ell,\ell)\Big].$$

We next formalize the claim made in our discussion after Theorem 4 that the same worst-case payoff remains possible under an adversarially chosen learning process whenever the agent is using an undominated stopping rule.

Proposition 8 (Worst-case under undominated stopping rules)**.** *For any mechanism ϕ such that $U^\phi = g \circ V$ for some convex function g:*

$$\inf_{\mathcal{F}\in\mathbb{F}} \sup_{\ell\in \mathsf{USR}[U^\phi,\mathcal{F}]} \mathbb{E}\Big[V(\mu_\ell,\ell)\Big] = \inf_{\mathcal{F}\in\mathbb{F}} \mathbb{E}\Big[V\Big(\mu_{\ell^*},\ell^*\big(\phi,\boldsymbol{\mu},U\big)\Big)\Big]$$

i.e., the worst-case principal payoffs under her best undominated stopping rule is identical to that under the optimal stopping level analyzed in the main text.

This follows quite quickly by taking the construction of weak optimism (the worst-case learning process from Theorem 4) and verifying that at each interim history, the agent finds it optimal to continue under her present belief so long as conclusive bad news has not arrived. The only subtlety is that this, by itself, does not guarantee that all undominated stopping rules generate this behavior because of ties—since weak optimism is constructed to keep the agent indifferent, it is also an undominated stopping rule to break ties in favor of stopping. But we can suitably perturb the learning process to make continuation incentives strict, and verify that the principal's payoff is continuous in this perturbation to yield Proposition 8.

V BRIDGING ROBUSTNESS AND BAYES

In Proposition 2 we made the (informal) claim that in the Bayesian problem with a sufficiently diffuse prior $f \in \Delta(\mathbb{F} \times \mathbb{U})$, Bayes-optimal mechanisms required hard limits. We now state a more precise version of this result that requires us to clarify what it means for a prior to be 'diffuse'.

There, of course, some conceptual difficulties since the space of learning processes $\mathbb{F}$ and preferences $\mathbb{U}$ are both large and difficult to describe. Indeed, this is sometimes held as a reason for robustness. We will try nonetheless, keeping in mind that by describing a class of learning processes and preferences that we can put a prior over, we will necessarily rule out many others. Thus, our terminology 'diffuse' is only with respect to the class we have chosen to describe. At the same time, the reasoning behind Proposition 2 is straightforward and in the interest of transparency, we will pick particularly simple parametric classes to put priors over; we can do the same exercise with fancier classes to obtain the same qualitative result.

Continue with the setting of Example 1: the principal has CRRA preferences over consumption with risk-aversion parameter $\gamma_P > 0$. Consumption is given by

$$c(\theta, l) = \begin{cases} \exp\Big(a \cdot l\Big) & \text{if } \theta = 1 \\ \exp\Big(-b \cdot l\Big) & \text{if } \theta = 0 \end{cases}$$

Suppose that $\tilde{\mathbb{U}}$ consists of the set of CRRA preferences over consumption with *level* λ and risk-aversion parameter $\gamma_A \in [0, \gamma_P]$ so that for an agent of preference with parameter (λ, γ) utility is

$$u(\theta, l) = \lambda^{1-\gamma} \cdot \frac{c(\theta, l)^{1-\gamma}}{1-\gamma}.$$

we have thus written down a simple two-parameter family of agent preferences. Let $\widetilde{\mathbb{F}}$ denote the set of Levy signal processes

$$dX_l = \theta \cdot dl + \sigma \cdot dB_l + J \cdot dN_l$$

where $\sigma \geq 0$ is the degree of Brownian interference and N_l is a Poisson process with intensity $\lambda_\theta \geq 0$ in state θ, and $\lambda_0 \neq \lambda_1$. We have thus written down the simplest three-parameter family $(\sigma, \lambda_0, \lambda_1)$ we can put a prior over.

