ROBUST TECHNOLOGY REGULATION*

Andrew Koh
MIT

Sivakorn Sanguanmoo
MIT

This version: August 23, 2025

Abstract

We analyze how uncertain technologies should be robustly regulated, and how regulation should respond to evolving information. An *adaptive sandbox* comprising of a zero marginal tax on R&D up to an evolving hard 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 mechanisms evaluated at any learning process and any agent preference; (iii) *time-consistent*: it is the only robust mechanism that can be implemented without commitment; and (iv) *apt*: absent robust regulation, worst-case payoffs can be arbitrarily poor and is induced by weak but growing optimism that encourages excessive risk-taking. Our results offer optimality foundations for existing policy as well as guidance for future implementation.

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

We are especially grateful to Stephen Morris, Drew Fudenberg, and Daron Acemoglu for guidance, support, and insightful conversations. We also thank Ian Ball, Simon Board, Roberto Corrao, Théo Durandard, Piotr Dworczak, Matt Elliott, Andrea Galeotti, Joshua Gans, Alexis Ghersengorin, Gillian Hadfield, Jiangtao Li, Zihao Li, Jonathan Libgober, Parag Pathak, Jean Tirole, Mark Whitmeyer, Alex Wolitzky, Frank Yang, Weijie Zhong, as well as participants 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

It is sometimes held that worry over new technologies are overwrought because those developing them will discover and forestall potential dangers. Thus writes Sam Altman: "the best way to make an AI system safe is by iteratively and gradually releasing it... learning from experience" This view is now commonplace across Silicon Valley[1]—what are we to make of it? Should the prospect that *they* can learn about these new and perhaps transformative technologies reassure *us*? And if not, how should we regulate R&D, and how should regulation adapt to new information? We develop a theory of adaptive technology regulation to answer these questions. Our model takes three stylized facts about R&D seriously:

1. **Technology is uncertain.** There is uncertainty about whether a new technology is beneficial or harmful.[2] A real-valued *technology level* reflects the extent of irreversible R&D, more of which is beneficial if the technology is safe, but harmful if it is dangerous.

2. **Externalities & incentive wedges.** A firm irreversibly develops the technology. The firm is less risk-averse than a regulator whose preferences reflects those of wider society. This differential risk aversion might be intrinsic, or might arise if the firm does not fully internalize the potential harms, stands to gain disproportionately from the potential benefits, or is caught up in an R&D arms race.

3. **Private sector learns more.** As the technology is developed, both regulator and firm learn privately about its harms and benefits. The firm learns more than the regulator, and might do so in quite varied ways e.g., extra independent information, extra correlated information, or simply learn the same information more quickly. The agent's R&D choices are jointly shaped by (i) *adaptive regulation*—the (stochastic) path of taxes or subsidies that evolve with the regulator's beliefs; and the (ii) *learning process*—what the agent has learnt thus far, what she might yet learn, and how she expects future regulation to unfold.

Our model is simple but expressive, capturing how R&D in different states is jointly shaped by the interaction between learning, preferences, and policy. We establish broad principles for how technology regulation should be conducted, and how it should evolve with new information—this allows us

[1]Demis Hassabis, the CEO of Google Deepmind has likewise advocated for "carefully testing AI models for dangerous capabilities and only gradually releasing them to more users with effective guardrails."; see https://time.com/7277608/demis-hassabis-interview-time100-2025/ and https://blog.samaltman.com/reflections.

[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.

to speak to the rich middle ground between ex ante and ex post regulation (Laffont and Tirole, 1993).

Economically, we argue that an adaptive sandbox mechanism which imposes a zero marginal tax on R&D up to an evolving quantity limit on R&D that tightens and slackens with the regulator's information is (i) *robust* to the agent's learning process and/or preferences; (ii) *undominated* and *dominant* among robust and regular mechanisms; (iii) *time-consistent* so can be implemented without commitment; and (iv) *apt*—sans robustness, the firm's worst-case learning process is natural, and the regulator's payoffs are poor.

Practically, our results speak directly to existing debates around how new financial technologies and artificial intelligence should be regulated. The adaptive sandbox in our model qualitatively resembles real-world policies ('regulatory sandboxes') that have recently been introduced to regulate new financial technologies,[3] and are increasingly deployed to regulate artificial intelligence.[4] 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, as well as offering optimality foundations for the oft-articulated goal of sandboxes as providing a *"dynamic, evidence-based approach to regulation"* (OECD, 2023).

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 notion of adaptive robustness is new, requiring optimal mechanisms to 'learn under ambiguity'. But it also reduces to more familiar robustness notions when the regulator does not learn. In so doing, we offer a novel perspective on how policy should evolve to avoid pitfalls associated with ex ante regulation which fails to exploit interim information, and ex post regulation (e.g., liability regimes) which places excessive faith in verification and/or tort law.[5]

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 have. *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 zero marginal tax on R&D up to an evolving hard limit—beyond

[3] As of 2020, regulators in over 50 jurisdictions have adopted some form of sandbox to regulate new financial technologies (Appaya et al., 2020). A more recent OECD report states that approximately 100 sandbox initiatives had been implemented around the world (OECD, 2023).

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

[5] Our framework and results also offer analytical precision and guidance for for a recent trend in environmental and AI policy advocating for 'anticipatory regulation' (Bennear and Wiener, 2019; Reuel and Undheim, 2024).

which further R&D is disallowed—is both learning- and dually-robust. Its robustness is driven by two key properties:

- *Aligned sensitivity* to information before the sandbox limit: the zero marginal tax on R&D minimally distorts the shape of relative incentives between regulator and firm, guaranteeing that under any learning process, the sandbox positively correlates the events that both firm and regulator wish to halt R&D.
- *Minimal sensitivity* to information at the sandbox limit safeguards against *weak optimism*: learning processes under which the agent grows progressively optimistic to push on with R&D, but not quickly enough—if the regulator held the same belief, they would prefer to stop. Hard limits safeguard against this possibility that the learning process anti-correlates the firm and regulator's desire continue R&D.

The adaptive sandbox is simple and easily computable, which stands in contrast to the Bayesian problem whose solutions are complicated and sensitive to the prior. It is also behaviorally robust, continuing to deliver optimal payoff guarantees whenever the agent fulfills a minimal rationality assumption that accommodates myopia, misoptimization, or misspecification about her own learning process.

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 more R&D. That is, for any learning process and any agent preference—even those away from the worst case—the adaptive sandbox delivers higher regulator payoffs.[6] Dominance is driven by a third key property:

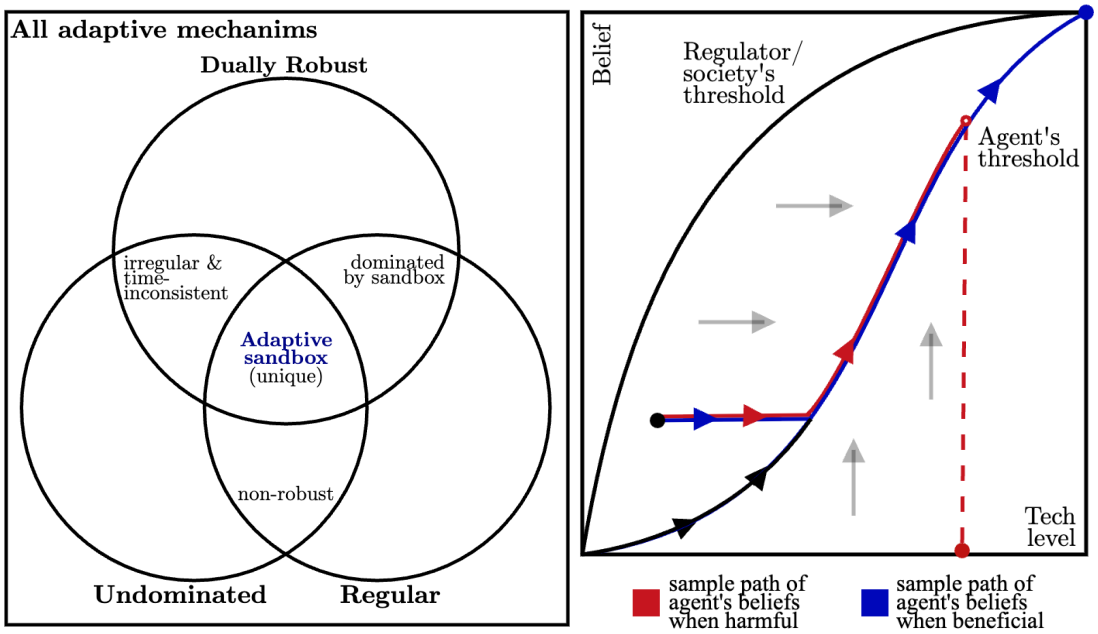
- *Maximal sensitivity* to information: among all mechanisms that induce aligned sensitivity to information (a necessary condition for robustness), our adaptive sandbox keeps the firm maximally sensitive to interim adverse information: under any learning process and any firm preference, sandboxes do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

Our result formalizes the sense in which regulatory sandboxes are intended to *"foster business learning, i.e. the development and testing of innovations in a real-world environment"* (European Parliament Briefing, 2024): it does so by keeping the firm's R&D choices sensitive to new adverse information. Indeed, a recent 10-year legal retrospective of sandboxes for financial technologies notes that when sandboxes performed poorly, it was because regulators acted as *"cheerleaders and sponsors"* (Allen, 2025). Our results lend precision to such claims as well as guidance for future policy: before the sandbox limit,

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

regulators should neither penalize nor subsidize R&D—the former induces stopping too soon which gives up expected gains from innovation; the latter induces stopping too late because the agent is made less sensitive to adverse information that might arrive in the interim. The zero marginal tax threads this needle.

