Contracting with Unknown Complementarity: Ratchets and Replacement Risk*

Job Market Paper. Click here for the latest version.

Natacha Spitzer†

October 28, 2026

Abstract

How do a firm and a worker contract while learning how productive their relationship is? A worker's effort simultaneously generates value for the firm and reveals information about the fit between the worker and his assignment. But the prospect of future wage adjustments and replacement distorts his incentives to experiment. I study these forces in a two-period principal–agent model with moral hazard, limited commitment, and limited liability. I first establish existence of a pure-strategy equilibrium and characterize the optimal spot contract: when there is a chance that effort is counterproductive, the agent may be rewarded following both sufficiently strong and sufficiently poor performance, as both are evidence of effort. After the first period, the principal decides whether to replace the agent and, conditional on retention, offers a new wage contract. A forward-looking agent therefore has an incentive to influence her belief, and because he captures limited-liability rents in each period, replacement is costly to him. This incentive decomposes into an extensive retention force and an intensive wage force, which acts as a ratchet when good news raises the bar for pay. Relative to a benchmark without belief manipulation, the retention force dominates when the parties are initially pessimistic about the fit, leading to over-experimentation, while the ratchet dominates when they are initially optimistic and effort is not too costly, leading to under-experimentation. Dynamic incentives thus raise experimentation and learning when the agent's job is at risk and lower it once it is secure.

*I would like to thank my advisor George Mailath and my committee members Kevin He and Rakesh Vohra for their support and advice throughout this project. I am also grateful to Eduardo Azevedo, Aislinn Bohren, Olivier Compte, Li Guo, Michelle Hyun, Byunghoon Kim, Alfonso Maselli, Matthew Murphy, Alex Whitefield, Andrew Postlewaite, Juuso Toikka, and Maren Vairo for helpful conversations, and to the Penn Theory group and participants in the Penn Micro Theory Lunch Seminar for their useful feedback and comments. All errors are my own.

†Department of Economics, University of Pennsylvania (email: nspitzer@sas.upenn.edu).

1 Introduction

When a worker has just been hired, moved to a different task, or given a new technology to work with, neither he nor his employer necessarily knows at the outset how productive the pairing will be. That is, there is common uncertainty about *the complementarity* between the worker and his assignment. They learn it over time by observing the worker's performance, a noisy signal of the complementarity. And when effort is not directly observable, that performance is also the sole basis for his pay.

Effort therefore matters beyond today's performance and pay. The harder the worker works, the more his performance reveals about the complementarity: in the absence of effort, performance is pure noise, and as effort rises, performance becomes more indicative of how well the worker fits his assignment. Therefore, a worker who exerts more effort effectively experiments more. Although both parties learn simultaneously from performance, it is the firm's belief that determines whether the worker keeps the assignment and on which terms. A forward-looking worker thus chooses effort anticipating how the firm will respond tomorrow to what it sees today. Since the firm does not observe effort directly, it must form a conjecture about the worker's choice. This gives the worker leeway to influence the firm's belief in the hope of better terms of employment in the future. Altogether, effort plays three roles at once: it affects today's performance and with it his pay, determines how much is learned, and is a strategic tool the worker can use to shift the firm's belief.

A first question arises: how should the firm design contracts when the complementarity is commonly unknown? If there is a chance that the worker is a poor fit for the assignment, effort can make a bad signal more likely. Either way, effort pushes the signal away from what noise alone would produce. A mediocre performance is then evidence that the worker did not try, while a disastrous one is evidence that he did but is simply a poor fit for the assignment. The optimal contract then rewards catastrophe as well as success. For instance, a thorough report that only reveals that the worker is out of his depth is a worse signal than no report at all, yet it shows that he tried.

A second question arises when the firm cannot commit today to the terms it will offer tomorrow: how does the worker's dynamic incentive to influence the firm's future belief affect how much they learn about the complementarity? Good news may lead the firm to raise the bar that the signal must meet to receive a bonus tomorrow, a ratchet, while bad news may get him fired from the assignment. If he is at high risk of being replaced, this dynamic incentive pushes him to work harder. But if his position is secure, it pushes him to work less, so as to face more lenient terms in the future and escape the ratchet. Whether this incentive leads him to experiment more or less thus depends on how secure his position is at the start of the relationship.

I model the relationship as follows. A principal and an agent, both risk neutral, interact over two periods. They share a common prior over the complementarity between the agent's effort and the task, the marginal product of effort. In each period, the principal observes a noisy signal of the agent's value in the task. Three key frictions are present. First, effort is privately chosen, so there is moral hazard. Second, the principal can commit only to spot contracts, so commitment is limited. And third, wages are bounded: the agent cannot be made to pay the principal, so liability is limited, and pay is capped by a budget set in advance. After the first signal is realized, the principal updates her belief, decides whether to retain or replace the agent, and conditional on retention, proposes a new wage contract. Her belief thus determines the agent's

future terms of employment, and a forward-looking agent has a reason to influence it. Because she does not observe effort, she reads the signal under her conjecture about the agent's choice, and by deviating from that conjecture the agent can lead her to update in his favor. This is the belief-manipulation channel. Limited liability generates rents for the agent in each period, and hence he strictly prefers to be retained rather than replaced.

The setting differs from the classical ratchet problem (Freixas et al., 1985; Laffont and Tirole, 1987, 1988) in what the agent knows at the outset. There, his productivity is his private information, which he is tempted to conceal today in order to avoid full rent extraction tomorrow following revelation. Here, neither party knows the complementarity, and there are no types: a divergence in beliefs arises only if the agent deviates from the principal's conjecture, off the equilibrium path. And since the principal has an outside option, the agent faces replacement risk in addition to the ratchet.

In each period, the optimal contract is a bonus contract. Depending on the support of the complementarity, to implement strictly positive effort via the wage, the principal pays only when the signal falls in one or two tail regions of its distribution. When effort is productive at every complementarity in the support, I recover a standard bonus contract, in which the agent is paid whenever the realized signal exceeds a single threshold. The two-tail shape arises exactly when effort is *counterproductive* with strictly positive probability. The monotone likelihood ratio property of the signal in effort then fails in a specific way: the likelihood score of the signal in effort is not monotone but crosses any level at most twice, falling and then rising, so that extreme signals in either tail are evidence of effort. Whether the principal pays one tail or two then depends on how much wage incentive the target effort requires and on whether the likelihood score is bounded. With Gaussian noise the score is unbounded, and both tails are always paid. The two tail shape does not stem from the dynamic learning itself but from the possibility that effort reveals a poor fit in that it makes poor signals more likely. The value of information may affect which effort the principal implements, but not the form of the contract. The same form arises in the second period, where there is nothing left to learn, as it would in a one-shot version of the model.

To isolate how the agent's dynamic incentive to influence the principal's belief affects how much they learn about the complementarity, I compare equilibrium first-period effort with that of a manipulation-free benchmark. In this benchmark, the agent is myopic and therefore does not internalize the effect of his first-period effort on his continuation value, while the principal remains forward-looking. I say that the agent *over-experiments* when equilibrium first-period effort exceeds the benchmark's, since more effort makes the signal more informative about the complementarity, and that he *under-experiments* when it falls short.

The comparison calls for a specification of the general model: Gaussian noise, quadratic cost, and a family of priors on the complementarity, generated by exponentially tilting a strongly log-concave base density on a fixed support. The tilt parameter orders its members from pessimistic to optimistic in the likelihood ratio order. Fixing an effort conjecture, under Gaussian noise the posteriors are also indexed by a single parameter, affine in the prior parameter and in the realized signal. The index orders the posteriors in the likelihood ratio order, as the prior parameter orders the priors. The results hold when the signal is sufficiently noisy, so that beliefs do not respond too strongly to a single observation, and the prior is sufficiently concentrated, so that both parties have some idea of the complementarity at the outset. Noise keeps the agent's and the principal's problems well behaved. Concentration makes the principal's continuation value

and the agent's continuation rent track those of the second-period problem in which the complementarity is known and equal to the posterior mean, which provides tractability.

In this regime, an optimal retention rule is a single cutoff in the signal realization, above which the agent is retained. The agent's dynamic incentive then decomposes into an *extensive* retention force and an *intensive* wage force. The retention force pushes towards more effort today, since the agent strictly prefers to be retained rather than replaced, and increasing effort at the margin pushes strictly positive probability mass past the firing threshold. The sign of the wage force is ambiguous: good news raises tomorrow's surplus at any given effort, but whether the agent keeps part of the increase depends on how the principal adjusts the effort she targets, and the contract that implements it, to her belief. If the complementarity were known and low, the principal would buy little effort with a bonus that is rarely paid, and the agent's rent would be small. As the complementarity increases, she implements more effort, which initially requires lowering the bonus bar to provide incentives, and the agent's rent rises. Once the bonus is paid with sufficiently high probability, further improvements call for an increasingly high bar to maintain incentives, and the agent's rent falls: this is the ratchet. Under sufficient concentration the same holds at the posterior mean, so the wage force changes sign at most once along the posterior index, from positive to negative.

When effort is not too costly, the ratchet appears when beliefs are sufficiently optimistic and the two forces balance at a single threshold in the prior: compared to the manipulation-free benchmark, the agent over-experiments below the threshold and under-experiments above it, in every equilibrium. When the prior is just optimistic enough for the principal to prefer the agent to the outside option in the first period, the retention force necessarily dominates: at initiation, the agent is fired in expectation, so his priority is to keep his job tomorrow. Because this incentive constitutes a subsidy on the wage that the principal has to pay the agent in the first period, the relationship also forms at more pessimistic priors than in the benchmark. As the prior becomes more optimistic, retention becomes more and more likely, and the wage force dominates. The retention force fades rapidly once replacement becomes unlikely, whereas the wage force fades much more slowly as the prior becomes more optimistic, which is why the two cross only once. When effort is costly enough, the wage force reinforces the retention force and he over-experiments at every prior at which the principal hires.

Answering these economic questions involves first answering a more technical one: does a pure-strategy equilibrium exist? In settings that combine a hidden action with limited commitment and sequential timing, where the principal responds to what she infers after the action is taken, the only equilibria are often in mixed strategies. In a similar environment, Bhaskar and Roketskiy (2023) show that full rent extraction in the second period leaves the agent's objective with a convex kink exactly at the conjectured effort, so that no interior effort can be implemented in pure strategies. Here, limited liability and a zero outside option for the agent ensure that his continuation rent is continuously differentiable in his first-period effort, so the kink is absent. Sufficient curvature of the effort cost then validates the first-order approach, and a pure-strategy equilibrium exists.

Two further results complete the analysis. First, the switch from over- to under-experimentation is not vacuous: when effort can backfire and the signal is noisy enough, there is a nonempty interval of intermediate cost curvatures for which the theorem applies and the two forces oppose. Second, I apply the result to a principal deciding whether to adopt a new technology. Adoption is an irreversible investment made before

the relationship starts. It buys the principal both an outside option, since she can run the technology on her own, and the chance to pair it with the agent, whose fit with it is commonly unknown and revealed over time as he exerts effort. Adoption is again governed by a threshold in the prior. Relative to the benchmark, she adopts weakly more at pessimistic priors, where the agent's dynamic incentive acts as a subsidy to the principal and the technology is worth more to her, and weakly less at optimistic ones.

Related literature. In the classical ratchet problem the agent knows his productivity and the principal does not, so the asymmetric information is exogenous and the agent is tempted to conceal it. Weitzman (1980) names the effect, and Freixas et al. (1985) and Laffont and Tirole (1987, 1988) establish the tension: revealing a good type invites tougher terms, so separation is hard to sustain. Gerardi and Maestri (2020) show that under limited commitment, as the parties become patient, the principal either learns nothing or dismisses the low type at once. Here there are no types: both parties start from the same belief about the complementarity, and their beliefs diverge only when the agent deviates from the principal's conjecture. Ekmekci et al. (2022) are closest in mechanism: below a firing cutoff their agent mimics the good type to save himself, and the incentive to save the relationship destroys learning. The agent near the retention cutoff here also raises effort to save the relationship, but since effort makes the signal more informative rather than imitating a type, the same incentive increases what the principal learns.

When uncertainty is symmetric, both parties learn the parameter from the signal and effort moves the principal's belief without revealing a type, as in the signal-jamming of Fudenberg and Tirole (1986) and the career-concerns literature (Holmström, 1999; Dewatripont et al., 1999). Cisternas (2018) derives a learning-driven ratchet principle in a continuous-time game against a market. Prat and Jovanovic (2014) and DeMarzo and Sannikov (2017) characterize the optimal long-term contract under full commitment, with symmetric learning and an additive hidden action. In the latter, shirking lowers the principal's estimate of profitability and earns the agent information rents, which the contract controls by distorting the termination decision. Gao and Xu (2020) study spot contracts in a discrete setting in which the contract can commit to firing the agent after certain outcomes. Three things differ here: the principal cannot commit, so the firing decision must be sequentially rational, and retention and the second-period contract are reoptimized after she learns. Effort is multiplicative, so the agent controls both *how informative* the signal is and *what* the principal infers from it, since she interprets it through her conjecture of his effort. Limited liability leaves a rent that the retention decision puts at risk, which makes the manipulation incentive two-sided.

Kim (2026) studies replacement risk in a career-concerns setting with *observable* effort and obtains the opposite sign on the same margin: since the principal sees effort, a failure after high effort is attributed to the agent's ability, so the agent shirks to make failure uninformative and avoid replacement. Here effort is hidden, so the agent cannot shift the blame and instead works to stay.

The closest model is Bhaskar and Roketskiy (2023): a principal who cannot commit, symmetric learning about the difficulty of the task with finitely many states and signals, and a continuum of efforts that determine the mixture of experiments the signal is drawn from, so that effort moves both where beliefs land and how fast they resolve, as here. The difference is limited liability, which removes the kink discussed above and creates the extensive margin. The same non-existence appears in hold-up with unobservable investment (Gul, 2001) and in moral hazard with interim renegotiation (Fudenberg and Tirole, 1990), where the only

equilibria are in mixed strategies. In a similar setting with finite actions, Bhaskar and Mailath (2019) show that effort becomes costlier to implement as the horizon lengthens.

A separate literature studies contracts for experimentation on a project whose quality is revealed over time: Bergemann and Hege (2005), Manso (2011) and Halac et al. (2016) under full commitment, and Hörner and Samuelson (2013) without it, with limited liability and termination, as here. One main difference lies in the information structure: in these papers the agent chooses whether to experiment, and the signal is a success or a failure, whereas here the signal is continuous and informative in both directions, and effort itself is the experiment, since it scales how informative the signal is. One feature is shared: the lower tail of the contract here resembles the tolerance for early failure in Manso (2011), in that both pay for outcomes the principal would rather not see but are evidence of the action she wants, exploration there and effort here. Saeedi and Shourideh (2026) make a related point in rating design: taking as given that effort increases tail risk, they find that the optimal rating pools poor outcomes rather than punishing them. Here that property arises from common uncertainty about the sign of the complementarity, and the instrument is a wage under limited liability, which pays both tails.

The optimal contract is the limited-liability bonus of Innes (1990). The second threshold is new, and arises because, when the complementarity can be negative, a very low signal is itself evidence of effort. The derivation of the contract involves replacing the agent's incentive constraint with his first-order condition, and the classical conditions under which that condition is sufficient (Mirrlees, 1999; Rogerson, 1985; Jewitt, 1988) fail here: Gaussian noise violates the convexity condition and counterproductive effort breaks the likelihood ratio order. The static moral hazard problem with bounded payments has been solved without the approach (Chi and Olsen, 2025), and Azevedo and Wolff (2025) restore it through a high reservation utility, a route unavailable when the agent's utility is the endogenous limited-liability rent. Here validity comes from curvature of the effort cost, established directly.

The remainder of the paper proceeds as follows. Section 2 introduces the model and the agent's and the principal's problems in each period. Section 3 validates the first-order approach, establishes existence of equilibrium, and characterizes the optimal contract, all for general noise. Section 4 works with a specification of the model, Gaussian noise, quadratic cost and a family of priors, introduces the manipulation-free benchmark, decomposes the agent's dynamic incentive into a retention force and a wage force, and compares equilibrium effort with the benchmark's. Section 5 extends the analysis to the adoption of a new technology. Section 6 concludes. Appendix A proves the general results, and Appendix B the results under the specification of Section 4.

2 Model

2.1 Primitives

Players and the signal. A principal (she) and an agent (he) interact over two periods, $t = 1, 2$. There is no discounting and both parties are risk-neutral. In each period, the principal observes a signal y_t of the agent's

performance, which measures the value he generates for her and is the only basis for his pay:

$$y_t = \theta\, e_t + \varepsilon_t,$$

where $\theta \in \Theta := [\underline{\theta}, \bar{\theta}]$ is a commonly unknown complementarity parameter, $e_t \in \mathbb{R}_+$ is the agent's private effort choice, and $\varepsilon_t \in \mathbb{R}$ is an unobserved noise term. θ has prior density g_θ and ε_t is i.i.d. across periods and independent of θ and of effort, with density g_ε and distribution function G_ε. Whether effort is productive and by how much depends on the degree of complementarity θ, which may be positive or negative.

Assumption 1 (Distributional regularity)**.**

(i) (Noise.) *The noise density g_ε is $C^2(\mathbb{R})$, has full support on $\mathbb{R}$ and mean zero. Its location kernel $(y, x) \mapsto g_\varepsilon(y - x)$ is strictly totally positive of order three, so in particular g_ε is log-concave. Its Fisher information $\int_{\mathbb{R}} (g_\varepsilon')^2 / g_\varepsilon$ is finite.*[1]

(ii) (Complementarity.) $g_\theta \in C^2(\Theta)$ *is log-concave with compact support $\Theta = [\underline{\theta}, \bar{\theta}]$, where $\bar{\theta} > 0$.*

Outside options. The principal has an outside option worth $\pi_0 > 0$ per period, commonly known and independent of θ. Her outside option is therefore $2\pi_0$ at the start of the first period and π_0 at the start of the second period. The agent's outside option is zero. At the beginning of each period, the principal may decide to fire the agent and take the outside option, and similarly, upon observing the contract, the agent has the option to exit the relationship. Denote by $\rho_t \in \{0, 1\}$ the principal's retention decision, with $\rho_t = 1$ when the agent is retained. The first-period decision is whether to form the relationship at all and the second whether to fire the agent after observing his signal, and since only one such decision is taken in each period the subscript is dropped where no ambiguity arises.

Wages. The wage schedule $w_t : \mathbb{R} \to \mathbb{R}$ is subject to limited liability and a uniform cap: for every signal $y_t \in \mathbb{R}$,

$$0 \;\le\; w_t(y_t) \;\le\; \bar{w}.$$[2]

Limited liability generates positive rents for the agent. The cap is a bound on the reward the principal can promise. It is also what makes the optimal contract well defined, since without it the principal could buy

[1] Strict total positivity of order three means $\det[g_\varepsilon(y_i - x_j)]_{i,j \le n} > 0$ for $y_1 < \cdots < y_n$, $x_1 < \cdots < x_n$ and $n \le 3$. It implies strict log-concavity, hence unimodality and $g_\varepsilon' \in L^1(\mathbb{R})$ with $\int |g_\varepsilon'| = 2 \max g_\varepsilon$. Together with finite Fisher information, log-concavity also gives $g_\varepsilon'' \in L^1(\mathbb{R})$ with $\int |g_\varepsilon''| \le 2 \int (g_\varepsilon')^2 / g_\varepsilon$ (see the proof of Proposition 1 in Appendix A). See Bagnoli and Bergstrom (2005) for more properties of log-concave densities. Order three itself is used only for the two-tail contract of Theorem 2(ii). Densities of order three are characterized by Weinberger (1983): writing $g_\varepsilon = e^{-\psi}$, and given the other requirements of the assumption, the kernel is strictly totally positive of order three if and only if $\log \psi'' - 2\psi$ is strictly concave. By Schoenberg's characterization of such densities (see Karlin, 1968), the Gaussian, logistic, Laplace and exponential densities, and the gamma density with integer shape, are all totally positive of every order. The Gaussian and logistic densities satisfy all of these requirements, the Gaussian with $-(\log g_\varepsilon)'' = 1/\sigma^2$. The others fail them: the Laplace and exponential densities are not C^2, and the exponential and gamma densities do not have full support.

[2] The cap is uniform rather than tied to the realized signal as in Innes (1990). There the setting is a financial relationship and output is a nonnegative profit: the entrepreneur's payment comes out of it, so he cannot pay more than there is, and the optimum is a debt contract. Here, y_t is a performance signal rather than a cash flow the agent hands over, and since the signal can be negative, such a cap would read $w_t \le \max\{0, y_t\}$. Such a cap would rule out payments at low signals, which Theorem 2 shows are part of the optimal contract when effort can be counterproductive.

any effort with a payment made with vanishing probability far in the tails.[3]

An *admissible* contract in period t is a wage schedule w_t satisfying these bounds. Beyond limited liability and the wage cap the principal is unrestricted. The retention decision is separate, taken at the start of the period before any contract is proposed. For every public history (w_1, y_1), we get a retention decision $\rho_2(w_1, y_1)$ and wage schedule $w_2(\cdot; w_1, y_1)$.

Limited commitment. The principal's commitment power is limited: although she can commit to the current wage schedule upon retention, she cannot commit to future terms, nor can she commit to retain the agent in the future. Therefore, the principal uses what she observes in the first period to reoptimize the second-period retention choice and proposed wage schedule based on her updated belief about θ.

Costs. The agent's effort cost $c : \mathbb{R}_+ \to \mathbb{R}_+$ is $C^2(\mathbb{R}_+)$, strictly convex, with $c(0) = 0$, $c'(0) = 0$, and c'' nondecreasing.

Timing. In each period, the principal first decides whether to retain the agent. If she retains the agent, she proposes a wage schedule. Then, the agent accepts or rejects the contract. If he accepts, he privately chooses effort. Finally, the signal is realized and both parties update their beliefs. If either party walks away at any point, each gets their outside option and they do not have the chance to interact again in the future.[4]

2.2 Equilibrium

The solution concept is perfect Bayesian equilibrium, and I restrict attention to pure strategies. The restriction is without loss rather than a selection. Under the cost curvature maintained for the first-order approach, the agent's objective is strictly concave in effort, so his best response is unique at every history. For the principal, since the signal has a density, resolving her ties deterministically, signal by signal, generates every continuation incentive a mixture could. The principal's belief over the agent's first-period effort is accordingly characterized by a single conjecture $\hat{e}_1 \in \mathbb{R}_+$ rather than a distribution over effort levels. Section 3.2 shows that such an equilibrium exists.

Beliefs. Updating beliefs about the complementarity depends on the effort one conditions on. For any effort level $e \geq 0$, Bayes' rule gives the posterior density

$$p(\theta \mid y_1, e) = \frac{g_\varepsilon(y_1 - e\,\theta)\, g_\theta(\theta)}{\int_\Theta g_\varepsilon(y_1 - e\,\theta')\, g_\theta(\theta')\, d\theta'}. \tag{1}$$

Both players update by this formula and differ only in the effort they substitute: the agent knows the effort he exerted and substitutes the true e_1, while the principal, who cannot observe effort directly, substitutes her conjecture $\hat{e}_1$. On the equilibrium path, $e_1 = \hat{e}_1$, and the two coincide.

[3] This is the reward counterpart of the problem in Mirrlees (1999), where unbounded penalties let the principal approach the first best by punishing severely on a vanishing set of signals, so that no optimum exists.

[4] This may be interpreted as a property of the outside options, rather than additional commitment power by the principal.

Stage 2: the agent's problem. Fix a public history (w_1, y_1). Conditional on retention and acceptance, an agent who exerted e_1 observes the contract and solves

$$\max_{e_2 \geq 0} \int w_2(y_2; y_1)\, f_2(y_2 \mid e_2; y_1, e_1)\, dy_2 \;-\; c(e_2), \tag{2}$$

where the second-period signal density is

$$f_2(y_2 \mid e_2; y_1, e_1) = \int_\Theta g_\varepsilon(y_2 - e_2\,\theta)\, p(\theta \mid y_1, e_1)\, d\theta.$$

Let $e_2^*(w_2; y_1, e_1)$ denote a solution to (2). Write the agent's indirect utility from (2) as

$$U_2(w_2; y_1, e_1) = \int w_2(y_2; y_1)\, f_2(y_2 \mid e_2^*; y_1, e_1)\, dy_2 - c(e_2^*).$$

Stage 2: the principal's problem. The principal's second-period problem at history (w_1, y_1) and conjecture $\hat{e}_1$ is

$$\max_{0 \leq w_2(\cdot\,; y_1) \leq \bar{w},\; e_2 \geq 0} \int \big[y_2 - w_2(y_2; y_1)\big]\, f_2(y_2 \mid e_2; y_1, \hat{e}_1)\, dy_2$$

$$\text{s.t. } e_2 \;\in\; \arg\max_{\tilde{e}_2 \geq 0} \int w_2(y_2; y_1)\, f_2(y_2 \mid \tilde{e}_2; y_1, \hat{e}_1)\, dy_2 - c(\tilde{e}_2), \tag{IC$_2$}$$

$$\int w_2(y_2; y_1)\, f_2(y_2 \mid e_2; y_1, \hat{e}_1)\, dy_2 - c(e_2) \;\geq\; 0. \tag{IR$_2$}$$

By the observation above, (IR$_2$) is implied by the wage bounds together with (IC$_2$): limited liability makes the payoff from zero effort nonnegative, and incentive compatibility makes the agent's equilibrium payoff at least as high. The first-period contract w_1 enters her second-period problem only through the conjecture $\hat{e}_1$ it induces, and the signal only through the posterior $p(\theta \mid y_1, \hat{e}_1)$, so her choice depends on y_1 and $\hat{e}_1$ alone. Denote by $w_2^*(\cdot\,; y_1, \hat{e}_1)$ the solution she offers, one of them if there are multiple, and $\Pi_2^*(y_1, \hat{e}_1)$ the resulting value for the principal. The indirect utility of the agent is then $U_2^*(y_1, e_1, \hat{e}_1) = U_2(w_2^*(\cdot\,; y_1, \hat{e}_1); y_1, e_1)$.

Retention. At the beginning of the second period, unable to commit to future terms, the principal compares her continuation value from retaining the agent with that of replacing him. An optimal retention rule therefore satisfies

$$\rho^*(y_1, \hat{e}_1) \begin{cases} = 1 & \text{if } \Pi_2^*(y_1, \hat{e}_1) > \pi_0, \\ \in \{0, 1\} & \text{if } \Pi_2^*(y_1, \hat{e}_1) = \pi_0, \\ = 0 & \text{if } \Pi_2^*(y_1, \hat{e}_1) < \pi_0. \end{cases} \tag{3}$$

Where she is indifferent, either choice is optimal, and an equilibrium may resolve these ties differently from one history to another.

Stage 1: the agent's problem. Fixing the contract w_1 and the continuation, which runs through the conjecture $\hat{e}_1$, the agent solves

$$\max_{e_1 \geq 0} \int \left[w_1(y_1) + \rho^*(y_1, \hat{e}_1)\, U_2^*(y_1, e_1, \hat{e}_1) \right] f_1(y_1 \mid e_1)\, dy_1 \;-\; c(e_1),$$

with

$$f_1(y_1 \mid e_1) = \int_\Theta g_\varepsilon(y_1 - e_1\,\theta)\, g_\theta(\theta)\, d\theta.$$

Let $e_1^*(w_1, \hat{e}_1)$ denote a solution. Note that the agent's deviation affects both the density of the signal $f_1(\cdot \mid e_1)$ and his own posterior inside U_2, but does not affect the retention rule nor the stage-2 contract, both chosen by the principal under the conjecture $\hat{e}_1$.

Stage 1: the principal's problem. The principal's full first-period program is

$$\max_{0 \leq w_1(\cdot) \leq \bar{w},\; e_1 \geq 0} \int \left[y_1 - w_1(y_1) + \rho^*(y_1, e_1)\, \Pi_2^*(y_1, e_1) + (1 - \rho^*(y_1, e_1))\, \pi_0 \right] f_1(y_1 \mid e_1)\, dy_1$$

$$\text{s.t. } e_1 \in \arg\max_{\tilde{e}_1 \geq 0} \int \left[w_1(y_1) + \rho^*(y_1, e_1)\, U_2^*(y_1, \tilde{e}_1, e_1) \right] f_1(y_1 \mid \tilde{e}_1)\, dy_1 - c(\tilde{e}_1), \tag{IC$_1$}$$

$$\int \left[w_1(y_1) + \rho^*(y_1, e_1)\, U_2^*(y_1, e_1, e_1) \right] f_1(y_1 \mid e_1)\, dy_1 - c(e_1) \;\geq\; 0. \tag{IR$_1$}$$

Again (IR$_1$) is implied by the wage bounds together with (IC$_1$): zero effort yields nonnegative wages and a nonnegative continuation rent at zero cost, and the agent's optimal choice does at least as well. Constraint (IC$_1$) embeds a *fixed-point problem* within the principal's program. In equilibrium, a target effort is implemented under w_1 if and only if it is a fixed point of the agent's best-response correspondence: a conjecture $\hat{e}_1$ is consistent with Bayes' rule if and only if the agent would indeed choose $\hat{e}_1$ given w_1, anticipating the optimal retention and contract in the second period. Denote by Π_1^* the value of the principal's first-period problem.

3 Existence of equilibrium and the optimal contract

This section contains the paper's general results. It establishes that a pure-strategy equilibrium exists and characterizes the optimal contract. I first establish in Section 3.1 that the first-order approach is valid for sufficient cost curvature, allowing us to replace the agent's incentive-compatibility constraint with his first-order condition in the principal's problem. Section 3.2 then establishes that a pure-strategy equilibrium exists, and Section 3.3 characterizes the cheapest contract implementing strictly positive effort.

3.1 Regularity and the first-order approach

I use the first-order approach, meaning that it is without loss for the principal to replace the agent's incentive-compatibility constraint with the agent's first-order condition when designing the contract.[5] To establish the

[5]The alternative is to impose incentive compatibility globally, a continuum of constraints comparing the target to every other effort level, which is generally intractable. The validity of the first-order approach is therefore a question commonly faced in

validity of this approach, I show that for a sufficiently convex cost of effort, the agent's objective in each period is strictly concave in effort, at every contract and retention rule he may face. I first establish that, conditional on the strict concavity of the agent's second-period objective across all admissible contracts, the agent's continuation rent is continuously differentiable everywhere and twice continuously differentiable almost everywhere in first-period effort. The agent's first-period objective includes his continuation rent $U_2^*(y_1, e_1, \hat{e}_1)$, which depends on his first-period effort through his own posterior and his resulting second-period effort. Strict concavity in the second period ensures that the optimal second-period effort choice of the agent is unique, and hence continuous in his first-period effort choice. And by the envelope theorem, the continuation rent is smooth.[6]

Fix any admissible second-period contract w_2 and define

$$u_2(w_2; \theta, y_1, e_1) := \int_{\mathbb{R}} w_2(y_2; y_1)\, g_\varepsilon(y_2 - e_2^*(w_2; y_1, e_1)\,\theta)\, dy_2 - c(e_2^*(w_2; y_1, e_1)),$$

the state-conditional continuation rent at that contract, evaluated at the agent's second-period response, so that $U_2 = \mathbb{E}_p[u_2]$. The second-period objective is twice continuously differentiable in e_2 for every admissible contract, with second derivative $\int_{\mathbb{R}} w_2\, \partial_{e_2}^2 f_2\, dy_2 - c''(e_2)$, so the concavity condition of the next lemma is a statement about a well-defined object.

Lemma 1 (Own-effort regularity and the continuation incentive)**.** *Suppose the second-period objective is strictly concave in e_2 at every admissible contract. Then for every admissible contract and retention rule, $U_2(w_2; y_1, \cdot) \in C^1(\mathbb{R}_+)$, and it is C^2 almost everywhere on $\mathbb{R}_+$, and*

$$\frac{d}{de_1}\,\mathbb{E}[\rho\, U_2] = -\int_\Theta \theta \Big[\int_{\mathbb{R}} \rho(w_1, y_1)\, u_2(w_2; \theta, y_1, e_1)\, g_\varepsilon'(y_1 - e_1\theta)\, dy_1\Big]\, g_\theta(\theta)\, d\theta.$$

By the envelope theorem, conditional on the signal, a marginal change in first-period effort has no first-order effect on his continuation rent through his second-period choice: that choice is optimal given his posterior, and the contract he faces is fixed, because the principal cannot directly observe his effort. What remains is the effect of his effort on the distribution of the signal in each state of the complementarity, which is what the formula captures. As discussed in the introduction, this differentiability is a key feature that distinguishes this model from the related one of Bhaskar and Roketskiy (2023), where full extraction in the second period leaves the agent's continuation with a kink exactly at the conjectured effort. Here, limited liability ensures the non-negativity of the agent's second-period payoff after any deviation, so the quit option never binds and the continuation is smooth.[7]

The analysis characterizes the agent's optimal effort at a given contract by his first-order condition, replacing (IC_2) and (IC_1) in the principal's programs. In general, whether the first-order condition identifies a global maximum may depend on the proposed contract. The classical sufficient conditions (Rogerson, 1985; Hart and Holmström, 1987; Jewitt, 1988; Mirrlees, 1999) do not apply here: log-concave noise fails the

contracting models with a continuum of action choices.

[6] Without strict concavity, his second-period effort could jump from one optimum to another as his first-period effort changes, introducing kinks into the agent's first-period objective.

[7] The agent's continuation at a retained history is $\max\{0, U_2^*\}$, since he may decline the second-period contract. Under limited liability, $U_2^* \geq 0$ pointwise, so the quit option is never exercised.

convexity-of-the-distribution-function condition even at a known complementarity, effort shifts the signal's location rather than mixing two fixed distributions, and when effort can be counterproductive the monotone likelihood ratio property of the distribution of the signal with respect to effort fails as well, since more effort raises the likelihood of very low signals.[8] The solution to static problems with bounded payments has been characterized without the approach (Chi and Olsen, 2025), and Azevedo and Wolff (2025) restore it through a high reservation utility when the agent is risk averse. Neither route is directly available in this model with dynamic incentives, unknown complementarity, and risk neutrality. I instead directly require the agent's objective to be strictly concave in effort in each period, uniformly over the contract space, and show that this is satisfied when the cost has sufficient curvature.