Our prior is thus over the smaller parametric class

$$P \in \Delta\left(\widetilde{\mathbb{U}} \times \widetilde{\mathbb{F}}\right)$$

which is equivalently over the family of parameters $\langle(\lambda, \gamma), (\sigma, \lambda_0, \lambda_1)\rangle$. Say it is *diffuse* if it is fully supported on this parameter space. Proposition 2 follows.

VI EXISTENCE AND SELECTION OF STOPPING LEVELS

In the main text we made a specific selection of equilibrium experimentation—fixing the mechanism ϕ, the agent's learning process $\mathcal{F}$ and preferences U, we focused on the largest stopping level $\ell^*(\phi, \mathcal{F}, U)$ that solved the agent's problem (O). In so doing, we presumed (i) the existence of *some* stopping level; (ii) the existence of a *largest* stopping level; and, perhaps most substantively, (iii) that our selection was not driving results on instrument performance. We deal with each of these concerns in turn.

VI.1 Existence of optimal stopping levels. The set of optimal stopping levels is:

$$\mathcal{O}(\phi, \mathcal{F}, U) := \left\{\ell : \ell \ \mathcal{F}\text{-adapted, solves (O)}\right\}$$

denote the set of optimal stopping levels.

We first claim $\mathcal{O}(\phi, \mathcal{F}, U) \neq \emptyset$. Recall the agent solves

$$\sup_{\ell} \mathbb{E}\Big[\underbrace{U(\mu_\ell, \ell) - \phi_\ell}_{:=X_\ell}\Big] \quad \text{s.t. } \ell \text{ is } \mathcal{F}\text{-adapted.}$$

The process $(X_l)_l$ is (i) a.s. upper-semicontinuous since $U(\mu, \cdot)$ is continuous and ϕ_l is lower-semicontinuous; and (ii) class (D) since U is bounded and ϕ_ℓ is class (D) so their sum must also be class (D). This implies that the Snell envelope $(Z_t)_t$ exists where $Z_l := \sup_{\ell \geq l} \mathbb{E}[X_\ell | \mathcal{F}_l]$, and so too does an optimal stopping level (these arguments are standard; see El Karoui (1979)).

We next claim that there exists a largest optimal stopping level

$$\ell^* := \operatorname{ess\,sup} \mathcal{O}^*(\phi, \mathcal{F}, U)$$

where here we take the essential supremum because the set of optimal stopping

times is potentially uncountable.

It will be helpful to define a helpful property for stochastic processes—*upper-semicontinuous in expectation along stopping levels* (USCE)—as follows:

Definition 6. The stochastic process $(Z_l)_l$ is right-USCE (left-USCE) if for any stopping level ℓ and for all sequences of stopping levels (ℓ_n) such that $\ell_n \downarrow \ell$ ($\ell_n \uparrow \ell$):

$$\mathbb{E}[Z_\ell] \geq \limsup_{n\to\infty} \mathbb{E}[Z_{\ell_n}].$$

Z is USCE if it is both right- and left-USCE.

We now check that $X := V - \phi$ is USCE. Recall that we have already argued that it is of class (D), and a.s. i.e., pathwise upper-semicontinuous. We show both right- and left-USCE togeher.

Fix the stopping level ℓ and take any sequence of stopping levels $(\ell_n)_n$ such that $\ell_n \to \ell$. This holds for either $\ell_n \uparrow \ell$ or $\ell_n \downarrow \ell$. Since paths are upper-semicontinuous

$$\limsup_{s\uparrow l} X_s \leq X_l \quad \text{a.s. for all } l \in \mathcal{L}.$$

which means that along our sequence of stopping levels $\ell_n \uparrow \ell$:

$$\limsup_{n\to\infty} X_{\ell_n} \leq X_\ell \quad \text{a.s.}$$

and from Fatou:

$$\limsup_{n\to\infty} \mathbb{E}[X_{\ell_n}] \leq \mathbb{E}[\limsup_{n\to\infty} X_{\ell_n}] \leq \mathbb{E}[X_\ell].$$

We have shown that $X := V - \phi$ is USCE. Now apply Theorem 2.15 of Kobylanski and Quenez (2012) to conclude that ℓ^* is, in fact, an $\mathcal{F}$-stopping level.