Figure 1: Outline of results



(a) Desiderata for mechanisms

(b) Weak optimism sans robust regulation

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.

Time-consistency. Beyond evaluating mechanisms according to their *ex ante* performance, 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.[7] The adaptive sandbox is indeed time-consistent: whenever the agent halts or continues R&D, whatever the regulator has learnt thus far, she has no incentive to revise the mechanism. Time-consistency is driven by positive selection induced by the sandbox's aligned sensitivity: if an agent stops, the regulator must infer that her own optimal R&D level has already passed. This, in turn, prevents a Coase conjecture-like logic—in which the agent strategically halts R&D to induce more lenient regulation—from taking hold. 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.

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

The aptness of robustness. We do not think robustness is always (or even often) a good guide for policy. But we argue that it is especially *apt* for risky technologies: sans robust regulation, the worst-case learning process is both natural ('qualitatively important') and delivers poor regulator payoffs ('quantitatively important'). To make this precise, we fully characterize 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 R&D. We call such learning processes *weak optimism* and they are illustrated in Figure 1 (b).

Weak optimism hurts welfare by maximizing the distribution of false positives i.e., excess R&D when the technology is harmful (red sample path in Figure 1 (b)), while inducing no false negatives i.e., R&D is always maximized when the technology is beneficial (blue sample path). In the absence of robust regulation, this generates unboundedly poor regulator payoffs when the agent's preferences are also chosen adversarially. But, as we emphasize, preferences need not be particularly extreme—even small wedges in risk aversion can be substantially magnified by weak optimism. 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 Dario Amodei: "We always monitor the models... so that we have a continuous process where we don't get taken by surprise." 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 R&D 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,[8] but nonetheless persist with R&D—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. Acemoglu (2021) surveys potential socioeconomic harms posed by AI systems used in prediction; Hadfield and Koh (2025) surveys risks distinctive to AI agents that can plan and execute

[8] See, e.g., a CNN interview of Dario Amodei in July 2025 in which he warns of a 'white-collar bloodbath'.

complex tasks over long horizons; Feyen et al. (2021) surveys the promises and perils of new financial technologies. Our focus is how policy should be designed when the private sector has misaligned incentives[9] 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.[10] 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. Thus, our optimal policies 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-based' regulations like Pigouvian taxes perform poorly.

Our first desiderata is robustness and so relate 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 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 that dominates 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 influencing the *joint* distribution between equilibrium R&D 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 desiderata 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). But we show that

[9] See Jones (2016, 2024); Aschenbrenner and Trammell (2024) for analysis of the welfare-maximizing solution without misaligned incentives.

[10] In Guerreiro et al. (2025) the agent's beliefs are dogmatic and independently drawn from a known distribution so the agent does not have superior private information.

[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. Riedel (2009) studies optimal stopping where the agent is ambiguity averse; in our setting, it is the principal who has a concern for robustness. Our results on robustness extends when the agent has a generalized version of such preferences; see Appendix [X].

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 the static environments of project choice and persuasion. In our context, dominance 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 holds 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 evolving information; and (ii) agent's experimentation problem is dynamic. When the regulator's information is fixed, our sandbox has a fixed boundary and thus resembles the static 'interval delegation' solutions of Holmström (1977); Amador and Bagwell (2013) though the forces driving our results are quite distinct. 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). Different from this literature, 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, 2024a).[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

[13] If the principal were concerned about either robustness or dominance (but not both), not all optimal policies are 'delegation-like'.

[14] The literature on dynamic information design in stopping problems has focused on a designer with preferences over stopping times (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, 2024a).

boundary to implement a target stopping time under a known diffusion.[15] 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.

Organization. The rest of the paper is organized as follows. Section 2 details the model. Section 3 establishes the robustness, time-consistency, and dominance of adaptive sandboxes. Section 4 establishes the qualitative and quantitative importance of robustness. Section 5 discusses policy implications, extensions, and generalizations of our results. Section 6 concludes.

2 MODEL

2.1 Technology and uncertainty. There is a binary state taking values in $\Theta = \{0, 1\}$ where $\theta = 1$ corresponds to the technology being beneficial; $\theta = 0$ corresponds to harmful.[16] The *technology space* is the compact set $\mathscr{L} \subseteq \mathbb{R}$ which can be either discrete or continuous, and we write $\max \mathscr{L} =: L$. The technology level $l \in \mathscr{L}$ might correspond to the degree of R&D, state of the technological frontier, extent of deployment, and so on. Importantly, we assume that technology is irreversible.[17]

2.2 Learning. There is a principal representing a regulator, and an agent representing a (technological) firm. They share a common prior $\mu_0 \in \Delta(\Theta) := [0, 1]$. Both principal and agent learn about θ as the technology is developed.[18] The principal's learning process is given by a right-continuous filtration $\mathscr{G} := (\mathscr{G}_l)_{l \in \mathscr{L}}$ generated by some real-valued signal process. The agent learns (weakly) more: her learning process is represented by the filtration $(\mathscr{F}_l)_l$ that satisfies the relation:

$$\mathscr{G}_l \subseteq \mathscr{F}_l \text{ for all } l \in \mathscr{L}.^{19} \tag{R}$$

The set of the agent's potential learning processes is given by

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

[15] That is, a Markov process with continuous sample paths. By contrast, our ambiguity set includes non-Markov right-continuous belief processes.

[16] Almost every result in this paper extends to a finite number of ordered states, though at the cost of extra notation.

[17] 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., build 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, as with the deployment of AI agents which might resist being shut down (Russell, 2019).

[18] In the background there is an appropriately filtered probability space that we suppress in the main text to focus on the economics; details are in Appendix A.

[19] This clearly fulfills a continuous-time version of sequential sufficiency (Greenshtein, 1996) which, in turn, is the dynamic analog of the familiar Blackwell (1951) ordering over static information structures.

and use $\mu_l := \mathbb{E}[\theta \mid \mathscr{F}_l]$ to denote the agent's beliefs about the state.[20] The agent's beliefs $(\mu_l)_l$ evolve (i) as a martingale by Bayes' rule; and (ii) right-continuously since $(\mathscr{F}_l)_l$ is right continuous.

The set of agent learning processes $\mathbb{F}$ is rich, specifying not just how her information about the state evolves ('first-order' information), but also about the informativeness of the agent's and principal's future learning process ('higher-order' information). This allows us to capture various ways in which private sector firms might be more informed—they might have extra independent information, extra correlated information, or simply learn the same information sooner. The following examples illustrate:

1. **No principal learning.** $\mathscr{G}_l = \mathscr{G}_0$ a.s. We have already noted that agent's beliefs $\mu_l = \mathbb{E}[\theta \mid \mathscr{F}_l]$ evolves as a càdlàg martingale. Moreover, any càdlàg martingale—potentially non-Markov or non-continuous—can be induced by some learning process $\mathscr{F} \in \mathbb{F}$.

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. **Adapted 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 conditional distribution adapted to the natural filtration $\mathscr{G}$ 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 lead of $\gamma \geq 0$: $Y_l = X_{l+\gamma \wedge L}$.

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 $v(\theta, l)$.[21] It will be helpful to work with indirect utilities by defining

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

and we assume throughout that for each belief $\mu \in [0,1]$, $V(\mu, \cdot)$ is continuous and quasiconcave. More R&D is increases payoffs when the technology is beneficial i.e., $V(1, \cdot)$ is increasing, and decreases payoffs when the technology is harmful i.e., $V(0, \cdot)$ is decreasing. We normalize payoffs from no R&D to zero so $V(\mu, 0) = 0$. These assumptions imply that the principal prefers more R&D when she is more optimistic.[22] This is illustrated in Figure 2 (a) where the principal's indirect utility corresponding to belief $\mu \in (0,1)$ is given by the

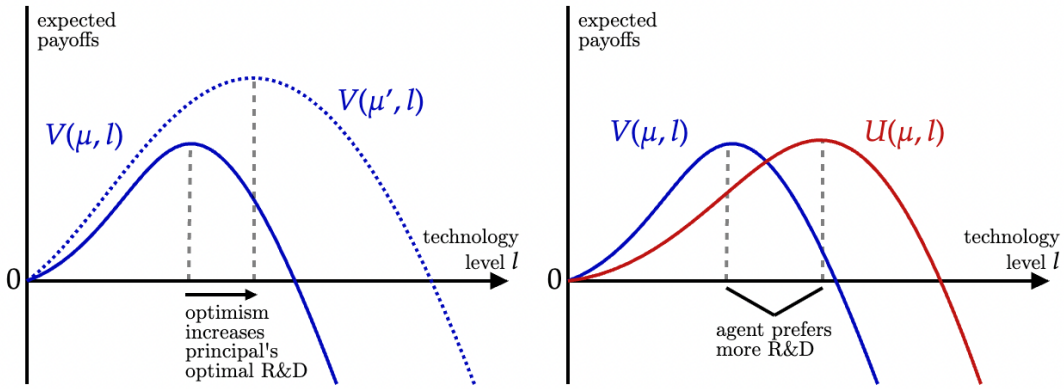
[20] The common prior implies $\mu_0 = \mathbb{E}[\theta \mid \mathscr{G}_0] = \mathbb{E}[\theta \mid \mathscr{F}_0]$.