Assumption 2 (First-order approach)**.** *For $t \in \{1, 2\}$ and all admissible contracts and retention rules, the agent's period-t objective is strictly concave in e_t. In the second period, at every first-period signal y_1 and for all $e_1, e_2 \geq 0$,*

$$c''(e_2) \;>\; \int_{\mathbb{R}} w_2 \, \frac{\partial^2 f_2}{\partial e_2^2} \, dy_2, \tag{4}$$

and in the first period, for all $e_1 \geq 0$,

$$c''(e_1) \;>\; \int_{\mathbb{R}} \left[(w_1 + \rho\, U_2) \frac{\partial^2 f_1}{\partial e_1^2} + 2\rho \, \frac{dU_2}{de_1} \frac{\partial f_1}{\partial e_1} + \rho \, \frac{d^2 U_2}{de_1^2} \, f_1 \right] dy_1. \tag{5}$$

Proposition 1 (Cost curvature validates the first-order approach)**.** *There exists $\alpha^{\mathrm{FOA}} \in (0, +\infty)$ such that Assumption 2 holds for all costs with $c'' > \alpha^{\mathrm{FOA}}$. In particular it holds for the quadratic cost $\frac{\alpha}{2}e^2$ whenever $\alpha > \alpha^{\mathrm{FOA}}$.*

The proposition rests on two observations. First, because payments are bounded, the wage can add only limited curvature to the agent's problem: at most the wage cap times the curvature of the density of the signal in effort, integrated over the signal. That integral is bounded uniformly in effort, because the complementarity is bounded and the noise has finite Fisher information. The same bound covers the direct effect of first-period effort on the continuation, since the agent's rent is also bounded by the wage cap. Second, his continuation rent depends on first-period effort through what the agent learns and how he responds tomorrow, and cost curvature dampens that response: his second-period effort moves with his posterior at a rate inversely proportional to the concavity of his second-period problem, so this channel shrinks as the cost becomes more convex. The left-hand sides of (4) and (5) therefore grow without bound in the curvature while the right-hand sides are bounded or falling, so sufficient curvature dominates.

3.2 Existence of equilibrium

In Section 2, I note that the principal's first-period program involves solving a fixed point problem for the agent: solving for the second period and imposing Bayes' rule, the optimal effort choice at the first-period contract must be consistent with the principal's conjecture going into the second period. Showing that

[8] At a known complementarity θ, $\partial_e^2 F(y \mid e) = \theta^2 g_\varepsilon'(y - e\theta)$, which is positive below the mode of the noise and negative above it, so the distribution function is convex in effort on one side of the mode and concave on the other.

a fixed point exists is the main obstacle to existence of pure-strategy equilibrium, which fails in similar environments without limited liability (Bhaskar and Roketskiy, 2023).

Theorem 1 (Existence)**.** *Under Assumptions 1 and 2, a pure-strategy perfect Bayesian equilibrium exists.*

By Proposition 1, Assumption 2 holds whenever $c'' > \alpha^{\text{FOA}}$. At wage schedule w_1 and conjecture $\hat{e}_1$, the agent's effort is characterized by his first-order condition,

$$c'(e_1) = \int_{\mathbb{R}} w_1(y_1)\, \partial_{e_1} f_1(y_1 \mid e_1)\, dy_1 + T(e_1, \hat{e}_1), \tag{6}$$

where $T(e_1, \hat{e}_1)$ is the continuation incentive $\frac{d}{de_1}\mathbb{E}[\rho^* U_2^*]$ of Lemma 1, evaluated at the contract and retention rule selected under the conjecture $\hat{e}_1$ and effort choice e_1. Given the principal's second-period behavior, its solution is unique by Assumption 2. But depending on the conjecture and the realized signal, the principal may be indifferent between retaining and replacing the agent, or between contracts, and her selection among optimal choices may generate different incentives for the agent in the first period. Her choices can be made measurable in first-period signal by a measurable selection theorem, and looking across all possible selection rules, the agent's best response is a correspondence. Ties may be broken signal by signal, and because Lebesgue measure on the signal has no atoms, the continuation incentives she may generate through such selections form a compact interval, by Aumann's theorem on integrals of correspondences (Aumann, 1965). So does the set of best responses, which varies upper hemicontinuously with the conjecture. A consistent effort therefore exists for some measurable selection, by Kakutani's theorem.

The principal chooses among all wage schedules between 0 and $\bar{w}$, a set that is compact in the weak-* topology of L^∞. Because both parties are risk neutral, expected payments and incentives are integrals of the contract against integrable densities, hence continuous in that topology. Moreover, the set of consistent efforts across all selection rules is upper hemicontinuous in the contract. If the agent plays the consistent effort that the principal prefers across all selections, along with the selection by the principal that supports it, the principal's value is upper semicontinuous on the compact contract space, and an optimal contract exists.

Because the principal may be indifferent and consistent efforts may be multiple, there may be multiple equilibria, so a comparative static must say which equilibrium it describes. Section 4 addresses this under its specification: there, the comparison holds for every equilibrium, and the contract the principal offers uniquely pins down the agent's effort (see the proof of Theorem 3).

3.3 The optimal contract

I now characterize the optimal contract implementing strictly positive effort. The cheapest contract implementing zero effort is the null contract, which pays zero after every signal: under a conjecture of zero, the continuation incentive is null, and exerting effort would cost the agent something without raising his expected payment. In the second period, a strictly positive wage incentive is the only way to implement positive effort, whereas in the first period the continuation incentive may implement positive effort even without one. The result therefore concerns efforts that require a strictly positive wage incentive. In the specification of Section 4, I show that the principal strictly gains by providing a strictly positive wage incentive in the first period, so that there the theorem characterizes the contract offered in equilibrium in both periods.

As in the canonical model of Innes (1990), with limited liability and risk-neutral parties, the optimal contract is a bonus contract: the principal pays either the full cap or nothing, depending on the realized signal. Following the classic characterization of incentives (Mirrlees, 1999; Holmström, 1979), the contract depends on the signal only through the score of effort, the likelihood ratio of a marginal increase in effort, and with both a floor and a cap on pay, it pays the cap exactly where the score exceeds a level. When effort is always productive, the principal pays when the signal is high enough. When effort is counterproductive with strictly positive probability, she pays on up to two tails, above one threshold and below another. The second tail is the novel feature, and it comes from a specific failure of the monotone likelihood ratio property: when the support extends below $\theta = 0$, a very low signal is evidence of hard work by an agent whose complementarity with the assignment is negative, and since the contract is designed before the complementarity is known, rewarding that tail becomes part of the cheapest incentive scheme.

Theorem 2 (Shape of the optimal contract)**.** *Fix a period $t \in \{1, 2\}$ and a strictly positive implementable effort that requires strictly positive wage incentive.*[9] *The cheapest contract implementing it pays either $\bar{w}$ or nothing at almost every signal, and:*

(i) *if $\underline{\theta} \geq 0$, so that effort is never counterproductive, it has exactly one threshold z_t, paying $\bar{w}$ on $[z_t, +\infty)$ and nothing below;*

(ii) *if $\underline{\theta} < 0$, so that effort can be counterproductive with strictly positive probability, it pays $\bar{w}$ on a union of at most two tails, $(-\infty, z_t^-] \cup [z_t^+, +\infty)$ with $z_t^- < z_t^+$, one of which may be empty, and nothing in between. If the noise is strongly log-concave, both tails are always paid.*[10]

The presence of lower-tail payments at the optimal contract is not a learning phenomenon. It also arises in the one-period problem, where what the signal reveals about the complementarity is never used: the complementarity is unknown, and when its support extends below $\theta = 0$ the distribution of the signal fails the monotone likelihood ratio property, so the cheapest incentive scheme is non-monotonic in the signal. Learning affects the contract only through the target effort. The value of information may change how much effort the principal implements in the first period. And, when the score is bounded, as under logistic noise, this may then affect how many tails are paid. But the characterization holds at every target that requires strictly positive wage incentive.[11]

The proof has three steps. First, with risk neutrality and pay between zero and the cap, the principal's Lagrangian is linear in the wage. So the cheapest contract pays the cap exactly where the score of effort, $\partial \log f_t / \partial e_t$, exceeds $1/\lambda_t$, where $\lambda_t > 0$ is the Lagrange multiplier on the agent's first-order condition in period t, and nothing elsewhere. Second, the thresholds are the sign changes of $\partial f_t / \partial e_t - f_t / \lambda_t$, which I show is the convolution of the location kernel generated by the noise density, $K(y_t, \theta e_t) \equiv g_\varepsilon(y_t - \theta e_t)$, and an

[9] In period 2 this is the only way to implement positive effort, since there is no continuation. In period 1, the agent's first-order condition reads $c'(\tilde{e}_1) = M_w + T$, with M_w the marginal incentive delivered by the wage and T by the continuation. The hypothesis is $M_w > 0$, which holds at every target above a continuation-only effort e_0, the highest consistent effort the agent would choose at the null contract.

[10] That is, $-(\log g_\varepsilon)'' \geq m$ for some $m > 0$.

[11] Logistic noise, $g_\varepsilon(u) = e^{-u/s}/(s(1 + e^{-u/s})^2)$, satisfies Assumption 1, with Fisher information $1/(3s^2)$, but its likelihood score is bounded, since $(\log g_\varepsilon)'$ takes values in $(-1/s, 1/s)$. When $\underline{\theta} < 0$ the cheapest contract may then pay on one tail only, and if $|\underline{\theta}| > \bar{\theta}$, so that effort can hurt more than it can help, for some targets that tail is the left one: the principal pays only for a very low signal.

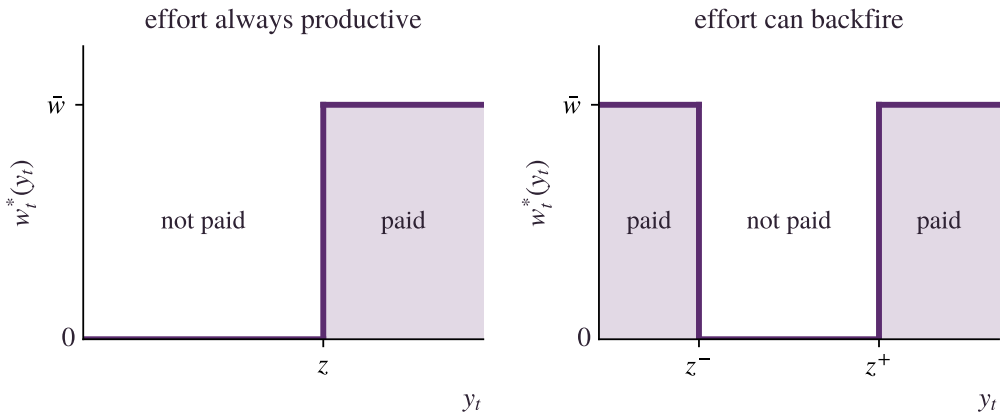


Figure 1: The optimal contract of Theorem 2. Left: when effort is productive at every complementarity, the principal pays the cap above a single threshold z, where a high signal is the evidence of effort. Right: when effort can be counterproductive and the noise is strongly log-concave, she pays on both tails, $(-\infty, z^-] \cup [z^+, +\infty)$, and nothing on the middle interval. A very low signal is then evidence that effort was exerted and backfired.

explicit weight function of θe_t. The weight changes sign at most once when $\underline{\theta} \geq 0$, and at most twice when $\underline{\theta} < 0$. The noise is variation diminishing at the order Assumption 1 provides, so smoothing cannot add sign changes, and there are at most that many thresholds. Finally, there is at least one crossing, since otherwise there would be no wage incentive, and depending on the limit of the score in the tails of the distribution a second crossing may appear. When noise is strongly log-concave, the score is unbounded in the tails, so necessarily there are two crossings when $\underline{\theta} < 0$.[12]

Corollary 1 (Gaussian noise)**.** *Suppose the noise is Gaussian, and fix a period and a target effort as in Theorem 2. If $\underline{\theta} \geq 0$, the cheapest contract implementing the target has exactly one threshold and pays $\bar{w}$ above it. If $\underline{\theta} < 0$, it has exactly two, paying $\bar{w}$ on $(-\infty, z_t^-] \cup [z_t^+, +\infty)$ with both tails nonempty.*

This follows directly from the strong log-concavity of the Gaussian density, for which $-(\log g_\varepsilon)'' = 1/\sigma^2$. Since Section 4 maintains Gaussian noise, this is the form the contract takes throughout the analysis of the belief-manipulation channel.

4 The belief-manipulation channel

In this section, I compare the equilibria of the game described in Section 2 to those of a manipulation-free benchmark. All proofs for this section are in Appendix B. In particular, I study how much effort is exerted in the first period as the common prior becomes more optimistic: since effort multiplies the

[12] Strong log-concavity makes the likelihood score unbounded in both tails when $\underline{\theta} < 0$: $(\log g_\varepsilon)'(u) \leq -m(u - m_\varepsilon)$ to the right of the mode m_ε, and symmetrically to the left. Under log-concavity alone the score can have finite tail limits, and a cutoff above the smaller of them leaves only one tail paying, the one with the higher limit. That the two tails are the only paid regions uses the order-three total positivity of Assumption 1, as the footnote there explains.

underlying parameter in the signal, the informativeness of the signal is increasing in effort.[13] I say that there is *over-experimentation* when first-period effort in every equilibrium exceeds the benchmark's, and *under-experimentation* when it falls short of the benchmark's in every equilibrium.

The comparison is driven by the manipulation term T, the part of the agent's incentive that comes from the continuation, evaluated at the benchmark's equilibrium effort. Its sign indicates whether the agent would prefer to exert more or less effort at the benchmark contract, and hence whether implementing effort is cheaper or costlier for the principal under dynamic incentives. The agent's concern about retention gives him a reason to exert more effort, whereas his concern about future rents can push him in either direction, depending on how he expects the principal to adjust the contract after good news.

Section 4.1 introduces the specification of the general model under which I run the analysis. Section 4.2 presents the manipulation-free benchmark. Section 4.3 decomposes the agent's continuation incentive into these two forces, and Section 4.4 states the comparative static.

4.1 The specification

The comparison is made under a family of strongly log-concave prior densities, indexed by a single parameter. I first fix a base density whose log-concavity parameter ς captures how concentrated it is around its mean. I then tilt the base exponentially, with tilt parameter γ, which measures prior optimism. Varying the tilt shows whether the agent's dynamic incentive leads to more or less learning, depending on how optimistic the common prior is at the start of the relationship. The priors in the family are completely ordered by their likelihood ratios.[14]

Definition 1 (MLR-ordered, strongly log-concave family on a fixed support)**.** *Fix a base density g_0 on Θ that is strongly log-concave, $-(\log g_0)'' \geq \varsigma > 0$. For $\gamma \in \mathbb{R}$, define the exponentially tilted prior*

$$g_\theta(\theta \mid \gamma) = \frac{e^{\gamma\theta} g_0(\theta)}{\int_\Theta e^{\gamma\theta'} g_0(\theta')\, d\theta'}.$$

For $\gamma_2 > \gamma_1$ the ratio $g_\theta(\theta \mid \gamma_2)/g_\theta(\theta \mid \gamma_1)$ is proportional to $e^{(\gamma_2-\gamma_1)\theta}$ and hence strictly increasing in θ, so the family is strictly ordered by the likelihood ratio. Intuitively, as γ moves from $-\infty$ to $+\infty$, the prior is initially concentrated near $\underline{\theta}$, then spreads out, and eventually concentrates near $\bar{\theta}$.[15] The priors drawn in Figure 2 below are members of this family.

Beyond restricting priors to this family, I specialize the noise and the cost: the noise is Gaussian, indexed by its standard deviation σ, and the cost is quadratic, indexed by its curvature α.

Assumption 3 (Prior family, Gaussian noise, quadratic cost)**.** *The following are maintained for the remainder of the paper and are not restated in the results below.*

[13] At effort $e > 0$ the statistic $y_1/e = \theta + \varepsilon_1/e$ is sufficient for θ, a Gaussian location experiment with noise variance σ^2/e^2. For $e' > e$ the experiment at e is obtained from the one at e' by adding Gaussian noise of variance $\sigma^2/e^2 - \sigma^2/e'^2$, drawn independently of θ, so the experiment at e' is more informative in the Blackwell order.

[14] This is the order in which one belief is more favorable than another in the sense of Milgrom (1981). It implies first-order stochastic dominance but is not implied by it.

[15] I adopt an exponential-tilt family rather than a location-shift family because the complementarity parameter space is compact: a location shift would move the support, whereas the tilt preserves it.

(i) Prior family: *the prior g_θ belongs to the family of Definition 1, with tilt parameter $\gamma \in \mathbb{R}$.*

(ii) Gaussian noise: $\varepsilon_t \sim N(0, \sigma^2)$.

(iii) Quadratic cost: $c(e) = \alpha e^2/2$.

The specification increases tractability and makes the mechanism of the general model visible. In general, posteriors after two different signal realizations need not be comparable in how they affect the principal's continuation value or the agent's rent, since a change in the signal changes the shape of the likelihood in addition to its location, and the continuation incentive is an integral that in general can be neither separated nor signed. With the conjectured effort fixed and Gaussian noise, each posterior is instead indexed by a single parameter, which orders posteriors by likelihood ratio and is affine in the prior parameter and the signal.[16]

Lemma 2 (The posterior depends on a single index)**.** *Fix the principal's conjecture $\hat{e}_1$, and define the modified base*

$$\tilde{g}_0(\theta) := g_0(\theta)\, e^{-\hat{e}_1^2\theta^2/2\sigma^2}$$

and the index

$$b := \gamma + \frac{\hat{e}_1}{\sigma^2}\, y_1.$$

The posterior at prior γ after the first-period signal y_1 is the tilt of $\tilde{g}_0$ at b:

$$p(\theta \mid y_1; \gamma) \;\propto\; e^{b\,\theta}\, \tilde{g}_0(\theta). \tag{7}$$

At each conjecture, each posterior generated by every prior in the family is an exponential tilt of one base density. Posteriors are thus also indexed by a scalar and themselves completely ordered according to their likelihood ratios.

The results below hold when the noise is large enough given the curvature, and the prior is concentrated enough given both. Noise that is large enough given the curvature is large in two senses: relative to the other primitives, which are held fixed throughout, and relative to the curvature, $\alpha\sigma^2$, as Section 4.4 makes precise. Large noise means that a single signal is not too informative, so learning is gradual: the agent's dynamic incentive is small relative to today's wage incentive. Curvature relative to the noise makes the agent's problem concave in effort, since the second derivative of his expected pay in effort is of order $1/\sigma^2$, which the cost curvature dominates once $\alpha\sigma^2$ is large enough, as in Proposition 1. Under concentration, the principal's continuation value and the agent's rent, and how they move with news, track those of the problem at a known complementarity equal to the posterior mean. Noise and concentration together also make retention a cutoff in the signal: at a known complementarity, large noise makes good news raise tomorrow's expected value faster than its wage bill, and concentration makes the same hold at every posterior, so the principal's continuation value increases with good news. Together, these conditions make the principal's and the agent's problems well behaved and tractable enough to compare the equilibria with the benchmark.

[16] Among additive noise structures, the Gaussian family is the only one whose posteriors are exponential tilts in θ: a single index requires $\partial^2 \log g_\varepsilon(y_1 - \hat{e}_1\theta)/\partial y_1 \partial\theta$ to factor, which forces the log-curvature to be of the form $-Ce^{ku}$, and only $k = 0$, the Gaussian, makes the tilt linear in θ. With $k \neq 0$, a density of Gumbel type, the posteriors form a one-parameter family in $e^{-k\hat{e}_1\theta}$ instead.

They do not remove either force: Corollary 2 shows that both the retention threat and the ratchet are present in this regime.

Since a higher signal raises the posterior index b, and so makes the posterior more favorable, the retention decision takes a simple form: the principal keeps the agent exactly when the first-period signal is high enough.

Lemma 3 (Retention is a threshold rule)**.** *Fix $\alpha > 0$. There exists $\sigma^R(\alpha) \in (0, +\infty)$ such that for all $\sigma \geq \sigma^R(\alpha)$ there exists $\varsigma^R(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma^R(\alpha, \sigma)$ and all conjectures $\hat{e}_1 > 0$, every optimal retention rule is a cutoff rule in the first-period signal:*

$$\rho^*(y_1, \hat{e}_1) \begin{cases} = 1 & \text{if } y_1 > \underline{y}(\hat{e}_1), \\ \in \{0, 1\} & \text{if } y_1 = \underline{y}(\hat{e}_1), \\ = 0 & \text{if } y_1 < \underline{y}(\hat{e}_1), \end{cases}$$

where $\underline{y}(\hat{e}_1) := \inf\{y_1 : \Pi_2^*(y_1, \hat{e}_1) \geq \pi_0\}$ and $\Pi_2^*(\cdot, \hat{e}_1)$ is strictly increasing on a neighborhood of $\underline{y}(\hat{e}_1)$.

The level $\sigma^R(\alpha)$ is the larger of a level σ^{sc} set by the other primitives and the noise level at which the curvature floor $\alpha^{\text{FOA}}(\sigma)$ of Proposition 1 falls below α. The principal is indifferent only at the single signal value $\underline{y}(\hat{e}_1)$, an event of probability zero, so how she resolves that tie has no effect on first-period incentives. For definiteness I write $\rho^*(y_1, \hat{e}_1) = \mathbf{1}\{y_1 \geq \underline{y}(\hat{e}_1)\}$.

To see why retention is a cutoff, take the effort the principal implements in the second period at some signal y_1, and consider a slightly higher signal, which makes her belief more favorable in the likelihood ratio order. She can keep implementing the same effort as under the lower signal by moving the bar so that the agent's incentive is unchanged at the new belief. This contract is feasible, and her optimal contract can only do better. At that fixed effort, the better belief raises tomorrow's expected value by the effort times the increase in the posterior mean, while the expected wage bill changes by the cost of maintaining incentives. Large noise keeps that cost small relative to the gain, and concentration makes the same comparison hold at the actual posterior, so her value rises with the belief. At pessimistic beliefs it is below her outside option, so when replacement is credible, it crosses the outside option once.

Outside this regime, this monotonicity can fail. Good news raises the expected value that a given effort produces, but it also changes the terms on which effort can be bought, which can work against the principal. The second tail of the contract makes this concrete. When effort can backfire, the principal also pays for very low signals, which are evidence that effort was exerted in a state where it is counterproductive. Good news makes such states less likely, so these payments become a weaker instrument, and a contract that relied on them would have to shift incentives to the upper tail, possibly at a higher cost. Under a concentrated prior the lower tail carries almost no incentive, so this loss is negligible. With a dispersed prior that puts substantial weight on counterproductive complementarities, the loss would have to be weighed against the rise in tomorrow's expected value from good news, and it could outweigh it.

4.2 The manipulation-free benchmark

The benchmark is identical to the model of Section 2, except that the agent does not account for the effect of his first-period effort on his continuation value. The principal remains forward-looking: she values the first-period signal and what it reveals for the second period exactly as in the model. The benchmark thus isolates the agent's dynamic incentive to manipulate the principal's belief, while preserving the three frictions of the model: moral hazard, limited liability, and limited commitment.[17] Since the second period is the same in the model and in the benchmark, Lemma 3 applies to both.

In the first period, given the contract $w_1(\cdot)$, the agent's problem is[18]

$$\max_{e_1 \geq 0} \int w_1(y_1) f_1(y_1|e_1)\, dy_1 - c(e_1).$$

As in Theorem 1, existence relies on sufficient convexity of the agent's effort cost. Uniqueness is also easier to obtain here: without the dynamic incentive, the agent's best response does not depend on the principal's conjecture, so each contract pins down effort, and with enough noise and concentration the principal's problem in the target is strictly concave.[19]

Proposition 2 (Benchmark equilibrium)**.** *Fix $\alpha > 0$. There exists $\sigma^B(\alpha) \in (0, +\infty)$ such that for all $\sigma \geq \sigma^B(\alpha)$ a pure-strategy perfect Bayesian equilibrium of the benchmark game exists, and there exists $\varsigma^B(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma^B(\alpha, \sigma)$ it is unique up to the principal's hiring and retention choices at indifference.*

Existence needs only a curvature floor $\alpha^B(\sigma)$ of order σ^{-2}, and uniqueness a noise level of its own and enough concentration, so $\sigma^B(\alpha)$ is the larger of the two noise levels. The benchmark's conditions are implied by the model's: the benchmark agent faces the same wage incentive without the continuation, so the model's first-order approach implies the benchmark's, $\alpha^B(\sigma) \leq \alpha^{\mathrm{FOA}}(\sigma)$. For uniqueness, the principal's problem in the target she implements must be strictly concave. At a known complementarity there is nothing to learn, and her problem is concave because the wage bill is convex in the implemented effort. A concentrated prior keeps small what first-period effort reveals, and with it how her continuation value depends on that effort, so her problem at the actual prior is concave as well.

4.3 The two forces

The comparison with the benchmark is driven by the continuation incentive T of (6), the part of the agent's marginal incentive that comes from the second period. To see what drives it, write $u_2^*(\theta', y_1; \hat{e}_1)$ for the

[17] Other natural benchmarks include the following. Under full commitment, the principal designs the agent's dynamic incentive ex ante and does not reoptimize, so there is no interim belief to manipulate. Observable effort removes the agent's ability to manipulate beliefs, as well as the rents that generate the retention incentive. A planner removes all frictions at once.

[18] This formulation has multiple interpretations. The agent could simply be myopic. He could be forward-looking but naive, in that he does not internalize the effect of his effort today on his future contract terms and on his retention. A third interpretation, which I do not develop here, is that he is forward-looking and sophisticated but one of many similar agents who share the same complementarity, so that he is informationally small and his effort does not move the principal's belief, whereas the single agent of the model is informationally large.

[19] The principal's problem is simpler here, since removing the agent's dynamic incentive eliminates the fixed-point problem.

second-period rent of an agent whose complementarity is θ', after the first-period signal y_1, under the contract and the retention rule the principal sets at conjecture $\hat{e}_1$. Substituting the cutoff rule of Lemma 3 into Lemma 1 and integrating by parts in y_1 splits T into two terms:[20]

$$T(e_1, \hat{e}_1) = \int_\Theta \theta' \Big[u_2^*(\theta', \underline{y}(\hat{e}_1); \hat{e}_1)\, g_\varepsilon(\underline{y}(\hat{e}_1) - e_1\theta') + \int_{\underline{y}(\hat{e}_1)}^{+\infty} \frac{\partial u_2^*(\theta', y_1; \hat{e}_1)}{\partial y_1}\, g_\varepsilon(y_1 - e_1\theta')\, dy_1 \Big] g_\theta(\theta')\, d\theta'. \tag{8}$$

The first term is the *retention force*, an extensive effect. In state θ', a unit of effort moves the signal by θ', and so pushes probability mass across the retention cutoff, where the agent gains his whole rent. It is the value of being pushed just into the retention region. In the regime studied, where the principal's beliefs put most weight on complementarities at which effort raises the signal, the retention force is positive: the threat of replacement gives the agent a reason to work harder. Its size depends on how likely the signal is to land near the cutoff. It is large at pessimistic priors, where the cutoff sits in the bulk of the signal's distribution, and it dies out like a Gaussian tail as the prior becomes optimistic and the cutoff recedes into the lower tail.

The second term is the *wage force*, an intensive effect: given retention, how the agent's rent moves when a higher signal makes the principal more optimistic and she adjusts the second-period contract. By Lemma 2, the signal moves the posterior only through the index b, so $\partial u_2^*/\partial y_1 = (\hat{e}_1/\sigma^2)\, \partial u_2^*/\partial b$: the wage force prices good news in terms of tomorrow's rent. Its sign is ambiguous, because good news raises the value of the relationship but also changes the bar the agent must clear to earn the bonus. Under concentration, the rent moves as it would at a known complementarity equal to the posterior mean, and there the bar is U-shaped in the complementarity. At pessimistic beliefs, good news lowers the bar and raises the agent's rent. At optimistic beliefs, it raises the bar and lowers his rent, so good news today costs him rent tomorrow and gives him a reason to reduce effort. This is the ratchet, and it arises only when effort is cheap enough, since with costly effort the principal never brings the bar close enough to expected value for good news to raise it. Unlike the retention force, the wage force fades slowly, in proportion to the posterior variance, which shrinks as the prior becomes more optimistic and concentrates near the top of the support.

The two forces also respond differently to the conjecture, which the proof of Theorem 3 uses: the conjecture enters the retention force through the cutoff $\underline{y}(\hat{e}_1)$, while the wage force barely moves with it. Which force dominates, and where, is the subject of Section 4.4: the retention force dominates at pessimistic priors and, when effort is cheap enough, the ratchet dominates at optimistic ones, with a single switch in between.

4.4 The comparative static in the prior

The comparative static follows the posterior along the family through the index of Lemma 2. Fixing the conjecture, the retention rule is a threshold in the index: the agent is retained if and only if $b \geq \bar{b}$, with $\bar{b}$ fixed by π_0, and the second-period contract at a retained history depends on the history only through b.

[20] By Lemma 1 the agent's own posterior contributes nothing to the derivative, so effort enters his continuation value $\int_\Theta \int_{\underline{y}}^{\infty} u_2^*\, g_\varepsilon(y_1 - e_1\theta')\, dy_1\, g_\theta(\theta')\, d\theta'$ only through the location of the signal's density, and $T = -\int_\Theta \theta' \int_{\underline{y}}^{\infty} u_2^*\, g_\varepsilon'(y_1 - e_1\theta')\, dy_1\, g_\theta(\theta')\, d\theta'$. Integrating by parts on $[\underline{y}, \infty)$, the boundary term at infinity vanishes because the rent is at most $\bar{w}$ and the density vanishes there.

Retention regimes. Let $\pi^- := \Pi_2(\delta_{\underline{\theta}})$ and $\pi^+ := \Pi_2(\delta_{\bar{\theta}})$ denote the principal's second-period profit when the complementarity is known with certainty to be the lowest and the highest possible, written $\pi^{\pm}(\alpha, \sigma)$ when the dependence on the curvature and the noise matters. Where the outside option sits relative to $[\pi^-, \pi^+]$ determines which of three retention regimes obtains.

(P1) $\pi_0 \leq \pi^-$ *(always retain)*: replacement is never worthwhile at any posterior, and there is no retention cutoff.

(P2) $\pi^- < \pi_0 < \pi^+$ *(credible replacement)*: replacement is worthwhile at poor posteriors and not at good ones, so the retention decision depends on what the principal has learned.

(P3) $\pi_0 \geq \pi^+$ *(never retain)*: the outside option exceeds the retention value at every posterior and the relationship does not form.

Write γ_L for the prior at which the relationship forms: the smallest γ at which the principal's value from initiating reaches $2\pi_0$. The theorem is stated for (P2), where both forces are present. (P1) is discussed afterward, and in (P3) the relationship does not form, so there is no equilibrium effort to compare. Since π^- and π^+ are second-period profits at the extreme complementarities, they depend on the noise and the curvature, and only through the product $\alpha\sigma$: π^+ is large when $\alpha\sigma$ is small, since cheap effort and a precise signal make the best complementarity valuable, and falls to zero as $\alpha\sigma$ grows, and π^- behaves the same way at $\underline{\theta}$ when $\underline{\theta} > 0$ and is zero otherwise. For a given outside option and curvature, (P2) therefore holds on a range of noise levels, bounded above by the level at which even the best complementarity is not worth retaining. The theorem's conditions are to be read on that range.

The benchmark effort e^B is the effort the benchmark principal implements if she hires, which is defined at every prior. The benchmark relationship forms only at more optimistic priors, $\gamma_L^B > \gamma_L$, so on (γ_L, γ_L^B) the agent is hired only in the model, and there is trivially more effort than in the benchmark.

For a fixed curvature of the cost of effort, with enough noise and a concentrated enough prior, the comparison with the benchmark either switches exactly once as the prior improves or never does. The theorem distinguishes the two cases by a curvature $\bar{\alpha}(\sigma)$, defined at the most optimistic belief: when good news arrives there, the principal rewrites the second-period contract, and the rewrite lowers the rent of any given agent when the curvature is below $\bar{\alpha}(\sigma)$ and raises it when the curvature is above. So below $\bar{\alpha}(\sigma)$ the ratchet appears at optimistic beliefs, and above it good news benefits the agent.

The levels $\sigma^*(\alpha)$ and $\varsigma^*(\alpha, \sigma)$ of the theorem below include those of Propositions 1 and 2 and Lemma 3, as well as those under which the offered contract pins down effort, established in the proof of the theorem. The noise level $\sigma^*(\alpha)$ collects the levels of the results above. Each requires the noise to be large relative to the curvature, stated either through the floor $\alpha\sigma^2 \geq \underline{K}$, which keeps the agent's and the principal's problems well behaved, or as a bound σ_0 uniform in the curvature, under which the benchmark equilibrium is unique and the offered contract pins down the agent's effort. The level $\sigma^*(\alpha)$ is the larger of the two. As the curvature falls, only the first becomes more demanding. In every equilibrium the offered contract then determines the agent's effort, in the model as in the benchmark, and the comparison between their equilibrium efforts is well defined.

Theorem 3 (Over- and under-experimentation)**.** *Fix $\alpha > 0$. There exists $\sigma^*(\alpha) \in (0, +\infty)$ such that for all $\sigma \geq \sigma^*(\alpha)$ with $\pi^-(\alpha, \sigma) < \pi_0 < \pi^+(\alpha, \sigma)$ there exists $\varsigma^*(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma^*(\alpha, \sigma)$, the conditions of Propositions 1 and 2 and Lemma 3 hold at every prior at which the relationship forms, and:*

(i) *if $\alpha < \bar{\alpha}(\sigma)$, there is a unique γ^* such that every equilibrium over-experiments for all $\gamma \in (\gamma_L, \gamma^*)$ and under-experiments for all $\gamma \in (\gamma^*, +\infty)$;*

(ii) *if $\alpha \geq \bar{\alpha}(\sigma)$, every equilibrium over-experiments for all $\gamma \in (\gamma_L, +\infty)$.*

The theorem holds for every equilibrium at a prior, even when there are multiple. At a given prior the model's principal and the benchmark's face the same problem except for one thing: when the manipulation motive is positive the agent supplies part of the incentive himself, so she buys less with the wage, and when it is negative she buys more. So her objective is the benchmark's plus a term that is increasing in the target where the manipulation term is positive and decreasing where it is negative. Adding an increasing function to an objective can only move its maximizers up, so wherever the manipulation term has one sign across her targets, every equilibrium of hers sits on that side of the benchmark's target, whether or not her problem has a single optimum.

The switch happens at a single prior because of the shapes of the two forces of Section 4.3. The retention force is large at pessimistic priors, where the expected signal is close to the cutoff, and dies out like a Gaussian tail as the prior improves. The wage force becomes a ratchet only once most retained signals lie past the point at which good news starts to raise the bar, and it is then of the order of the posterior variance and fades only at a polynomial rate. So the two cross only once. Where the retention force is large, the principal may have multiple optimal contracts, but there the manipulation term is positive at every target and every equilibrium over-experiments, so uniqueness is not needed. Near the switch the retention force is small, her problem is strictly concave, and the equilibrium is unique and moves continuously with the prior, which is what pins down a single crossing.