VI.2 Adversarial selection of optimal stopping levels. Recall $\mathcal{O}(\phi, \mathcal{F}, U)$ denotes the set of stopping levels that solve (O).

Consider an alternative selection of optimal stopping levels in which nature also adversarially choses the agent's optimal stopping levels:

$$\sup_{\phi\in\Phi} \inf_{\substack{\mathcal{F}\in\mathbb{F} \\ \ell^*\in\mathcal{O}(\phi,\mathcal{F},U)}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*\Big)\Big] \tag{L-ADV}$$

$$\sup_{\phi\in\Phi} \inf_{\substack{(\mathcal{F},U)\in\mathbb{F}\times\mathbb{U} \\ \ell^*\in\mathcal{O}(\phi,\mathcal{F},U)}} \mathbb{E}\Big[V\Big(\mu_{\ell^*}, \ell^*\Big)\Big] \tag{D-ADV}$$

Note that the proof of robustness (Theorem 1) does not rely on any selection of stopping levels hence ϕ^* remains learning- and dually-robust under this alternate selection.

Next, we argue that the proof of Proposition 1 carries through for adversarial selections of optimal stopping levels. For sufficiency of limit with option mechanisms, observe that the proof of Lemma 7 only used (i) the fact that the agent was choosing *a* optimal stopping level; and (ii) Lemma 2 that applies to any pair of stopping times (optimal or not). For necessity, the argument from Lemmas 4 and 5 on the necessity of the option and limit goes through as stated by the same construction of the agent's learning process and preferences since an adversarially chosen stopping time continues to do strictly worse than the payoff guarantee.

We thus obtain the following versions of Theorem 1 and Proposition 1:

Proposition 9. *The adaptive sandbox ϕ^* is learning-robust and dually-robust under adversarial selections of optimal stopping time i.e., it solves* (L-ADV) *and* (D-ADV)*. ϕ is dually-robust under adversarial selection of stopping levels if and only if it is a limit with option mechanism induced by $\overline{\ell}$.*

Next, we discuss alternative selections of stopping levels for dominance and undominated mechanisms. For two adaptive mechanism $\phi, \phi' \in \Phi$, say ϕ dominates ϕ' under adversarial selection of stopping levels (ASSL) if, for any learning process $\mathcal{F} \in \mathbb{F}$ and any agent preference $U \in \mathbb{U}$,

$$\inf_{\ell^* \in \mathcal{O}(\phi, \mathcal{F}, U)} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*)\Big] \geq \inf_{\ell^* \in \mathcal{O}(\phi', \mathcal{F}, U)} \mathbb{E}\Big[V(\mu_{\ell^*}, \ell^*)\Big].$$

with strict inequality for some $(\mathcal{F}, U) \in \mathbb{F} \times \mathbb{U}$. A mechanism is undominated under adversarial selection if it is not dominated under ASSL.

The proof of Theorem 2 holds as is under adversarial selection of the agent's stopping level which yields:

Proposition 10. *ϕ^* is undominated under ASSL. ϕ^* dominates under ASSL all other dually-robust and regular mechanisms.*

VII STATIC VS DYNAMIC REGULATION

In the main text we analyzed a setting in which the principal faced uncertainty about the agent's learning process. This uncertainty encodes *both* uncertainty over the *law* of the signal process, as well as its *realizations* over time. In this appendix, we

consider variations of our setting. This will clarify what is distinctive—and difficult—about dynamic regulation in the face of uncertain learning, and connect our results to the work of Kruse and Strack (2015, 2019).