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

[22] Directly from Milgrom and Shannon (1994).

solid blue line, and the same for a more optimistic belief $\mu' > \mu$ is given by the dotted blue line.

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 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 this assumption is that for each belief μ, the principal's static choice of R&D (without 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.

Discussion of differential risk aversion. We have chosen to focus on risk aversion as the only wedge between the principal and agent's preferences. In particular, Assumption 1 implies alignment at certainty: if the state were known, optimal R&D levels coincide. We think this is a reasonable benchmark for many technologies e.g., artificial intelligence in which tech firms do not wish to develop AI systems which are obviously harmful but, in the presence of uncertainty, are willing to take on more risk than society. It is this interplay between risk-aversion and private learning that generates stark policy implications. Of course, we might also worry that tech firms are misaligned even at certainty. For instance, profit incentives might lead them to deploy technologies they know to be harmful—our results straightforwardly extend in the presence of these wedges. But, as we will see, slight differences in risk aversion can, on their own, generate unboundedly poor payoffs when the learning process is chosen adversarially so our weaker assumptions on the degree of misalignment yields stronger arguments for robust regulation. The following example illustrates different channels through which differential risk appetites might arise:

Example 1 (Constant absolute 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.[23] It is 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.

Consumption 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 $\gamma_P\neq 1$.[24] The following examples of agent preferences all fulfill Assumption 1:[25]

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., $\gamma_P>\gamma_A$.

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 constraints. This can be represented by $U(0,l)=\alpha\cdot V(0,l)$ for $\alpha<1$.

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

4. **R&D arms race.** There are n identical firms playing a continuous-time stopping 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 $U(1,l)=\beta_1\cdot V(1,l)$.[26] ♢

[23] Jones (2024) focuses on existential risk which he models as the event that consumption is zero for all future periods.

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

[25] See Appendix [X] for details.

[26] Details are in Appendix [X].

2.4 Adaptive mechanisms. A mechanism is a real-valued stochastic process $(\phi_l)_l$ that specifies, if the agent stops at R&D technology level l, the tax/subsidy ϕ_l. Mechanisms are adapted to the principal's learning process:

$$\Phi := \left\{ \phi : \mathscr{L} \to \mathbb{R} \cup \{+\infty, -\infty\} : \begin{array}{c} \phi \text{ is (i) } (\mathscr{G}_l)_l\text{-adapted;} \\ \text{(ii) a.s. lower semi-continuous; and} \\ \text{(iii) uniformly integrable on } \{|\phi| < \infty\} \end{array} \right\}$$

is the set of all adaptive mechanisms. Condition (i) states that current regulation cannot 'look into the future' and must only be conditioned on the information regulators know thus far; (ii) and (iii) are mild regularity conditions to guarantee the existence of solutions to the agent's optimal stopping problem. Stopping R&D at level l under belief μ yields an expected payoff of

$$U^{\phi}(\mu, l) := U(\mu, l) - \phi(l) \quad \text{and} \quad V(\mu, l)$$

for the agent and principal respectively. We let the mechanism take values in the extended reals to capture both 'price' instruments ($\phi \in \mathbb{R}$), 'quantity' instruments ($\phi \in \{-\infty, +\infty\}$), as well as mixtures thereof. We note that neither additional randomization nor elicitation of type reports—either static or dynamic—can improve upon the principal's payoff under robustness concerns.[27] We are primarily interested in efficiency and thus focus on the case in which ϕ does not feature directly in the designer's payoff: if ϕ is a tax, this might be paired with a lump sum rebate; alternatively, ϕ might be a non-monetary instrument e.g., legal, administrative, or regulatory barriers. Our results can be straightforwardly modified to accommodate revenue motives for the regulator; see Appendix [X].

Discussion of ex-{ante, interim, post} regulation. By making the regulator's learning process explicit, our framework offers a unified umbrella for analyzing regulatory timing (Laffont and Tirole, 1993) in terms of what kinds of information policy can be conditioned on. Table 1 illustrates.

	ex ante	ex interim	ex post
Requisite information	Policy is $\mathscr{G}_0$-adapted	Policy is $(\mathscr{G}_l)_l$-adapted	Policy is $\mathscr{G}_\infty$-adapted
Feasibility	Feasible	Feasible	Infeasible
Performance	Suboptimal	Second-best	First-best

Table 1: Comparison of regulatory timing

Under ex ante regulation, the policy specifies a rigid path of taxes in advance which fails to make use of information generated as the technology is developed and is, in general, suboptimal. Conversely, ex post regulation conditions policy on the true state—for instance, stipulating that if harms are generated ($\theta = 0$), firms should be accordingly penalized. Such policies are typically thought of as 'backwards looking' liability regimes (Shavell, 1984) which has

[27] This is formalized in Online Appendix [X].

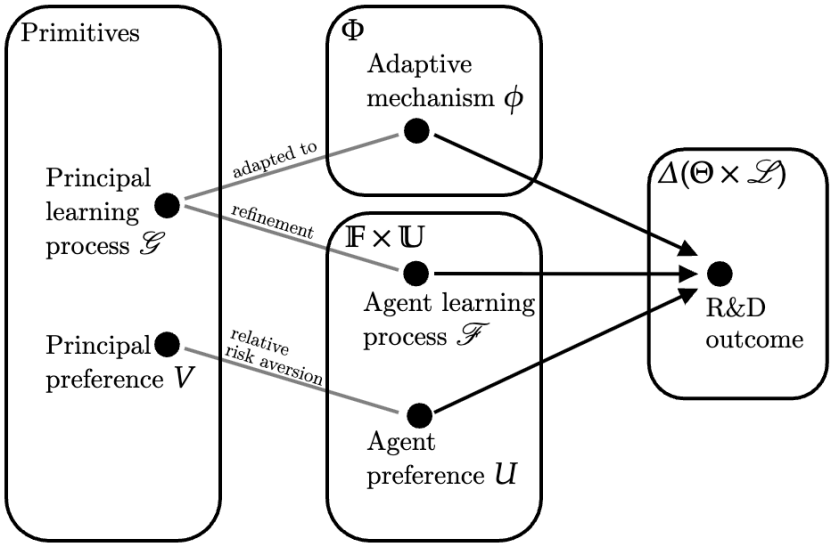
been recently proposed for regulating artificial intelligence (Gans, 2025). But looking backwards from where? Our framework highlights a conceptual tension when learning is endogenous to policy: state-contingent regulation is feasible only if it always induces enough R&D for the principal to learn the state—but this is self-defeating since point of regulation is to induce different R&D levels in different states.[28] On this view, regulation that evolves with current information is the best we can do.

2.5 Agent's problem. The agent with preferences U, facing mechanism ϕ and learning process $\mathscr{F}$, first decides whether or not to participate. If she does not, the technology is not developed and the principal and agent receive payoffs $V(\mu_0, 0)$ and $U(\mu_0, 0)$ respectively. If she participates, the agent solves the following optimal stopping problem:

$$\sup_{l} \mathbb{E}\Big[U(\mu_l, l) - \phi_l\Big] \qquad \text{(O)}$$
$$\text{s.t. } l \text{ is } \mathscr{F}\text{-adapted.}$$

Notice that the agent's optimal R&D choices depend jointly on the adaptive mechanism ϕ, the learning process $\mathscr{F}$, and her preferences U. Let $l^*(\phi, \mathscr{F}, U)$ be the largest stopping level that solves (O) i.e., breaking indifferences in favor of more R&D.[29] Figure 3 takes stock of how variables interact to jointly shape R&D outcomes.

Figure 3: How policy, learning, and preferences shape R&D outcomes



2.6 Desiderata. We outline three desiderata for adaptive mechanisms.

[28] There are, of course, other standard objections to ex post regulation e.g., the weakness of tort law, limited liability, or regulator moral hazard.

[29] Observe that (i) the compactness of the level space $\mathscr{L}$; (ii) right-continuity of càdlàg belief processes $\boldsymbol{\mu}$; and (iii) lower semicontinuity of the adaptive mechanism ϕ guarantees existence of a bounded optimal stopping level for the agent.