When the prior is sufficiently pessimistic, so that the relationship just forms, the retention force dominates, leading to *more* experimentation than in the benchmark: the agent will likely not be retained in the second period, so he places more weight on getting something at all than on how his rent will change conditional on retention. The relationship forms at such priors for two reasons. The expected signal lies below the retention cutoff, so the principal hires an agent she expects to replace, but she can do better on good news, which is the option value of learning. And his incentive to work harder there makes the first period worth more to her than it would be in the benchmark. The over-experimentation there is first order, the atom at the cutoff, whereas the under-experimentation at optimistic priors is of the order of the posterior variance. When the cost curvature is not too high and the prior sufficiently optimistic, good news shifts the signal's density so that the bonus is paid with high probability, and maintaining incentives then requires *raising* the bar. In that case the two forces pull in opposite directions. As the prior becomes increasingly optimistic, concentrating on the upper endpoint of the support, the agent is replaced with vanishingly small probability, and the ratchet effect leads to *less* experimentation than in the benchmark. When the cost curvature is very high, incentives are costlier, and the principal optimally *loosens* the contract with good news at every prior in the family. There, the wage force constitutes a *reverse ratchet effect*, and both forces push in the same direction, leading to *more* experimentation at *every* prior. The result is depicted in Figure 2.

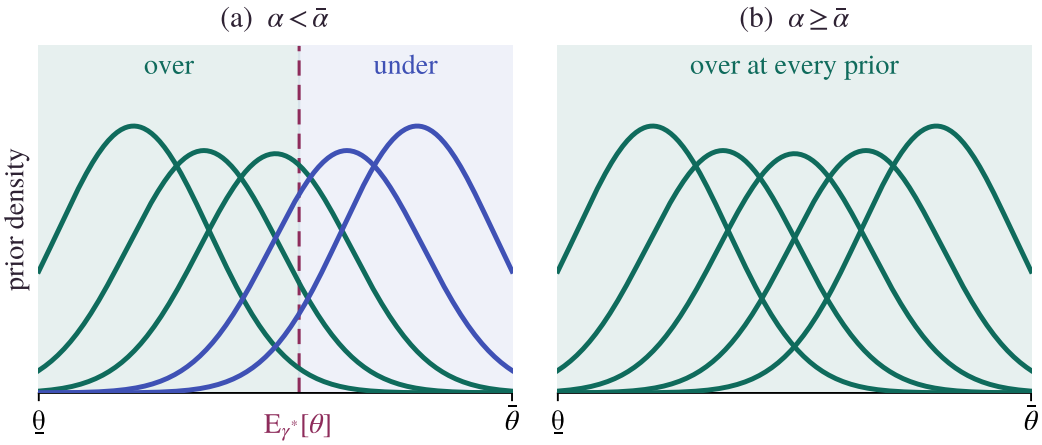


Figure 2: The two cases of Theorem 3. Each curve is a prior from the tilted family, ordered from pessimistic to optimistic. A prior drawn in teal induces over-experimentation and one drawn in indigo induces under-experimentation. In (a), the two-sided case $\alpha < \bar{\alpha}(\sigma)$, the sign changes once, at the threshold γ^* whose mean is marked. In (b), effort is costly, $\alpha \geq \bar{\alpha}(\sigma)$, and the agent over-experiments at every prior.

In regime (P1) the retention force is absent and the manipulation term is the wage force alone, so whether there is over- or under-experimentation is driven by the shape of the rent. The single peak of that rent, established in the proof of Theorem 3, then gives at most one crossing, from over- to under-experimentation.

Corollary 2 (Case (i) is nonvacuous)**.** *Suppose $\underline{\theta} \leq 0$. There exist $\sigma_{\min} \in (0, +\infty)$ and a curvature floor $\underline{\alpha}(\sigma)$ such that for all $\sigma \geq \sigma_{\min}$, $\underline{\alpha}(\sigma) < \bar{\alpha}(\sigma)$, and every $\alpha \in (\underline{\alpha}(\sigma), \bar{\alpha}(\sigma))$ satisfies $\sigma \geq \sigma^*(\alpha)$ and $\pi^-(\alpha, \sigma) < \pi_0 < \pi^+(\alpha, \sigma)$, so that Theorem 3 applies to it in its case (i).*

With $\underline{\theta} \leq 0$ the worst complementarity is worth nothing, $\pi^- = 0$, so the first inequality in (P2) holds at every noise level. When even the worst complementarity is productive, the same holds on a bounded range of noise levels if $\underline{\theta}$ is small enough relative to $\bar{\theta}$. At higher noise, cheap effort makes the worst complementarity worth keeping, and replacement stops being credible.

The theorem's two conditions bound the curvature from opposite sides. The first-order approach, existence, the benchmark, unique effort and the hiring threshold each require the curvature not to be too low relative to the noise, $\alpha \geq \underline{\alpha}(\sigma)$, where $\underline{\alpha}(\sigma) = \underline{K}\sigma^{-2}$ for a constant $\underline{K}$ depending only on $\bar{w}$, Θ and π_0: when effort is cheap relative to how noisy the signal is, the agent's effort is highly sensitive to the contract, the first-order approach fails, and neither the agent's nor the principal's problem is well behaved. With the curvature fixed, these are the requirement $\sigma \geq \sigma^*(\alpha)$. Case (i), by contrast, needs the curvature low enough for the ratchet to appear, $\alpha < \bar{\alpha}(\sigma)$, and $\bar{\alpha}(\sigma)$ is of order $\sigma^{-3/2}$: when effort is expensive, the wage buys little of it, good news never lowers the agent's rent, and there is no ratchet to oppose retention. The two are compatible because they scale differently in the noise: the floor comes from comparing curvatures and falls as σ^{-2}, the ceiling from the belief at which the bonus bar starts to rise with good news, and falls as $\sigma^{-3/2}$. The same region can be read with the curvature fixed instead: fix a curvature that is not too high. These results need $\sigma \geq (\underline{K}/\alpha)^{1/2}$, and the ratchet needs $\sigma < (\overline{K}/\alpha)^{2/3}$, beyond which the signal is so noisy that

the bonus is rarely paid at every belief, good news always lowers the bar, and $\bar{\alpha}(\sigma)$ falls below α. For small α the first level lies well below the second, so the window of noise levels is nonempty and widens as the curvature falls.

With no noise, the first-period signal reveals the complementarity exactly and the second-period contract is a forcing contract, paying the cap only if the signal meets its target, which leaves the agent no rent and extracts the whole surplus. That is the full rent extraction under which a pure-strategy equilibrium fails to exist, as in Bhaskar and Roketskiy (2023). With very noisy signals, a bounded wage buys almost no effort. The comparison here needs a signal that is informative about the complementarity without revealing it: case (i) is the regime of sufficiently noisy signals and intermediate effort costs, or, with the curvature fixed, of moderately noisy signals.

5 Application: technology adoption

One application is to the adoption of a new technology. Suppose that before the game starts, the principal must decide whether to adopt the technology. The investment is irreversible and costs the principal $k > 0$. If the principal adopts, then the game is as described in Section 2. If she does not adopt, the game is identical except that the principal's outside option is zero and the complementarity is known and equal to one:

$$y_t = e_t + \varepsilon_t.$$

So the investment gives the principal an outside option, running the technology on her own for π_0 per period, and the chance to pair it with the agent, whose complementarity with it is unknown once it is installed. Without the technology there is nothing to learn, so the agent has no belief incentive and the model and the benchmark coincide.

Write $\Pi(\gamma)$ for the principal's equilibrium value in the adopted game at prior γ, $\Pi^B(\gamma)$ for her value in the manipulation-free benchmark, and Π^0 for her value in the game without the technology, which is the same in both since there is nothing to learn. She adopts if and only if $k \leq \Pi(\gamma) - \Pi^0$, and the benchmark principal if and only if $k \leq \Pi^B(\gamma) - \Pi^0$. So there is weakly more adoption than in the benchmark, meaning adoption at every cost at which the benchmark principal adopts, exactly when $\Pi(\gamma) \geq \Pi^B(\gamma)$, and weakly less, meaning adoption only at costs at which she also adopts, exactly when $\Pi(\gamma) \leq \Pi^B(\gamma)$: what matters is whether the technology is worth more or less to the principal in the model than in the benchmark.

Proposition 3 (Technology adoption)**.** *Fix $\alpha > 0$. With $\sigma^*(\alpha)$ and $\varsigma^*(\alpha, \sigma)$ as in Theorem 3, for all $\sigma \geq \sigma^*(\alpha)$ with $\pi^-(\alpha, \sigma) < \pi_0 < \pi^+(\alpha, \sigma)$ and all $\varsigma \geq \varsigma^*(\alpha, \sigma)$:*

(i) *if $\alpha < \bar{\alpha}(\sigma)$, there is a unique γ^{**} such that there is weakly more adoption than in the benchmark at every prior below γ^{**} and weakly less at every prior above it;*

(ii) *if $\alpha \geq \bar{\alpha}(\sigma)$, there is weakly more adoption than in the benchmark at every prior.*

The threshold γ^{**} is the prior at which the principal's value crosses her benchmark value. It need not coincide with γ^*: the comparison of efforts is signed by the continuation incentive corrected for how the

conjecture moves with the target, while the comparison of values is signed by the continuation incentive itself, and the proof bounds the distance between the two zeros. Below it, the continuation delivers part of the effort incentive for free, so the principal implements the effort she wants with a lower wage bill, and the technology is worth more to her than in the benchmark. Above it, the continuation works against her, every target costs more, and the technology is worth less. The agent's dynamic incentive thus raises the principal's willingness to invest in the technology when his job is at risk, through the retention force, and lowers it when his position is secure, through the ratchet.

6 Conclusion

This paper studies dynamic contracting when the complementarity between the agent and a task is commonly unknown, under moral hazard, limited commitment, and limited liability. There are three main findings. First, a pure-strategy equilibrium exists for sufficient curvature of the effort cost, which the limited-liability rent ensures by keeping the agent's objective smooth in his effort, and, under Gaussian noise and quadratic cost with the noise not too precise, the contract the principal offers pins down the agent's effort in every equilibrium. Second, optimal spot contracts are one- or two-tailed: the principal pays a bonus on a single tail of the signal when effort is always productive, and on two tails when effort can be counterproductive, with no payment in between. Third, the agent has a dynamic incentive to manipulate the principal's belief about the complementarity through his private action choice, and this pushes towards more learning about the underlying state at pessimistic priors and less learning once the prior becomes sufficiently optimistic. Whether the channel raises or lowers learning, and with it the value of a new technology to the principal, therefore depends on how secure the agent using it is.

A natural extension would let the relationship run over an infinite horizon, with the principal choosing when to end it and draw a new agent. Replacement is then a stopping problem, and because it is observed, continuation play can condition on the principal's past replacements, which may sustain behavior that the two-period game cannot. The questions are which contracts and replacement policies can be sustained, and how much is learned about each agent compared to the single-agent case.

Another direction would be to grant the principal additional commitment power. Even without commitment to a particular retention rule, replacement risk works in her favor: it is what pushes the agent to work harder at pessimistic priors. If she could fix this rule in advance, she could further exploit this channel, for instance by promising not to fire the agent after an early failure in order to encourage experimentation. Since a very poor signal is evidence of effort in a poor fit rather than of shirking, the optimal rule may even retain the agent after sufficiently poor as well as sufficiently strong signals, the termination analogue of the two-tailed wage contract. At the other extreme, with no option to replace at all, as under strict employment protection, only the ratchet remains. The comparison between full and limited commitment in this setting, where effort determines how much is learned, remains to be characterized.

References

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

Robert J. Aumann. Integrals of set-valued functions. *Journal of Mathematical Analysis and Applications*, 12(1):1–12, 1965.

Eduardo M. Azevedo and Ilan Wolff. Broad validity of the first-order approach in moral hazard. Working paper, Wharton; arXiv:2506.18873, 2025.

Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic Theory*, 26(2): 445–469, 2005.

Dirk Bergemann and Ulrich Hege. The financing of innovation: Learning and stopping. *RAND Journal of Economics*, 36(4):719–752, 2005.

V. Bhaskar and George J. Mailath. The curse of long horizons. *Journal of Mathematical Economics*, 82: 74–89, 2019.

V. Bhaskar and Nikita Roketskiy. The ratchet effect: A learning perspective. *Journal of Economic Theory*, 214:105762, 2023.

Christer Borell. Complements of Lyapunov's inequality. *Mathematische Annalen*, 205:323–331, 1973.

Herm Jan Brascamp and Elliott H. Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *Journal of Functional Analysis*, 22(4):366–389, 1976.

Chang Koo Chi and Trond E. Olsen. Optimal incentive contracts with bonus caps. *RAND Journal of Economics*, 56(1):55–73, 2025.

Gonzalo Cisternas. Two-sided learning and the ratchet principle. *Review of Economic Studies*, 85(1):307–351, 2018.

Peter DeMarzo and Yuliy Sannikov. Learning, termination, and payout policy in dynamic incentive contracts. *Review of Economic Studies*, 84(1):182–236, 2017.

Mathias Dewatripont, Ian Jewitt, and Jean Tirole. The economics of career concerns, part I: Comparing information structures. *Review of Economic Studies*, 66(1):183–198, 1999.

Mehmet Ekmekci, Leandro Gorno, Lucas Maestri, Jian Sun, and Dong Wei. Learning from manipulable signals. *American Economic Review*, 112(12):3995–4040, 2022.

Xavier Freixas, Roger Guesnerie, and Jean Tirole. Planning under incomplete information and the ratchet effect. *Review of Economic Studies*, 52(2):173–191, 1985.

Drew Fudenberg and Jean Tirole. A "signal-jamming" theory of predation. *RAND Journal of Economics*, 17(3):366–376, 1986.

Drew Fudenberg and Jean Tirole. Moral hazard and renegotiation in agency contracts. *Econometrica*, 58 (6):1279–1319, 1990.

Hong Gao and Haibo Xu. Learning, belief manipulation and optimal relationship termination. *Economics Letters*, 190:109108, 2020.

Dino Gerardi and Lucas Maestri. Dynamic contracting with limited commitment and the ratchet effect. *Theoretical Economics*, 15(2):583–623, 2020.

Faruk Gul. Unobservable investment and the hold-up problem. *Econometrica*, 69(2):343–376, 2001.

Marina Halac, Navin Kartik, and Qingmin Liu. Optimal contracts for experimentation. *Review of Economic Studies*, 83(3):1040–1091, 2016.

Oliver Hart and Bengt Holmström. The theory of contracts. In Truman F. Bewley, editor, *Advances in Economic Theory: Fifth World Congress*, pages 71–155. Cambridge University Press, Cambridge, 1987.

Bengt Holmström. Moral hazard and observability. *Bell Journal of Economics*, 10(1):74–91, 1979.

Bengt Holmström. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182, 1999.

Johannes Hörner and Larry Samuelson. Incentives for experimenting agents. *RAND Journal of Economics*, 44(4):632–663, 2013.

Robert D. Innes. Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, 52(1):45–67, 1990.

Ian Jewitt. Justifying the first-order approach to principal-agent problems. *Econometrica*, 56(5):1177–1190, 1988.

Samuel Karlin. *Total Positivity, Volume I*. Stanford University Press, Stanford, CA, 1968.

Byunghoon Kim. Signal jamming by shirking under replacement risk. Working paper, University of Pennsylvania, 2026.

Jean-Jacques Laffont and Jean Tirole. Comparative statics of the optimal dynamic incentive contract. *European Economic Review*, 31(4):901–926, 1987.

Jean-Jacques Laffont and Jean Tirole. The dynamics of incentive contracts. *Econometrica*, 56(5):1153–1175, 1988.

David G. Luenberger. *Optimization by Vector Space Methods*. Wiley, New York, 1969.

Gustavo Manso. Motivating innovation. *Journal of Finance*, 66(5):1823–1860, 2011.

Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(1):157–180, 1994.

Paul R. Milgrom. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12(2):380–391, 1981.

James A. Mirrlees. The theory of moral hazard and unobservable behaviour: Part I. *Review of Economic Studies*, 66(1):3–21, 1999.

Julien Prat and Boyan Jovanovic. Dynamic contracts when the agent's quality is unknown. *Theoretical Economics*, 9(3):865–914, 2014.

William P. Rogerson. The first-order approach to principal-agent problems. *Econometrica*, 53(6):1357–1367, 1985.

Maryam Saeedi and Ali Shourideh. Optimal rating design under moral hazard. Working paper, Carnegie Mellon University, 2026.

H. F. Weinberger. A characterization of the Pólya frequency functions of order 3. *Applicable Analysis*, 15 (1-4):53–69, 1983.

Martin L. Weitzman. The "ratchet principle" and performance incentives. *Bell Journal of Economics*, 11 (1):302–308, 1980.

A Proofs of the general results

This appendix proves the results of Sections 2 and 3, which hold under Assumptions 1 and 2 alone: Lemma 1, Proposition 1, Theorems 1 and 2, and Corollary 1. Appendix B proves the results of Section 4, which hold under the specification.

Proof of Lemma 1 (Own-effort regularity and the continuation incentive)

Efforts are nonnegative, so derivatives in effort at $e = 0$ are taken from the right.

Part (i): smoothness. Since Θ is compact and $g_\varepsilon \in C^2(\mathbb{R})$, the density f_2 and its first two e_1-derivatives are obtained by differentiating $g_\varepsilon(y_1 - e_1\theta)$ under the θ-integral in (1), the derivatives $-\theta g_\varepsilon'(y_1 - e_1\theta)$ and $\theta^2 g_\varepsilon''(y_1 - e_1\theta)$ being dominated on bounded sets of e_1 by integrable functions of θ.[21] Hence the second-period objective $\int w_2 f_2 \, dy_2 - c(e_2)$ is C^2 in (e_1, e_2). By the strict concavity assumed in the lemma, $e_2^*(w_2; y_1, e_1)$ is unique. Where it is interior, it is the unique solution of the stage-2 first-order condition

$$c'(e_2) = \int w_2(y_2; y_1)\, \partial_{e_2} f_2 \, dy_2, \tag{9}$$

[21] Explicitly, by $\bar{h}_k(\theta) := |\theta|^k \sup_{|u| \le \delta|\theta|} |g_\varepsilon^{(k)}(y_1 - e^0\theta - u)|\, g_\theta(\theta)$ for $k = 1, 2$ on $|e_1 - e^0| < \delta$. The same domination justifies every differentiation under a θ-integral in the appendix, and is not repeated.

whose e_2-derivative $\int w_2\, \partial^2_{e_2} f_2\, dy_2 - c''(e_2)$ is strictly negative, so by the implicit function theorem e_2^* is C^1 in e_1 there, with derivative bounded on bounded sets of e_1, the objective being C^2 and its e_2-derivative bounded away from zero on them. It is at the corner $e_2^* = 0$ exactly when the right side of (9) at $e_2 = 0$, which is $\mathbb{E}_{p(\cdot|y_1;e_1)}[\theta] \int w_2\, (-g'_\varepsilon(y_2))\, dy_2$, is nonpositive, and where that quantity is negative it is locally constant. So e_2^* is continuous in e_1, and C^1 except where it switches between the corner and the interior. The agent's maximized continuation $U_2(w_2; y_1, e_1)$ is the objective evaluated at $e_2^*(w_2; y_1, e_1)$. By the envelope theorem, which holds at the corner as well because the constraint $e_2 \geq 0$ does not depend on e_1, its derivative is the e_1-derivative of the objective evaluated at e_2^*, which is continuous. So $U_2 \in C^1(\mathbb{R}_+)$, and U_2 is C^2 wherever e_2^* is. The switch happens only where the agent's posterior mean changes sign, hence only when $\underline{\theta} < 0$, and since that mean is increasing in the signal it happens at a single signal value for each e_1. So U_2 is C^2 almost everywhere on $\mathbb{R}_+$, as the statement claims. When $\underline{\theta} \geq 0$ no switch occurs and $U_2 \in C^2(\mathbb{R}_+)$.

Part (ii): the incentive identity. Write the continuation state-conditionally, using $U_2 = \mathbb{E}_p[u_2]$, the posterior (1), and Fubini's theorem, which applies since u_2 is bounded by $\bar{w}$:

$$\mathbb{E}[\rho\, U_2] = \int_\Theta \Big[\int_{\mathbb{R}} \rho(y_1; \hat{e}_1)\, u_2(w_2; \theta, y_1, e_1)\, g_\varepsilon(y_1 - e_1\theta)\, dy_1 \Big]\, g_\theta(\theta)\, d\theta.$$

Differentiating in e_1 under the integral is justified by a dominating function of the same form as $\bar{h}_1$. Since the integral is now over $y_1 \in \mathbb{R}$ rather than the compact Θ, that function must be integrable in y_1, and it is, because $|g'_\varepsilon|$ is integrable and, by log-concavity, eventually monotone in each tail, so its supremum over a bounded window is attained at an endpoint. Effort e_1 enters twice, through $g_\varepsilon(y_1 - e_1\theta)$ and, inside u_2, through $e_2^*(w_2; y_1, e_1)$. The e_2^*-channel contributes

$$\int \frac{\partial e_2^*}{\partial e_1}\, \rho\, f_1(y_1)\, \mathbb{E}_{p(\cdot|y_1;e_1)}\Big[\frac{\partial u_2}{\partial e_2}\Big]\, dy_1 = 0,$$

by the envelope theorem: where e_2^* is interior, the inner posterior expectation is the agent's stage-2 first-order condition (9), where it is at the corner $\partial e_2^*/\partial e_1 = 0$, and the switches occur on a null set of signals. The remaining term, from differentiating $g_\varepsilon(y_1 - e_1\theta)$, is the displayed expression. □

Proof of Proposition 1 (Cost curvature validates the first-order approach)

The agent's first-period objective is twice continuously differentiable in e_1: it averages U_2 over the signal, U_2 is C^2 almost everywhere by Lemma 1, and the bounds of Claims 4 and 5 allow differentiating under the integral twice. Strict concavity is the usual condition that its second derivative be negative.

Two elementary bounds are used throughout.

Claim 1. *For any probability density q on Θ, $|\mathbb{E}_q\theta| \leq \theta_{\max}$ and $\mathrm{Var}_q\theta \leq \tau^2_{\max}$. In particular this holds for the prior and for every posterior, which share the support Θ because the noise has full support.*

Proof. A mean lies in the convex hull of the support. The variance bound is Popoviciu's inequality. □

Claim 2. *Let q be a probability density on Θ and let $f(y \mid e) = \int_\Theta g_\varepsilon(y - e\tilde{\theta})\, q(\tilde{\theta})\, d\tilde{\theta}$. For every $e \geq 0$,*

1. $\int_{\mathbb{R}} \big|\frac{\partial f}{\partial e}\big| dy \leq \nu_1 \theta_{\max}$,

2. $\int_{\mathbb{R}} \left|\frac{\partial^2 f}{\partial e^2}\right| dy \le \nu_2[\theta_{\max}^2 + \tau_{\max}^2]$.

Proof. Differentiating f with respect to e,

$$\frac{\partial f}{\partial e} = -\int_\Theta \tilde{\theta}\, g_\varepsilon'(y - e\tilde{\theta})\, q(\tilde{\theta})\, d\tilde{\theta},$$
$$\frac{\partial^2 f}{\partial e^2} = \int_\Theta \tilde{\theta}^2\, g_\varepsilon''(y - e\tilde{\theta})\, q(\tilde{\theta})\, d\tilde{\theta},$$

differentiating under the integral as in the proof of Lemma 1. Taking absolute values, integrating in y, and switching the order of integration by Tonelli's theorem, translation invariance gives $\int_{\mathbb{R}} |g_\varepsilon^{(k)}(y - e\tilde{\theta})|\, dy = \nu_k$ for every $\tilde{\theta}$, so

$$\int_{\mathbb{R}} \left|\frac{\partial f}{\partial e}\right| dy \le \nu_1 \int_\Theta |\tilde{\theta}|\, q(\tilde{\theta})\, d\tilde{\theta},$$
$$\int_{\mathbb{R}} \left|\frac{\partial^2 f}{\partial e^2}\right| dy \le \nu_2 \int_\Theta \tilde{\theta}^2\, q(\tilde{\theta})\, d\tilde{\theta}.$$

Finally $\mathbb{E}_q[\tilde{\theta}^2] = (\mathbb{E}_q\tilde{\theta})^2 + \mathrm{Var}_q\tilde{\theta} \le \theta_{\max}^2 + \tau_{\max}^2$ and $\mathbb{E}_q|\tilde{\theta}| \le \theta_{\max}$ by Claim 1. □

Claim 3. *For every $e_1, e_2 \ge 0$, $y_1 \in \mathbb{R}$, and all admissible w_2,*

$$\int w_2\, \frac{\partial^2 f_2}{\partial e_2^2}\, dy_2 \;\le\; \bar{w}\, \nu_2[\theta_{\max}^2 + \tau_{\max}^2] =: \bar{M}_2' < \infty.$$

Proof. Since $0 \le w_2 \le \bar{w}$, the left side is at most $\bar{w} \int |\partial_{e_2}^2 f_2|\, dy_2$, and Claim 2 with $q = p(\cdot \mid y_1, e_1)$ bounds the integral by $\nu_2[\theta_{\max}^2 + \tau_{\max}^2]$. □

Claim 4. *Suppose $c''(e_2) - \int w_2 \partial_{e_2}^2 f_2\, dy_2 \ge \underline{D} > 0$, uniformly in e_1, e_2, w_2. Then:*

1. *for all y_1, $0 \le U_2(w_2; y_1, e_1) \le \bar{w}$;*
2. *for all e_1, $\int \left|\frac{dU_2}{de_1}\right| |\partial_{e_1} f_1|\, dy_1 \le \bar{w}\, J_1 < \infty$, where $J_1 := \sup_{e_1} \int (\int |\partial_{e_1} f_2| dy_2) |\partial_{e_1} f_1|\, dy_1$;*
3. *for all e_1, $\int \left|\frac{d^2 U_2}{de_1^2}\right| f_1\, dy_1 \le \bar{w}\, J_2 + \frac{\bar{w}^2\, \Lambda_{\mathrm{FOA}}}{\underline{D}} < \infty$, where $J_2 := \sup_{e_1} \int (\int |\partial_{e_1}^2 f_2| dy_2) f_1\, dy_1$ and $\Lambda_{\mathrm{FOA}} := \sup_{e_1} \int (\int |\partial_{e_1}\partial_{e_2} f_2|\, dy_2)^2 f_1\, dy_1$.*

Proof. The first statement is immediate from limited liability and $c(0) = 0$: retention yields $U_2 \ge 0$, and $U_2 \le \bar{w}$ since $w_2 \le \bar{w}$. At a bonus contract the tighter $U_2 \le \bar{w} P_A$ holds, with P_A the payment probability.

For the derivative bounds, the claim's hypothesis

$$c''(e_2) - \int w_2\, \partial_{e_2}^2 f_2\, dy_2 \;\ge\; \underline{D} > 0$$

is the stage-2 problem's strict second-order condition, which Part (i) below establishes with $\underline{D} = \alpha - \bar{M}_2'$. Under it the best response $e_2^*(w_2; y_1, e_1)$ is unique. If the marginal incentive at zero effort is nonpositive, $e_2^* = 0$. The signal at zero effort is then pure noise, f_2 does not depend on the posterior, and U_2 does not

depend on e_1, so the bounds below hold trivially. Otherwise e_2^* is interior, the envelope theorem applies, and by the implicit function theorem e_2^* is differentiable in e_1 with the denominator below bounded away from zero. By the envelope theorem,

$$\partial_{e_1} U_2 = \int w_2 \, \partial_{e_1} f_2 \, dy_2.$$

Differentiating again along $e_2^*(w_2; y_1, e_1)$,

$$\partial_{e_1}^2 U_2 = \int w_2 \, \partial_{e_1}^2 f_2 \, dy_2 + \Big(\int w_2 \, \partial_{e_1} \partial_{e_2} f_2 \, dy_2 \Big) \frac{\partial e_2^*}{\partial e_1},$$

where

$$\frac{\partial e_2^*}{\partial e_1} = \frac{\int w_2 \, \partial_{e_1} \partial_{e_2} f_2 \, dy_2}{c''(e_2^*) - \int w_2 \, \partial_{e_2}^2 f_2 \, dy_2}.$$

Bounding $w_2 \le \bar{w}$ and $|\partial e_2^* / \partial e_1| \le (\int |\partial_{e_1} \partial_{e_2} f_2| dy_2) \bar{w} / \underline{D}$ (the last by the stage-2 SOC), so that the second term of $\partial_{e_1}^2 U_2$ is at most $\bar{w}^2 (\int |\partial_{e_1} \partial_{e_2} f_2| dy_2)^2 / \underline{D}$, then integrating against $|\partial_{e_1} f_1|$ (resp. f_1) over y_1 yields the stated forms with $J_1, J_2, \Lambda_{\mathrm{FOA}}$.

It remains to show these are finite. Effort e_1 enters f_2 only through the posterior $p(\cdot \mid y_1; e_1)$, and differentiating the posterior gives

$$\partial_{e_1} p(\tilde{\theta} \mid y_1) = p(\tilde{\theta} \mid y_1) \, [\mathsf{s}(\tilde{\theta}) - \bar{\mathsf{s}}],$$

where

$$\mathsf{s}(\tilde{\theta}) := -\tilde{\theta} \, (\log g_\varepsilon)'(y_1 - e_1 \tilde{\theta})$$

is the score of the stage-1 likelihood in e_1 and $\bar{\mathsf{s}} := \mathbb{E}_p[\mathsf{s}]$ is its posterior mean, arising from differentiating the normalizing constant. Since $|\tilde{\theta}| \le \theta_{\max}$ on the compact support, each inner y_2-integral is bounded by a constant multiple of

$$\int_\Theta \big| (\log g_\varepsilon)'(y_1 - e_1 \tilde{\theta}) \big| \, p(\tilde{\theta} \mid y_1) \, d\tilde{\theta}$$

for the first derivative, and of its square together with $\int_\Theta |(\log g_\varepsilon)''(y_1 - e_1 \tilde{\theta})| \, p(\tilde{\theta} \mid y_1) \, d\tilde{\theta}$ for the second, since differentiating the score once more produces $\tilde{\theta}^2 (\log g_\varepsilon)''$. Integrating against f_1 (resp. $|\partial_{e_1} f_1|$) over y_1 therefore yields finite expected-score and expected-squared-score quantities, the square in Λ_{FOA} being handled by Jensen's inequality: the Fisher information $\int (\log g_\varepsilon)'^2 g_\varepsilon$ is finite by Assumption 1, and

$$\int |(\log g_\varepsilon)''| g_\varepsilon = \int (-(\log g_\varepsilon)'') g_\varepsilon = \int (\log g_\varepsilon)'^2 g_\varepsilon,$$

the first equality by log-concavity and the second an integration by parts. Its boundary term g_ε' vanishes at $\pm\infty$. To see this, write $g_\varepsilon = e^{-\psi}$ with ψ convex. Beyond the mode, ψ' is positive and increasing and $\int_x^\infty \psi' g_\varepsilon = g_\varepsilon(x)$, so $|g_\varepsilon'(x)| = \psi'(x) g_\varepsilon(x) \le \int_x^\infty (\psi')^2 g_\varepsilon$, a tail of the Fisher integral, and the same argument applies below the mode. Since $g_\varepsilon'' = ((\psi')^2 - \psi'') g_\varepsilon$, the equality also gives $\int |g_\varepsilon''| \le 2 \int (g_\varepsilon')^2 / g_\varepsilon$, the bound stated in the footnote to Assumption 1. The marginal product $\tilde{\theta}$ is bounded on the compact Θ. Hence $J_1, J_2, \Lambda_{\mathrm{FOA}} < \infty$. □

Claim 5. *For every $e_1 \geq 0$, every retention rule ρ with values in $[0,1]$, and every admissible first-period contract w_1, the right-hand side of (5) is at most*

$$\bar{M}_1 := 2\bar{w}\,\nu_2[\theta_{\max}^2 + \tau_{\max}^2] + 2\bar{w}\,J_1 + \bar{w}\,J_2 + \frac{\bar{w}^2\,\Lambda_{\mathrm{FOA}}}{\underline{D}} < \infty,$$

with $J_1, J_2, \Lambda_{\mathrm{FOA}}, \underline{D}$ the finite constants of Claim 4, none of which depends on w_1 or ρ.

Proof. Bound the three terms of (5).

- Term I: $w_1 + \rho U_2 \leq 2\bar{w}$ by limited liability and Claim 4, and $\int |\partial_{e_1}^2 f_1| dy_1 \leq \nu_2[\theta_{\max}^2 + \tau_{\max}^2]$ by Claim 2 with $q = g_\theta$.
- Term II: by Claim 4, with $\rho \leq 1$, $\int 2\rho \left| dU_2/de_1 \right| |\partial_{e_1} f_1|\, dy_1 \leq 2\bar{w}J_1$.
- Term III: by Claim 4, with $\rho \leq 1$, $\int \rho \left| d^2U_2/de_1^2 \right| f_1\, dy_1 \leq \bar{w}J_2 + \frac{\bar{w}^2\,\Lambda_{\mathrm{FOA}}}{\underline{D}}$.

The second-period objects entering $J_1, J_2, \Lambda_{\mathrm{FOA}}$ are those of arbitrary admissible second-period contracts, for which part (i) supplies the slack $\underline{D} \geq \alpha - \bar{M}_2'$. □

Part (i). By Claim 3 the right-hand side of (4) is at most $\bar{M}_2'$ for every admissible w_2, independent of e_2 and of the history, so $\alpha_2^{\mathrm{FOA}} := \bar{M}_2'$ works. At this curvature the stage-2 problem at every history has a unique agent best response.

Part (ii). By Claim 5 the right-hand side of (5) is at most $\bar{M}_1$, independent of e_1.[22] The threshold is not $\max\{\bar{M}_1, \bar{M}_2'\}$, however, because $\bar{M}_1$ is not free of the curvature: Term III contains $\bar{w}^2\Lambda_{\mathrm{FOA}}/\underline{D}$, and $\underline{D}$ is the stage-2 slack, at least $\alpha - \bar{M}_2'$ once (4) holds. Writing A for the part of $\bar{M}_1$ that does not involve $\underline{D}$, the stage-1 requirement $\alpha > \bar{M}_1$ reads

$$\alpha \;>\; A + \frac{\bar{w}^2\Lambda_{\mathrm{FOA}}}{\alpha - \bar{M}_2'}.$$

Since $\alpha > \bar{M}_2'$, this is equivalent to

$$(\alpha - A)(\alpha - \bar{M}_2') \;>\; \bar{w}^2\Lambda_{\mathrm{FOA}},$$

a quadratic inequality in α, and the larger root of the quadratic is

$$\alpha^{\mathrm{FOA}} := \frac{(A + \bar{M}_2') + \sqrt{(A - \bar{M}_2')^2 + 4\bar{w}^2\Lambda_{\mathrm{FOA}}}}{2},$$

finite and exceeding $\bar{M}_2' = \alpha_2^{\mathrm{FOA}}$, so that $\alpha_1^{\mathrm{FOA}} := \alpha^{\mathrm{FOA}}$ delivers part (ii) and, being larger, part (i) as well. The quadratic cost $c(e) = \alpha e^2/2$ has $c'' = \alpha$, so it qualifies if and only if $\alpha > \alpha^{\mathrm{FOA}}$. This proves Proposition 1. □

[22] The bounds in Claims 3 and 5 use only the wage cap. On the contracts the analysis produces, bonuses on at most two tails, integrating by parts on each tail bounds the wage terms in both periods by $2\bar{w}\theta_{\max}^2 \sup|g_\varepsilon'|$, which lowers both thresholds when the cutoffs lie in the tails of the noise.