Static information with unknown law. We focus on the case in which the principal does not learn i.e., $\mathcal{G}_l = \mathcal{G}_{0-}$ for all $l \in \mathcal{L}$, and the agent's preference $U \in \mathbb{U}$ is known. Consider a special case of our model in which the agent observes information *only* at level-0. That is, write:

$$\mathbb{F}^{\mathsf{s}} := \left\{ \mathcal{F} \in \mathbb{F} : \mathcal{F}_l = \mathcal{F}_0 \text{ for all } l \in \mathcal{L} \right\}$$

recalling that we allow for information to arrive at level $l = 0$ since the filtration is right-continuous i.e., we can have $\mathcal{F}_{0-} \neq \mathcal{F}_0$. We will make the mild assumption that $\mu \in \text{Int}\Delta(\Theta)$ implies $l_V^*(\mu) \in \text{Int}\mathcal{L}$. That is, for any interior belief the principal chooses an interior technology level.

Observe that a necessary and sufficient condition for ϕ^{s} to correct the externality belief-by-belief is for each $\mu \in \Delta(\Theta)$:

$$U\Big(\mu, l_V^*(\mu)\Big) - \phi^{\mathsf{s}}\Big(l_V^*(\mu)\Big) = \underbrace{\max_l \Big\{ U(\mu, l) - \phi^{\mathsf{s}}(l) \Big\}}_{=:U^*(\mu)}.$$

If this holds, the envelope theorem gives:

$$\frac{dU^*(\mu)}{d\mu} = U(1, l_V^*(\mu)) - U(0, l_V^*(\mu))$$

so for each belief $\mu \in \Delta(\Theta)$ choose

$$\phi^{\mathsf{s}}\Big(l_V^*(\mu)\Big) = U\Big(\mu, l_V^*(\mu)\Big) - \int_0^{\mu} \Big[U(1, l_V^*(\nu)) - U(0, l_V^*(\nu)) \Big] d\nu.$$

We can write ϕ^{s} entirely in terms of l by changing variables:

$$\phi^s(l) = U\Big((l_V^*)^{-1}(l), l\Big) - \int_0^{l} \Big[U(1, s) - U(0, s) \Big] d\Big((l_V^*)^{-1}\Big)(s).$$

It remains to verify that choosing $l_V^*(\mu)$ under belief μ is optimal for the agent.

Consider that

$$\begin{aligned}\frac{d\phi^{\mathsf{s}} \circ l_V^*}{d\mu} &= \mu \cdot \frac{\partial U}{\partial l}\Big(1, l_V^*(\mu)\Big) \cdot \frac{dl_V^*(\mu)}{d\mu} + (1-\mu) \cdot \frac{\partial U}{\partial l}\Big(0, l_V^*(\mu)\Big) \cdot \frac{dl_V^*(\mu)}{d\mu} \\ &= \frac{dl_V^*(\mu)}{d\mu} \cdot \frac{\partial U}{\partial l}\Big(\mu, l_V^*(\mu)\Big),\end{aligned}$$

where the first equality is from noting that the terms $U(1, l_V^*(\mu)) - U(0, l_V^*(\mu))$ cancel out with itself; the second is from the definition of $U(\mu, \cdot)$. However,

$$\begin{aligned}&\frac{d\phi^{\mathsf{s}} \circ l_V^*}{d\mu} = \frac{dl_V^*(\mu)}{d\mu} \cdot (\phi^s)'\Big(l_V^*(\mu)\Big) \\ &\implies (\phi^s)'\Big(l_V^*(\mu)\Big) = \frac{\partial U}{\partial l}\Big(\mu, l_V^*(\mu)\Big) \\ &\implies (\phi^s)'(l) = \frac{\partial U}{\partial l}\Big((l_V^*)^{-1}(l), l\Big).\end{aligned}$$

The first-order condition of the agent's payoff net of the mechanism ϕ^{s} is thus:

$$\begin{aligned}\frac{\partial}{\partial l}\Big(U(\mu, l) - \phi^s(l)\Big) &= \frac{\partial U}{\partial l}(\mu, l) - \frac{\partial U}{\partial l}\Big((l_V^*)^{-1}(l), l\Big) \\ &= \underbrace{\left(\frac{\partial U}{\partial l}(1, l) - \frac{\partial U}{\partial l}(0, l)\right)}_{>0} \cdot \Big(\mu - (l_V^*)^{-1}(l)\Big).\end{aligned}$$

Fact. l_V^* is strictly increasing.