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

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

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

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

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

Call mechanisms that achieve (D) *dually-robust*.[30]

Our robustness notions recover, as special cases, some of those studied in the literature. When the principal does not learn (i.e., $\mathscr{G}_l=\mathscr{G}_0$ 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 $\mathscr{F}_0$-measurable), robustness to the agent's preferences is similar to that studied by (Armstrong, 1995; Frankel, 2014).

Economically, 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 (i) what might be learnt via more R&D ($\mathscr{F}\in\mathbb{F}$)—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))—and (ii) what preference the private sector has ($U\in\mathbb{U}$)—for instance, Jones (2024) wonders: *"do the tech entrepreneurs developing AI have low risk aversion?"* Conceptually, our robustness notion is novel and requires mechanisms to 'learn under ambiguity' (Epstein and Schneider, 2007). In particular, the regulator's filtration $(\mathscr{G}_l)_l$ contains both 'first-order' information about the state θ but could also allow the regulator to rule out agent learning processes or preferences, thereby narrowing the set from $\mathbb{F}\times\mathbb{U}$.

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

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

[30] See Armstrong (1995) for an early analysis of robustness to the agent's preferences.

with strict inequality for some $(\mathscr{F}, U) \in \mathbb{F} \times \mathbb{U}$. 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 a similar spirit to work by Wald (1950); Savage (1972) and variants of it have been recently employed to rank robust mechanisms for project choice and persuasion in static environments (Guo and Shmaya, 2023; Dworczak and Pavan, 2022).[31]

We say that the mechanism ϕ is *undominated* if it is not dominated by another adaptive mechanism i.e., $\phi' \not\succ \phi$ for all $\phi' \in \Phi$. This notion of undominated static mechanisms has been recently analyzed by Borgers, Li, and Wang (2024).[32]

Time-consistency. A final desiderata is for the regulator to have incentives to follow-through with the mechanism, both on and off the equilibrium path. We emphasize that this is a feasibility rather than optimality criterion—time consistency matters because it determines whether a mechanism can be credibly run. Indeed, the problem of time-inconsistency 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.

3 ROBUST, DOMINANT, AND TIME-CONSISTENT REGULATION

We begin by defining a simple family of adaptive mechanisms.

Definition 1 (Adaptive sandboxes)**.** ϕ is an adaptive sandbox induced by the (random) technology level $\widetilde{l}$ if:

$$\phi_l = \begin{cases} 0 & \text{if } l \leq \widetilde{l} \\ +\infty & \text{if } l > \widetilde{l}. \end{cases}$$

Adaptive sandboxes prescribe a zero marginal tax on R&D until a level $\widetilde{l}$, beyond which R&D is disallowed.

[31] Guo and Shmaya (2023) study regret minimization; Dworczak and Pavan (2022) rank robust information structures away from the worst case with respect to a single alternate conjecture—the 'base case payoff' that nature delivers no additional information.

[32] 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 càdlàg paths, all examples of dominance in this paper is 'strong dominance' with respect to the Wiener measure.

Let ϕ^* be the adaptive sandbox induced by the smallest stopping level l_* that solves

$$\sup_{l} \mathbb{E}\Big[V(\mu_l, l)\Big]$$
$$\text{s.t. } l \text{ is } \mathscr{G}\text{-adapted.}$$

The adaptive sandbox ϕ^* is illustrated by Figure 4.

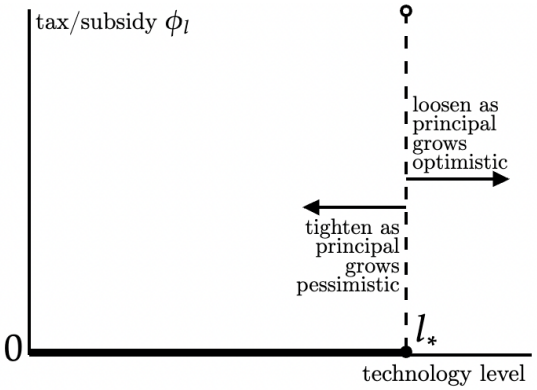


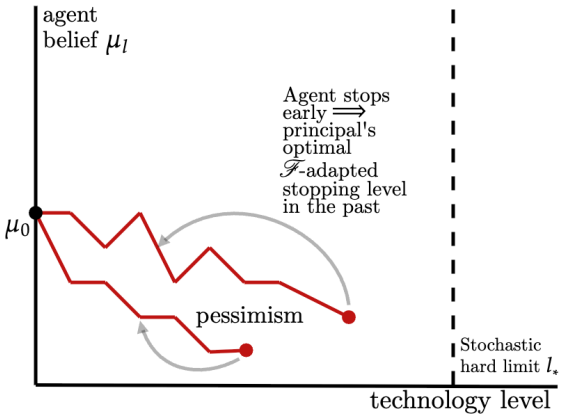
Figure 4: Adaptive sandbox

3.1 Robustness.

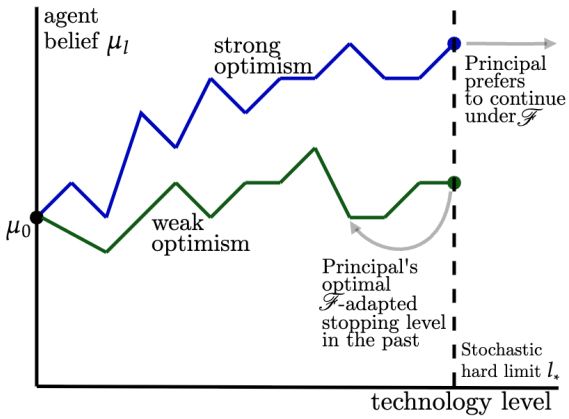
Theorem 1 (Robustness)**.** *The adaptive sandbox ϕ^* is learning-robust and dually-robust i.e., it solves* (L) *and* (D)*.*

Theorem 1 establishes the learning- and dual-robustness of ϕ^* This is driven by several features. First and most immediately, it delivers a *participation guarantee*: if the agent does not learn any more than the principal ($\mathscr{F}_l = \mathscr{G}_l$ a.s.), her participation constraint is fulfilled since always doing R&D until the sandbox limit—a feasible R&D strategy—is better than not participating. But since her information is a refinement of the principal's,[33] her participation payoff must weakly increase under any learning process $\mathscr{F} \in \mathbb{F}$.

Figure 5



(a) Aligned sensitivity before limit



(b) Minimal sensitivity at limit

Second, the zero marginal tax induces *aligned but asymmetric sensitivity to information*. If the agent becomes sufficiently pessimistic and decides to stop R&D (red sample belief paths in Figure 5 (a)), early stopping must also be preferred by the principal. Thus, by minimally distorting the shape of relative incentives, our mechanism generates positive correlation between the

[33] See, e.g., Greenshtein (1996).

events that both principal and agent prefer to stop. But this alignment is also asymmetric since the opposite is not true—that the agent is optimistic enough to continue R&D is no guarantee the principal wishes the same.

Third, the hard limit induces *minimal sensitivity to information* at the sandbox limit l_* since whatever the agent's beliefs, it is strictly dominated to push past it.[34] Consider, for instance, the sample belief path depicted in blue (labeled 'strong optimism') and green (labeled 'weak optimism') in Figure 5 (b). On both of these paths, the agent grows more optimistic that the technology is beneficial and so finds it optimal to continue until the hard limit, upon which she stops R&D.

Although this limit is stringent, this is exactly what robustness demands. Consider the realization of the blue belief path ('strong optimism') in Figure 5 (b) in which at the sandbox boundary l_* the agent has privately become optimistic and if the principal knew this, she would like to slacken the sandbox limit. Under the green belief path ('weak optimism'), however, the principal prefers stopping although the agent is willing to take on more risk. But since the agent's learning process is chosen adversarially, adaptive mechanisms cannot distinguish between these kinds of optimism to guarantee positive correlation between the events that both agent and principal prefer more R&D.[35] Indeed, we will see in Section 4 that in the absence of robustness, the worst-case learning process negative correlates these events by inducing a specific kind of weak optimism that maximizes risk-taking—hard limits safeguard against exactly this.

A few further comments:

Simplicity. Although the space of adaptive mechanisms is large, comprising all $\mathscr{G}$-adapted real-valued stochastic processes, the adaptive sandbox ϕ^* is quite simple. This echoes a broader theme from robust mechanism design (Bergemann and Morris, 2005; Carroll, 2019) that robust mechanisms are particularly appealing when they are also 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 belief process is (i) known; (ii) continuous; and (iii) Markov—then, results from Kruse and Strack (2019) can be adapted to solve for the optimal mechanism (which is quite sensitive to the belief process). But the specificity of this exception suggests that the Bayesian problem is more generally hard and detail-dependant. On this view, Theorem 1 represents the best a boundedly rational regulator can do.[36]

Behavioral robustness. The adaptive sandbox continues to deliver optimal worst-case guarantees whenever the agent employs any *undominated stopping rule*. Undominated stopping rules impose minimal assumptions on

[34] An equivalent interpretation is that the agent simply cannot push past it i.e., the hard limit constrains the agent's technology choices.