Proof of Theorem 1 (Existence)

Step 1: contracts. A contract is a measurable $w : \mathbb{R} \to [0, \bar{w}]$, identified with every contract that agrees with it at almost every signal, since the signal has a density. Let $\mathcal{A}$ be the set of contracts, with the weak-* topology of $L^\infty(\mathbb{R})$, in which $w_n \to w$ when $\int w_n h \to \int wh$ for every integrable h. The closed ball of radius $\bar{w}$ in $L^\infty(\mathbb{R})$ is compact by the Banach–Alaoglu theorem. If $w_n \in \mathcal{A}$ and $w_n \to w$, then for every bounded measurable set B, $\int_B w = \lim_n \int_B w_n$ lies in $[0, \bar{w}\,|B|]$, so $0 \le w \le \bar{w}$ almost everywhere and $w \in \mathcal{A}$. So $\mathcal{A}$ is weak-* closed in the ball, hence compact, and it is metrizable because $L^1(\mathbb{R})$ is separable. At belief p and effort e, the signal has density $f_p(y \mid e) := \int_\Theta g_\varepsilon(y - \theta e)\, p(\theta)\, d\theta$, and the expected payment of contract w is

$$W_p(w; e) := \int_{\mathbb{R}} w(y)\, f_p(y \mid e)\, dy.$$

Differentiating in e, the marginal incentive it delivers is

$$M_p(w; e) := \int_{\mathbb{R}} w(y)\, \partial_e f_p(y \mid e)\, dy,$$

with $\partial_e f_p(y \mid e) = -\int_\Theta \theta\, g'_\varepsilon(y - \theta e)\, p(\theta)\, d\theta$.

Write $\theta_{\max} := \max\{|\underline{\theta}|, \bar{\theta}\}$ for the bound on the complementarity, and $\nu_1 := \int |g'_\varepsilon|$ and $\nu_2 := \int |g''_\varepsilon|$, which are finite by the footnote to Assumption 1. For shifts a and b, $g_\varepsilon(y-a) - g_\varepsilon(y-b)$ is the integral of $-g'_\varepsilon(y-x)$ over x between b and a, so integrating over y and exchanging the order of integration,

$$\int_{\mathbb{R}} \left| g_\varepsilon(y - a) - g_\varepsilon(y - b) \right| dy \;\le\; |a - b|\, \nu_1.$$

With $a = \theta e$ and $b = \theta e'$, averaging over p gives

$$\left\| f_p(\cdot \mid e) - f_p(\cdot \mid e') \right\|_1 \;\le\; \theta_{\max} \nu_1\, |e - e'|,$$

and the same argument applied to g'_ε gives

$$\left\| \partial_e f_p(\cdot \mid e) - \partial_e f_p(\cdot \mid e') \right\|_1 \;\le\; \theta_{\max}^2 \nu_2\, |e - e'|.$$

Only continuity in effort is needed for what follows, and the constant ν_1 also bounds every incentive. Since $\int g_\varepsilon = 1$ and $\int |g'_\varepsilon| = \nu_1$, changing the belief from p' to p moves the two densities by at most $\|p - p'\|_{L^1(\Theta)}$ and $\theta_{\max}\nu_1 \|p - p'\|_{L^1(\Theta)}$ in $L^1(\mathbb{R})$. So $f_p(\cdot \mid e)$ and $\partial_e f_p(\cdot \mid e)$ are Lipschitz in $L^1(\mathbb{R})$ in the effort and in the belief.

Now let $w_n \to w$ in $\mathcal{A}$, $e_n \to e$ and $\|p_n - p\|_{L^1(\Theta)} \to 0$, and write $f_n := f_{p_n}(\cdot \mid e_n)$ and $f := f_p(\cdot \mid e)$. By the triangle inequality and the bounds above, $\|f_n - f\|_1 \le \theta_{\max}\nu_1 |e_n - e| + \|p_n - p\|_{L^1(\Theta)} \to 0$. Writing

$$\int w_n f_n - \int wf \;=\; \int w_n\, (f_n - f) + \int (w_n - w)\, f,$$

the first term is at most $\bar{w}\, \|f_n - f\|_1$ in absolute value, since $0 \le w_n \le \bar{w}$, and the second goes to zero because

$w_n \to w$ in the weak-* topology and f is integrable. So $W_{p_n}(w_n; e_n) \to W_p(w; e)$. The same argument with $\partial_e f$ in place of f gives $M_{p_n}(w_n; e_n) \to M_p(w; e)$. Hence W_p and M_p are jointly continuous in the contract, the effort and the belief. Finally, $|M_p| \le \bar{w}\theta_{\max}\nu_1$.

A second-period behavior s of the principal specifies, measurably in first-period signal y_1, whether she keeps the agent, $\rho^s(y_1) \in \{0, 1\}$, and which contract $w_2^s(\cdot\,; y_1) \in \mathcal{A}$ she offers him. Step 3 shows that optimal behaviors exist and collects them in a set $\mathcal{S}(\hat{e})$. For every s, the continuation incentive $T(e, \hat{e}; s)$ of Lemma 1 satisfies $|T(e, \hat{e}; s)| \le \bar{w}\theta_{\max}\nu_1$, since the agent's state-conditional rents lie in $[-\bar{w}, \bar{w}]$: payments lie in $[0, \bar{w}]$, and his second-period effort costs at most $\bar{w}$, because zero effort is available. Let

$$\bar{e} := \frac{\bar{w}\theta_{\max}\nu_1}{\inf_{e\in\mathbb{R}_+} c''(e)}.$$

For $e > 2\bar{e}$, $c'(e) \ge e \inf c'' > 2\bar{w}\theta_{\max}\nu_1 \ge M_{g_\theta} + T(e, \hat{e}; s)$, so no effort above $2\bar{e}$ is a best response to any contract, and it is without loss to restrict first-period efforts to $[0, 2\bar{e}]$.

Step 2: the agent's first-period best response. Given the principal's second-period behavior s, the agent's first-order condition at the first-period contract w_1 and conjecture $\hat{e}$ is

$$M_{g_\theta}(w_1; e) + T(e, \hat{e}; s) - c'(e) = 0. \tag{10}$$

Its left side is the derivative of the agent's first-period objective, which is differentiable and, by Assumption 2, strictly concave. The derivative of a differentiable concave function of one variable is continuous, so the left side is continuous and strictly decreasing in e. Either it is negative at every $e \ge 0$, and the best response is $e^* = 0$, or it vanishes at a unique $e^* \in [0, 2\bar{e}]$, which is the best response.

Step 3: the second period. The posterior $p(\cdot \mid y_1, \hat{e})$ is continuous in $(y_1, \hat{e})$ pointwise in θ, and hence in $L^1(\Theta)$ by dominated convergence, since the posterior densities are bounded when $(y_1, \hat{e})$ ranges over a compact set. By Assumption 2, the agent's second-period best response to the contract w_2 at belief p is unique. Write it $e_2^*(w_2; p)$. His objective $W_p(w_2; e) - c(e)$ is continuous in the effort, the contract and $(y_1, \hat{e})$, and the wage alone keeps his effort in $[0, \bar{e}]$. So by Berge's maximum theorem, applied on $[0, \bar{e}]$, and since e_2^* is unique, it is continuous in the contract and in $(y_1, \hat{e})$. The principal's continuation value is

$$\Pi_2^*(y_1, \hat{e}) = \max_{w_2\in\mathcal{A}} \left\{\mathbb{E}_p[\theta]\, e_2^* - W_p(w_2; e_2^*)\right\},$$

where $p := p(\cdot \mid y_1, \hat{e})$ and $e_2^* := e_2^*(w_2; p)$. The objective is continuous in $(w_2, y_1, \hat{e})$ and $\mathcal{A}$ is compact. By Berge's maximum theorem, Π_2^* is continuous, and the set $\mathcal{A}^*(y_1, \hat{e})$ of optimal contracts is nonempty, compact, and upper hemicontinuous on $\mathbb{R} \times \mathbb{R}_+$. Since Π_2^* is continuous, the optimal retention decisions $\rho^*(y_1, \hat{e})$ of (3) also form an upper hemicontinuous correspondence with nonempty compact values, and so does the correspondence of optimal pairs $(\rho^*, \mathcal{A}^*)$. Its values lie in $\{0, 1\} \times \mathcal{A}$, which is compact and metrizable, hence Polish. An upper hemicontinuous correspondence with compact values is weakly measurable, so by the Kuratowski–Ryll-Nardzewski selection theorem (Aliprantis and Border, 2006, Theorem 18.13), a selection of $(\rho^*, \mathcal{A}^*)$ that is measurable in y_1 exists. Such a selection is an optimal second-period behavior in the sense of Step 1. Denote by $\mathcal{S}(\hat{e})$ the set of all of them at conjecture $\hat{e}$. The principal's continuation

value

$$\Psi_1(\hat{e}) = \mathbb{E}[\max\{\Pi_2^*(\hat{e}\theta + \varepsilon, \hat{e}), \pi_0\}],$$

the expectation taken over θ and ε, is continuous in $\hat{e}$ by dominated convergence and does not depend on the selection.

Step 4: the continuation incentives. By Lemma 1, under a selection $s \in \mathcal{S}(\hat{e})$ the agent's continuation incentive is

$$T(e, \hat{e}; s) = -\int_\Theta \theta' \Big[\int_{\mathbb{R}} \rho^s(y_1)\, u_2^s(\theta', y_1; e, \hat{e})\, g_\varepsilon'(y_1 - \theta' e)\, dy_1 \Big] g_\theta(\theta')\, d\theta', \tag{11}$$

where $\rho^s(y_1)$ is the retention decision under s and $u_2^s(\theta', y_1; e, \hat{e})$ is the agent's state-conditional rent at the contract s picks, at his optimal second-period effort after first-period effort e. To separate what the principal chooses from what the agent does, write a choice of the principal after the signal y_1 as $x = (\rho, w_2) \in \{0, 1\} \times \mathcal{A}$, a retention decision and a second-period contract. Let $u_2(\theta', y_1; w_2, e)$ be the agent's state-conditional rent under w_2, at his optimal second-period effort given his posterior after first-period effort e and the signal y_1, and for $x = (\rho, w_2)$ let

$$k(y_1, x; e) := -\rho \int_\Theta \theta'\, u_2(\theta', y_1; w_2, e)\, g_\varepsilon'(y_1 - \theta' e)\, g_\theta(\theta')\, d\theta'$$

be the contribution of the signal y_1 to (11) when the principal chooses x there. Exchanging the order of integration in (11),

$$T(e, \hat{e}; s) = \int_{\mathbb{R}} k(y_1, s(y_1); e)\, dy_1.$$

The function k is continuous in (y_1, x, e), so in particular $(y_1, x) \mapsto k(y_1, x; e)$ is a Carathéodory function. The state-conditional payment $\int w_2(y)\, g_\varepsilon(y - \theta' e_2)\, dy$ is continuous in the contract and in e_2 by the shift bound of Step 1 and the argument that follows it, the agent's second-period effort e_2 is continuous in the contract and in his posterior after first-period effort e by Step 3, and $g_\varepsilon'(y_1 - \theta' e)$ is continuous in (y_1, e). Continuity in y_1 holds because the agent's posterior after the signal y_1, and hence his second-period effort, is continuous in y_1 by Step 3. Let $C^*(y_1, \hat{e}) \subset \{0, 1\} \times \mathcal{A}$ be the set of the principal's optimal choices after y_1, upper hemicontinuous with nonempty compact values by Step 3, and therefore weakly measurable in y_1. Let

$$\mathcal{T}(e, \hat{e}) := \{T(e, \hat{e}; s) : \ s \in \mathcal{S}(\hat{e})\}.$$

Fix $(e, \hat{e})$. At each signal, let

$$K(y_1) := \{k(y_1, x; e) : \ x \in C^*(y_1, \hat{e})\}$$

be the set of contributions the principal can give the agent there through her optimal choices. Lebesgue measure on signals has no atoms, so the integrals $\int \varphi\, dy_1$ of the measurable functions φ with $\varphi(y_1) \in K(y_1)$ at every signal form a convex and compact set (Aumann, 1965). This set is $\mathcal{T}(e, \hat{e})$. Every $s \in \mathcal{S}(\hat{e})$ gives such a function, $\varphi(y_1) = k(y_1, s(y_1); e)$. Conversely, for every such φ, Filippov's implicit function theorem (Aliprantis and Border, 2006, Theorem 18.17) gives a measurable s with $s(y_1) \in C^*(y_1, \hat{e})$ and $k(y_1, s(y_1); e) = \varphi(y_1)$ at every signal, so $s \in \mathcal{S}(\hat{e})$ and its incentive is $\int \varphi\, dy_1$. Write $\mathcal{T}(e, \hat{e}) = [a(e, \hat{e}), b(e, \hat{e})]$, where $a, b : [0, 2\bar{e}]^2 \to \mathbb{R}$ give its smallest and largest elements. The largest is attained by a selection that picks, af-

ter each signal, an optimal choice with the largest contribution, which is measurable by the measurable maximum theorem (Aliprantis and Border, 2006, Theorem 18.19). So

$$b(e,\hat{e}) = \int_{\mathbb{R}} \max_{x \in C^*(y_1,\hat{e})} k(y_1, x; e)\, dy_1,$$

and $a(e,\hat{e})$ is the same integral with a minimum.

Let $(e_n, \hat{e}_n) \to (e, \hat{e})$. At each signal, upper hemicontinuity of C^* and continuity of k give

$$\limsup_{n\to\infty} \max_{x \in C^*(y_1,\hat{e}_n)} k(y_1, x; e_n) \;\le\; \max_{x \in C^*(y_1,\hat{e})} k(y_1, x; e).$$

Every contribution satisfies $|k(y_1, x; e)| \le h_e(y_1) := \bar{w}\theta_{\max} \int_\Theta |g'_\varepsilon(y_1 - \theta' e)|\, g_\theta(\theta')\, d\theta'$. The functions h_{e_n} converge pointwise to h_e, and they all have the same integral $\bar{w}\theta_{\max}\nu_1$, since $\int |g'_\varepsilon| = \nu_1$ is finite by log-concavity and does not change under translation. Fatou's lemma applied to the nonnegative functions $h_{e_n} - \max_{x \in C^*(y_1,\hat{e}_n)} k(y_1, x; e_n)$, together with the inequality above, gives

$$\limsup_{n\to\infty} b(e_n, \hat{e}_n) \le b(e, \hat{e}),$$

and symmetrically

$$\liminf_{n\to\infty} a(e_n, \hat{e}_n) \ge a(e, \hat{e}).$$

So the graph of $(e, \hat{e}) \mapsto \mathcal{T}(e, \hat{e})$ is closed. The same argument applied to a single selection shows that each $T(\cdot, \hat{e}; s)$ is continuous. So $b(\cdot, \hat{e})$, a maximum of continuous functions, is lower semicontinuous, and $a(\cdot, \hat{e})$ is upper semicontinuous. With the inequalities above, $a(\cdot, \hat{e})$ and $b(\cdot, \hat{e})$ are continuous.

Fix a first-period contract w_1 and write

$$N_{w_1}(e) := c'(e) - M_{g_\theta}(w_1; e).$$

By Assumption 2, $N_{w_1}(e) - T(e, \hat{e}; s)$ is strictly increasing in e for every s. Hence so are $N_{w_1}(e) - a(e, \hat{e})$ and $N_{w_1}(e) - b(e, \hat{e})$, because a and b are attained. If $e < e'$ and s' attains $b(e', \hat{e})$,

$$N_{w_1}(e') - b(e', \hat{e}) = N_{w_1}(e') - T(e', \hat{e}; s') > N_{w_1}(e) - T(e, \hat{e}; s') \ge N_{w_1}(e) - b(e, \hat{e}),$$

and the argument for a uses a selection attaining $a(e, \hat{e})$.

Step 5: consistent efforts. Fix w_1 and $\hat{e}$. By Step 2, the agent has a single best response to each $s \in \mathcal{S}(\hat{e})$: zero if $N_{w_1}(0) \ge T(0, \hat{e}; s)$, and otherwise the effort at which $N_{w_1}(e) = T(e, \hat{e}; s)$. Let $\mathcal{B}(\hat{e})$ be the set of these best responses over $s \in \mathcal{S}(\hat{e})$. Then

$$\mathcal{B}(\hat{e}) = [e_{\min}(\hat{e}),\, e_{\max}(\hat{e})],$$

where $e_{\min}(\hat{e})$ is 0 if $N_{w_1}(0) \ge a(0, \hat{e})$ and otherwise the effort at which $N_{w_1}(e) = a(e, \hat{e})$, and $e_{\max}(\hat{e})$ is 0 if $N_{w_1}(0) \ge b(0, \hat{e})$ and otherwise the effort at which $N_{w_1}(e) = b(e, \hat{e})$. Each of these equations has a unique solution in $[0, 2\bar{e}]$ by Steps 1 and 4. Every best response lies in this interval because $a(e, \hat{e}) \le T(e, \hat{e}; s) \le b(e, \hat{e})$. Conversely, every positive e in the interval satisfies $a(e, \hat{e}) \le N_{w_1}(e) \le b(e, \hat{e})$, so some $s \in \mathcal{S}(\hat{e})$ has

$T(e, \hat{e}; s) = N_{w_1}(e)$ and e is its best response, and 0 lies in the interval exactly when it is the best response to a selection attaining $a(0, \hat{e})$.

Since the graph of $\mathcal{T}$ is closed and N_{w_1} is continuous, the graph of $\hat{e} \mapsto \mathcal{B}(\hat{e})$ is closed, and $\mathcal{B}$ maps $[0, 2\bar{e}]$ into itself with nonempty closed intervals as values. By Kakutani's fixed-point theorem there is $\hat{e} \in \mathcal{B}(\hat{e})$, that is, a selection $s \in \mathcal{S}(\hat{e})$ to which $\hat{e}$ is the agent's best response. So $\hat{e}$ is a consistent effort at w_1, and the principal's second-period behavior s is optimal and pure. The set $\mathcal{E}(w_1)$ of consistent efforts is therefore nonempty, and it is compact as a closed subset of $[0, 2\bar{e}]$. Its graph is closed in the contract, since T does not depend on the first-period contract, which enters only through M_{g_θ}, continuous on $\mathcal{A} \times \mathbb{R}_+$.

Step 6: an optimal contract. The principal's profit when she offers w_1 and the agent plays $e \in \mathcal{E}(w_1)$ is $\mu_0 e + \Psi_1(e) - W_{g_\theta}(w_1; e)$, continuous in (w_1, e) by Steps 1 and 3. Let

$$V(w_1) := \max \{\mu_0 e + \Psi_1(e) - W_{g_\theta}(w_1; e) : \ e \in \mathcal{E}(w_1)\}.$$

Since $\mathcal{E}$ has closed graph and compact values, V is upper semicontinuous on the compact set $\mathcal{A}$. So it attains its maximum at some w_1^*, with some $e^* \in \mathcal{E}(w_1^*)$ attaining $V(w_1^*)$.

Consider the following profile.

(i) The agent accepts every contract, which gives him a nonnegative payoff since zero effort is available.

(ii) At every first-period contract, the agent plays a consistent effort that maximizes the principal's profit there.

(iii) In the second period, the agent plays his best response under his own posterior, including after his own deviations.

(iv) The principal conjectures the agent's first-period effort and resolves her second-period ties by the selection that makes it consistent.

(v) The principal offers w_1^*.

(vi) The principal forms the relationship if $V(w_1^*) \geq 2\pi_0$. At equality either choice is optimal, and either gives an equilibrium.

Each player's strategy is optimal given the other's, and the principal's beliefs are consistent with the strategies, so the profile is a pure-strategy perfect Bayesian equilibrium. □

Proofs of Theorem 2 (Shape of the optimal contract) and Corollary 1

Step 1 derives the bang-bang form of the wage. Steps 2–4 write the function whose zeros are the thresholds as a smoothing of an explicit weight, so that its crossings are bounded by the sign changes of that weight. Step 5 removes a restriction on the belief by approximation, and Step 6 makes the counts exact.

Fix a period $t \in \{1, 2\}$, a strictly positive target effort, and the conjecture at which the continuation objects are built, conditional on retention when $t = 2$. All period-2 objects depend on the realized first-period signal, which is suppressed in the notation. Write p_t for the period-t belief over Θ, the prior at $t = 1$

and the posterior at $t = 2$, which is log-concave and C^1 on the interior of Θ;[23] write $e_t > 0$ for the target effort and f_t, $\partial_e f_t$ for the density of the signal of Section 3.3 and its effort derivative, both evaluated at e_t and at the belief p_t.

Step 1: the wage is bang-bang on the score.

Consider period 1. Period 2 is identical with the continuation terms absent. The cheapest contract implementing the target e_1 minimizes the expected wage subject to the agent's first-order condition and the wage bounds, with $\rho(y_1) \in \{0, 1\}$ the retention rule and $U_2(y_1)$ the agent's continuation rent when retained, both with their remaining arguments suppressed. Attaching multiplier λ_1 to the first-order condition and $\underline{\eta}(\cdot) \geq 0$, $\overline{\eta}(\cdot) \geq 0$ to the liability floor and the cap,[24]

$$\begin{aligned}\mathcal{L}(w_1; \lambda_1, \underline{\eta}, \overline{\eta}) = &- \int w_1(y_1)\, f_1(y_1)\, dy_1 \\ &+ \lambda_1 \left[\int \left(w_1(y_1) + \rho(y_1)\, U_2(y_1)\right) \partial_e f_1(y_1)\, dy_1 - c'(e_1) \right] \\ &+ \int \underline{\eta}(y_1)\, w_1(y_1)\, dy_1 + \int \overline{\eta}(y_1)[\bar{w} - w_1(y_1)]\, dy_1.\end{aligned}$$

Because w_1 enters $\mathcal{L}$ only through the integrals above, the first-order condition may be taken pointwise in y_1: differentiating the integrand with respect to the value $w_1(y_1)$ and setting the result to zero gives

$$-f_1(y_1) + \lambda_1\, \partial_e f_1(y_1) + \underline{\eta}(y_1) - \overline{\eta}(y_1) = 0,$$

with complementary slackness $\underline{\eta}(y_1)\, w_1(y_1) = 0$ and $\overline{\eta}(y_1)[\bar{w} - w_1(y_1)] = 0$. The multiplier is strictly positive. Given λ_1, the Lagrangian is maximized by paying $\bar{w}$ wherever $f_1 - \lambda_1 \partial_e f_1 < 0$ and nothing elsewhere. If $\lambda_1 = 0$ this pays nowhere, and if $\lambda_1 < 0$ it pays only where $\partial_e f_1 < 0$, so the wage delivers zero or negative marginal incentive. Both contradict the hypothesis that it must deliver a strictly positive one, so $\lambda_1 > 0$. Set $\kappa_1 := 1/\lambda_1 > 0$ and

$$D_1(y_1) := \partial_e f_1(y_1) - \kappa_1 f_1(y_1),$$

so that stationarity reads $\lambda_1 D_1(y_1) = \overline{\eta}(y_1) - \underline{\eta}(y_1)$. Write $\ell_1 := \partial_e f_1 / f_1$ for the likelihood score of first-period signal in effort, and ℓ_2 for its second-period counterpart, so that $D_1 = f_1(\ell_1 - \kappa_1)$.

Since $\bar{w} > 0$, the floor and the cap cannot bind at the same y_1. Where $D_1(y_1) > 0$, necessarily $\overline{\eta}(y_1) > \underline{\eta}(y_1) \geq 0$, so the cap binds and $w_1(y_1) = \bar{w}$. Where $D_1(y_1) < 0$, the floor binds and $w_1(y_1) = 0$. At an interior wage both multipliers are slack, forcing $D_1(y_1) = 0$. In each case the constraint that binds is the one that helps: the principal pays the most it may where the coefficient on the wage is positive and the least where it is negative.

On the level set $\{D_1 = 0\} = \{\ell_1 = \kappa_1\}$ the principal is indifferent pointwise, so the wage there is undetermined by stationarity alone. Step 6 shows this set is null, so that the cheapest contract is unique up to a null set and pays $\bar{w}$ on $\{D_1 \geq 0\}$ and nothing elsewhere.

[23] The prior by Assumption 1. The posterior because it is proportional to $g_\varepsilon(y_1 - e_1\theta)\, g_\theta(\theta)$, a product of functions log-concave in θ, the first by composition of a log-concave function with an affine map.

[24] Participation does not appear: the constraint never binds, so its multiplier is zero by complementary slackness.

The level κ_1. A multiplier exists by the Lagrange multiplier theorem for convex programs (Luenberger, 1969, Section 8.3), since some admissible contract delivers strictly more than the required incentive, and the constraint pins down its value κ_1: the incentive of the level-set contract is continuous and decreasing in the level, from its maximum at $\kappa = 0$ to zero, so a unique level delivers exactly the incentive the target requires. The same holds in period 2.

Hence the payment set is $\{D_1 \geq 0\}$ and the number of thresholds is $S^-(D_1)$, where for a function $\xi : \mathbb{R} \to \mathbb{R}$ the number of sign changes is

$$S^-(\xi) := \sup\Big\{ n \geq 0 : \ \exists\, y_0 < y_1 < \cdots < y_n \text{ with } \xi(y_{i-1})\xi(y_i) < 0 \text{ for } i = 1, \ldots, n \Big\}.$$

This count ignores zeros and is the functional the variation-diminishing theorem bounds.

The same argument at $t = 2$ gives the payment set $\{D_2 \geq 0\}$ with $D_2 := \partial_e f_2 - \kappa_2 f_2$, $\kappa_2 = 1/\lambda_2 > 0$. From here on the periods are treated uniformly, writing D_t.

In terms of the score ℓ_t, the payment set is $\{\ell_t \geq \kappa_t\}$, so the cutoff is a level of the score.

Write $\underline{\ell}_t := \inf_{y \in \mathbb{R}} \ell_t(y)$ and $\overline{\ell}_t := \sup_{y \in \mathbb{R}} \ell_t(y)$ for the range of the score.

Claim 6 (Interior cutoff)**.** *The payment set $\{\ell_t \geq \kappa_t\}$ is a nonempty proper subset of $\mathbb{R}$ if and only if $\underline{\ell}_t < \kappa_t \leq \overline{\ell}_t$, in which case ℓ_t takes values on both sides of κ_t. Outside that range the wage is constant and delivers zero marginal incentive.*

In period 2 the condition is implied by positive effort, since the wage is then the only source of incentive. In period 1 it is the theorem's hypothesis: the target lies above the continuation-only effort e_0.

Proof. The set is nonempty exactly when some y has $\ell_t(y) \geq \kappa_t$, that is when $\kappa_t \leq \overline{\ell}_t$, and proper exactly when some y has $\ell_t(y) < \kappa_t$, that is when $\kappa_t > \underline{\ell}_t$. If it is empty the wage is identically 0. If it is all of $\mathbb{R}$ the wage is identically $\bar{w}$, and then

$$M_w = \bar{w} \int \partial_e f_t \, dy = \bar{w}\, \partial_e \int f_t \, dy = 0.$$

Either way the wage delivers no marginal incentive. In period 2 there is no continuation, so the agent's first-order condition reads $c'(e_2) = 0$ and $c'(0) = 0$ gives $e_2 = 0$, contradicting the hypothesis. Since ℓ_t is continuous its range is an interval, so on the displayed range it takes values on both sides of κ_t. □

No condition on the tail behavior of the score is used here: the claim is about where κ_t sits relative to the range of ℓ_t, whether that range is bounded or not.

Step 2: the sign function.

Explicitly,

$$D_t(y) = \int_\Theta \Big[-\theta\, g'_\varepsilon(y - e_t\,\theta) - \kappa_t\, g_\varepsilon(y - e_t\,\theta) \Big] p_t(\theta)\, d\theta.$$

Step 3: reduction to a single convolution.

Reparametrize by the location $x = e_t\theta$, strictly increasing in θ since $e_t > 0$, so sign counts are preserved. The reparametrized belief is the density of x,

$$n_t(x) = \frac{1}{e_t}\, p_t\!\left(\frac{x}{e_t}\right)$$

on $[\underline{x}_t, \overline{x}_t] := [e_t\,\underline{\theta},\ e_t\,\overline{\theta}]$, which is log-concave because $\log n_t(x) = \log p_t(x/e_t) - \log e_t$ is a concave function composed with an affine map, plus a constant, and it is C^1 on the interior because p_t is. In these variables $\theta = x/e_t$ and

$$f_t(y) = \int_{\underline{x}_t}^{\overline{x}_t} g_\varepsilon(y-x)\, n_t(x)\, dx = (g_\varepsilon * n_t)(y),$$
$$\partial_e f_t(y) = -\frac{1}{e_t}\int_{\underline{x}_t}^{\overline{x}_t} x\, g_\varepsilon'(y-x)\, n_t(x)\, dx,$$

so the density of the signal is the convolution of the noise with the reparametrized belief. Suppose first that the belief equals zero at the endpoints of its support, $n_t(\underline{x}_t) = n_t(\overline{x}_t) = 0$. Step 5 removes this restriction. Using $g_\varepsilon'(y-x) = -\partial_x[g_\varepsilon(y-x)]$ and integrating by parts in x, the boundary terms vanishing by the endpoint condition,

$$D_t(y) = (g_\varepsilon * h_t)(y), \tag{12}$$

where

$$h_t(x) := -\frac{1}{e_t}\big(x\, n_t(x)\big)' - \kappa_t\, n_t(x),$$

so D_t itself is a single convolution: the noise density against the explicit weight h_t.

Step 4: the crossing count.

The variation-diminishing theorem (Karlin, 1968, Ch. 5, Thm. 3.1(i)) states that if a kernel is totally positive of order r and a weight has at most $r-1$ sign changes, then their transform has no more sign changes than the weight. Under Assumption 1 the location kernel $(y,x) \mapsto g_\varepsilon(y-x)$ is totally positive of order three, so the theorem applies to weights with at most two sign changes, which covers both cases. Case (i), with one sign change, needs only the order-two property that log-concavity provides.[25] Applied to (12),

$$S^-(D_t) \;\le\; S^-(h_t). \tag{13}$$

Writing $\psi_t := (\log n_t)'$ for the score of the reparametrized belief,

$$h_t(x) = -\frac{n_t(x)}{e_t}\Big[\underbrace{(1 + e_t\kappa_t) + x\,\psi_t(x)}_{=:\,\chi_t(x)}\Big].$$

Since $n_t > 0$ on the interior, h_t has the same number of sign changes as χ_t, with the opposite sign pattern: a pattern $-,+,-$ for χ_t is $+,-,+$ for h_t. Display (13) transfers only the count. The sign pattern of D_t is de-

[25] Order three cannot be dropped for case (ii): for the noise $g_\varepsilon \propto \exp(-u^2/2 - 0.03\cos 5u)$, which is strongly log-concave with curvature in $[0.25, 1.75]$ but not of order three, the belief $p \propto (1-\theta^2)^2$ on $[-1,1]$ and target 0.3, the score has four local minima, and for every cutoff in $(0.046, 0.054)$ the payment set has four components.

termined in Step 6. Log-concavity of n_t enters from here on through the single fact that ψ_t is nonincreasing. Because n_t is positive on the interior and equals zero at both endpoints, its mode m_t is interior, $\psi_t > 0$ on $(\underline{x}_t, m_t)$, $\psi_t < 0$ on $(m_t, \overline{x}_t)$, and $\psi_t \to +\infty$ at $\underline{x}_t$, $\psi_t \to -\infty$ at $\overline{x}_t$, since the slope of a concave function diverges where the function falls to $-\infty$.

Claim 7. *Let n_t be as above.* (i) *If $\underline{x}_t \geq 0$, then χ_t changes sign exactly once, from + to −.* (ii) *If $\underline{x}_t < 0 < \overline{x}_t$, then χ_t changes sign exactly twice, in the pattern −, +, −.*

Proof. Since $1 + e_t\kappa_t > 0$,

$$x\,\psi_t(x) \geq 0 \implies \chi_t(x) \;\geq\; 1 + e_t\kappa_t \;>\; 0, \tag{14}$$

so χ_t can equal zero only where x and $\psi_t(x)$ have strictly opposite signs: on one of

$$P := \{x \in (\underline{x}_t, \overline{x}_t) : x > 0,\ \psi_t(x) < 0\},$$
$$N := \{x \in (\underline{x}_t, \overline{x}_t) : x < 0,\ \psi_t(x) > 0\}.$$

Both are intervals: P is the part of the support to the right of both 0 and m_t, and N the part to the left of both.

Exactly one crossing on P. The product $x\,\psi_t(x)$ is strictly decreasing on P: for $x_1 < x_2$ in P,

$$x_2\,\psi_t(x_2) \;\leq\; x_2\,\psi_t(x_1) \;<\; x_1\,\psi_t(x_1),$$

the first inequality because ψ_t is nonincreasing and $x_2 > 0$, the second because $\psi_t(x_1) < 0$ and $x_2 > x_1$. Hence χ_t is strictly decreasing on P. At the left endpoint of P, $x\,\psi_t(x) \geq 0$ by continuity, so $\chi_t \geq 1 + e_t\kappa_t > 0$ there by (14). As $x \uparrow \overline{x}_t$,

$$\chi_t(x) \;\longrightarrow\; -\infty,$$

because $\overline{x}_t > 0$ and $\psi_t \to -\infty$. A continuous, strictly decreasing function that starts strictly positive and falls to $-\infty$ equals zero exactly once, crossing from + to −.

Exactly one crossing on N, when N is nonempty. Symmetrically, $x\,\psi_t(x)$ is strictly increasing on N: for $x_1 < x_2$ in N,

$$x_1\,\psi_t(x_1) \;\leq\; x_1\,\psi_t(x_2) \;<\; x_2\,\psi_t(x_2),$$

the first inequality because ψ_t is nonincreasing and $x_1 < 0$, the second because $\psi_t(x_2) > 0$ and $x_1 < x_2$. Hence χ_t is strictly increasing on N. When $\underline{x}_t < 0$, as $x \downarrow \underline{x}_t$,

$$\chi_t(x) \;\longrightarrow\; -\infty,$$

because there $\psi_t \to +\infty$ and $x < 0$, while $\chi_t \geq 1 + e_t\kappa_t > 0$ at the right endpoint of N by (14). So χ_t equals zero exactly once on N, crossing from − to +.