To see this, observe that a necessary condition for $l_V^*(\mu)$ is that it must solve

$$\left|\frac{\partial V/\partial l(1, l_V^*(\mu))}{\partial V/\partial l(0, l_V^*(\mu))}\right| = \frac{\mu}{1-\mu}.$$

But since the function $x \mapsto \frac{x}{1-x}$ is strictly increasing, then for any $\nu \neq \mu$, $l_V^*(\nu) \neq l^*(\mu)$, But since l_V^* is weakly increasing in μ it must also be strictly increasing.

Since l_V^* is strictly increasing so is $(l_V^*)^{-1}$. Hence, the agent's first-order condition is weakly greater than 0 if and only if $\mu \geq (l_V^*)^{-1}(l)$. This is equivalent to $l \leq l_V^*(\mu)$ since the principal's optimal technology level $l_V^*(\mu)$ is increasing in μ. Hence, the agent's payoff net of the mechanism is quasiconcave in l and maximized at $l = l_V^*(\mu)$. We have thus found a nonlinear tax schedule that is ex post optimal—it corrects the externality belief-by-belief. The following proposition formalizes this:

Proposition 11. *For the mechanism ϕ^{s} constructed above:*

$$\mathbb{E}\Big[V(\mu_{\ell^*},\ell^*(\phi^{\mathsf{s}},\mathcal{G},\mathcal{F})\Big] = \sup_{\ell \text{ s.t. } \ell \text{ is } \mathcal{F}\text{-adapted}} \mathbb{E}\Big[V(\mu_\ell,\ell)\Big] \qquad \text{for all } \mathcal{F}\in\mathbb{F}^{\mathsf{s}}$$

i.e., the principal fully internalizes the externality belief-by-belief.

Proposition 11 states that the principal can pick a nonlinear tax carefully to fully correct the externality for any realization of the agent's beliefs. What is more, this mechanism does not depend on the information structure—ϕ^{s} depends only on U and V. This means that (static) informational robustness has no bite in our static environment.

Dynamic information with known law. Now suppose that $\mathcal{F}$ is Markov:

$$\mathbb{F}^{\mathsf{m}} := \Big\{\mathcal{F}\in\mathbb{F} : \big(\mathbb{E}[\theta|\mathcal{F}_l]\big)_{l\in\mathcal{L}} \text{ has the strong Markov property}\Big\}.$$

If $\mathcal{F}\in\mathbb{F}^{\mathsf{m}}$ is *known* to the principal, we can once again achieve the first-best outcome—this is the insight of Kruse and Strack (2015) who establish this in discrete time, and Kruse and Strack (2019) in continuous time when the belief process is a diffusion. We expect the same is true of the whole set $\mathbb{F}^{\mathsf{m}}$ (including those without continuous sample belief paths e.g., jump processes) but we will not attempt a proof since the present paper is already far too long.

	$\mathcal{F}$ known	$\mathcal{F}$ unknown
Static	First-best (Proposition 11)	
Dynamic	First-best if Markov (Kruse and Strack, 2015, 2019)	**Our focus**

Table 2: A taxonomy of regulation problems

Table 2 puts these observations together. We further note that if $\mathcal{F}$ is known but non-Markov, the learning process can contain information about the informativeness of *future* signals so beliefs are no longer a sufficient statistic for the principal's optimal stopping decisions—this prevents the designer from carefully choosing the mechanism to internalize the externality. Likewise if $\mathcal{F}$ is unknown then simply correcting externalities belief-by-belief does not mean that we have closed the wedge between the agent's and principal's stopping levels.