[35] This remains true if we allowed for dynamic elicitation of type reports; see Appendix E.

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

the agent's rationality by *only* ruling out stopping at a given pair of beliefs and technology levels (μ, l) when more R&D is strictly better in the absence of further information. These include stopping levels which are optimal i.e., those that solve (O), but also allows for agents to be misspecified or ambiguity averse about the future learning process, misoptimize, or be myopic.[37] For any such undominated stopping rule, the adaptive sandbox ϕ^* continues to deliver optimal worst-case guarantees.

Connection to real-world policy. Finally, Theorem 1 offers optimality foundations for real-world policy. The World Bank reports that as of 2020, regulators in over 50 jurisdictions have adopted some form of sandboxes to regulate new financial technologies (Appaya et al., 2020). Sandboxes have likewise been proposed to regulate artificial intelligence (European Parliament, 2024). Our adaptive sandbox ϕ^* reflects qualitative features of such real-world policies:

1. *Zero marginal tax as fostering innovation*. In the first use of regulatory sandboxes for fintech in the UK, regulators describe 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.[38] 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 associated with R&D.[39]

2. *Limits as safeguards*. Real-world implementation of sandboxes for fintech have imposed different kinds of limits such as (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). A common theme across such policies is some form of 'quantity limits' on research or deployment.

3. *Adaptivity.* To push beyond the sandbox boundary, firms must convince the regulator to extend the sandbox—either by producing evidence that

[37] A formal definition of undominated stopping rules and statement of the result that ϕ^* remains learning- and dually-robust is given in Appendix [X].

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

[39] An influential legal perspective by Allen (2019) describes it as an environment in which firms can conduct *"limited tests of their innovations with fewer regulatory constraints, real customers, less risk of enforcement action."*

[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.

(i) the technology is safe (so the principal grows optimistic about θ); or (ii) further R&D might be informative (so the principal grows optimistic about the informativeness of additional R&D)[41]—failing which the firm is forced to stop R&D. Indeed, these 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 generate pessimism 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 Parliament Briefing, 2024). Our adaptive sandbox lends precision to such claims: by loosening the boundary whenever the principal turns more optimistic, this carves out extra room for the agent to experiment. But if the agent then halts R&D before the boundary, aligned sensitivity to information ensures that the principal also prefers to halt.

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.

Regularity requires mechanisms to not vacillate between penalizing and rewarding more R&D. This is in part motivated by practical considerations: regular mechanisms are "not too complex" and so implementable (Diamond and Saez, 2011).[43] 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. These are illustrated in panels (a)-(c) of Figure 6. Panel (d) illustrates an irregular mechanism which oscillates between a low and a high tax. Regularity is more permissive than the monotonicity condition typically imposed in security design (Nachman

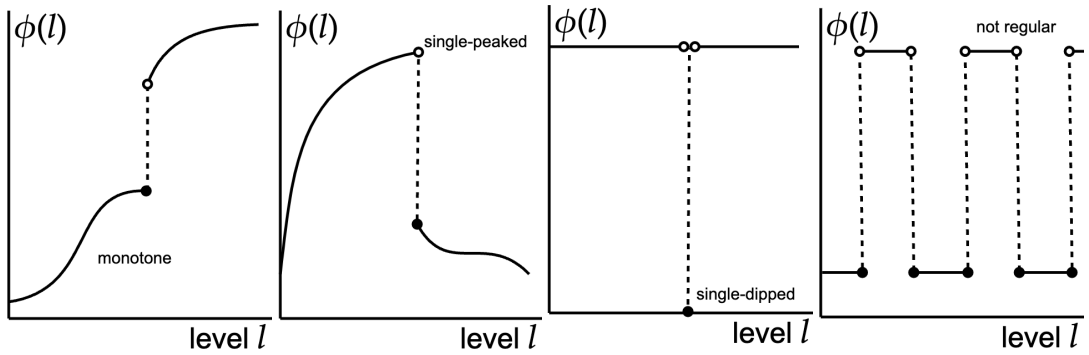
[41] At the sandbox limit, firms can 'apply to remove the restrictions on... permissions to become a fully authorised firm.' (Financial Conduct Authority, 2023). A similar policy is in place in Australia (Australian Securities and Investments Commission, 2024).

[42] For an outline of current regulation FDA clinical trials see https://www.fda.gov/patients/drug-development-process/step-3-clinical-research. Junod (2008) surveys the history of FDA clinical trials.

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

and Noe, 1994; DeMarzo and Duffie, 1999)[44] since any tax schedule that penalizes more R&D is increasing and thus regular.

Figure 6: Illustration of regularity



(a) Nonlinear tax
Regular

(b) Tax-then-subsidize
Regular

(c) Quota
Regular

(d) Oscillatory
Not regular

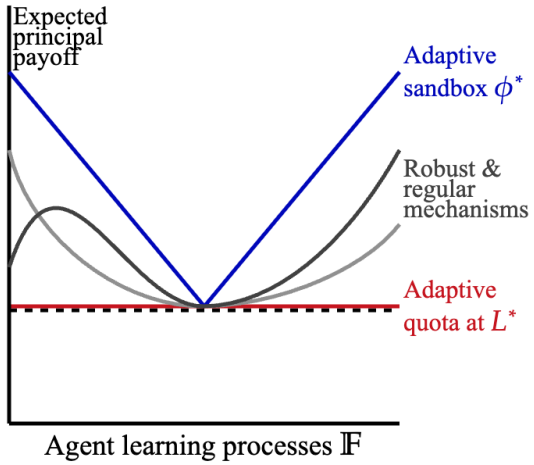
Theorem 2. *The adaptive sandbox ϕ^* is undominated. Furthermore, it dominates all other dually-robust and regular mechanisms.*

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 $\mathscr{F}$ and preference U so that ϕ^* delivers strictly higher regulator payoffs than ϕ. The second part of Theorem 2 further states that ϕ^* dominates all other mechanisms ϕ which are regular and dually-robust.[45] Thus, if our designer has lexicographic attitudes toward uncertainty about the agent's learning process and preferences and, perhaps for practical reasons, is restricted to regular mechanisms which do not vacillate between punishing and rewarding more R&D, then ϕ^* is uniquely optimal.

[44]There, the asset payout (as a function of the underlying cash flow) is typically required to be monotone or 'doubly monotone' so that the asset value less the payout is also monotone.

[45]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.

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



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. It turns out that among regular mechanisms which induce aligned sensitivity, the zero marginal tax on R&D induces *maximal sensitivity* to new adverse information—this drives its dominance.

But this is not always how real-world sandboxes have been implemented. 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 R&D even if they turn pessimistic, thereby leading to the over-proliferation of harmful technologies. In this regard, Theorem 2 refines and unifies different intuitions about why real-world sandboxes might not have worked well, as well as offers concrete guidance for how they should be operationalized.

Outline of proof ideas. The proof of Theorem 2 hinges on several ideas we briefly sketch here; the proof is in Appendix A.

Step 1: Comparative static for experimentation. We first develop a simple but powerful comparative static result relating risk aversion to optimal experimentation:

Lemma 1. *Fix any agent preference $U \in \mathbb{U}$ and any learning process $\mathscr{F} \in \mathbb{F}$. Any $\mathscr{F}$-adapted stopping level solving $\sup_l \mathbb{E}[V(\mu_l, l)]$ is almost-surely smaller than the largest $\mathscr{F}$-adapted stopping level solving $\sup_l \mathbb{E}[U(\mu_l, l)]$.*

Lemma 1 nests and unifies previous work analyzing how different kinds of risk-aversion can either prolong or shorten experimentation (Chancelier et al., 2009; Keller et al., 2019). What is more, it holds path-by-path for

any learning process, including those that induce non-Markov and/or non-continuous beliefs, and without requiring explicit solutions to the optimal stopping problem—this generalizes past work that have worked within indexable or exponential bandit frameworks.

Step 2: Characterization of dual-robustness. We next develop a complete characterization of dually-robust mechanisms:

Lemma 2. *ϕ is dually-robust if and only if (i) it imposes a hard limit at l_*; and (ii) $\phi(l) \geq \phi(l_*)$ a.s. for all $l \leq l_*$.*

This characterization is simple and easily verified. Condition (i) states that all dually-robust mechanisms ϕ must impose a hard limit at l_*—which, we recall, is the $\mathscr{G}$-adapted stopping level solving $\sup_l \mathbb{E}[V(\mu_l, l)]$. The necessity of such a hard limit is jointly driven by concerns that the agent might grow just a little more optimistic, but turn out to be quite risk seeking. Hard limits safeguard against these kinds of joint adversarial deviations by nature, and is why 'quantity-like' instruments must emerge under robustness concerns.

Condition (ii) states that before the limit l_*, the mechanism is an adapted process that is conditioned to remain above $\phi(l_*)$. By ensuring that the agent always has the *option value* of pushing R&D until l_* without being penalized (though she might be heavily punished for stopping in the interim), this safeguards against the possibility that regulation might stifle innovation by distorting R&D downwards.