If $\underline{x}_t \geq 0$ the support contains no $x < 0$, so N is empty and χ_t has exactly one sign change, + to −. If $\underline{x}_t < 0 < \overline{x}_t$, each of N and P contributes exactly one crossing, and $\chi_t > 0$ between them by (14): the pattern is −, +, −. □

Since $e_t > 0$, the case split depends only on primitives: $\underline{x}_t \geq 0 \iff \underline{\theta} \geq 0$, and $\underline{x}_t < 0 < \overline{x}_t \iff \underline{\theta} <$

$0 < \bar{\theta}$, the second inequality automatic since $\bar{\theta} > 0$. Combining (13) with Claim 7: for beliefs that equal zero at the endpoints, $S^-(D_t) \le 1$ when $\underline{\theta} \ge 0$ and $S^-(D_t) \le 2$ when $\underline{\theta} < 0$.

Step 5: general log-concave beliefs.

Let p_t be an arbitrary log-concave belief on Θ, possibly positive at the endpoints (e.g. uniform, or a truncated Gaussian). For $n \in \mathbb{N}$ define

$$p_{t,n}(\theta) \;:=\; \frac{1}{C_n}\, p_t(\theta)\, w(\theta)^{1/n},$$

with $w(\theta) := (\theta - \underline{\theta})(\bar{\theta} - \theta)$, where $C_n := \int_\Theta p_t\, w^{1/n}\, d\theta$ is the normalizing constant. Each $p_{t,n}$ is log-concave, since its log adds the concave terms $\frac{1}{n}\log(\theta - \underline{\theta}) + \frac{1}{n}\log(\bar{\theta} - \theta)$ to $\log p_t$. It is C^1 on the interior, and equals zero at both endpoints, so Steps 3–4 apply to it: writing $D_{t,n}$ for the associated function in (12), $S^-(D_{t,n}) \le 1$ when $\underline{\theta} \ge 0$ and $S^-(D_{t,n}) \le 2$ when $\underline{\theta} < 0$, for every n.

Two elementary bounds control the approximation: with $\bar{W} := \max\{1, (\bar{\theta} - \underline{\theta})^2/4\}$ and $c_0 := \int_\Theta p_t \min\{w, 1\}\, d\theta > 0$, one has $p_{t,n} \le (\bar{W}/c_0)\, p_t$ uniformly in n, and $p_{t,n} \to p_t$ pointwise on the interior.[26] The same argument gives $D_{t,n}(y) \to D_t(y)$ for every y.[27]

Crossings can disappear in the limit but cannot appear. If $S^-(D_t) \ge k+1$, choose points $y_0 < \cdots < y_{k+1}$ at which D_t takes strictly alternating signs. For n large, $D_{t,n}$ takes the same strict signs at those points, so $S^-(D_{t,n}) \ge k+1$. Applying this with $k = 1$ and $k = 2$ transfers the ceilings of Step 4 to D_t with no restriction on the log-concave belief.

Step 6: exact counts and the shape of the payment set.

First, by Claim 6 the payment set is neither empty nor all of $\mathbb{R}$, so $S^-(D_t) \ge 1$ in both cases.

Case (i), $\underline{\theta} \ge 0$. Steps 4–5 give $S^-(D_t) \le 1$, so the count is exactly one. For the direction, let m_ε denote a mode of g_ε. Log-concave densities are unimodal, so $g'_\varepsilon \ge 0$ on $(-\infty, m_\varepsilon]$. For any $y \le m_\varepsilon$ and any $\theta \in \Theta$, the argument satisfies $y - e_t\,\theta \le y \le m_\varepsilon$, because $\theta \ge 0$ on the support and $e_t > 0$. Hence $g'_\varepsilon(y - e_t\,\theta) \ge 0$ and

$$\partial_e f_t(y) = -\int_\Theta \theta\, g'_\varepsilon(y - e_t\,\theta)\, p_t(\theta)\, d\theta \;\le\; 0 \;<\; \kappa_t\, f_t(y),$$

so $D_t < 0$ on $(-\infty, m_\varepsilon]$. The single sign change therefore runs from $-$ to $+$, and the payment set is a single upper interval $[z_t, +\infty)$. This argument needs no tail asymptotics, and it covers the boundary case $\underline{\theta} = 0$: the state $\theta = 0$ has zero marginal product and drops out of the integrand, so only productive complementarities with positive probability still yield the one-sided bonus.

Case (ii), $\underline{\theta} < 0$.

Claim 8. *Let $\underline{\theta} < 0$. Then $D_t(y) > 0$ for all $|y|$ sufficiently large.*

Proof. Since f_t is the mixture over states of $g_\varepsilon(y - e_t\theta)$, the score of the mixture is the posterior-weighted

[26] For every $n \ge 1$, $\min\{w(\theta), 1\} \le w(\theta)^{1/n} \le \bar{W}$, so $c_0 \le C_n \le \bar{W}$, which gives the uniform bound. For each interior θ, $w(\theta)^{1/n} \to 1$, so the integrands $p_t w^{1/n}$ converge pointwise to p_t while dominated by $\bar{W} p_t$. Dominated convergence gives $C_n \to 1$.

[27] Write $K_t(y, \theta) := -\theta g'_\varepsilon(y - e_t\theta) - \kappa_t g_\varepsilon(y - e_t\theta)$ for the integrand of D_t over states, continuous in θ on the compact Θ, so $B_t(y) := \max_\Theta |K_t(y, \theta)| < \infty$. Then $|K_t(y, \theta)| p_{t,n}(\theta) \le B_t(y)(\bar{W}/c_0) p_t(\theta)$, an integrable dominating function independent of n, and $K_t(y, \cdot) p_{t,n} \to K_t(y, \cdot) p_t$ pointwise.

average of the state-conditional scores,

$$\ell(y) = \frac{\partial_e f_t(y)}{f_t(y)} = \mathbb{E}_{p_t(\cdot|y)}\Big[-\theta\,(\log g_\varepsilon)'(y - e_t\,\theta)\Big],$$

so the sign of the score in a tail is governed by which states the posterior puts weight on there. It is *not* enough to bound the average below by the smallest state-conditional score: as $y \to +\infty$ the argument $y - e_t\theta$ diverges to $+\infty$ for every state, so $(\log g_\varepsilon)' \to -\tau_\varepsilon^+$, where $\tau_\varepsilon^+ := -\lim_{u\to+\infty}(\log g_\varepsilon)'(u) \in (0, +\infty]$ and $\tau_\varepsilon^- := \lim_{u\to-\infty}(\log g_\varepsilon)'(u) \in (0, +\infty]$ are the tail slopes of the noise, and the state-θ term tends to $\theta\tau_\varepsilon^+$, which is negative for the counterproductive states $\theta < 0$. The average is nonetheless positive because the posterior concentrates on the productive extreme.

Fix $\theta < \bar\theta$. By log-concavity, $\log g_\varepsilon$ is concave, and since $e_t\,\bar\theta > e_t\,\theta$ the argument at $\bar\theta$ is smaller, so as $y \to +\infty$,

$$\log g_\varepsilon(y - e_t\bar\theta) - \log g_\varepsilon(y - e_t\theta) \;\longrightarrow\; +\infty,$$

the increment of a concave function over a fixed gap growing without bound once its slope diverges. Hence $p_t(\theta \mid y) \to 0$ for every $\theta < \bar\theta$: the posterior concentrates on $\bar\theta$, and

$$\ell(y) \;\longrightarrow\; -\bar\theta\,(\log g_\varepsilon)'(y - e_t\bar\theta) \;\longrightarrow\; \bar\theta\,\tau_\varepsilon^+,$$

strictly positive since $\bar\theta > 0$. Symmetrically, as $y \to -\infty$ the posterior concentrates on $\underline{\theta}$, whose marginal product is negative when $\underline{\theta} < 0$, and $\ell(y) \to |\underline{\theta}|\,\tau_\varepsilon^- > 0$.

With the noise score unbounded, $\tau_\varepsilon^\pm = \infty$, so $\ell > \kappa_t$ and hence $D_t = f_t(\ell - \kappa_t) > 0$ for $|y|$ large, whatever the cutoff. This is the only place unboundedness is used. □

Steps 4–5 give $S^-(D_t) \le 2$, and Claim 6 gives $S^-(D_t) \ge 1$. If the noise is strongly log-concave, with $-(\log g_\varepsilon)'' \ge m > 0$, the score is unbounded: $(\log g_\varepsilon)'(u) \le -m(u - m_\varepsilon)$ for $u \ge m_\varepsilon$ and $(\log g_\varepsilon)'(u) \ge m(m_\varepsilon - u)$ for $u \le m_\varepsilon$, so $\tau_\varepsilon^\pm = \infty$. Claim 8 then gives $D_t > 0$ in both tails. The payment set is not all of $\mathbb{R}$, so a middle region is unpaid and $S^-(D_t) \ge 2$, hence exactly two, with pattern $+, -, +$. The payment set is $(-\infty, z_t^-] \cup [z_t^+, +\infty)$ with $z_t^- < z_t^+$, the ordering strict because the unpaid middle interval is nonempty. Under log-concavity alone, the two tail limits $\bar\theta\tau_\varepsilon^+$ and $|\underline{\theta}|\tau_\varepsilon^-$ can be finite. A cutoff below both still gives two thresholds, while a cutoff between them gives D_t opposite signs at the two ends, an odd count, hence one threshold, with the paid tail on the side with the higher limit. In either case the payment set is a union of at most two tails. Together with Steps 1–2 this establishes both cases of the theorem. *The level set.* In both cases h_t has finitely many sign changes, one or two, and by Assumption 1 the kernel $(y, x) \mapsto g_\varepsilon(y - x)$ is strictly totally positive of order three. The strict form of the variation-diminishing theorem (Karlin, 1968, Ch. 5, Thm. 3.1(ii)) then bounds the sign changes of D_t by those of h_t even when each zero of D_t is counted as a sign change in whichever direction produces more. A function vanishing on an interval would have infinitely many such changes, so D_t vanishes at finitely many points and the level set $\{D_t = 0\}$ of Step 1 is null.

□

Proof of Corollary 1. Gaussian noise is log-concave and totally positive of every order, and its Fisher

information is $1/\sigma^2$, so it satisfies Assumption 1. It is strongly log-concave, since $-(\log g_\varepsilon)'' = 1/\sigma^2$. Part (i) of Theorem 2 then gives exactly one threshold when $\underline{\theta} \geq 0$, with $\bar{w}$ paid above it, and part (ii) gives exactly two, with both tails paid, when $\underline{\theta} < 0$. In either case the upper tail is paid. □

B Proofs of the results under the specification

This appendix proves the results of Section 4, under Assumption 3. The curvature levels α^{FOA}, α^B and $\alpha^U(\gamma)$ depend on σ, as in the text; their σ-argument is suppressed in the proofs. After Lemma 2, a preliminary section collects the objects shared by the remaining proofs: the concentration error, the known-complementarity problem, the score and its level sets, and the concavity of the principal's reduced problem. The proof of Lemma 3 follows, then the regularity of her continuation value, which uses it, then Proposition 2, Theorem 3 with Corollary 2, and the adoption application, in the order of the text.

Proof of Lemma 2 (The posterior depends on a single index)

Proof. Fix the conjecture $\hat{e}_1$ and the prior index γ. By (1) and Definition 1, the posterior at the signal y_1 is proportional in θ to $g_\varepsilon(y_1 - \hat{e}_1\theta)\, e^{\gamma\theta} g_0(\theta)$. With $\varepsilon_t \sim N(0, \sigma^2)$,

$$g_\varepsilon(y_1 - \hat{e}_1\theta) \;\propto\; \exp\Big(-\frac{(y_1 - \hat{e}_1\theta)^2}{2\sigma^2}\Big) \;\propto\; \exp\Big(\frac{\hat{e}_1 y_1}{\sigma^2}\,\theta - \frac{\hat{e}_1^2\theta^2}{2\sigma^2}\Big),$$

the factor $\exp(-y_1^2/2\sigma^2)$ being free of θ and absorbed into the normalizing constant. Multiplying by $e^{\gamma\theta} g_0(\theta)$ and collecting the terms linear in θ,

$$p(\theta \mid y_1; \gamma) \;\propto\; \exp\Big(\Big[\gamma + \frac{\hat{e}_1}{\sigma^2} y_1\Big]\theta\Big)\, g_0(\theta)\, e^{-\hat{e}_1^2\theta^2/2\sigma^2} \;=\; e^{b\,\theta}\, \tilde{g}_0(\theta),$$

which is (7). □

Preliminaries under the specification

Assumption 3 holds throughout, and $V \leq 1/\varsigma$ denotes the bound on the variance of every posterior at a retained history.

The concentration error. Throughout, ϵ is the largest relative error, over retained posteriors and conjectures in $[0, 2\bar{e}]$, from replacing a posterior by the point mass at its mean, of order $\sqrt{V}$, and from shifting a cutoff by the effort $e\theta \leq 2\bar{e}\theta_{\max}$, of order $\bar{e}\theta_{\max}/\sigma$. So $\epsilon = O(\sqrt{V} + \bar{e}\theta_{\max}/\sigma)$, and $(1 \pm \epsilon)$ denotes a factor between $1 - \epsilon$ and $1 + \epsilon$. In particular, every effort is at most of order $1/(\alpha\sigma)$, so once α exceeds a level of order σ^{-2}, the shift θe of the signal by effort is small relative to σ. The wage incentive of an upper-tail contract evaluated at the agent's effort then equals its value at zero effort, which is explicit since the signal is pure noise there, up to the factor $1 \pm \epsilon$.

Tilt identities.

Let q be a density on Θ and let $q_b \propto e^{b\theta} q$ be its tilts. Write $\mathbb{E}_b$, Var_b and Cov_b for expectations, variances and covariances under q_b, and μ_b, V_b and $\kappa_k(b)$ for its mean, variance and k-th cumulant, with

$\bar{\kappa}_k := \kappa_k(b)/V_b^{k/2}$ the standardized cumulants, so that $\bar{\kappa}_3$ is its skewness and $\bar{\kappa}_4$ its excess kurtosis. By Lemma 2, at a fixed conjecture $\hat{e}_1$ every posterior is a tilt of the modified base $\tilde{g}_0$: the posterior after first-period signal y_1 is q_b at the index $b = \gamma + \hat{e}_1 y_1/\sigma^2$, which is increasing in y_1. The same identities apply to the prior family $e^{\gamma\theta} g_0$, with γ in place of b. Write $\Pi_2(b) := \Pi_2(q_b)$.

If $K(x) := \log \int e^{x\theta} q$ is the cumulant generating function of q, then $q_b = e^{b\theta - K(b)} q$ and $K'(b) = \mu_b$, so $\frac{d}{db} \log q_b(\theta) = \theta - \mu_b$. Hence, for every bounded measurable $h : \Theta \to \mathbb{R}$, differentiating $\mathbb{E}_b[h] = \int h\, q_b$ under the integral gives

$$\frac{d}{db}\, \mathbb{E}_b[h] \;=\; \mathrm{Cov}_b(\theta, h), \tag{15}$$

and in particular $\frac{d}{db}\mu_b = V_b > 0$. The cumulant generating function of q_b is $x \mapsto K(b + x) - K(b)$, so $\kappa_k(b) = K^{(k)}(b)$ and

$$\frac{d}{db}\, \kappa_k(b) = \kappa_{k+1}(b).$$

For $h \in C^1(\Theta)$, Hoeffding's covariance identity gives

$$\mathrm{Cov}_b(\theta, h) = \int_\Theta h'(\theta)\, H_b(\theta)(1 - H_b(\theta))\, d\theta,$$

with H_b the distribution function of q_b and $\int_\Theta H_b(1-H_b)\, d\theta = V_b$. For any distribution of θ on Θ with mean μ and variance V, and any differentiable h, the mean value theorem applied to $\mathrm{Cov}(\theta, h) = \mathbb{E}[(\theta-\mu)(h(\theta)-h(\mu))]$ gives

$$|\,\mathrm{Cov}(\theta, h)| \;\le\; V \sup_\Theta |h'|. \tag{16}$$

Finally, if $q = e^{-\psi}$ with $\psi'' \ge \varsigma$, then $q_b = e^{-(\psi - b\theta) - K(b)}$ and $\psi - b\theta$ has the same second derivative as ψ, so every posterior is ς-strongly log-concave, the Brascamp–Lieb inequality (Brascamp and Lieb, 1976) gives $V_b \le 1/\varsigma$, and $\mathbb{E}_b|\theta - \mu_b|^3 \le C_3' V_b^{3/2}$ for a universal constant C_3' of log-concave laws.

The score, its level sets, and the price of incentives.

The score of the first-period signal density in effort is $\ell_{\tilde{e}}(y) = \mu(y)\, y\, \sigma^{-2}(1 \pm \epsilon)$, with $\mu(y)$ the posterior mean at the signal y, and its slope satisfies

$$\ell'_{\tilde{e}}(y) = \frac{\mu(y)}{\sigma^2}(1 \pm \epsilon) \;\ge\; \underline{s} := \frac{\underline{\mu}}{2\sigma^2} \tag{17}$$

on retained histories, where the posterior mean is at least $\underline{\mu} := \sqrt{2\alpha\pi_0}$, since the second-period value at a belief with mean μ is at most the first-best surplus $\mu_+^2/(2\alpha)$ and it is at least π_0 where she retains. So the score is increasing on the range of retained signals, and its level sets are upper tails there. Let $\underline{\kappa}$ denote the lowest level at which this holds, so that the levels used below are those at or above it. Since the score is positive and increasing on retained signals, every nonnegative level qualifies, and $\underline{\kappa} := 0$. The levels that matter are in fact strictly positive: at every optimal contract the level is positive, by Step 4 of the proof of Proposition 2 for the benchmark and Step 2 of the proof of Claim 20 for the model.

For a target $\tilde{e}$ and a level κ, let $A(\kappa, \tilde{e}) := \{\ell_{\tilde{e}} \ge \kappa\}$ be the level set of the score computed at $\tilde{e}$. For a set of signals A and an effort e, write $M_w(A; e) := \bar{w} \int_A \partial_e f_1(y \,|\, e)\, dy$ for the marginal incentive of the contract that

pays $\bar{w}$ on A, and $W(A;e) := \bar{w}\int_A f_1(y\,|\,e)\,dy$ for its expected wage bill. Define the incentive and the bill of the level set at the effort it is computed at,

$$\mathcal{I}(\kappa,\tilde{e}) := M_w(A(\kappa,\tilde{e});\tilde{e})$$

and

$$\mathcal{W}(\kappa,\tilde{e}) := W(A(\kappa,\tilde{e});\tilde{e}),$$

and the density of the score at level κ,

$$\mathcal{D}(\kappa,\tilde{e}) := \sum_{y:\,\ell_{\tilde{e}}(y)=\kappa} \frac{f_1(y\,|\,\tilde{e})}{|\ell'_{\tilde{e}}(y)|}.$$

Claim 9 (Level sets of the score)**.** *For $\kappa > 0$,*

$$\partial_\kappa \mathcal{I} = \kappa\,\partial_\kappa \mathcal{W} = -\kappa\,\bar{w}\,\mathcal{D} < 0.$$

So the level delivering a given $r \in (0, \mathcal{I}(0,\tilde{e}))$ is unique, and along it $d\mathcal{W}/d\mathcal{I} = 1/\kappa$. More generally, at fixed effort $\tilde{e}$, any movement of the cutoffs of $A(\kappa,\tilde{e})$ changes $\mathcal{I}$ by κ times the change in $\mathcal{W}$.

Proof. The Leibniz rule produces boundary terms at $z^\pm$ only. The region is the level set of the score of the density being integrated, so at the cutoffs $\partial_e f_1(z^\pm\,|\,\tilde{e}) = \kappa\, f_1(z^\pm\,|\,\tilde{e})$, which converts each boundary term of $\mathcal{I}$ into κ times the corresponding term of $\mathcal{W}$. And $\partial_\kappa \mathcal{W} = -\bar{w}\mathcal{D} < 0$ by full support. □

The known-complementarity problem.

Take the complementarity known and equal to $\mu > 0$. A bonus paid above a cutoff z delivers, at zero effort, the wage incentive $\bar{w}\mu\phi(z/\sigma)/\sigma$ at level $\kappa = \mu z/\sigma^2$. At a cutoff at standardized position v above the expected signal, the agent's first-order condition gives $\alpha e = \bar{w}\mu\phi(v)/\sigma$, so the principal's value is $\bar{w}\mu^2\phi(v)/(\alpha\sigma) - \bar{w}\bar{\Phi}(v)$. Its derivative, $\bar{w}\phi(v)(1-\mu^2 v/(\alpha\sigma))$, vanishes at

$$v_\mu := \frac{\alpha\sigma}{\mu^2},$$

a maximum. So the optimal contract asks for effort $e^*(\mu) := \bar{w}\mu\phi(v_\mu)/(\alpha\sigma)$, has cutoff $z^*(\mu) := e^*(\mu)\mu + \sigma v_\mu$, and earns $\Pi(\mu) = \bar{w}[\phi(v_\mu)/v_\mu - \bar{\Phi}(v_\mu)]$. Its *incentive yield* is $\varkappa_\mu := \bar{w}\phi(v_\mu)/\sigma$, the incentive the bonus delivers per unit of complementarity, which is also the level of its cutoff, α/μ, times the effort it implements. The yield falls with the noise: incentives are expensive when the signal is noisy. At a concentrated belief with mean μ these values hold up to ϵ. At the prior mean μ_0, write

$$v_1 := v_{\mu_0} = \frac{\alpha\sigma}{\mu_0^2}, \tag{18}$$

so that the known-complementarity cutoff is $z_1 = v_1\sigma$, and the incentive of the bonus at this cutoff is

$$I_1 := \bar{w}\mu_0\frac{\phi(v_1)}{\sigma}.$$

The bonus with incentive $m < I_1$ has cutoff $z(m) = v(m)\sigma$ with $v(m) = (v_1^2 + 2\log(I_1/m))^{1/2} > v_1$. The Gaussian Mills ratio bounds, which hold for every $v > 0$,

$$\frac{\phi(v)}{v}\Big(1 - \frac{1}{v^2}\Big) \;\le\; \bar{\Phi}(v) \;\le\; \frac{\phi(v)}{v}, \tag{19}$$

relate the probability that the bonus is paid, $\bar{\Phi}(v)$, which sets the bill, to the density at the cutoff, $\phi(v)$, which sets the incentive. In particular the bill of a bonus above the cutoff $v\sigma$ is at most its incentive divided by the level of its cutoff, $\bar{w}\bar{\Phi}(v) \le \bar{w}\phi(v)/v$, up to ϵ.

Her reduced problem. The principal chooses a target $\tilde{e}$, the cheapest contract delivering the incentive $r(\tilde{e}) = c'(\tilde{e}) - \tau(\tilde{e})$ it requires is the level set of the score at a level $\kappa(\tilde{e})$, and her payoff is a benefit $G(\tilde{e})$ minus the bill $\mathcal{W}(\kappa(\tilde{e}), \tilde{e})$. The next two claims describe this family and give a condition for her payoff to be concave in the target.

Claim 10 (The family of cheapest contracts)**.** *Fix a belief with positive mean, under which the score is computed, an endpoint $\underline{e} \in [0, 2\bar{e})$, and a constant $\bar{s} \ge 0$. Let $r = c' - \tau$ with $\tau \in C^1[\underline{e}, 2\bar{e}]$ and $|\tau'| \le \bar{s}$. For $\tilde{e} \in [\underline{e}, 2\bar{e}]$ with $0 < r(\tilde{e}) \le \mathcal{I}(0, \tilde{e})$, let $\kappa(\tilde{e})$ solve $\mathcal{I}(\kappa(\tilde{e}), \tilde{e}) = r(\tilde{e})$, and let $\mathcal{K} := \{\tilde{e} \in [\underline{e}, 2\bar{e}] : \kappa(\tilde{e}) \ge \underline{\kappa}\}$. There is a constant $\varsigma_I < \infty$, depending only on the belief and the noise, such that if $\alpha > \bar{s} + \varsigma_I$, then:*

(i) $\mathcal{K}$ is an interval on which κ is C^1 and strictly decreasing;

(ii) along the family, with $\partial_e I(A; e) := \int_A \partial_e^2 f_1$,

$$\frac{d}{d\tilde{e}}\,\mathcal{W}(\kappa(\tilde{e}), \tilde{e}) = r(\tilde{e}) + \frac{r'(\tilde{e}) - \bar{w}\,\partial_e I(A(\kappa(\tilde{e}), \tilde{e}); \tilde{e})}{\kappa(\tilde{e})}. \tag{20}$$

Proof. (i) The level is monotone. On $\mathcal{K}$ the level sets are upper tails, and at their cutoffs $\ell' \ge \underline{s}$ by (17). So the cutoffs, $\mathcal{I}$ and $\mathcal{W}$ are C^1 in $(\kappa, \tilde{e})$ by the implicit function theorem, and by Claim 9

$$\kappa'(\tilde{e}) = \frac{r'(\tilde{e}) - \partial_{\tilde{e}}\mathcal{I}(\kappa, \tilde{e})}{-\kappa\,\bar{w}\,\mathcal{D}(\kappa, \tilde{e})}\bigg|_{\kappa = \kappa(\tilde{e})}. \tag{21}$$

Here $\partial_{\tilde{e}}\mathcal{I}$, at fixed level, collects the movement of the cutoff with the effort the score is computed at,

$$-\bar{w}\kappa\, f_1(z^+)\,\partial_{\tilde{e}} z^+,$$

where

$$\partial_{\tilde{e}} z^+ = -\frac{\partial_{\tilde{e}}\ell_{\tilde{e}}(z^+)}{\ell'_{\tilde{e}}(z^+)},$$

a polynomial in $|z^+|$ times the density there, and the movement of the density, $\bar{w}\partial_e I = \bar{w}\mathbb{E}[\theta^2(-g'_\varepsilon(z^+ - \tilde{e}\theta))]$. Hence $\varsigma_I := \sup_{\mathcal{K}} |\partial_{\tilde{e}}\mathcal{I}| < \infty$. Since $r' \ge \alpha - \bar{s}$, the numerator of (21) is positive and $\kappa' < 0$ on $\mathcal{K}$.

Being continuous, κ cannot cross the level $\underline{\kappa}$ upward, so $\mathcal{K}$ is an interval. It is $[\underline{e}, \tilde{e}_1]$ if $\kappa(\underline{e}) \ge \underline{\kappa}$. It is $(0, \tilde{e}_1]$ when $\underline{e} = 0$ and $\tau(0) = 0$, since then $r(\tilde{e}) \to 0$ while $\mathcal{I}(K, \tilde{e}) \ge \bar{w}K\Pr(\ell_{\tilde{e}} \ge K)$ stays bounded below for every finite K, the score being unbounded above under a belief with positive mean, so that $\kappa(\tilde{e}) \to \infty$ as $\tilde{e} \downarrow 0$. Otherwise it is empty.

(ii) The marginal bill. Write $W(\tilde{e}) := \mathcal{W}(\kappa(\tilde{e}), \tilde{e})$. Then $W' = \partial_\kappa \mathcal{W} \cdot \kappa' + \partial_{\tilde{e}} \mathcal{W} = (r' - \partial_{\tilde{e}} \mathcal{I})/\kappa + \partial_{\tilde{e}} \mathcal{W}$ by (21) and Claim 9. In $\partial_{\tilde{e}} \mathcal{I}/\kappa$ and $\partial_{\tilde{e}} \mathcal{W}$ the cutoff-movement parts cancel by the last sentence of Claim 9, and the density parts are $\bar{w} \partial_e I$ and $\bar{w} \int_A \partial_e f_1 = r$. This is (20). □

Claim 11 (Concavity of the reduced problem in the target)**.** *In the setting of Claim 10, let $G \in C^2[\underline{e}, 2\bar{e}]$, let τ' be Lipschitz with constant $L \geq 0$, and set*

$$\tilde{\Pi}(\tilde{e}) := G(\tilde{e}) - \mathcal{W}(\kappa(\tilde{e}), \tilde{e})$$

on a compact subinterval of $\mathcal{K}$ on which κ is bounded away from zero, with $\kappa_{\mathcal{K}} := \inf \kappa > 0$ there and $\bar{d} := \sup |G''| + L/\kappa_{\mathcal{K}}$. There are constants $\varsigma_2 < \infty$ and $\underline{\Xi} > 0$, depending only on the belief and the noise, such that if $\alpha > \bar{s} + \varsigma_I$ and

$$\underline{\Xi}(\alpha - \bar{s})(\alpha - \bar{s} - \varsigma_I) + (\alpha - \bar{s}) \;>\; \bar{d} + \varsigma_2, \tag{22}$$

then $\tilde{\Pi}$ is strictly concave there. Both conditions hold in particular when $\alpha > \bar{s} + \varsigma_I + \sqrt{(\bar{d} + \varsigma_2)/\underline{\Xi}}$.

Proof. The reduced profit is $\tilde{\Pi} = G - W$ with $W(\tilde{e}) := \mathcal{W}(\kappa(\tilde{e}), \tilde{e})$, so strict concavity means that the marginal bill W' rises faster than the marginal benefit G'. By (20), $W' = r + (r' - \bar{w}\partial_e I)/\kappa$: the marginal bill is the required incentive, which rises at rate at least $\alpha - \bar{s}$, plus a term that rises because $1/\kappa$ rises along the family, the level falling as the target rises (Claim 10(i)). The proof bounds, term by term, how fast W' rises between two targets, against the constants that could work the other way.

The constants. Let

$$\underline{\Xi} := \inf_{\mathcal{K}} (\kappa^3 \bar{w} \mathcal{D})^{-1}$$

and

$$\varsigma_2 := \sup_{\mathcal{K}} \Big| \frac{d}{d\tilde{e}} \Big[\frac{\bar{w} \partial_e I}{\kappa} \Big] \Big|.$$

$\underline{\Xi}$ is positive because at a cutoff at standardized position v, $\kappa = \mu_0 v/\sigma\,(1 \pm \epsilon)$ and $\mathcal{D} \leq f_1(z)/\underline{s} \leq \phi(v)/(\sigma \underline{s})$, so $\kappa^3 \mathcal{D}$ is bounded by a multiple of $v^3 \phi(v)$, which is bounded. ς_2 is finite because in $\bar{w} \partial_e I \cdot \kappa'/\kappa^2$ the density at the cutoff in $\partial_e I$ cancels against $\mathcal{D} \geq f_1(z^\pm)/|\ell'(z^\pm)|$, leaving a quadratic in $|z^\pm|$ over κ^3.

The increment of the marginal bill. Take $\tilde{e} < \tilde{e}'$ in $\mathcal{K}$, $\Delta := \tilde{e}' - \tilde{e}$, $\kappa := \kappa(\tilde{e})$, $\kappa' := \kappa(\tilde{e}')$. By (20) with $r' = c'' - \tau'$,

$$W'(\tilde{e}') - W'(\tilde{e}) = [r(\tilde{e}') - r(\tilde{e})] + \frac{c''(\tilde{e}') - c''(\tilde{e})}{\kappa'} + (c''(\tilde{e}) - \tau'(\tilde{e}))\Big(\frac{1}{\kappa'} - \frac{1}{\kappa}\Big)$$
$$- \frac{\tau'(\tilde{e}') - \tau'(\tilde{e})}{\kappa'} - \Big[\frac{\bar{w}\partial_e I(\tilde{e}')}{\kappa'} - \frac{\bar{w}\partial_e I(\tilde{e})}{\kappa}\Big].$$

Term by term:

- the first is at least $(\alpha - \bar{s})\Delta$, since $r' = c'' - \tau' \geq \alpha - \bar{s}$.
- the second is zero, c'' being constant.

- the third is at least $\underline{\Xi}(\alpha-\bar{s})(\alpha-\bar{s}-\varsigma_I)\Delta$: its coefficient is at least $\alpha-\bar{s}$, and by (21)

$$\frac{d}{d\tilde{e}}\Big(\frac{1}{\kappa}\Big)=\frac{r'-\partial_{\tilde{e}}\mathcal{I}}{\kappa^3\bar{w}\mathcal{D}}\ \geq\ (\alpha-\bar{s}-\varsigma_I)\,\underline{\Xi}.$$

- the fourth is at least $-L\Delta/\kappa_{\mathcal{K}}$, since τ' is Lipschitz with constant L.
- the fifth is at least $-\varsigma_2\Delta$, by the definition of ς_2.

The first and third terms are the two sources of concavity: the required incentive rises with the target, and the level falls, so each unit of incentive costs more.

Conclusion. With $G'(\tilde{e})-G'(\tilde{e}')\geq-\sup|G''|\,\Delta$,

$$\tilde{\Pi}'(\tilde{e})-\tilde{\Pi}'(\tilde{e}')\ \geq\ \Big[\underline{\Xi}(\alpha-\bar{s})(\alpha-\bar{s}-\varsigma_I)+(\alpha-\bar{s})-\bar{d}-\varsigma_2\Big]\Delta\ >\ 0$$

under (22). □

Priors at which she hires. The last claim bounds the standardized known-complementarity cutoff at every prior at which the relationship forms. Here v_A solves $\bar{w}[\phi(v)/v-\bar{\Phi}(v)]=\pi_0$, the standardized cutoff at the known complementarity whose second-period value equals the outside option.

Claim 12 (Priors at which she hires)**.** *For every $\delta>0$ there is a level of ς above which $v_1\leq v_A+\delta$ at every prior at which the relationship forms.*

Proof. Her value from initiating at prior γ is her first-period profit plus $\mathbb{E}[\max\{\Pi_2^*,\pi_0\}]$, and the relationship forms when it is at least $2\pi_0$. Suppose the prior's mean is below the mean of the belief at the retention cutoff by at least $\eta>0$. Under Assumption 3 the retention cutoff in the index then lies above the prior's expected index by at least η/V up to ϵ, so at every conjecture $\hat{e}\leq 2\bar{e}$ the retention cutoff in the signal lies at least $\sigma\eta/(2\bar{e}V)$ standard deviations above the mode. There the density of the signal is at most $\phi(\sigma\eta/(2\bar{e}V))/\sigma$, which vanishes faster than any power of V, so the retention term of the continuation, the saving it brings to her first-period bill, and the option value of learning, which is at most the probability of reaching the retention cutoff times $\Pi_2-\pi_0$ there, are all smaller than any power of V. Her first-period profit is then, up to such terms, her profit at the known complementarity μ_0, which is $\bar{w}[\phi(v_1)/v_1-\bar{\Phi}(v_1)](1+\epsilon)$, strictly decreasing in v_1, so the relationship forms only if this is at least π_0 up to terms that vanish with V, that is only if $v_1\leq v_A+\delta$ for ς large. If instead the prior's mean is within η of that belief's, then $v_1\leq v_A+O(\eta)$ directly, v_1 being a smooth function of the mean. Taking η small gives the claim. □

Proof of Lemma 3 (Retention is a threshold rule)

Assumption 3 holds, and the conjecture $\hat{e}_1>0$ is fixed. The proof shows that her second-period value rises with first-period signal near every signal at which she retains the agent, so it crosses π_0 once.