Step 3: Dominance via coupling. We now draw on Steps 1 and 2 to show dominance. Take any dually-robust and regular mechanism ϕ. From Lemma 2, ϕ must be almost-surely decreasing over $[0, l_*]$ and impose a hard limit at l_*. Now for any learning process $\mathscr{F} \in \mathbb{F}$ and any agent preference $U \in \mathbb{U}$, 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 $\mathscr{G}$ and $\mathscr{F}$.

Now observe that on this coupled problem, the agent stops sooner facing the adaptive sandbox ϕ^* than under the alternative mechanism ϕ:

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

This is depicted by Figure 7 which illustrates two sample paths of beliefs (red and dark red) in the support of the agent's learning process $\mathscr{F}$. On each of these paths, the agent stops sooner facing ϕ^* than the alternative mechanism ϕ.

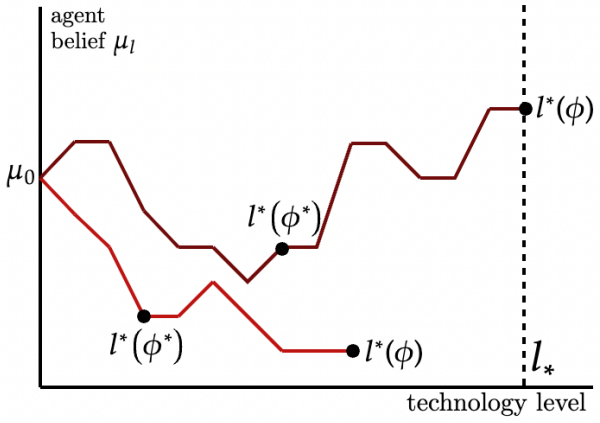


Figure 7: Pathwise comparison

This is the precise sense in which the regulatory sandbox induces *maximal sensitivity* to information. To finish, recall that our comparative static (Lemma 1) implies that whenever the agent prefers to stop, so does the principal as evaluated through the lens of the agent's private information:

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

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

3.3 Time consistency. Robustness and dominance are ex ante criteria that assess mechanism from before it is run. A separate concern is that regulators might be tempted to deviate to a different mechanism at interim histories, and this can render such mechanisms infeasible. Motivated by these concerns, we develop an analysis of time-consistency.

Time-consistency. Mechanism ϕ is *time-consistent* if for each filtration $\mathscr{G}_l$:

(i) **Stopping.** If the agent stops at l, ϕ is undominated from l over $\mathbb{F} \times \mathbb{U}\big|_{\mathscr{G}_l,\{l \geq l^*\}}$.

(ii) **Continuing.** If the agent continues at l, ϕ is undominated from l over $\mathbb{F} \times \mathbb{U}\big|_{\mathscr{G}_l,\{l < l^*\}}$,

where $\mathbb{F} \times \mathbb{U}\big|_{\mathscr{G}_l,\{l \geqslant l^*\}}$ is the set of agents' learning processes and preferences consistent with the principal's filtration $\mathscr{G}_l$ and the agent's R&D choices.

Time-consistency simply requires mechanisms to not be dominated following from at any interim history. There are two 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.[46] The second is because the regulator is learning—the set of agent's learning processes and preferences she regards as possible evolves as (i) the technology is developed; and (ii) the agent makes R&D choices. As this set shrinks, a mechanism that might have been previously rationalized for performing well at some learning-preference pair $(\mathscr{F}, U)$ might now become dominated as $(\mathscr{F}, U)$ becomes ruled out by regulator learning.[47] Time-consistency safeguards against both possibilities.

We emphasize that 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

[46] Here, the agent's belief is that the regulator will deviate to some dominant mechanism. Details of this extensive form game is given in Appendix [X].

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

objective, maximize expected utility, have a concern for misspecification, and so on.[48] Nonetheless, this minimal notion of time-consistency is enough to refine the set of robust mechnanisms:

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

We explain each part in turn.

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

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

2. **Continuing before sandbox limit.** If the agent chooses to continue R&D before the sandbox i.e., at technology level $l < l_*$, 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 remains large so maintaining the sandbox is conditionally undominated.

3. **Stopping at sandbox limit.** If the agent pushes R&D until the sandbox limit i.e., $l^* = l_*$, the regulator cannot rule out the possibility that the agent's information at l_* 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 $\phi(l) \geq \phi(l_*)$ a.s. for all $l \leq l_*$ (via our characterization from Lemma 2 above). But suppose that with positive probability this inequality is strict at filtration $\mathscr{G}_l$ i.e., ϕ almost surely decreases until l_*—supposing, for the moment, that l_* is fixed (degenerate) from the perspective of $\mathscr{G}_l$; the randomness of l_* can be handled with a more finesse.

[48] 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; see Appendix [X].

Our candidate mechanism ϕ is depicted in Figure 8 where the two blue paths reward R&D between l and l_*. Now consider the deviation to an alternate mechanism ϕ' continuing from l which 'irons out' the mechanism until l_*. This deviation is depicted in Figure 8, noticing that the resultant mechanism ϕ' is simply an adaptive sandbox continuing from l.

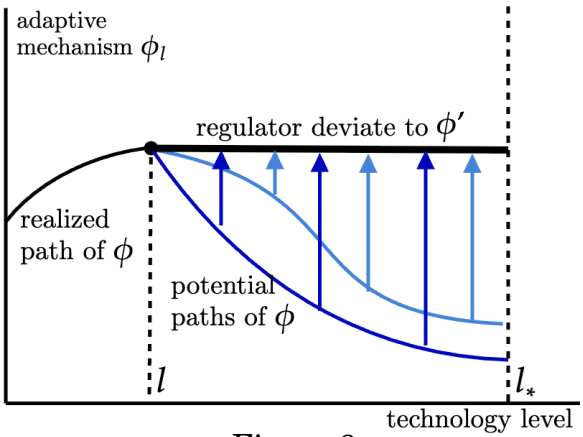


Figure 8

Now suppose that the agent stops R&D following this deviation to ϕ'. From the aligned sensitivity of the sandbox, this means that the principal's optimal technology level has already passed so this deviation delivers a strict improvement to the regulator's payoff.[49] It is precisely this temptation to screen out weakly optimistic agents that is in tension with time-consistency.

4 THE IMPORTANCE OF HARD LIMITS

We have shown that the adaptive sandbox mechanism is robust (Theorem 1), dominant (Theorem 4), and time-consistent (Theorem 3). But this leaves open the question of why, if at all, we should care about robustness.

Two notions of importance. Robustness can be important in two regards. First, it is *qualitatively important* if, under non-robust mechanisms, the worst-case learning process is plausible. If so, there is a sense in which robust mechanisms are safeguarding against something real. Second, it is *quantitatively important* if the difference in worst-case payoffs under robust and non-robust mechanisms is large. We will argue that in settings with risky R&D, robustness is important in both senses.

4.1 Benchmark without principal learning. We first consider the case in which the principal does not learn. In our notation, this corresponds to $\mathscr{G}_l = \mathscr{G}_0$ a.s. for all $l \in \mathscr{L}$. 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.[50] Let $\overline{\mathbb{F}}$ denote the set of all agent learning processes and let $\underline{\Phi}$ denote the set of mechanisms under no principal learning.

[49] We can likewise show that if the agent continues then under any agent preference $(\mathscr{F}, U) \in \mathbb{F} \times \mathscr{U}|_{\mathscr{G}, \{l<l^*\}}$ ϕ' performs weakly better than ϕ.

[50] Analogous to the Marshallian short run that holds capital fixed (Marshall, 1890).

The quantitative importance of learning-robustness for the mechanism ϕ is:

$$\Delta^L(\phi) := \underbrace{\sup_{\phi' \in \underline{\Phi}} \inf_{\mathscr{F} \in \overline{\mathbb{F}}} \mathbb{E}\Big[V(\mu_{l^*}, l^*(\phi', \mathscr{F}, U))\Big]}_{\substack{\text{Best payoff} \\ \text{guarantee over } \overline{\mathbb{F}}}} - \underbrace{\inf_{\mathscr{F} \in \overline{\mathbb{F}}} \mathbb{E}\Big[V(\mu_{l^*}, l^*(\phi, \mathscr{F}, U))\Big]}_{\substack{\text{Payoff guarantee} \\ \text{over } \overline{\mathbb{F}} \text{ under } \phi}}$$

The quantitative importance of dual-robustness for the mechanism ϕ is:

$$\Delta^D(\phi) := \underbrace{\sup_{\phi' \in \underline{\Phi}} \inf_{\substack{\mathscr{F} \in \overline{\mathbb{F}} \\ U \in \mathbb{U}}} \mathbb{E}\Big[V(\mu_{l^*}, l^*(\phi', \mathscr{F}, U))\Big]}_{\substack{\text{Best payoff} \\ \text{guarantee over } \overline{\mathbb{F}} \times \mathbb{U}}} - \underbrace{\inf_{\substack{\mathscr{F} \in \overline{\mathbb{F}} \\ U \in \mathbb{U}}} \mathbb{E}\Big[V(\mu_{l^*}, l^*(\phi, \mathscr{F}, U))\Big]}_{\substack{\text{Payoff guarantee} \\ \text{over } \overline{\mathbb{F}} \times \mathbb{U} \text{ under } \phi}}$$

We will show robustness is both qualitatively and quantitatively important. A few definitions are in order.

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

$$\operatorname{supp} \mathbb{E}[\theta | \mathscr{F}_l] \subseteq \Big\{0, \gamma(l)\Big\}$$

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

Bad news processes are such that, at each R&D 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.[51] It is illustrated in Figure 9 where the gray lines show different belief paths under different choices of F. The red region denotes the support the agent's stopping R&D level and belief.

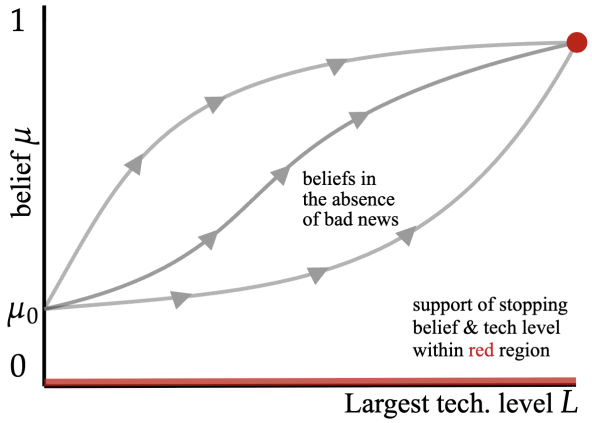


Figure 9: Bad news processes

The set of bad news processes is large. Let us construct a specific one. For

[51] 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.

the indirect utility $f : \Delta(\Theta) \times \mathscr{L} \to \mathbb{R}$, define:

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

$l_f^*(\mu)$ is the largest optimal 'one-shot' level of R&D under belief μ in the absence of further information; $\mu_f^*(l)$ is the *belief threshold*—the most pessimistic belief under which R&D until level l is optimal sans further information.

Definition 4 (Weak optimism)**.** The bad news process $\mathscr{F} \in \overline{\mathbb{F}}$ induces *weak optimism* for 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 R&D until 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 which 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:*

$$\lim_{L \to +\infty} \Delta^D(\phi) = +\infty \quad \textit{if} \quad \lim_{L \to +\infty} \left| \frac{V(0, L)}{V(1, L)} \right| = +\infty.$$

Theorem 4 pins down the learning processes which minimizes the principal's payoff under any mechanism ϕ that preserves the order of risk preferences.[52] The proof proceeds by recasting nature's problem as a dynamic information

[52] 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.

design problem to maximize $-V$, noting that here nature has state-dependent preferences which is novel to the literature[53]—this is fairly technical and we defer it to Appendix [X] for 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 R&D levels when the technology is harmful, subject to the constraint that the agent does, in fact, want to push R&D—i.e., by *magnifying false positives*.

This is illustrated by the red curve in Figure 10 which shows the CDF of R&D 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.

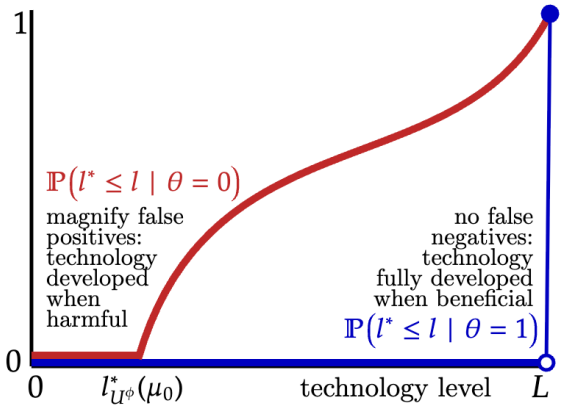


Figure 10: Worst-case joint dist.

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 incremental R&D can be heuristically decomposed as:

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

where the first term gives the value of additional R&D holding fixed the current belief, while the second gives the value of learning more about the technology which might inform future R&D decisions. For a typical learning process the agent can have positive value for future information and this incentivizes risk-taking i.e, the first term is negative while the second is positive. Weak optimism essentially holds both terms down to zero:[54] the agent's beliefs are dynamically steered such that she is continually kept indifferent between more R&D and halting, while her value from future information is kept at zero at all histories.

This has two further implications. First, fixing the mechanism ϕ, and agent

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

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

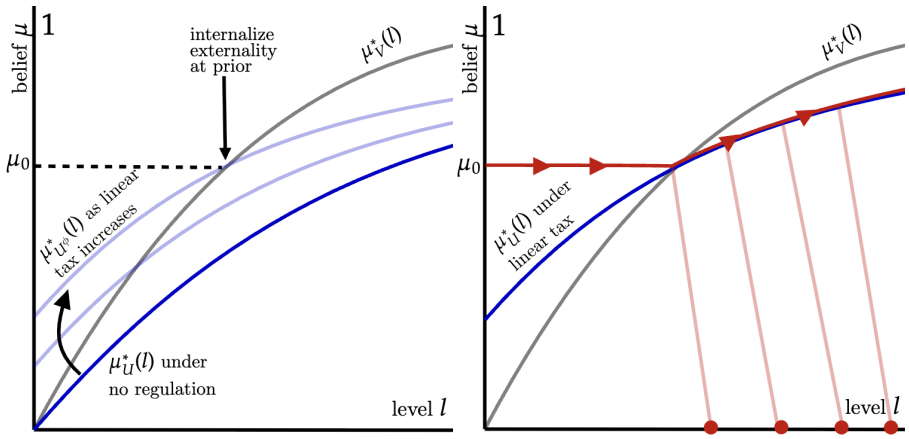
preference U, the worst-case learning process and attendant regulator payoff can be straightforwardly computed. We leverage this in Example 2 below to quantify the worst-case under CRRA utility. Second, the worst-case payoff remains possible whenever the agent employs any *undominated stopping rule*. 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.[55] Theorem 4 implies that tech firms' conservatism about future information per se is not enough to safeguard society.

Policy implications of weak optimism. Theorem 4 warns against correcting only externalities 'locally' at the prior. Consider, for instance, the linear Pigouvian tax $\phi(l) = \alpha \cdot l$ chosen such as to internalize externality under the prior. Panel (a) of Figure 11 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: $l^*_{U^\phi}(\mu_0) = l^*_V(\mu_0) = L^*$.[56] Panel (b) of Figure 11 depicts the worst-case learning process under this tax: the agent learns nothing about the safety of the technology until L^*, 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.

[55] Thus writes Cohen et al. (2024): *"empirical testing of sufficiently capable LTPAs* [long-term planning agents] *is unlikely to uncover these dangerous tendencies."* This is formalized in Appendix E.1. 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.

[56] For expositional simplicity, here we assume L^* is the unique maximizer of R&D under the prior.

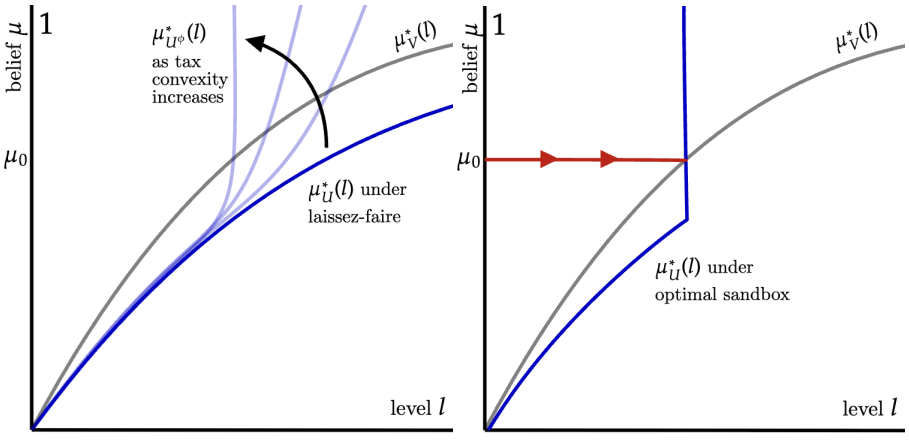
Figure 11: Weak optimism under linear taxes



(a) Increasing the linear tax (b) Worst-case learning process

Now consider panel (a) of Figure 12 which depicts the agent's belief thresholds under a sequence of mechanisms of 'increasing convexity'.[57] 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 less R&D than the principal, while at less optimistic beliefs the agent prefers more. Panel (b) of Figure 12 depicts a regulatory sandbox as the limit of such mechanisms—per 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 12: 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

[57] The idea that one can smoothly interpolate between price and quantity-like instruments dates back at least to Weitzman (1974).