The family of posteriors. By Lemma 2, the posterior after first-period signal y_1 is the tilt q_b of the modified base at the index $b=\gamma+\hat{e}_1y_1/\sigma^2$, which is increasing in y_1.

Retained posteriors. The agent's payoff is nonnegative, so the wage bill is at least $c(e)$ and

$$\Pi_2(b) \;\le\; \max_{e\ge 0}\{\mu_b\, e - \alpha e^2/2\} \;=\; \frac{(\mu_b)_+^2}{2\alpha},$$

the maximum being attained at $e = (\mu_b)_+/\alpha$. At a retained history $\Pi_2(b) \ge \pi_0 > 0$, so $\mu_b \ge \underline{\mu} = \sqrt{2\alpha\pi_0} > 0$, the bound of the preliminaries. Steps 1 to 3 work at every index b with $\mu_b \ge \underline{\mu}/2$, and Step 4 explains why that range suffices.

Step 1: a lower bound on the slope. Fix b with $\mu_b \ge \underline{\mu}/2$. Since $\mu_b > 0$, a bonus paid far enough above the expected signal implements a small positive effort at a profit, because its cost falls faster than the signal it induces as the cutoff moves out. So the optimal second-period contract at q_b implements an effort $e_2 > 0$, and its wage delivers positive marginal incentive. By Corollary 1, it pays $\bar{w}$ above a cutoff z_2^+ and, when $\underline{\theta} < 0$, below a cutoff z_2^-, with $\kappa_2 > 0$ the level of the score at the cutoffs. When $\underline{\theta} \ge 0$ there is no lower tail, which corresponds to $z_2^- = -\infty$.

Let

$$\xi^+(\theta) := \theta g_\varepsilon(z_2^+ - e_2\theta) - \kappa_2\, \bar{G}_\varepsilon(z_2^+ - e_2\theta)$$

and

$$\xi^-(\theta) := -\theta g_\varepsilon(z_2^- - e_2\theta) - \kappa_2\, G_\varepsilon(z_2^- - e_2\theta),$$

and let $\xi := \xi^+ + \xi^-$. In each, the first term is the incentive the tail delivers in state θ and the second is the probability that it pays, priced at κ_2, so $\xi^\pm(\theta)$ is the incentive the tail delivers in state θ net of its cost. Both terms of ξ^- vanish as $z_2^- \to -\infty$.

Let f_2 be the density of the signal at the belief q_b and effort e_2. Moving the upper cutoff up by dz changes the incentive by $-\partial_e f_2(z_2^+)\,dz$ and the payment probability by $-f_2(z_2^+)\,dz$. The cutoff lies where the likelihood score equals κ_2, so $\partial_e f_2(z_2^+) = \kappa_2 f_2(z_2^+)$, which is nonzero. For b' near b, keep the effort e_2 and the lower cutoff z_2^-, and move the upper cutoff to $z^a(b')$ so that the agent's first-order condition still holds at $q_{b'}$. By (15), the change in belief moves the incentive of the old contract at the rate of its covariance with θ, so the implicit function theorem gives a differentiable $z^a(b')$ with $z^a(b) = z_2^+$ and

$$\left.\frac{dz^a}{db'}\right|_{b'=b} = \frac{1}{\kappa_2 f_2(z_2^+)}\, \mathrm{Cov}_b\left(\theta,\; \theta g_\varepsilon(z_2^+ - e_2\theta) - \theta g_\varepsilon(z_2^- - e_2\theta)\right).$$

The contract so obtained is feasible at $q_{b'}$. Its value is

$$\widehat{\Pi}_b(b') := \mu_{b'}\, e_2 - \bar{w}\, \mathbb{E}_{b'}[\bar{G}_\varepsilon(z^a(b') - e_2\theta) + G_\varepsilon(z_2^- - e_2\theta)],$$

so $\Pi_2(b') \ge \widehat{\Pi}_b(b')$, with equality at $b' = b$. Here b is fixed and b' varies, and $\widehat{\Pi}_b$ is differentiable in b' since expectations move smoothly along the tilt and $z^a(b')$ is differentiable. Differentiating the expected payment in b' gives two terms, the direct effect of the change in belief, by (15), and the effect of moving the cutoff.

Substituting the derivative of the cutoff and the definition of ξ,

$$\begin{aligned}\frac{d}{db'}\,\mathbb{E}_{b'}\big[\bar{G}_\varepsilon(z^a(b')-e_2\theta)+G_\varepsilon(z_2^- - e_2\theta)\big]\Big|_{b'=b}&\\ = \mathrm{Cov}_b\left(\theta,\ \bar{G}_\varepsilon(z_2^+ - e_2\theta)+G_\varepsilon(z_2^- - e_2\theta)\right) - f_2(z_2^+)\,\frac{dz^a}{db'}\Big|_{b'=b}&\\ = -\frac{1}{\kappa_2}\,\mathrm{Cov}_b(\theta,\xi).&\end{aligned}$$

So each unit of incentive restored costs $\bar{w}/\kappa_2$ in expected payment. Since the derivative of $\mu_{b'}e_2$ at $b'=b$ is e_2V_b,

$$\frac{d}{db'}\,\widehat{\Pi}_b(b')\Big|_{b'=b} \;=\; e_2V_b + \frac{\bar{w}}{\kappa_2}\,\mathrm{Cov}_b(\theta,\xi). \tag{23}$$

Step 2: the upper tail. With Gaussian noise $g_\varepsilon'(x) = -(x/\sigma^2)g_\varepsilon(x)$, so

$$\xi^{+\prime}(\theta) = g_\varepsilon(z_2^+ - e_2\theta)\Big[\,(1-\kappa_2e_2) + \frac{e_2\,\theta\,(z_2^+ - e_2\theta)}{\sigma^2}\Big].$$

At the belief q_b, the optimal cutoff and the price of incentives equal, up to a factor $1\pm\epsilon$, their values when the complementarity is known to equal μ_b: $z_2^+ - e_2\mu_b = v_b\sigma(1\pm\epsilon)$ with $v_b := v_{\mu_b} = \alpha\sigma/\mu_b^2 > 0$, and $\kappa_2e_2 = \varkappa(1\pm\epsilon)$ with $\varkappa := \varkappa_{\mu_b} = \bar{w}\phi(v_b)/\sigma$, where v_μ, $\varkappa_\mu$ and the concentration error ϵ are those of the preliminaries. At the posterior, each quantity in the principal's problem, such as the payment probability or the incentive of a contract, is an average $\mathbb{E}_b[h(\theta)]$ of a smooth function h, while at the known complementarity it equals $h(\mu_b)$. By the mean value theorem, $h(\theta) - h(\mu_b) = h'(\tilde{\theta})\,(\theta-\mu_b)$ for some $\tilde{\theta}$ between μ_b and θ, so $|h(\theta)-h(\mu_b)| \le \sup_\Theta|h'|\,|\theta-\mu_b|$ for every θ. Averaging over θ,

$$\big|\mathbb{E}_b[h(\theta)] - h(\mu_b)\big| \;\le\; \mathbb{E}_b\big|h(\theta)-h(\mu_b)\big| \;\le\; \sup_\Theta|h'|\;\mathbb{E}_b|\theta-\mu_b| \;\le\; \sup_\Theta|h'|\;\sqrt{V_b},$$

the last step by the Cauchy–Schwarz inequality, $\mathbb{E}_b|\theta-\mu_b| \le (\mathbb{E}_b(\theta-\mu_b)^2)^{1/2}$. So the optimal cutoff and the price of incentives at the posterior are within a factor $1\pm\epsilon$ of their values at the known complementarity. Set $\sigma^{\mathrm{sc}} := \bar{w}/2$. For $\sigma \ge \sigma^{\mathrm{sc}}$, $\varkappa \le \bar{w}\phi(0)/\sigma \le 2\phi(0) < \frac{4}{5}$, so both terms of the bracket are positive at $\theta=\mu_b$ and $\xi^{+\prime}(\mu_b) \ge \frac{1}{5}\,\phi(v_b)/\sigma\,(1-\epsilon) > 0$. By Taylor's theorem with the remainder in Lagrange form,

$$\xi^+(\theta) = \xi^+(\mu_b) + \xi^{+\prime}(\mu_b)\,(\theta-\mu_b) + R(\theta),$$

with $|R(\theta)| \le \frac{1}{2}\sup_\Theta|\xi^{+\prime\prime}|\,(\theta-\mu_b)^2$. The constant has zero covariance with θ and the linear term contributes $\xi^{+\prime}(\mu_b)V_b$, so

$$\mathrm{Cov}_b(\theta,\xi^+) = \xi^{+\prime}(\mu_b)\,V_b + \mathbb{E}_b\big[(\theta-\mu_b)R(\theta)\big].$$

The last term is at most $\frac{1}{2}\sup_\Theta|\xi^{+\prime\prime}|\,\mathbb{E}_b|\theta-\mu_b|^3$ in absolute value, and $\mathbb{E}_b|\theta-\mu_b|^3 \le C_3'V_b^{3/2}$. Hence

$$\mathrm{Cov}_b(\theta,\xi^+) \;\ge\; \xi^{+\prime}(\mu_b)\,V_b - \tfrac{1}{2}\sup_\Theta|\xi^{+\prime\prime}|\;C_3'\,V_b^{3/2},$$

and $\sup|\xi^{+\prime\prime}|/\xi^{+\prime}(\mu_b)$ is bounded by a constant that depends on v_b only through a compact range, since

$\mu_b \in [\underline{\mu}/2, \bar{\theta}]$ gives $v_b \in [\alpha\sigma/\bar{\theta}^2, 4\alpha\sigma/\underline{\mu}^2]$. Since $V_b \le 1/\varsigma$, the remainder is at most half the main term once $\sqrt{\varsigma} \ge C_3' \sup|\xi^{+\prime\prime}|/\xi^{+\prime}(\mu_b)$. So there exists $\varsigma^R(\alpha,\sigma) < \infty$ such that $\mathrm{Cov}_b(\theta, \xi^+) \ge \frac{1}{2}\xi^{+\prime}(\mu_b)V_b$ for all $\varsigma \ge \varsigma^R(\alpha,\sigma)$.

Step 3: the lower tail, and the sign of the slope. When $\underline{\theta} \ge 0$, $\xi^- = 0$. Let $\underline{\theta} < 0$. The lower cutoff lies far below the expected signal. With Gaussian noise, the likelihood score at the signal y is

$$\ell(y) = \frac{1}{\sigma^2}\,\mathbb{E}_y[\theta\,(y - e_2\theta)],$$

where $\mathbb{E}_y$ is the expectation under the distribution of θ given the signal y, $p_y(\theta) \propto q_b(\theta)\, g_\varepsilon(y - e_2\theta)$. Since $e_2\theta^2 \ge 0$, $\ell(y) \le y\,\mathbb{E}_y[\theta]/\sigma^2$. In particular $\ell(0) < 0$, so the lower tail lies at negative signals, and at a negative signal the score is positive only if $\mathbb{E}_y[\theta] < 0$. As a function of θ, p_y is proportional to $e^{\tau\theta}\tilde{q}_b(\theta)$, with $\tau := e_2 y/\sigma^2$ and $\tilde{q}_b(\theta) \propto q_b(\theta)\, e^{-e_2^2\theta^2/(2\sigma^2)}$. Every member of this family has log-curvature at least ς, so by the argument of (15) its mean increases in τ at the rate of its variance, which is at most $1/\varsigma$ by the Brascamp–Lieb inequality. By (16), the Gaussian factor lowers the mean by at most $\bar{e}^2\theta_{\max}/(\sigma^2\varsigma)$, so the mean of $\tilde{q}_b$ is at least $\underline{\mu}/4$ once $\varsigma \ge 4\bar{e}^2\theta_{\max}/(\sigma^2\underline{\mu})$. So $\mathbb{E}_y[\theta] < 0$ requires $\tau < -\varsigma\underline{\mu}/4$, that is, $y < -\varsigma\underline{\mu}\sigma^2/(4e_2)$, and since $e_2 \le \bar{e}$,

$$z_2^- \;<\; -\frac{\varsigma\,\underline{\mu}\,\sigma^2}{4\bar{e}}.$$

Every term of $\xi^{-\prime}$ contains g_ε or g_ε' evaluated at $z_2^- - e_2\theta \le z_2^- + \bar{e}|\theta|$, where the Gaussian density and its derivative are smaller than any power of $1/\varsigma$. So $\sup_\Theta|\xi^{-\prime}|$ is smaller than any power of $1/\varsigma$, and by (16), choosing $\varsigma^R(\alpha,\sigma)$ larger if necessary,

$$|\,\mathrm{Cov}_b(\theta, \xi^-)| \;\le\; V_b \sup_\Theta |\xi^{-\prime}| \;\le\; \tfrac{1}{4}\,\xi^{+\prime}(\mu_b)\,V_b$$

for all $\varsigma \ge \varsigma^R(\alpha,\sigma)$, since $\xi^{+\prime}(\mu_b)$ is bounded below by a positive constant on the compact range of v_b. With Step 2 and (23), for all $\varsigma \ge \varsigma^R(\alpha,\sigma)$,

$$\frac{d}{db'}\,\widehat{\Pi}_b(b')\Big|_{b'=b} \;\ge\; V_b\Big[e_2 + \frac{\bar{w}}{4\kappa_2}\,\xi^{+\prime}(\mu_b)\Big] \;>\; 0.$$

In sum, $\varsigma^R(\alpha,\sigma) < \infty$ is chosen so that four requirements hold for all $\varsigma \ge \varsigma^R(\alpha,\sigma)$: ϵ is small enough for the bounds on $\kappa_2 e_2$ and $\xi^{+\prime}(\mu_b)$ in Step 2, the Taylor remainder is at most half the main term, $\varsigma \ge 4\bar{e}^2\theta_{\max}/(\sigma^2\underline{\mu})$, and the lower tail is at most a quarter of the main term. It depends on α and σ, and on the other primitives, which are fixed, but not on the prior or the conjecture.

Step 4: from the slope to the cutoff. Π_2 is continuous in b by the maximum theorem. Since μ_b is strictly increasing in b, the indices with $\mu_b \ge \underline{\mu}/2$ form an interval $[b_0, +\infty)$. At every b in it, the derivative of $\widehat{\Pi}_b$ at $b' = b$ is positive by Steps 1–3, so for b' slightly above b,

$$\Pi_2(b') \;\ge\; \widehat{\Pi}_b(b') \;>\; \widehat{\Pi}_b(b) \;=\; \Pi_2(b).$$

So Π_2 rises just to the right of every point of $[b_0, +\infty)$, and a continuous function with this property is

strictly increasing there. Indeed, if $\Pi_2(y) \le \Pi_2(x)$ for some $x < y$ in the interval, the maximum of Π_2 on $[x, y]$, which exists by continuity, is attained at some point below y, and Π_2 rises just to the right of that point, a contradiction. In particular, once Π_2 reaches π_0 it stays above it for every larger index. Below b_0, $\mu_b < \underline{\mu}/2$, so by the bound in the paragraph on retained posteriors, $\Pi_2(b) \le (\underline{\mu}/2)^2/(2\alpha) = \pi_0/4 < \pi_0$. So Π_2 can reach π_0 only on $[b_0, +\infty)$, where it is strictly increasing, and it crosses π_0 at most once as b increases, from below. Since b is increasing in y_1, retention is the cutoff rule in the signal, with $\Pi_2^*(\cdot, \hat{e}_1)$ strictly increasing on a neighborhood of the cutoff. □

Regularity of the continuation value

Write $\Pi_2(y_1, \hat{e})$ for $\Pi_2(p(\cdot \mid y_1, \hat{e}))$, and let

$$\Psi_1(\tilde{e}) := \int \max\{\Pi_2(y_1, \tilde{e}), \pi_0\} f_1(y_1 \mid \tilde{e})\, dy_1$$

and

$$G(\tilde{e}) := \tilde{e}\,\mu_0 + \Psi_1(\tilde{e}),$$

the principal's expected continuation, and her benefit, when effort and conjecture both equal $\tilde{e}$. At conjecture zero, the signal is uninformative and the posterior is the prior after every signal. At the one prior whose second-period value equals π_0 exactly, the principal is then indifferent between retaining and replacing after every signal, so her retention rule at conjecture zero is not pinned down. For Ψ_1 to be continuous there, it is taken to be the limit of the threshold rules at positive conjectures, which retain exactly when the signal exceeds its mode.

The claim below bounds the derivatives of Ψ_1, which enter her reduced problem through G: the slope bound is compared with μ_0 in (U2) and (S), and the curvature bound enters the concavity condition of Claim 11.

Claim 13 (Regularity of the continuation)**.** *$\Psi_1 \in C^1[0, 2\bar{e}]$ with $\Psi_1'(0) = 0$ and $\bar{\Psi}_1' := \sup|\Psi_1'| = O(1/(\alpha\sigma^2))$, and $\Psi_1 \in C^2(0, 2\bar{e}]$ with $\sup_{[\underline{e}, 2\bar{e}]}|\Psi_1''| < \infty$ for every $\underline{e} > 0$. Under Assumption 3, moreover, $\sup_{[\underline{e}, 2\bar{e}]}|\Psi_1''| = O(V)$ for every $\underline{e} > 0$, with V the prior variance.*

Proof. The second period is smooth in the history. Differentiating the posterior under the integral, $\partial_{\hat{e}} p(\theta \mid y_1, \hat{e}) = p[-\theta q(y_1 - \hat{e}\theta) + \mathbb{E}_p \theta q(y_1 - \hat{e}\theta)]$, so $\|\partial_{\hat{e}} p\|_{L^1} \le 2\theta_{\max}\, \mathbb{E}_p|q(y_1 - \hat{e}\theta)|$, which is integrable against f_1 and $|\partial_e f_1|$ by the Fisher-information argument in the proof of Claim 5. At a retained history the belief has positive mean, since $\pi_0 \le \Pi_2 \le e_2 \mathbb{E}_p\theta$, so the principal's second-period problem is Claim 11 with $G(\tilde{e}) = \tilde{e}\,\mathbb{E}_p\theta$ and $\tau \equiv 0$. Her first-order condition in the target puts her level at

$$\kappa_2^* = \frac{\alpha - \bar{w}\,\partial_{e_2} I}{\mathbb{E}_p\theta - c'(e_2)} \;\ge\; \frac{\alpha - \bar{w}J}{\bar{\theta}},$$

with $J := \theta_{\max}^2 \int |\partial_{e_2}^2 f_2|$, so for $\alpha > \bar{w}J$ her optimum is unique, lies in the tail family, and is C^1 in $(y_1, \hat{e})$ by the implicit function theorem. By the envelope theorem and Claim 9, the derivative of Π_2 when the belief moves in a direction h with $\int h = 0$ is $e_2^* \int \theta h - \bar{w} \int h\,[P_{A_2} - D_{A_2}/\kappa_2^*]$, which involves only first derivatives of

the maximizer, so Π_2 is C^2 and Lipschitz in the belief with constant $L_\Pi = O(1/\alpha)$: the payment probability is at most $\mathcal{I}/\kappa_2^{*2}$, with $\mathcal{I}$ the Fisher information of the signal in effort, since the score is at least κ_2^* on the payment set. The retention cutoff $\underline{y}(\hat{e})$ solves $\Pi_2(y_1, \hat{e}) = \pi_0$, and $\partial_{y_1}\Pi_2 > 0$ at the crossing by the rate bound in the proof of Lemma 3, so $\underline{y}$ is C^2 wherever finite.

First derivative. The boundary terms of Ψ_1' cancel because $\Pi_2 = \pi_0$ at the bar, so, using $\int \partial_e f_1 = 0$,

$$\Psi_1'(\tilde{e}) = \int_{\underline{y}}^{\infty} \partial_{\hat{e}}\Pi_2 \, f_1 \, dy_1 + \int_{\underline{y}}^{\infty} (\Pi_2 - \pi_0) \, \partial_e f_1 \, dy_1. \tag{24}$$

Here $|\partial_{\hat{e}}\Pi_2| \le 2L_\Pi\theta_{\max}(|y_1| + 2\bar{e}\theta_{\max})/\sigma^2$ and $\Pi_2 - \pi_0 \le \bar{\theta}^2/(2\alpha)$ on retained histories, and L_Π and $\bar{e}$ are $O(1/(\alpha\sigma))$, $\int |y_1| f_1 = O(\sigma)$ and $\nu_1 = O(1/\sigma)$, so both integrals are $O(1/(\alpha\sigma^2))$. As $\tilde{e} \downarrow 0$ the retained region tends to $\mathbb{R}$ or to $\emptyset$, because the posterior at bounded signals tends to the prior, whose value differs from π_0 except at the indifferent prior, where the convention above applies. And $\int \partial_{\hat{e}}\Pi_2(y_1, 0)\, g_\varepsilon \, dy_1 = 0$, because $\partial_{\hat{e}} p(\theta \,|\, y_1, 0) = -g_\theta(\theta)(\theta - \mu_0)q(y_1)$ and $\int q\, g_\varepsilon = 0$. So $\Psi_1'(0) = 0$.

Second derivative. Differentiating (24), the interior terms are dominated by the bounds above, and the boundary term is $(\partial_{\hat{e}}\Pi_2)^2 f_1/\partial_{y_1}\Pi_2$ at the bar. There $\partial_{y_1}\Pi_2 = (\hat{e}/\sigma^2)\,\partial_b\Pi_2 \ge (\hat{e}/\sigma^2)\, e_2 \mathrm{Var}_p(\theta)$ by Step 2 of the proof of Lemma 3, with $e_2 \ge \pi_0/\bar{\theta}$, and the posterior is, up to bounded factors, the tilt of the log-concave prior with parameter $\tau \le \hat{e}(|\underline{y}| + 2\bar{e}\theta_{\max})/\sigma^2$, whose variance is at least a constant times $\min\{1, \tau^{-2}\}$. On $\hat{e} \ge \underline{e}$ the boundary term is therefore a polynomial in $|\underline{y}|$ times $f_1(\underline{y})$, divided by $\underline{e}$, which is bounded because f_1 has Gaussian tails.

Order V under the specification. Under Assumption 3, Π_2 depends on the history through the posterior, whose mean moves at rate $V_b \le 1/\varsigma$ per unit of index by the tilt identities of Part (A) of the proof of Theorem 3, so the index-derivatives of Π_2, and with them $\partial_{\hat{e}}\Pi_2$ and $\partial_{\hat{e}}^2\Pi_2$, are of order V. Differentiating (24), two of the interior terms carry these derivatives and are of order V. The third, $\int_{\underline{y}}^{\infty}(\Pi_2 - \pi_0)\,\partial_{\hat{e}}^2 f_1 \, dy_1$, equals $-\int_{\underline{y}}^{\infty} \partial_{y_1}\Pi_2 \, \partial_{\hat{e}}^2 F_1 \, dy_1$ after an integration by parts, with F_1 the distribution function of first-period signal, because $\Pi_2 - \pi_0$ vanishes at the bar and $\partial_e^2 F_1$ vanishes at $+\infty$. Since $\partial_{y_1}\Pi_2 = (\hat{e}/\sigma^2)\,\partial_b\Pi_2$ is of order V and $\partial_e^2 F_1$ is integrable, it is of order V too. In the boundary term the numerator $(\partial_{\hat{e}}\Pi_2)^2$ is of order V^2 and the denominator $\partial_{y_1}\Pi_2 \ge (\hat{e}/\sigma^2)\, e_2 \mathrm{Var}_p(\theta)$ is at least of order V on retained histories, so it is of order V. Hence $\sup_{[\underline{e}, 2\bar{e}]} |\Psi_1''| = O(V)$. □

Proof of Proposition 2 (Benchmark equilibrium)

The benchmark.

Step 1: the agent's problem.

The benchmark agent's objective is $W(A; e) - \alpha e^2/2$, with no term in the conjecture. Its second derivative is at most $C_1 - \alpha < 0$ for $\alpha > C_1$, where $C_1 := \bar{w}\theta_{\max}^2 \nu_2 = 4\phi(1)\bar{w}\theta_{\max}^2/\sigma^2$ bounds $|\partial_e M_w|$; so his best response $e^B(A)$ is unique, lies in $[0, 2\bar{e}]$, and is continuous in the contract. The corner binds exactly when A *delivers no incentive*, that is, when $M_w(A; 0) \le 0$.

Step 2: existence, and the case $\mu_0 \le 0$.

Her payoff is $\Pi^B(A) = G(e^B(A)) - W(A; e^B(A))$, with $G = \tilde{e}\mu_0 + \Psi_1$ of Claim 13: the second period is

identical given the posterior. Π^B is continuous on the compact set of contracts, so a maximizer exists and the equilibrium exists. If $\mu_0 \le 0$, then $\Pi_2(g_\theta) \le \max_e[e\mu_0 - c(e)] = 0$, so $\Psi_1(0) = \pi_0$ and every contract yields at most $\pi_0 + 2\bar{\Psi}_1'\bar{e} < 2\pi_0$: the relationship is not formed, and the outcome is unique.

Step 3: for $\mu_0 > 0$, contracts that deliver no incentive are never optimal.

A contract that delivers no incentive implements zero effort and yields at most $\Psi_1(0)$. Let $A_\circ$ pay $\bar{w}$ above the cutoff $2v_1\sigma$. At the agent's response $e_\circ$, the incentive is $\alpha e_\circ = \bar{w}\mu_0\phi(2v_1)/\sigma\,(1 \pm \epsilon)$, and the bill is $\bar{w}\bar{\Phi}(2v_1)(1+\epsilon) \le (\bar{w}\phi(2v_1)/(2v_1))(1+\epsilon)$ by (19). Since $\alpha\sigma = v_1\mu_0^2$, the bill is at most $e_\circ\mu_0(v_1/(2v_1))(1+2\epsilon) = \frac{1}{2}e_\circ\mu_0(1+2\epsilon)$, and the first-period profit $e_\circ\mu_0$ minus the bill is at least $e_\circ\mu_0/3$. With $\Psi_1(e_\circ) - \Psi_1(0) \ge -\bar{\Psi}_1'e_\circ \ge -e_\circ\mu_0/12$ once $\bar{\Psi}_1' \le \mu_0/12$, $\Pi^B(A_\circ) \ge \Psi_1(0) + e_\circ\mu_0/4 > \Psi_1(0)$. So every optimal contract A^* implements $e^* := e^B(A^*) > 0$, and since $\Pi^B(A^*) \le \Psi_1(0) + (\mu_0 + \bar{\Psi}_1')e^*$,

$$e^* \;\ge\; \underline{e} := \frac{e_\circ\,\mu_0}{4(\mu_0 + \bar{\Psi}_1')} \;>\; 0.$$

Step 4: reduction to the target.

By the preliminaries the level sets of the score are upper tails, and the level delivering a given incentive is unique (Claim 9). For $\tilde{e} \in [\underline{e}, 2\bar{e}]$, let $A(\tilde{e})$ be the upper tail whose incentive at $\tilde{e}$ is $c'(\tilde{e}) = \alpha\tilde{e}$. The agent's first-order condition at $A(\tilde{e})$ holds at $\tilde{e}$ and his objective is strictly concave, so $e^B(A(\tilde{e})) = \tilde{e}$. An optimal A^* is the cheapest contract for its target, by Theorem 2: replacing it by $A(e^*)$ leaves the agent's response unchanged and lowers the bill. So every optimal contract is in the family, and her value over all contracts is the maximum of $\tilde{e} \mapsto \Pi^B(A(\tilde{e}))$ over $[\underline{e}, 2\bar{e}]$. Its first-order condition in the cutoff, with $\partial_e W = M_w = c'$ at the agent's optimum, puts the cutoff at the level

$$\kappa^* = \frac{\alpha - \bar{w}\,\partial_e I(A^*; e^*)}{G'(e^*) - c'(e^*)} \;>\; 0.$$

Step 5: every optimal target has a large wage incentive.

Existence holds for $\alpha > \alpha^B := C_1$, a level of order σ^{-2}, by Steps 1–4. The benchmark needs no more than this, since its agent faces no continuation term, and $C_1 \le \bar{M}_2' < \alpha^{\mathrm{FOA}}$ by the proof of Proposition 1, so the model's first-order-approach level implies it. Uniqueness is a statement about her reduced problem in the target, $\tilde{e} \mapsto \Pi^B(A(\tilde{e}))$ on $[\underline{e}, 2\bar{e}]$ (Step 4), which is the problem of Claim 11 with $G(\tilde{e}) = \tilde{e}\mu_0 + \Psi_1(\tilde{e})$, $r = c'$ and $\tau \equiv 0$. Two quantities of the preliminaries organize the argument: the standardized known-complementarity cutoff v_1, whose level is α/μ_0, and the wage incentive I_1 it delivers.

By Claim 10(ii),

$$\frac{d}{d\tilde{e}}\Pi^B(A(\tilde{e})) = \mu_0 + \Psi_1'(\tilde{e}) - \alpha\tilde{e} - \frac{\alpha - \bar{w}\partial_e I}{\kappa(\tilde{e})},$$

and $\partial_e I \ge 0$ here, because on $[\underline{e}, 2\bar{e}]$ the cutoff lies above $\tilde{e}\theta$ for every $\theta \in \Theta$ (it sits $v\sigma$ above the expected signal with $v \ge v_1$, and $2\bar{e}\theta_{\max} \le \epsilon\sigma$). At a target whose contract has incentive $\alpha\tilde{e} \le \frac{4}{5}I_1$, the cutoff is at a standardized position v with $\phi(v) \le \frac{4}{5}\phi(v_1)(1+\epsilon)$, that is $v \ge (v_1^2 + 2\log\frac{5}{4})^{1/2}$ up to ϵ, and $\kappa(\tilde{e}) = \mu_0 v/\sigma\,(1 \pm \epsilon)$,

so $\alpha/\kappa(\tilde{e}) \le \mu_0(v_1/v)(1+\epsilon) \le \mu_0(1-c_S)$ with

$$c_S := 1 - \frac{v_1}{(v_1^2 + 2\log\frac{5}{4})^{1/2}} - \epsilon.$$

The derivative is then at least $\mu_0 c_S - \bar{\Psi}_1' - \frac{4}{5}I_1$, which is positive under

$$\bar{\Psi}_1' + \tfrac{4}{5}I_1 \;<\; \mu_0\, c_S\,. \tag{S_B}$$

This is a level on $\sigma/\bar{w}$: $\frac{4}{5}I_1/\mu_0 = \frac{4}{5}\bar{w}\phi(v_1)/\sigma$ and $\bar{\Psi}_1' = O(1/(\alpha\sigma^2))$ by Claim 13, so it holds for $\sigma/\bar{w} \ge \frac{4}{5}\phi(v_1)/c_S$ up to the last term. Since $\phi(v)/c_S(v)$ is bounded over $v \ge 0$, $c_S(0)$ being close to one and $\phi(v)$ decaying faster than $c_S(v) \asymp 1/v^2$, one level $\sigma^B := \frac{4}{5}\bar{w}\sup_{v\ge 0}\phi(v)/c_S(v)$ serves for every α, which is why σ^B does not depend on α. So her profit is increasing at every target with $\alpha\tilde{e} \le \frac{4}{5}I_1$, and every optimal target lies in

$$\mathcal{K}_1^B := \{\tilde{e} \in [\underline{e}, e^\dagger] : \alpha\tilde{e} \ge \tfrac{4}{5}I_1\},$$

an interval, where $e^\dagger$ is the target whose cheapest contract has level $\alpha/(4\mu_0)$: beyond $e^\dagger$, $\alpha/\kappa \ge 4\mu_0(1-\epsilon)$ exceeds $\mu_0 + \bar{\Psi}_1'$, so the derivative is negative there.

Step 6: concavity on the band.

On $\mathcal{K}_1^B$ the levels lie in $[\kappa(e^\dagger), \kappa(\tilde{e}_1)]$, so $\kappa_{\mathcal{K}} \ge \frac{\alpha}{4\mu_0}(1-\epsilon)$, and Claim 11 applies with $\bar{s} = L = 0$ as soon as $\alpha > \varsigma_I$ and

$$\underline{\Xi}\,\alpha\,(\alpha - \varsigma_I) + \alpha \;>\; \sup_{\mathcal{K}_1^B} |\Psi_1''| + \varsigma_2.$$

Under Gaussian noise and a concentrated prior, these terms are explicit up to terms of order V. At a point-mass belief the score is linear in the signal, so at a fixed level the standardized position of the cutoff, and with it the incentive, does not move with the target, and $\varsigma_I = O(V)$. For the same reason $\bar{w}\partial_e I/\kappa$ equals the incentive $\mathcal{I}$ up to terms of order V, and along the family the incentive is the required one, $r = c'$, so $\varsigma_2 = \alpha + O(V)$. The two linear terms therefore cancel up to order V, and the condition reads

$$\underline{\Xi}\,\alpha^2 \;>\; \sup_{\mathcal{K}_1^B} |\Psi_1''| + O(V).$$

Its left side is at least $\alpha^2\sigma^2/(\bar{w}\bar{\theta}^2 \sup_v v^3\phi(v))$, by the expression of $\underline{\Xi}$ in the proof of Claim 11, uniformly in the prior, and $\sup|\Psi_1''| = O(V)$ by Claim 13. So the condition holds, at every v_1, for ς above a level depending on α and σ. Steps 5 and 6 both use the concentration of the prior, which is why uniqueness needs $\varsigma \ge \varsigma^B(\alpha,\sigma)$. Together with the level on σ in Step 5, this defines σ^B and $\varsigma^B(\alpha,\sigma)$.

Step 7: conclusion.

$\tilde{e} \mapsto \Pi^B(A(\tilde{e}))$ is strictly concave on $\mathcal{K}_1^B$, which contains every optimal target, so two optimal contracts have the same target, and they coincide up to null sets by Theorem 2. The second period is unique at every history, up to her decision at the retention cutoff, a single signal value. At the one prior at which her value from hiring equals $2\pi_0$, hiring and not hiring are both optimal, and the outcome is unique up to that choice. Claims 20 and 21, in the proof of Theorem 3, extend Steps 5 and 6 to the model, where the continuation

adds a slope and a curvature to her reduced problem. □

Proofs of Theorem 3 (Over- and under-experimentation) and Corollary 2

Overview. Part (0) shows that the contract the principal offers pins down the agent's effort (Lemma 4). Part (A) records how the realized index moves with effort. Part (B) shows that the wage effect changes sign at most once as the posterior improves (Lemma 5). Part (C) shows that the manipulation term T is positive at pessimistic priors and negative at optimistic priors when $\alpha < \bar{\alpha}$. Part (D) locates the principal's optimal targets and the priors at which her problem is concave and at which she hires. Part (E) turns the sign of T into the sign of $e^* - e^B$ (Lemma 9), and Part (F) assembles the cases. The levels at which σ and ς are taken large are collected at the end.