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 frontier of R&D on dangers of powerful AI systems (Greenblatt et al., 2024)—but not enough to induce halting R&D. 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 R&D limit, 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 2 (Worst-case under CRRA utility)**.** Suppose, as in Example 1, that both agent and principal have CRRA preferences over consumption with risk-aversion parameter $\gamma_P > \gamma_A \geq 1$ and that consumption is given by

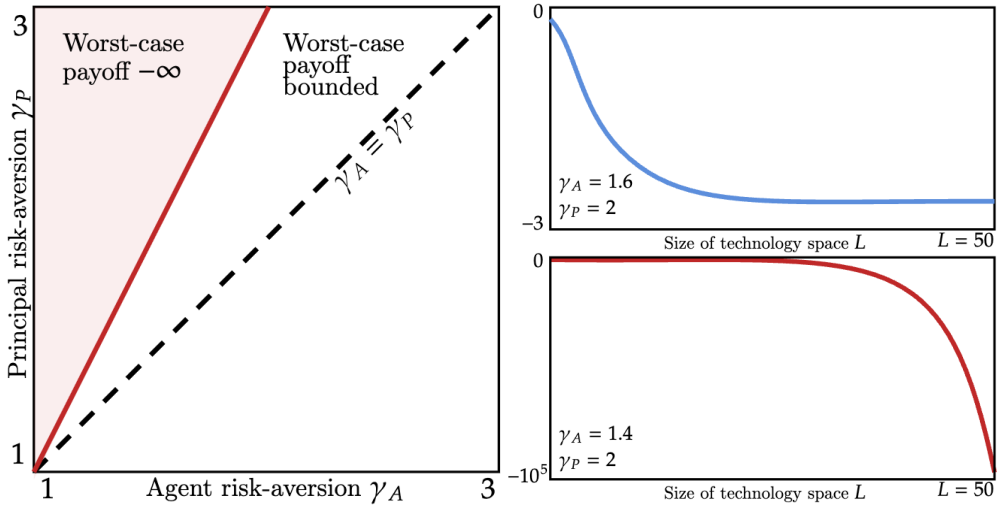
$$c(\theta, l) = \begin{cases} \exp\left(a \cdot l\right) & \text{if } \theta = 1 \\ \exp\left(-b \cdot l\right) & \text{if } \theta = 0 \end{cases}$$

with $a, b > 0$. Then for any linear Pigouvian tax $\phi_l = \alpha \cdot l$ for $\alpha \geq 0$,

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

Figure 13 illustrates worst-case payoffs as the principal and agent's risk-aversion coefficients vary, holding fixed $a = b = 1$. 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 13: Worst case (weak optimism) payoffs under CRRA utility



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 captured by our robustness notions. ♢

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

Proposition 1 (Informal)**.** *Suppose that a prior over learning processes and preferences $f \in \Delta(\mathbb{F} \times \mathbb{U})$ is 'rich'. Then:*

$$\operatorname*{argmax}_{\phi \in \Phi} \int_{\substack{(U,\mathscr{F}) \\ \in \mathbb{U} \times \mathbb{F}}} \mathbb{E}\Big[V\big(\mu_{l^*}, l^*(\phi, \mathscr{F}, U)\big)\Big] df \subseteq \Phi^{\mathsf{H}}$$

where Φ^{H} is the set of adaptive mechanisms that impose a hard limit at a finite level.

Proposition 1 states that when the principal's prior f is sufficiently 'rich' (in a sense made precise in Appendix [X]), Bayesian optimality requires hard limits. The intuition is straightforward: sans hard limits, the richness of the prior allows us to find a positive-measure set of learning processes and preferences under which payoffs are undboundedly poor per Theorem 4 (ii).

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 2. *Fix two mechanisms ϕ and ϕ' such that there exists a strictly increasing and 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 $l \in \mathscr{L}$. Then, $\Delta^L(\phi) < \Delta^L(\phi')$.*

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 insights into 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 $\mathscr{G}$ comprises of iid signals resembling the Wald problem (Wald, 1947). Indeed, this is in essence what the FDA 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[58] and for such transparently learnable technologies, the challenge regulators face is to monitor the agent's experimentation such as to close the gap between $\mathscr{G}$ and $\mathscr{F}$ e.g., through enforcing the reporting of clinical trials, pre-analysis plans, and detecting fraud.

Other technologies are *opaque*, beset by difficulties not just around learning about the technology, but also uncertainty over 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 AI systems 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

[58] See McClellan (2022) for an analysis of approval mechanisms. Key differences are that in our framework (i) our mechanism space contains all stochastic processes rather than just approval mechanisms; (ii) private information can arrive while learning rather than ex ante; and (iii) the cost of experimentation is state-dependent i.e., it is risky, as opposed to paying a constant flow cost that grows with l.

are developed, but might nonetheless invest in *active learning* i.e., choosing $\mathscr{G}$ 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 $\mathscr{G}$ and optimizes over $\mathscr{G}$-adapted mechanisms. This paves the way to analyze the 'outer problem' of how $\mathscr{G}$ 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., $\mathscr{G}$ 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 $\mathscr{F}$ 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., $\mathscr{G}$ 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[59] is now upper-bounded by the performance of mechanisms that impose hard limits on R&D (see Proposition 1) 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 model in which equilibrium R&D is jointly shaped by learning, preferences, and policy. But there is quite a lot we can add to enrich the model along various dimensions—we detail some possibilities here and collect details in Online Appendix [X].

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 con-

[59] Here for a sufficiently 'rich' prior over $\mathbb{F} \times \mathbb{U}$.

trast 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 R&D until level l in state θ is: $V(\theta, l) + \lambda \cdot \phi_l$ for some $\lambda \geq 0$. The adaptive sandbox ϕ^* augmented with a carefully chosen fixed participation tax is both learning- and dually-robust.

3. **Endogenous agent learning.** Suppose that the agent's learning process $\mathscr{F}$ 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. **R&D 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 R&D that could make a technology safe. The sandbox mechanism ϕ^* remains robustly optimal. In particular, the regulator's beliefs about the state $\mathbb{E}[\theta \mid \mathscr{G}_l]$ exhibits a positive drift even in the absence of any information, and the sandbox limit l_* takes this into account by loosening accordingly.

5. **Multidimensional Technology.** Suppose that our technology space $\mathscr{L}^{\mathsf{M}} \subseteq \mathbb{R}^d$ 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 $\boldsymbol{l} \in \mathscr{L}^{\mathsf{M}}$. If the regulator does not learn, a *multidimensional sandbox* that imposes a zero tax on R&D, and a hard limit on R&D at a manifold $\mathscr{O}_* \subset \mathscr{L}^{\mathsf{M}}$ (see Figure 14) is robustly optimal, undominated, and time-consistent.

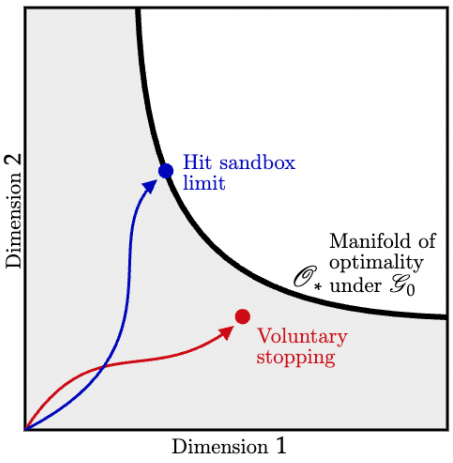


Figure 14: Multidim. sandbox

6 CONCLUSION

We have developed what is perhaps the simplest model in which learning, preferences, and policy jointly shape equilibrium R&D outcomes. We showed that adaptive sandboxes mechanisms comprising of a zero marginal tax on R&D 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 R&D such as real-world 'regulatory sandboxes' or the structure of FDA clinical trials work well, as well as guidance for how they might be improved. Taken together, our results offer simple but concrete guidance for policymakers who, in the coming years, face difficult choices around how they should regulate new and perhaps transformative technologies like artificial intelligence.

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 here it is both quantitatively and qualitatively important. This, we think, are reasons to take it seriously.

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*.

ALLEN, H. J. (2019): "Regulatory sandboxes," *Geo. Wash. L. Rev.*, 87, 579.

AMADOR, M. AND K. BAGWELL (2013): "The theory of optimal delegation with an application to tariff caps," *Econometrica*, 81, 1541–1599.

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.

BLACKWELL, D. (1951): "Comparison of experiments," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, vol. 1, 26.

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.
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.
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*.
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. (2024): "Regulating the Direction of Innovation," *Available on SSRN*.
——— (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.
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.
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.
KOH, A. AND S. SANGUANMOO (2022): "Attention capture," *arXiv preprint arXiv:2209.05570*.
KOH, A., S. SANGUANMOO, AND W. ZHONG (2024a): "Persuasion and Optimal Stopping," *arXiv preprint arXiv:2406.12278*.
——— (2024b): "Persuasion and Optimal Stopping," *Working Paper*.
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.
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 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.
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," .
SAVAGE, L. J. (1972): *The foundations of statistics*, Courier Corporation.
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.
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.