Setup. Fix the conjecture $\hat{e}_1$. By Lemma 2, the posterior after first-period signal y_1 is the tilt $q_b \propto e^{b\theta}\tilde{g}_0$ of the modified base $\tilde{g}_0$ at the index $b = \gamma + \omega y_1$, where $\omega := \hat{e}_1/\sigma^2$. Its precision is at least ς. Write $\Pi_2(q)$ for the principal's second-period value at belief q, so that $\Pi_2^*(y_1, \hat{e}_1) = \Pi_2(q_b)$. By Lemma 3, retention is the cutoff $\{b \geq \bar{b}\}$ with $\Pi_2(q_{\bar{b}}) = \pi_0$. The theorem's range is $\pi_0 \in (\pi^-, \pi^+) = (\Pi_2(\delta_{\underline{\theta}}), \Pi_2(\delta_{\bar{\theta}}))$, which is $(0, \pi^+)$ when $\underline{\theta} \leq 0$. In the second period at belief q the optimal wage is $\bar{w}\mathbf{1}\{y \geq z^+\} + \bar{w}\mathbf{1}\{y \leq z^-\}$, the lower tail present iff $\mu_-(q) > 0$, where $\mu_-(q)$ is the mass q places on $\theta < 0$. Retained posteriors have mean at least $\underline{\mu} = \sqrt{2\alpha\pi_0}$, the bound of the preliminaries, so $\mu_-(q_b) \leq C_0 e^{-\varsigma\underline{\mu}^2/2}$. Throughout the proof, $O(x)$ denotes a quantity bounded in absolute value by a constant times x, where the constant depends on the primitives but not on the prior, the index or the conjecture. Every quantity below that comes from the lower tail is a bounded function times μ_-, written $O(\mu_-)$. These terms vanish exponentially fast as ς grows, so each requirement below that ς be large enough also makes them negligible.

Notation. The proofs use the following objects, all built from the conjecture $\hat{e}_1$, the noise σ and the prior γ.

- *The index and its scales.* $b := \gamma + \omega y_1$ is the realized posterior index, with weight $\omega := \hat{e}_1/\sigma^2$ on the signal. The index has standard deviation $s := \hat{e}_1/\sigma = \omega\sigma$ given the state, and $\lambda := 1/s^2 = \sigma^2/\hat{e}_1^2$ is the inverse informativeness of a period's signal. The three are one quantity, the signal-to-noise ratio $\hat{e}_1/\sigma$, in different powers.
- *Beliefs.* $q_b \propto e^{b\theta}\tilde{g}_0$ is the posterior after index b, a tilt of the modified base $\tilde{g}_0$. $L_\gamma(b)$ is the density of the index under prior γ. For a density q on Θ, $q_t \propto e^{t\theta}q$ are its tilts, with mean m_t, variance V_t and cumulants $\kappa_k(t)$.
- *Cutoffs.* $\bar{b}$ is the retention margin, $\Pi_2(q_{\bar{b}}) = \pi_0$. $z^\pm$ are the second-period bonus cutoffs and $u_\theta^\pm := (\theta e - z^\pm)/\sigma$ their standardized positions in state θ. At a known complementarity μ, $v := (z^+ - e\mu)/\sigma$ is the bar in units of the noise, and $\varkappa := \bar{w}\phi(v)/\sigma$ is the bonus yield, the marginal incentive the cap buys at the bar.
- *Forces.* $T(\gamma)$ is the continuation incentive at prior γ, split into the retention force T_R and the wage force T_W. $\beta(b) := -\mathbb{E}_{q_b}[\theta\,\partial_b u_2^*]$ is the wage effect at index b, positive when better news lowers the agent's

rent.

- *Levels.* σ_0 collects the noise levels that do not depend on the curvature, and $\underline{K}$, $\overline{K}$ are the constants in the curvature floor $\underline{\alpha}(\sigma) = \underline{K}\sigma^{-2}$ and the ceiling $\bar{\alpha}(\sigma)$ of order $\overline{K}\sigma^{-3/2}$, as defined with the levels at the start of the proof of Theorem 3.

Part (0): a unique implemented effort

Parts (D) and (E) use that the contract the principal offers pins down the agent's effort. This part establishes it.

Lemma 4 (A unique implemented effort)**.** *There exist σ^U, $\alpha^U(\gamma,\sigma) \geq \alpha^{\mathrm{FOA}}(\sigma)$ at each prior γ, and $\varsigma^U(\alpha,\sigma)$, all in $(0,+\infty)$, such that for all $\sigma \geq \sigma^U$, $\alpha > \alpha^U(\gamma,\sigma)$ and $\varsigma \geq \varsigma^U(\alpha,\sigma)$, in every equilibrium the contract offered by the principal implements a unique consistent effort.*

Proof.
At a contract whose wage incentive is at least the level I_u of (27), the consistent effort is unique (Step 1); a contract with less wage incentive is dominated (Step 2); so the offered contract determines the effort (Step 3); and the conditions used hold for $\sigma \geq \sigma^U$ and $\alpha > \alpha^U$ (Step 4).

The case $\mu_0 \leq 0$. If $\mu_0 \leq 0$ then $\Psi_1(0) = \pi_0$ and every contract yields at most $\pi_0 + 2\bar{\Psi}_1'\bar{e} < 2\pi_0$ under any consistent effort, so no equilibrium forms the relationship and there is nothing to prove. From here on $\mu_0 > 0$.

The continuation and its derivatives.

Write the two terms of (8) as $T = T_R + T_W$: the retention term $T_R(e,\hat{e}) = \int \theta\, u_2^*(\theta; p_{\tau^*})\, g_\varepsilon(\underline{y}(\hat{e}) - e\theta)\, g_\theta(\theta)d\theta$, where p_{τ^*} is the belief at the retention cutoff, the same for every conjecture, and the interior wage term. Neither depends on the contract. Three quantities describe them:

$$
\begin{aligned}
t_R &:= \mathbb{E}_{g_\theta}[|\theta|\, u_2^*(\theta; p_{\tau^*})]\frac{\phi(0)}{\sigma},\\
t_- &:= \sup_{e,\hat{e}} \big(-T(e,\hat{e})\big)_+,\\
C_* &:= \sup\{|\partial_e T|,\ |\partial_{\hat{e}} T_W|,\ |\partial_{\hat{e}}^2 T_W|,\ |\partial_e \partial_{\hat{e}} T|,\ |(\Psi_1'')_{\mathrm{int}}|\},
\end{aligned}
$$

the retention scale, the size of the negative part of T, and one bound on every derivative of the continuation in which no power of $1/\hat{e}$ appears, with $(\Psi_1'')_{\mathrm{int}}$ the interior part of Ψ_1'' in the proof of Claim 13. The retention term is at most its scale: the density at the retention cutoff is at most the peak density, $|T_R| \leq t_R(1+\epsilon)$. Its slope in the conjecture carries a factor $1/\hat{e}$, because the retention cutoff $\underline{y}(\hat{e}) = \sigma^2\tau^*/\hat{e}$ moves at rate $-\underline{y}/\hat{e}$ while the rent at it does not move: $\partial_{\hat{e}} T_R = (\underline{y}/(\hat{e}\sigma^2)) \int \theta u_2^*(\underline{y} - e\theta) g_\varepsilon(\underline{y} - e\theta) g_\theta$, and $\sup_x\, x(x + 2\bar{e}\theta_{\max})\phi(x - 2\bar{e}\theta_{\max})/\sigma^2$ over standardized x gives

$$
|\partial_{\hat{e}} T_R(e,\hat{e})| \ \leq\ \frac{K_g\, t_R}{\hat{e}}, \tag{25}
$$

where $K_g := 2\phi(\sqrt{2})(1+\epsilon)/\phi(0)$. $|\partial_{\hat{e}} T_W| \leq C_*$ with no power of $1/\hat{e}$, because the rent's sensitivity to the signal contains a factor $\hat{e}$ that cancels the one in $\underline{y}'$. Finally, the retention scale against the incentive at the

known-complementarity cutoff: the rent at the retention cutoff is at most $\bar{w}\bar{\Phi}(v_2)$ with v_2 its standardized second-period cutoff, and $v_2 \geq v_1(1-\epsilon)$, since the belief at the retention cutoff is more pessimistic than the prior whenever the prior is retained, and within $o(1)$ of it at initiation as ς grows (Claim 12); so, the Mills ratio $\bar{\Phi}/\phi$ being decreasing,

$$\frac{t_R}{I_1} \;\leq\; \rho(v_1)(1+\epsilon), \tag{26}$$

where $\rho(v) := \bar{\Phi}(v)\phi(0)/\phi(v) \leq \frac{1}{2}$.

The condition (U). Steps 1–3 hold under the condition (U) that the continuation be a small correction to the wage at the known-complementarity cutoff. The scale for "small" is the slack c_w by which a bonus paid just beyond the known-complementarity cutoff beats any contract that relies on the continuation, and the tolerance is one tenth of it, $\delta := c_w/10$. With $m := \frac{1}{2}(I_u + I_1)$ and I_u defined in (27),

$$c_w := 1 - \frac{v_1}{v(m)}(1+\epsilon) \;\geq\; 1 - \frac{v_1}{(v_1^2 + 2\log\frac{10}{9})^{1/2}} - \epsilon.$$

The condition (U) has four parts:

(U1) $2C_*/\alpha \leq \delta$: the continuation's curvature is small next to the cost's.

(U2) $(I_1 + \bar{\Psi}_1')/\mu_0 \leq \delta$: the incentive at the known-complementarity cutoff and her continuation's slope are small next to μ_0.

(U3) $50\,C_1 \leq \alpha$: the wage incentive moves little with effort next to the cost's curvature.

(U4) $I_u \leq \frac{4}{5}I_1$: the wage level above which effort is unique lies below the incentive at the known-complementarity cutoff.

Step 4 shows that each holds for $\sigma \geq \sigma^U$ and $\alpha > \alpha^U$.

Step 1: uniqueness at contracts with enough wage. Fix a contract A with $I := M_w(A;0) > 0$, write $F(e,\hat{e}) := M_w(A;e) + T(e,\hat{e}) - c'(e)$ for the left side of the agent's first-order condition (10), and let $h(\hat{e}) := F(\hat{e},\hat{e})$, whose zeros are the consistent efforts. At a zero, $h'(\hat{e}) = \partial_e F + \partial_{\hat{e}} T$, and by (25), $|\partial_{\hat{e}} T| \leq K_g t_R/\hat{e} + C_*$, while $-\partial_e F \geq \alpha - C_1 - C_*$. Every zero satisfies $c'(\hat{e}) = M_w(A;\hat{e}) + T(\hat{e},\hat{e}) \geq I - C_1\hat{e} - t_-$, hence $\hat{e} \geq (I - t_-)/(\alpha + C_1)$. So $h' < 0$ at every zero, that is, every zero is a downcrossing, as soon as

$$I \;\geq\; I_u := (1+2\epsilon)\Big[t_- + K_g\, t_R\, \frac{\alpha + C_1}{\alpha - C_1 - 2C_*}\Big], \tag{27}$$

the factor $1 + 2\epsilon$ leaving room for the relative change of an upper-tail contract's incentive with effort (the preliminaries of Appendix B), which Claim 20 uses. Since $h(0) \geq I - t_- > 0$ and $h < 0$ for large $\hat{e}$, a continuous function all of whose zeros are downcrossings has exactly one: $\mathcal{E}(A)$ is a single point.

Step 2: contracts with little wage are dominated. A first-period contract is a level set of the score (Theorem 2), and the wage incentive of a level-set contract is nondecreasing in effort: $\partial_e M_w = \bar{w}\int_A \mathbb{E}[\theta^2 g_\varepsilon''(y - e\theta)]dy$, and over an upper tail $\{y \geq z^+\}$ the inner integral is $-g_\varepsilon'(z^+ - e\theta) \geq 0$ since the cutoff lies beyond the mode, and symmetrically over a lower tail. So the incentive a contract delivers at a consistent effort is at

least its incentive at zero effort, and at most that plus C_1 times the effort. The comparison below is made in delivered incentive, which is where the C_1 terms would otherwise multiply the slack c_w.

Let A' be a level-set contract with $I' := M_w(A'; 0) < I_u$, let $e' \in \mathcal{E}(A')$ be any consistent effort at it, and let $J' := M_w(A'; e') \in [I', I' + C_1 e']$ be the incentive it delivers there. Let $m := \frac{1}{2}(I_u + I_1)$, so that $I_u < m \le \frac{9}{10} I_1$ by (U4), and let A_m be the upper-tail bonus that delivers incentive exactly m at its own consistent effort e_m: the delivered incentive of the bonus with cutoff z at its consistent effort is continuous and decreasing in z, so such a bonus exists. Its incentive at zero effort is at least $m - C_1 e_m \ge I_u$, so e_m is unique (Step 1): the continuation delivers at most $|T| \le t_R(1+\epsilon) + O(V) I_1 \le I_1$, so $(\alpha - C_1) e_m \le m + I_1 \le 2 I_1$ and $C_1 e_m \le 2 C_1 I_1 / (\alpha - C_1) \le I_1/20 \le m - I_u$ by (U3). We show $\Pi(A_m, e_m) > \Pi(A', e')$.

Efforts. Subtracting the two first-order conditions,

$$c'(e_m) - c'(e') = (m - J') + \big[T(e_m, e_m) - T(e', e')\big].$$

The diagonal continuation $e \mapsto T(e, e)$ has derivative $\partial_e T + \partial_{\hat{e}} T_R + \partial_{\hat{e}} T_W \ge -2C_*$, because the retention term is nondecreasing in the conjecture: the retention cutoff moves toward the mode as the conjecture rises, and the rent there does not move. And $m - J' \ge m - I_u - C_1 e' \ge \frac{1}{4}(I_1 - I_u) > 0$, since likewise $(\alpha - C_1) e' \le I_u + I_1 \le 2 I_1$ and $C_1 e' \le I_1/20 \le \frac{1}{4}(I_1 - I_u)$ by (U3) and (U4). If $e' \ge e_m$, then $c'(e') - c'(e_m) \ge \alpha(e' - e_m) \ge 2C_*(e' - e_m)$ while the right side is at most $-(m - J') + 2C_*(e' - e_m)$, a contradiction. So $e_m > e'$, and

$$(\alpha + 2C_*)(e_m - e') \;\ge\; m - J'. \tag{28}$$

Wage bills. Write $C(J; e)$ for the bill of the cheapest contract delivering wage incentive J at effort e, the level-set contract of Theorem 2. Any contract delivering J at e costs at least $C(J; e)$, so $W(A'; e') \ge C(J'; e')$; A_m is a level-set contract, so its bill is exactly $C(m; e_m)$. By the envelope theorem $\partial_e C(J; e) = M_w = J$, and by the pricing identity of Claim 9, $\partial_J C(J; e) = 1/\kappa(J; e)$ with $\kappa(J; e)$ the level of the score paid, decreasing in J and hence at least κ_m for $J \le m$. So

$$\begin{aligned} W(A_m; e_m) - W(A'; e') \le C(m; e_m) - C(J'; e') &= \big[C(m; e_m) - C(J'; e_m)\big] + \big[C(J'; e_m) - C(J'; e')\big] \\ &\le \frac{m - J'}{\kappa_m} + J'(e_m - e'). \end{aligned}$$

The comparison. With $|\Psi_1(e_m) - \Psi_1(e')| \le \bar{\Psi}_1'(e_m - e')$,

$$\Pi(A_m, e_m) - \Pi(A', e') \;\ge\; (e_m - e')[\mu_0 - J' - \bar{\Psi}_1'] - \frac{m - J'}{\kappa_m},$$

and by (28) and the definition of c_w, $\kappa_m = \mu_0 v(m)/\sigma\,(1 \pm \epsilon)$ being the level at the cutoffs of incentive at most m, so that $\mu_0 \kappa_m \ge \alpha/(1 - c_w)$,

$$\frac{\Pi(A_m, e_m) - \Pi(A', e')}{m - J'} \;\ge\; \frac{\mu_0}{\alpha}\Big[\frac{\alpha}{\alpha + 2C_*}\Big(1 - \frac{J' + \bar{\Psi}_1'}{\mu_0}\Big) - (1 - c_w)\Big].$$

Since $J' \le I_1$, by (U1) and (U2) each of the two factors in the first term is at least $1 - \delta$ and the bracket is

at least $1-2\delta-(1-c_w)=\frac{4}{5}c_w>0$. So the switch to A_m is profitable, whatever consistent effort the agent plays at A'.

Step 3: the offered contract determines the effort. In any equilibrium the offered contract A^* satisfies $M_w(A^*;0)\geq I_u$, since otherwise A_m would be a profitable deviation by Step 2 (her payoff at A_m is unambiguous, $\mathcal{E}(A_m)$ being a single point by Step 1). By Step 1, $\mathcal{E}(A^*)$ is a single point: the offered contract determines the agent's effort. This proves the proposition.

Step 4: the condition (U). Each part of (U) is a level on $\sigma/\bar{w}$, on ς, or on α of order σ^{-2}, with a constant that depends on v_1 only through c_w. Under Assumption 3, $g_\varepsilon=\phi(\cdot/\sigma)/\sigma$, $m=M=\sigma^{-2}$, $-q(z)=z/\sigma^2$, and with a concentrated belief the known-complementarity cutoff of (18) is $z_1=v_1\sigma$, so that $I_1=\bar{w}\mu_0\phi(v_1)/\sigma$. The rent at the retention cutoff is at most $\bar{w}\bar{\Phi}(v_2)$, with v_2 its standardized second-period cutoff. And $K_g=2e^{-1}(1+O(\bar{e}\theta_{\max}/\sigma))$ in (25), since $2\phi(\sqrt{2})/\phi(0)=2e^{-1}$.

The belief at the retention cutoff. For a prior at which the relationship is retained without learning, $\Pi_2(g_\theta)\geq\pi_0$, the belief at the retention cutoff is more pessimistic than the prior, so $v_2\geq v_1$ and $\bar{\Phi}(v_2)\leq\bar{\Phi}(v_1)$. For the other priors at which she hires, the belief at the retention cutoff is more optimistic than the prior, and Claim 12 of the preliminaries bounds how far: its cutoff is v_A, with $v_1\leq v_A+\delta$. In the first case the retention scale satisfies $t_R\leq\bar{w}\bar{\Phi}(v_1)\mu_0\phi(0)/\sigma\,(1+O(V))$, and since the Mills ratio $\bar{\Phi}/\phi$ is decreasing,

$$\frac{K_g t_R}{I_1}\;\leq\;2\phi(\sqrt{2})\,\frac{\bar{\Phi}(v_1)}{\phi(v_1)}\,(1+O(V))\;\leq\;2\phi(\sqrt{2})\,\frac{\bar{\Phi}(0)}{\phi(0)}\,(1+O(V))\;=\;e^{-1}(1+O(V)).$$

The condition (U). Its four parts, with $\delta=c_w/10$, hold as follows. (U1): C_* is a continuation-curvature bound of order σ^{-2} at moderate v_1 and decays like $\bar{\Phi}(v_1)$ at large v_1, so $2C_*/\alpha\leq\delta$ is a curvature level of order σ^{-2}/c_w and holds at every larger curvature. $I_1=\bar{w}\mu_0\phi(v_1)/\sigma$ and $\bar{\Psi}_1'=O(1/(\alpha\sigma^2))$ by Claim 13, so (U2), $(I_1+\bar{\Psi}_1')/\mu_0\leq\delta$, is the level $\sigma\geq 10\bar{w}\phi(v_1)/c_w$ together with a curvature level of order σ^{-2}/c_w. At large v_1 both sides decay, $\phi(v_1)$ faster than $c_w\asymp 1/v_1^2$. $C_1=\bar{w}\theta_{\max}^2\nu_2$ with $\nu_2=\int|g_\varepsilon''|=4\phi(1)/\sigma^2$, so (U3), $50C_1\leq\alpha$, is the curvature level $\alpha\geq 200\phi(1)\bar{w}\theta_{\max}^2/\sigma^2$, of order σ^{-2} and holding at every larger curvature. And $I_u=(1+2\epsilon)[t_-+K_gt_R(\alpha+C_1)/(\alpha-C_1-2C_*)]\leq(1+2\epsilon)[t_-+e^{-1}(1+\epsilon)(1+\delta)I_1]$ by (26), since $K_g=2e^{-1}(1+\epsilon)$ and $\rho\leq\frac{1}{2}$, with $t_-\leq\sup|T_W|=O(V)I_1$ at a concentrated prior, so (U4), $I_u\leq\frac{4}{5}I_1$, holds for ς large. Finally c_w itself: with $m=M=\sigma^{-2}$, the score's slope at a cutoff shifted by the effort differs from $-q(z(m))$ by a relative $O(\bar{w}/\sigma)$, so $\kappa_m=\mu_0(-q(z(m)))(1-O(\bar{w}/\sigma))$ and $z(m)\geq\sigma(v_1^2+2\log\frac{10}{9})^{1/2}$, so

$$c_w\;\geq\;1-\frac{v_1}{(v_1^2+2\log\frac{10}{9})^{1/2}}-O(\bar{w}/\sigma),$$

which is positive at every v_1, with $c_wv_1^2\to\log\frac{10}{9}$ as $v_1\to\infty$. So every level in (U) is a curvature level of order σ^{-2} with a constant that depends on v_1 only through c_w, or a level on $\sigma/\bar{w}$ or on ς: these define σ^U and α^U. □

Part (A): derivatives along the tilt family

At a fixed conjecture $\hat{e}_1$, every posterior is a tilt of one modified base. Claim 14 describes how the realized index moves with effort.

Claim 14 (Score of the realized index)**.** *Let $L_\gamma(b)$ be the density, under the prior γ, of the index b at which beliefs land after first-period signal, and let $s := \hat{e}_1/\sigma = \omega\sigma$ be the standard deviation of the index given the complementarity. Then*

$$\partial_\gamma \log L_\gamma(b) = \frac{b-\gamma}{s^2} - \mu_\gamma,$$

where μ_γ is the mean of the prior. Equivalently, $\partial_\gamma \log L_\gamma(b) = (b - \gamma - \omega\hat{e}_1\mu_\gamma)/s^2$, where $\gamma + \omega\hat{e}_1\mu_\gamma$ is the expected index.

Proof. Given θ, the index $b = \gamma + \omega\hat{e}_1\theta + \omega\varepsilon$ is Gaussian with mean $\gamma + \omega\hat{e}_1\theta$ and standard deviation s. With $Z(\gamma) := \int_\Theta e^{\gamma\theta} g_0(\theta)\, d\theta$ the normalizer of the prior,

$$L_\gamma(b) = \frac{1}{Z(\gamma)} \int_\Theta e^{\gamma\theta} g_0(\theta)\, \frac{1}{s}\, \phi\Big(\frac{b-\gamma-\omega\hat{e}_1\theta}{s}\Big)\, d\theta.$$

Expanding the square in the exponent of ϕ, and using $\omega\hat{e}_1/s^2 = 1$ and $\omega^2\hat{e}_1^2/s^2 = \hat{e}_1^2/\sigma^2$,

$$-\frac{(b-\gamma-\omega\hat{e}_1\theta)^2}{2s^2} = -\frac{(b-\gamma)^2}{2s^2} + (b-\gamma)\,\theta - \frac{\hat{e}_1^2\theta^2}{2\sigma^2}.$$

The last two terms combine with $e^{\gamma\theta} g_0(\theta)$ into $e^{b\theta}\tilde{g}_0(\theta)$, so, with $\tilde{Z}(b) := \int_\Theta e^{b\theta}\tilde{g}_0(\theta)\, d\theta$ the normalizer of the posterior,

$$L_\gamma(b) = \frac{1}{\sqrt{2\pi}\, s}\, e^{-(b-\gamma)^2/(2s^2)}\, \frac{\tilde{Z}(b)}{Z(\gamma)}.$$

Differentiating the logarithm in γ, with $Z'(\gamma)/Z(\gamma) = \mu_\gamma$ as in Part (A), gives the claim. □

Part (B): the wage effect and the cutoff

Up to a smaller term, the wage effect β is the marginal cost of effort times the movement of the bonus cutoff with good news (Claim 15). At a known complementarity the cutoff first falls and then rises as the complementarity increases (Claim 16), a concentrated posterior keeps that shape (Claims 17 and 18), and so β changes sign at most once as the posterior improves (Lemma 5).

At index b the second-period contract pays $\bar{w}$ above the cutoff $z^+(b)$, and below $z^-(b)$ when $\mu_-(q_b) > 0$, by Corollary 1, which describes every optimal contract and so restricts nothing, and the agent supplies effort $e_2(b)$. The state-conditional rent is $u_2^*(\theta, b) = \bar{w}\,\bar{G}_\varepsilon(z^+ - e_2\theta) - \alpha e_2^2/2 + O(\mu_-)$, and the agent's first-order condition is $\alpha e_2 = \bar{w}\,\mathbb{E}_b[\theta g_\varepsilon(z^+ - e_2\theta)] + O(\mu_-)$.

Claim 15 (Wage effect and the cutoff)**.**

$$\beta(b) = \alpha e_2(b)\, z^{+\prime}(b) - \bar{w}\, e_2'(b)\ \mathrm{Cov}_b\left(\theta,\ \theta g_\varepsilon(z^+ - e_2\theta)\right) + O(\mu_-).$$

In particular $\beta(b) = \alpha e_2(b)\, z^{+\prime}(b) + O(V_b^2) + O(\mu_-)$.

Proof. At a fixed state, $\partial_b u_2^* = -\bar{w} g_\varepsilon(z^+ - e_2\theta)(z^{+\prime} - e_2'\theta) - \alpha e_2 e_2' + O(\mu_-)$. Multiplying by $-\theta$ and taking

expectations at index b,

$$\beta = \bar{w} z^{+\prime}\, \mathbb{E}_b[\theta g_\varepsilon] - \bar{w} e_2'\, \mathbb{E}_b[\theta^2 g_\varepsilon] + \alpha e_2 e_2' \mu_b + O(\mu_-).$$

By the agent's first-order condition $\bar{w}\mathbb{E}_b[\theta g_\varepsilon] = \alpha e_2 + O(\mu_-)$, which turns the first term into $\alpha e_2 z^{+\prime}$ and the last two into $-\bar{w} e_2'(\mathbb{E}_b[\theta^2 g_\varepsilon] - \mu_b \mathbb{E}_b[\theta g_\varepsilon]) = -\bar{w} e_2'\, \mathrm{Cov}_b(\theta, \theta g_\varepsilon)$. The contract depends on b only through the posterior at index b, so by the tilt identities of Part (A) its derivatives in b are of order V_b, and so is the covariance, by Hoeffding's identity. □

Take the complementarity known and equal to $\mu > 0$, and recall from the preliminaries the optimal standardized cutoff $v_\mu = \alpha\sigma/\mu^2$, the effort $e^*(\mu) = \bar{w}\mu\phi(v_\mu)/(\alpha\sigma)$, the cutoff $z^*(\mu) = e^*(\mu)\mu + \sigma v_\mu$ and the incentive yield $\varkappa := \varkappa_\mu = \bar{w}\phi(v_\mu)/\sigma$. Write $R(\mu) := v_\mu^2(1-\varkappa) - \varkappa$.

Claim 16 (The cutoff is decreasing, then increasing)**.** *The slope of the cutoff has the sign of $-R$, and R has a unique zero μ°. So z^* is strictly decreasing on $(0, \mu^\circ]$ and strictly increasing on $[\mu^\circ, \infty)$. Moreover, $\mu^\circ \geq \bar{\theta}$ if and only if $\alpha \geq \bar{\alpha}$. The threshold $\bar{\alpha}$ exists, is unique, and equals $\bar{\theta}^2 \sqrt{\bar{w}\phi(0)}\, \sigma^{-3/2}(1 + O(\sigma^{-1}))$.*

Proof. Since $v_\mu' = -2v_\mu/\mu$ and $\phi'(v_\mu) = -v_\mu\phi(v_\mu)$, $e^{*\prime} = e^*(1 + 2v_\mu^2)/\mu$, so

$$z^{*\prime} = e^*(2 + 2v_\mu^2) - \frac{2\sigma v_\mu}{\mu} = -\frac{2\sigma}{\mu v_\mu} R(\mu),$$

using $e^*\mu/\sigma = \varkappa/v_\mu$. Since $d\varkappa/dv_\mu = -v_\mu\varkappa$, $dR/dv_\mu = v_\mu(2 - \varkappa + \varkappa v_\mu^2) > 0$. So R increases with v_μ, which falls as μ rises and rises with the curvature. Hence R strictly decreases in μ, from $+\infty$ as $\mu \to 0$ to $-\bar{w}\phi(0)/\sigma < 0$ as $\mu \to \infty$, and $R(\bar{\theta})$ increases with the curvature and vanishes at $\bar{\alpha}$. So $\mu^\circ \geq \bar{\theta}$ exactly when $\alpha \geq \bar{\alpha}$. The threshold $\bar{\alpha}$ is where $R(\bar{\theta}) = 0$, that is, where $v_\mu^2(1-\varkappa) = \varkappa$ at $v_\mu = \alpha\sigma/\bar{\theta}^2$. The left side increases with the curvature and the right side decreases, so $\bar{\alpha}$ exists and is unique. As $\sigma \to +\infty$, $\varkappa \leq \bar{w}\phi(0)/\sigma \to 0$, hence $v_\mu \to 0$ and $v_\mu^2 = \bar{w}\phi(0)/\sigma\,(1 + O(\sigma^{-1}))$. Substituting back, $\bar{\alpha} = v_\mu\bar{\theta}^2/\sigma$ has the stated order. □

Claim 17 (A concentrated posterior behaves like its mean)**.** *Fix an index b and let $\zeta(b) := z^{+\prime}(b)/V_b$, the slope of the cutoff per unit of posterior variance. For $\varsigma \geq \max\{8\bar{e}^2/\sigma^2, 16/\underline{\mu}^2\}$,*

$$\zeta = z^{*\prime}(\mu_b) + A(\mu_b)\,\bar{\kappa}_3 \sqrt{V_b} + R_1$$

and

$$\partial_b \zeta = V_b[z^{*\prime\prime}(\mu_b) + A(\mu_b)(\bar{\kappa}_4 - \bar{\kappa}_3^2)] + R_2,$$

where A is an explicit continuous function, and $|R_1| \leq K_1 V_b + O(\mu_-)$ and $|R_2| \leq K_2 V_b^{3/2} + O(\mu_-)$ for constants K_1, K_2 that do not depend on the prior.

Proof. Up to $O(\mu_-)$, the cutoff and the effort at index b solve the agent's first-order condition $\alpha e = \bar{w}\mathbb{E}_b[\theta g_\varepsilon(z - e\theta)]$ and the principal's,

$$e_z(\mu_b - \alpha e) + \bar{w}\, \mathbb{E}_b[g_\varepsilon(z - e\theta)] = 0,$$

where $e_z = \bar{w}\mathbb{E}_b[\theta g'_\varepsilon(z - e\theta)]/(\alpha + \bar{w}\mathbb{E}_b[\theta^2 g'_\varepsilon(z - e\theta)])$ is the agent's response to the cutoff. At a point mass they give the contract of Claim 16. Expanding each expectation around μ_b, $\mathbb{E}_b[F] = F(\mu_b) + F''(\mu_b)V_b/2 + F'''(\mu_b)\bar{\kappa}_3 V_b^{3/2}/6 + O(V_b^2)$, and solving the linearized conditions, whose Jacobian at the point mass is non-singular, gives

$$z^+(b) = z^*(\mu_b) + A(\mu_b)\, V_b + O(V_b^{3/2}),$$

with the term of order $V_b^{3/2}$ proportional to $\bar{\kappa}_3$. For A, work in units $\sigma = 1$, write δ_z and δ_e for the changes in the cutoff and the effort per unit of variance, and $\delta_v := \delta_z - \mu\delta_e$ for the change in the standardized cutoff $z - e\mu$, which equals $v_\mu = \alpha/\mu^2$ at the point mass, where $e = \varkappa/(\mu v_\mu)$. The agent's condition gives

$$\mu^3 v_\mu\, \delta_e = \varkappa\Big[-\mu^2 v_\mu\, \delta_v + \varkappa + \frac{\varkappa^2(v_\mu^2 - 1)}{2v_\mu^2}\Big],$$

the principal's, after substituting it, gives

$$\mu^2\, \delta_v = \frac{\varkappa}{v_\mu}\,(1 - 2v_\mu^2 + 2\varkappa),$$

and $A = \delta_z = \delta_v + \mu\delta_e$, with σ restored, is

$$A(\mu) = \frac{\sigma\varkappa}{\mu^2 v_\mu}\Big[1 - 2v_\mu^2(1 - \varkappa) + 2\varkappa(1 - \varkappa) - \frac{\varkappa^2(1 - v_\mu^2)}{2v_\mu^2}\Big].$$

The hypothesis on ς keeps every denominator within a factor two of its point-mass value. Differentiating in b with the tilt identities of Part (A), $d\mu_b/db = V_b$, $dV_b/db = \bar{\kappa}_3 V_b^{3/2}$ and $d(\bar{\kappa}_3\sqrt{V_b})/db = V_b(\bar{\kappa}_4 - \bar{\kappa}_3^2)$, and each derivative of a remainder is one order higher in $\sqrt{V_b}$, which gives the two displays. □

Claim 18 (At the minimizer of the cutoff)**.** *For $\sigma \geq \sigma^{\mathrm{sc}}$, at the minimizer μ° of the cutoff (Claim 16), $v_\mu < 1$, $z^{*\prime\prime}(\mu^\circ) = \frac{4\sigma v_\mu}{\mu^2}(2 - \varkappa + \varkappa v_\mu^2) > 0$, and*

$$\frac{A(\mu^\circ)}{z^{*\prime\prime}(\mu^\circ)} = \frac{2 + 3v_\mu^2 - v_\mu^4}{8(1 + v_\mu^2)^2(2 + v_\mu^2 + v_\mu^4)} \in \Big(0, \frac{1}{8}\Big].$$

Proof. At $R = 0$, $\varkappa = v_\mu^2/(1 + v_\mu^2)$. If $v_\mu \geq 1$, then $\varkappa \leq 2\phi(1) < \frac{1}{2}$ for $\sigma \geq \sigma^{\mathrm{sc}}$ while $v_\mu^2(1 - \varkappa) > \frac{1}{2}$, so $R > 0$; hence $v_\mu < 1$. Differentiating $z^{*\prime} = -(2\sigma/(\mu v_\mu))R$ at a zero of R, with $dR/dv_\mu = v_\mu(2 - \varkappa + \varkappa v_\mu^2)$ and $du/d\mu = -2v_\mu/\mu$, gives $z^{*\prime\prime}(\mu^\circ)$. Substituting $\varkappa = v_\mu^2/(1 + v_\mu^2)$ into A and $z^{*\prime\prime}$ gives the ratio. Its numerator is positive for $v_\mu < 1$ and at most $2 + 3v_\mu^2 \leq 2(1 + v_\mu^2)^2$, and its denominator is at least $16(1 + v_\mu^2)^2$, so the ratio is at most $\frac{1}{8}$. □

Universal bounds. Two facts about the shape of a distribution are used below. First, every distribution with a finite fourth moment has $\bar{\kappa}_4 \geq \bar{\kappa}_3^2 - 2$: with $Y := X - \mathbb{E}X$ and $V := \mathbb{E}Y^2$, the Cauchy–Schwarz inequality gives $\mathbb{E}[Y(Y^2 - V)]^2 \leq V\,\mathbb{E}[(Y^2 - V)^2]$, which is $\bar{\kappa}_3^2 \leq \bar{\kappa}_4 + 2$. Second, for a log-concave distribution every standardized cumulant is bounded by a constant that depends only on its order. Conditioned on its sign, each part of $X - \mathbb{E}X$ has a log-concave law, and for a nonnegative log-concave variable the normalized moments

$\mathbb{E}[Y^k]/k!$ are log-concave in k (Borell, 1973). Since a log-concave law places probability between $1/e$ and $1 - 1/e$ above its mean, this bounds every standardized central moment, and hence every standardized cumulant. Below, C_3, C_4, C_5 and C_{53} denote the bounds on $|\bar{\kappa}_3|$, $\bar{\kappa}_4$, $|\bar{\kappa}_5|$ and $|\bar{\kappa}_5 - \bar{\kappa}_3\bar{\kappa}_4|$.

Lemma 5 (The wage effect crosses zero once)**.** *Under Assumption 3 with $\sigma \geq \sigma^{\text{sc}}$, there exists $\varsigma_{\text{peak}}(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma_{\text{peak}}$: if $\alpha < \bar{\alpha}$ there is an index b° with $\beta < 0$ on $[\bar{b}, b^\circ)$ and $\beta > 0$ on $(b^\circ, +\infty)$, where nonempty, and $\partial_b\beta > 0$ at b°. If $\alpha \geq \bar{\alpha}$, then $\beta < 0$ at every retained index.*

Proof. By Claim 15, $\beta = \alpha e_2 V_b\, \zeta + O(V_b^2) + O(\mu_-)$, so it suffices to sign ζ. Fix $\delta > 0$ small.

Away from the minimizer of the cutoff. If $|R(\mu_b)| \geq \delta$, then $|z^{*\prime}(\mu_b)| \geq c_\delta > 0$, and by Claim 17 and the universal bounds, $|\zeta - z^{*\prime}(\mu_b)| \leq |A|\, C_3 \sqrt{V_b} + K_1 V_b + O(\mu_-) < c_\delta/2$ for ς large enough. So ζ, and with it β, has the sign of $z^{*\prime}(\mu_b)$, which is the sign of $-R(\mu_b)$.

Near the minimizer of the cutoff. If $|R(\mu_b)| \leq \delta$, then $0 < A \leq z^{*\prime\prime}/4$ and $z^{*\prime\prime} > 0$ by Claim 18 and continuity, for δ small, so the bound $\bar{\kappa}_4 - \bar{\kappa}_3^2 \geq -2$ gives $A(\bar{\kappa}_4 - \bar{\kappa}_3^2) \geq -2A$, and Claim 17 gives

$$\partial_b\zeta \;\geq\; \tfrac{1}{2}\, z^{*\prime\prime}(\mu_b)\, V_b - K_2 V_b^{3/2} - O(\mu_-) \;>\; 0$$

for ς large enough. So ζ crosses zero at most once there, and only from below. Every zero of β lies in this region: at a zero, $\zeta = O(V_b)$ by Claim 15, so $z^{*\prime}(\mu_b) = O(\sqrt{V_b})$ by Claim 17, and $|R(\mu_b)| < \delta$ for ς large enough. Since $\partial_b(\alpha e_2 V_b) = O(V_b^{3/2})$, $\partial_b\beta = \alpha e_2 V_b\, \partial_b\zeta + O(V_b^{3/2}|\zeta|) + O(V_b^{5/2}) + O(\mu_-)$, and at a zero of β the last three terms are of order $V_b^{5/2}$, below the first. So $\partial_b\beta > 0$ at every zero of β, and β crosses zero at most once. Together with the first step, β is negative below a single index b° and positive above it.

The case $\alpha \geq \bar{\alpha}$. If $\alpha > \bar{\alpha}$, then $R \geq R(\bar{\theta}) > 0$ on the retained range by Claim 16, and the first step gives $\beta < 0$ throughout. If $\alpha = \bar{\alpha}$, the minimum of the cutoff is at $\bar{\theta}$. Since $z^{*\prime\prime}$ and A are continuous and $z^{*\prime\prime}(\bar{\theta}) > 0$ (Claim 18), there is $\eta > 0$ such that $z^{*\prime\prime} \geq \frac{3}{4} z^{*\prime\prime}(\bar{\theta})$ and $0 < A \leq \frac{1}{4} z^{*\prime\prime}(\bar{\theta})$ on $[\bar{\theta} - \eta, \bar{\theta}]$. Below $\bar{\theta} - \eta$, $R \geq R(\bar{\theta} - \eta) > 0$, and the first step, with $\delta \leq R(\bar{\theta} - \eta)$, gives $\beta < 0$. Now let $\mu_b \geq \bar{\theta} - \eta$, and write $\bar{\theta} - \mu_b = c\sqrt{V_b}$. Since $\bar{\theta} - \theta$ is nonnegative with a log-concave law, its variance is at most its squared mean, by the log-concavity of its normalized moments (Borell, 1973), so $c \geq 1$. And $(\theta - \mu_b)_+ \leq c\sqrt{V_b}$, so $\mathbb{E}_b[(\theta - \mu_b)^3] \leq c\sqrt{V_b}\, V_b$, that is, $\bar{\kappa}_3 \leq c$. By the mean value theorem and $z^{*\prime}(\bar{\theta}) = 0$, $z^{*\prime}(\mu_b) \leq -\frac{3}{4} z^{*\prime\prime}(\bar{\theta})\, c\sqrt{V_b}$, and $A\bar{\kappa}_3\sqrt{V_b} \leq \frac{1}{4} z^{*\prime\prime}(\bar{\theta})\, c\sqrt{V_b}$. So, by Claim 17,

$$\zeta \;\leq\; -\tfrac{1}{2}\, z^{*\prime\prime}(\bar{\theta})\, c\sqrt{V_b} + K_1 V_b + O(\mu_-) \;<\; 0$$

for ς large enough, since $c \geq 1$. So $\beta < 0$ at every retained index. □

Part (C): the manipulation term

The sign of T. The manipulation term can be written in terms of the posterior index (Claim 19). The next two lemmas show that it is positive at pessimistic priors (Lemma 6) and negative at optimistic priors when $\alpha < \bar{\alpha}$ (Lemma 7). Both hold at any fixed conjecture in the range, with constants that do not depend on it.

Write $T(\gamma)$ for the continuation incentive (11) at prior γ, with the agent's effort and the principal's conjecture both equal to $\hat{e}_1$, and

$$\beta(b) := -\mathbb{E}_b[\theta\, \partial_b u_2^*]$$

for the wage effect, which is positive when better news lowers the agent's rent.

Claim 19 (Two forms)**.** *With the realized index of state θ Gaussian around $\hat{b}_\theta := \gamma + \omega\theta\hat{e}_1$,*

$$\begin{aligned} T(\gamma) &= \frac{1}{\omega\sigma^2}\mathbb{E}[\theta\, u_2^*(\theta, b)(b - \hat{b}_\theta)\mathbf{1}\{b \geq \bar{b}\}] \\ &= \omega\Big[r(\bar{b})L_\gamma(\bar{b}) - \int_{\bar{b}}^{\infty} \beta(b)L_\gamma(b)db\Big], \end{aligned}$$

where $r(\bar{b}) := \mathbb{E}_{\bar{b}}[\theta u_2^*]$.

Proof. For the Gaussian density f_θ of the realized index at state θ, with $s = \omega\sigma$ as in Claim 14,

$$\partial_e f_\theta = \frac{\omega\theta\,(b - \hat{b}_\theta)}{s^2}\, f_\theta.$$

The first form follows by differentiating the continuation in e at the conjecture and averaging over θ. The second follows by integrating by parts in b and applying Bayes' rule in the form

$$\int \psi\, f_\theta\, g_\theta(\theta \mid \gamma)\, d\theta = L_\gamma(b)\, \mathbb{E}_b[\psi].$$

It is the decomposition (8) in index coordinates. □

Lemma 6 (Retention dominates at pessimistic priors)**.** *Let $\bar{\gamma}$ be the prior whose expected index is the retention cutoff, $\mathbb{E}_{\bar{\gamma}}[b] = \bar{b}$. There exists $\varsigma_{\text{init}} \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma_{\text{init}}$:*

(i) *$T(\gamma) > 0$ for all $\gamma \leq \bar{\gamma}$.*

(ii) *$\gamma_L \leq \bar{\gamma}$.*

Hence $T(\gamma_L) > 0$.

Proof. Two conditions are used, and Step 6 verifies them. For some $k > 0$, the first is

$$\Big(\frac{k\hat{e}_1}{\sigma\sqrt{\varsigma}}\Big)^2 < \frac{2u_2^-}{u_2^+}, \tag{29}$$

and the second is

$$OV(\bar{\gamma}) > D_1(\bar{\gamma}) + \big|\Pi_2(g_\theta) - \pi_0\big|, \tag{30}$$

where $u_2^{\mp}$ are the minimum and maximum of u_2^* over retained index values within $k/\sqrt{\varsigma}$ of the mean, $OV := \mathbb{E}[(\Pi_2 - \pi_0)^+]$ is the option value of learning, D_1 the principal's learning distortion, and $\Pi_2(g_\theta)$ her value from a single period at the prior, all at $\bar{\gamma}$. The objects in the second condition are defined in Step 5.

Step 1: state by state. Fix $\gamma \le \bar{\gamma}$ with prior mean μ_γ, and use the first form of Claim 19: T is a sum over states θ of θ times the retained part of the derivative of the state's index density. With $c_\theta := \omega\hat{e}_1(\theta - \mu_\gamma)$, the state's expected index is

$$\hat{b}_\theta = \gamma + \omega\hat{e}_1\mu_\gamma + c_\theta \;\le\; \bar{b} + c_\theta,$$

since the expected index $\gamma + \omega\hat{e}_1\mu_\gamma$ is at most $\bar{b}$. In particular $\hat{b}_\theta \le \bar{b}$ for $\theta \le \mu_\gamma$.

Step 2: states near the mean, below it. For $\theta \le \mu_\gamma$ every retained index value satisfies $b - \hat{b}_\theta \ge 0$, so the derivative of the state's density is nonnegative on the retained set. Since $\partial_\theta u_2^* = \bar{w}e_2 g_\varepsilon(z^+ - e_2\theta) \le \bar{w}\bar{e}\phi(0)/\sigma$, for θ within $k/\sqrt{\varsigma}$ of μ_γ the state-conditional rent $u_2^*(\theta, \cdot)$ is within a bounded multiple of $k/\sqrt{\varsigma}$ of the average rent $U_2 > 0$, hence positive for ς large. Such states contribute positively.

Step 3: states near the mean, above it. For $\theta > \mu_\gamma$, with $x := b - \hat{b}_\theta$ and $s = \omega\sigma$ the standard deviation of the index given the state (Claim 14), the retained set contains index values on both sides of $\hat{b}_\theta$, and the state's contribution is at least

$$\frac{\theta}{\omega\sigma^2}\Big[\frac{u_2^- s}{\sqrt{2\pi}} - \frac{u_2^+ c_\theta^2}{2s\sqrt{2\pi}}\Big],$$

positive whenever $((\theta - \mu_\gamma)\hat{e}_1/\sigma)^2 < 2u_2^-/u_2^+$, which condition (29) gives for $\theta - \mu_\gamma \le k/\sqrt{\varsigma}$.

Step 4: states far from the mean. Steps 2 and 3 cover the states within $k/\sqrt{\varsigma}$ of the mean. Under a ς-strongly log-concave prior the mass of the states farther away, on either side, is at most $2e^{-k^2/2}$, and their contribution is bounded by a constant times that mass. Counterproductive states contribute at most a constant times their retained mass, $O(\mu_-)$. For k large both are dominated by Steps 2–3, which proves (i) at $\bar{\gamma}$. Lowering γ moves every $\hat{b}_\theta$ down, which enlarges the positive contributions and shrinks the negative ones, so (i) holds at all $\gamma \le \bar{\gamma}$.

Step 5: initiation. The principal's value from initiating is

$$\mathcal{V}(\gamma) = \Pi_1(\gamma) + \pi_0 + OV(\gamma),$$

and her first-period profit is $\Pi_1 = \Pi_2(g_\theta) + S_1 - D_1$. Here $S_1 \ge 0$ is the reduction in the wage bill for the implemented target relative to buying its whole marginal cost, nonnegative because $T > 0$ at every conjecture at $\bar{\gamma}$, Step 4 having been proved at a fixed conjecture without regard to which, and $D_1 \ge 0$ is the loss from moving the target away from her single-period optimum. Hence

$$\mathcal{V}(\bar{\gamma}) - 2\pi_0 \;\ge\; OV - D_1 - \big|\Pi_2(g_\theta) - \pi_0\big| \;>\; 0$$

under condition (30), and $\gamma_L \le \bar{\gamma}$ by the definition of γ_L as the smallest prior with $\mathcal{V} \ge 2\pi_0$.

Step 6: the conditions. The posterior at the retention cutoff has mean $\mu_{\bar{\gamma}} + O(V_\gamma^{3/2})$, since the posterior mean is a martingale in b and its second derivative in the index is κ_3, and its variance is below the prior's by $O(V_\gamma^2)$, so

$$|\Pi_2(g_\theta) - \pi_0| = |\Pi_2(g_\theta) - \Pi_2(q_{\bar{b}})| = O(V_\gamma^{3/2}).$$

Also $D_1 = O(\Pi_2'^2) = O(V_\gamma^2)$. For the option value, the proof of Lemma 3 gives $\Pi_2'(b) \ge c_0 V_b \ge c_0 V_\gamma/2$ on the retained range, with $c_0 > 0$ not depending on the prior, so $OV \ge \frac{1}{2}c_0 V_\gamma\,\mathbb{E}[(b - \bar{b})^+]$. At $\bar{\gamma}$ the retention cutoff is the mean of the index, and the index is its conditional mean given θ plus independent Gaussian

noise of standard deviation s, so $\mathbb{E}[(b-\bar{b})^+] \geq s/\sqrt{2\pi}$. Hence $OV \geq cV_\gamma$ with $c := c_0 s/(2\sqrt{2\pi}) > 0$. Fix k large enough for Step 4. The first condition (29) then holds once $\varsigma \geq k^2\bar{e}^2 u_2^+/(2\sigma^2 u_2^-)$, and the second once cV_γ exceeds the lower-order terms. Both hold for ς large enough, and ς_{init} is the level above which they do. □

Lemma 7 (The ratchet dominates at optimistic priors)**.** *If $\alpha < \bar{\alpha}$, then $T(\gamma) < 0$ for all γ large enough.*

Proof. By the second form of Claim 19, $T = \omega[r(\bar{b})L_\gamma(\bar{b}) - \int_{\bar{b}}^{\infty} \beta L_\gamma]$, and $\bar{b}$ is fixed by π_0. The proof bounds the retention term above and the wage term below.

The retention term. By Claim 14, the index has mean $\gamma + c$, with $c := \omega\hat{e}_1\mu_\gamma$ bounded, and standard deviation s given the state, so

$$r(\bar{b})\, L_\gamma(\bar{b}) \;\leq\; \bar{r}\, C\, e^{-(\gamma-\bar{b}-c)^2/(2s^2)}$$

for constants $\bar{r}$ and C, which is Gaussian in γ.

The wage term. By Claims 15 and 17, $\beta(b) = \alpha e_2 V_b(z^{*\prime}(\mu_b) + O(\sqrt{V_b}))$. For $\alpha < \bar{\alpha}$, $R(\bar{\theta}) < 0$ by Claim 16, so $z^{*\prime} > 0$ on an interval $(\mu_1, \bar{\theta}]$ with $\mu_1 < \bar{\theta}$. Let b_1 be the index at which the posterior mean equals μ_1, which does not depend on γ. For $b \geq b_1$ and ς large enough, $\beta(b) \geq c_1 V_b$ for a constant $c_1 > 0$. At the top of the family the base $g_0 = e^{-\psi}$, with $\psi'' \geq \varsigma$, is bounded and bounded away from zero on its compact support, so the tilt at index b has mean within $O(1/b)$ of $\bar{\theta}$ and variance $V_b \geq c_2 b^{-2}$ for a constant $c_2 > 0$. For γ large the interval $J_\gamma := [\gamma + c - s, \gamma + c + s]$ lies beyond b_1, carries index mass at least $\frac{1}{2}$, and on it $V_b \geq c_2(\gamma + c + s)^{-2}$. Below b_1, $|\beta| \leq K/\varsigma$ and the index mass is at most $e^{-(\gamma+c-b_1)^2/(2s^2)}$, by Claim 14. Hence

$$\int_{\bar{b}}^{\infty} \beta L_\gamma \, db \;\geq\; \frac{c_1 c_2}{2(\gamma + c + s)^2} - \frac{K}{\varsigma}\, e^{-(\gamma+c-b_1)^2/(2s^2)} \;\geq\; \frac{c_1 c_2}{4(\gamma + c + s)^2}$$

for γ large enough.

Combining the two bounds, $T \leq \omega[\bar{r}Ce^{-(\gamma-\bar{b}-c)^2/(2s^2)} - c_1 c_2/(4(\gamma + c + s)^2)]$, and the Gaussian term is below the polynomial one for all γ large enough, so $T < 0$ there. □

Part (D): the principal's problem

The principal's problem. Three facts are established: where her optimal targets lie (Claim 20), where her problem is concave (Claim 21), and that she hires on an upper interval of priors (Lemma 8).

Her problem near the known-complementarity cutoff. The next claims extend Steps 5 and 6 of the proof of Proposition 2 to the model, where the continuation adds a slope and a curvature to her reduced problem in the target. It uses the notation and Steps 1–3 of the proof of Lemma 4: in every equilibrium the offered contract has wage incentive at least I_u and a unique consistent effort. Three objects enter the statement.

- v_1 and I_1 are the standardized known-complementarity cutoff and the wage incentive it delivers, as in the preliminaries.
- $e_G := (I_u - t_-)/(\alpha + C_1)$ is the smallest effort at any contract with incentive at least I_u (Step 1 of that proof), and $e^\dagger$ is the target whose cheapest contract has level $\alpha/(4\mu_0)$, a quarter of the level at the known-complementarity cutoff.

- v_A solves $\bar{w}[\phi(v)/v - \bar{\Phi}(v)] = \pi_0$: the standardized cutoff at the known complementarity whose second-period value equals the outside option.

Besides (U), the claims below assume the bound

$$\bar{\Psi}_1' + \tfrac{4}{5} I_1 + \frac{2C_*\mu_0}{\alpha} \;<\; \mu_0\, c_S\,, \tag{S}$$

where $c_S := 1 - v_1/(v_1^2 + 2\log\frac{5}{4})^{1/2} - \epsilon$ as in Step 5 of the proof of Proposition 2. It is the bound (S_B) of Step 5 of the proof of Proposition 2 with the continuation's term $2C_*\mu_0/\alpha$ added: it compares the three quantities bounded in (U) with the gain from moving the cutoff until the density falls by a factor $\frac{4}{5}$, in the same form as the slack c_w of (U). On the left, $\frac{4}{5} I_1/\mu_0 = \frac{4}{5}\bar{w}\phi(v_1)/\sigma$, while $\bar{\Psi}_1'/\mu_0$ and $2C_*/\alpha$ are the quantities already bounded in (U). So (S) is the level $\sigma/\bar{w} \geq \frac{4}{5}\phi(v_1)/c_S$ up to the terms of (U). Since c_S decreases in v_1, (S) holds for σ above a level at every $v_1 \leq v_A + 1$, which covers every prior at which she hires, by Claim 12.

Claim 20 (Her problem near the known-complementarity cutoff)**.** *Under (U) and (S), With $r_w(\tilde{e}) := c'(\tilde{e}) - T(\tilde{e}, \tilde{e})$, every equilibrium target lies in*

$$\mathcal{K}_1 := \{\tilde{e} \in [e_G, e^\dagger] : r_w(\tilde{e}) \geq \tfrac{4}{5} I_1\},$$

an interval on which $|\bar{T}'| \leq \bar{s}$ for a constant $\bar{s} \leq (\frac{5}{4e} + \delta)\alpha$, where $\bar{T}(\tilde{e}) := T(\tilde{e}, \tilde{e})$.

Proof. **Step 1: the equilibrium contract is the cheapest one for its own target.**

Let A^* be the contract of some equilibrium and e^* its effort, unique by Steps 1–3 of the proof of Lemma 4, with $e^* \geq e_G$. The cheapest contract implementing e^* at conjecture e^* is $A(\kappa, e^*)$ with $\mathcal{I}(\kappa, e^*) = r_w(e^*)$ (Theorem 2, Claim 9). Its incentive at e^* is $r_w(e^*) = M_w(A^*; e^*) \geq M_w(A^*; 0) \geq I_u$, the first inequality because the cutoff lies above the expected signal, where an upper-tail bonus's incentive increases with effort, and its incentive at zero effort is at least $1 - \epsilon$ times that (the preliminaries of Appendix B), which is above the threshold of Step 1 of the proof of Lemma 4 without its factor $1 + 2\epsilon$, so it has a unique effort. It has e^* as a fixed point, hence implements e^*, at a bill no larger than that of A^*, so $A^* = A(\kappa^*, e^*)$. Its first-order conditions in the cutoffs, with $\partial_e W = M_w = r_w$ at the fixed point, put both cutoffs at the common level

$$\kappa^* = \frac{c''(e^*) - \bar{T}'(e^*) - \bar{w}\partial_e I}{G'(e^*) - r_w(e^*)}.$$

Step 2: every critical point has a large wage incentive.

Along the family, by Claim 10,

$$\frac{d}{d\tilde{e}}\Pi(A(\tilde{e})) = \mu_0 + \Psi_1'(\tilde{e}) - r_w(\tilde{e}) - \frac{c''(\tilde{e}) - \bar{T}'(\tilde{e}) - \bar{w}\partial_e I}{\kappa(\tilde{e})},$$

with κ strictly decreasing, and by (25) the slope of the continuation along the diagonal is at least $-2C_*$: the retention term is nondecreasing in the conjecture, as in Step 2 of the proof of Lemma 4. At a target whose contract has incentive $r_w(\tilde{e}) \leq \frac{4}{5} I_1$ the cutoff is at a standardized position v with $\phi(v) \leq \frac{4}{5}\phi(v_1)(1 + \epsilon)$, that is $v \geq (v_1^2 + 2\log\frac{5}{4})^{1/2}$ up to ϵ, so $\kappa(\tilde{e}) = \mu_0 v/\sigma\,(1 \pm \epsilon) \geq (\alpha/\mu_0)\, v/v_1\,(1 - \epsilon)$ and $\mu_0/\kappa \leq (\alpha/\mu_0)^{-1}\mu_0(1 - c_S)$.

The derivative is then at least

$$\mu_0 - \bar{\Psi}_1' - \tfrac{4}{5} I_1 - \frac{\alpha + 2C_*}{\alpha}\,\mu_0(1 - c_S) \;>\; 0$$

by (S). Hence her profit is increasing at every such target and every critical point has $r_w(\tilde{e}) > \frac{4}{5} I_1$, so every equilibrium target lies in $\mathcal{K}_1$, which is an interval since r_w is increasing. On $\mathcal{K}_1$, (25) gives

$$\bar{s} := \sup_{\mathcal{K}_1} |\bar{T}'| \;\le\; 2C_* + \frac{5K_g t_R}{4I_1}\,\alpha\,(1+\epsilon). \tag{31}$$

With the bound on $K_g t_R / I_1$ from Step 4 of the proof of Lemma 4,

$$\frac{\bar{s}}{\alpha} \;\le\; \frac{5}{4} \cdot 2\phi(\sqrt{2})\,\frac{\bar{\Phi}(v_1)}{\phi(v_1)}\,(1 + O(V)) + \frac{2C_*}{\alpha} \;\le\; \frac{5}{4e} + \delta,$$

since the Mills ratio $\bar{\Phi}(v)/\phi(v)$ is largest at $v = 0$, where it equals $\sqrt{\pi/2}$, so that the first term is at most $\frac{5}{4} \cdot \frac{2e^{-1}}{\sqrt{2\pi}} \cdot \sqrt{\pi/2} = \frac{5}{4e}$, and $2C_*/\alpha \le \delta$ by (U1). This proves the claim. In particular $\kappa^* \ge (\alpha - \bar{s} - C_1)/(\bar{\theta} + \bar{\Psi}_1') > 0$. □

Claim 21 (Concavity where the threat is small)**.** *In case (i), assume (U) and (S), and write $T = T_R + T_W$ for the split of T into the retention term and the wage term (proof of Lemma 4). There exists $\varsigma_{\mathrm{conc}}(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \ge \varsigma_{\mathrm{conc}}$, at every prior, one of the following holds.*

(a) *$T_R \ge 2 \sup_{\mathcal{K}_1} |T_W|$ at every target of $\mathcal{K}_1$, and then $T > 0$ on $\mathcal{K}_1$.*

(b) *Her reduced problem in the target is strictly concave on $\mathcal{K}_1$, increasing below it and decreasing beyond $e^\dagger$. All equilibria at that prior then share the first-period contract up to null sets and the first-period effort, and the second period is unique at every history.*

Alternative (b) holds at every zero of T along the equilibrium.

Proof. Concavity follows from Claim 11 once its condition holds with the continuation term $\bar{T}$ in the required incentive. The condition fails only if the curvature of the continuation is large, and the continuation's curvature comes almost entirely from the retention term, which is large only when the retention cutoff sits near the center of the distribution of the signal. The proof makes this precise: it states the condition, reduces it to the retention term, and shows that where the retention term is not already large enough to sign T on its own, it is small enough for the condition to hold.

The condition. On $\mathcal{K}_1$ the levels of the family lie in $[\kappa(e^\dagger), \kappa(\tilde{e}_1)]$, so $\kappa_{\mathcal{K}} \ge \frac{\alpha}{4\mu_0}(1 - \epsilon)$, and Claim 11 applies on $\mathcal{K}_1$ with $G = \tilde{e}\mu_0 + \Psi_1$, $\tau = \bar{T}$ and $\bar{s}$ of (31) as soon as (22) holds there. $\bar{T}''$ exists on $\mathcal{K}_1$, and by the computation of (25) taken one order further,

$$L := \sup_{\mathcal{K}_1} |\bar{T}''| \;\le\; \frac{K_3\, t_R}{e_G^2} + C_*,$$

where $K_3 = 2\phi(1)/\phi(0)$ is the supremum of the Gaussian shape factor that the second derivative of the

density at the retention cutoff produces, while $\sup_{\mathcal{K}_1} |\Psi_1''| = O(V)$ by Claim 13. Condition (22) reads

$$\underline{\Xi}(\alpha - \bar{s})(\alpha - \bar{s} - \varsigma_I) + (\alpha - \bar{s}) \;>\; \bar{d} + \varsigma_2, \tag{32}$$

with $\bar{d} := \sup_{\mathcal{K}_1} |\Psi_1''| + L/\kappa_{\mathcal{K}}$, and compares the continuation's curvature with the wage bill's.

Only the retention term matters. Without the continuation, the condition is the benchmark's, which Step 6 of the proof of Proposition 2 reduces to $\underline{\Xi}\alpha^2 > \sup |\Psi_1''| + O(V)$, holding for ς above a level depending on α and σ. The continuation adds $\bar{s}$ and $L/\kappa_{\mathcal{K}}$, and moves ς_2 by at most $\bar{s}$, the required incentive being now $r = c' - \bar{T}$. So (32) holds for ς above a level depending on α and σ as soon as $\bar{s}$ and L are small. The wage term T_W and its first two derivatives along the diagonal are of order V, as T_W is (proof of Claim 23), each derivative multiplying it by at most a constant over $\tilde{e}$. So only the retention term's contributions remain.

The dichotomy. Let $\nu := 2 \sup_{\mathcal{K}_1} |T_W|$, which is of order V (proof of Claim 23), and let $x(\tilde{e})$ be the standardized distance between the mean of the index and the retention cutoff at target $\tilde{e}$.

Case (a). If $T_R \geq \nu$ at every target of $\mathcal{K}_1$, then $T \geq T_R - |T_W| \geq \frac{1}{2} T_R > 0$ on $\mathcal{K}_1$, which is alternative (a).

Case (b). Otherwise $T_R(\tilde{e}_0) < \nu$ at some $\tilde{e}_0 \in \mathcal{K}_1$. The retention term is the retention scale times the density at the cutoff relative to its peak, $T_R = t_R\,(\phi(x)/\phi(0))(1 + o(1))$ as ς grows, and each derivative of it, in the effort or in the conjecture, multiplies it by at most a constant times $(1 + x^2)/\tilde{e}$, as in the proof of Claim 23. So its contributions to $\bar{s}$ and to L are at most a constant times $(1 + x^2)^2 \phi(x)\, t_R / e_G^2$, uniformly on $\mathcal{K}_1$. If $t_R \leq V^{1/2}$, these vanish as ς grows. If not, $\phi(x(\tilde{e}_0)) \leq C\nu/t_R \leq C' V^{1/2}$, so $x(\tilde{e}_0)^2 \geq \log(1/V) - C''$. The retention cutoff in the index does not depend on the conjecture, and the mean of the index moves by a bounded amount across $\mathcal{K}_1$, so $|x(\tilde{e})| \geq c\,|x(\tilde{e}_0)| - C$ on $\mathcal{K}_1$, hence $x^2 \geq c' \log(1/V)$ there for ς large, and $(1 + x^2)^2 \phi(x)$ vanishes as ς grows, uniformly on $\mathcal{K}_1$. In both cases the retention term's contributions vanish, and (32) holds for ς above a level depending on α and σ, which is ς_{conc}, and this is alternative (b).

At a zero of T along the equilibrium, $T_R = -T_W < \nu$ at the equilibrium target, so alternative (b) holds there.

Conclusion. In alternative (b), (32) holds, so $\tilde{e} \mapsto \Pi(A(\tilde{e}))$ is strictly concave on $\mathcal{K}_1$, which contains every equilibrium target by Step 2 of the proof of Claim 20. It is increasing below $\mathcal{K}_1$ by that step, and decreasing beyond $e^\dagger$: there the last term of the derivative displayed in that step is at least $4(1 - \bar{s}/\alpha - C_1/\alpha)\mu_0 \geq \frac{8}{5}\mu_0$, since $\bar{s}/\alpha + C_1/\alpha \leq \frac{5}{4e} + \delta + \frac{1}{50} < \frac{3}{5}$ by (31) and (U3), and this exceeds $\mu_0 + \bar{\Psi}_1' \leq \mu_0(1 + \delta)$. The targets of two equilibria both maximize it, hence coincide, and the contracts coincide up to null sets by Theorem 2. The second period is unique at every history by Step 3 of the proof of Lemma 4. This proves the claim. □

The hiring set. By Lemma 6, $T > 0$ up to the prior at which the relationship forms. The next lemma shows that it forms at every more optimistic prior, and at more pessimistic priors than in the benchmark.

Lemma 8 (Hiring is a threshold in the prior)**.** *Under Assumption 3, for $\sigma \geq \sigma_H(\alpha) := (K_H/\alpha)^{1/2}$, with K_H a constant of the primitives defined in the proof, there exists $\varsigma_H(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma_H$:*

(i) *$\mathcal{V}^B$ is nondecreasing in γ, so the benchmark principal hires exactly on $[\gamma_L^B, +\infty)$.*

(ii) *The model's principal hires exactly on $[\gamma_L, +\infty)$.*

(iii) $\gamma_L < \gamma_L^B$.

Proof. Write her value from initiating as a maximum over targets, $\mathcal{V}(\gamma) = \max_{\tilde{e}} \Pi(\tilde{e};\gamma)$, with

$$\Pi(\tilde{e};\gamma) = \tilde{e}\,\mu_\gamma - C(c'(\tilde{e}) - T(\gamma;\tilde{e});\tilde{e},\gamma) + \Psi_1(\tilde{e};\gamma),$$

where $C(I;\tilde{e},\gamma)$ is the least-cost bill delivering incentive I at target $\tilde{e}$ under prior γ, and Π^B is the same expression with $T \equiv 0$. Three facts about a fixed target are used.

(a) Expected signal rises. μ_γ is increasing, a higher γ being a likelihood-ratio improvement, with $d\mu_\gamma/d\gamma = V_\gamma$.

(b) Incentive gets cheaper. By the pricing identity of Claim 9, $C(I;\tilde{e},\gamma) = \int_0^I dI'/\kappa(I';\gamma)$, with $\kappa(I';\gamma)$ the level at which the I'-th unit of incentive is bought. A more optimistic prior delivers a given incentive at a higher cutoff, and the level $\kappa = \mu(z)z\sigma^{-2}(1 \pm \epsilon)$ is higher on both counts. So $\partial_\gamma C \le 0$ at a fixed incentive.

(c) The continuation rises. For a given the signal, the posterior under a higher prior has a higher index, and Π_2 is increasing in the index by Lemma 3. And the distribution of the signal improves in the likelihood-ratio order. So $\partial_\gamma \Psi_1 \ge 0$.

(i) By (a) to (c), $\Pi^B(\tilde{e};\cdot)$ is nondecreasing for every target, and a maximum of nondecreasing functions is nondecreasing.

(ii) below $\bar{\gamma}$. Here $T > 0$ at every conjecture, by Lemma 6. Differentiating $\Pi(\tilde{e};\gamma)$ in γ and using (a) to (c),

$$\partial_\gamma \Pi(\tilde{e};\gamma) \;\ge\; \tilde{e}\,V_\gamma + \frac{\partial_\gamma T(\gamma;\tilde{e})}{\kappa},$$

with $\kappa = (\alpha/\mu_0)(1 \pm \epsilon)$ the level of her first-period contract. By the second form of Claim 19 and the score of Claim 14,

$$\partial_\gamma T = \frac{\omega}{s^2}\Big[r(\bar{b})L_\gamma(\bar{b})\,(\bar{b} - \mathbb{E}_\gamma b) - \int_{\bar{b}}^{\infty} \beta L_\gamma\,(b - \mathbb{E}_\gamma b)\,db\Big].$$

The first term is nonnegative for $\gamma \le \bar{\gamma}$, since the retention cutoff lies above the expected index there, and so is the part of the integral on which $\beta \le 0$, since $b \ge \bar{b} \ge \mathbb{E}_\gamma b$ on the retained range. On the part on which $\beta > 0$, $R(\mu_b) < 0$ by Claims 15 to 17, so $|R(\mu_b)| \le \varkappa \le \bar{w}\phi(0)/\sigma$ and $0 < \beta \le 2\bar{e}\,\mu_b\,\varkappa\,V_b\,(1 + O(\sqrt{V_b}))$. Retained beliefs below $\bar{\gamma}$ are at least as optimistic as the prior up to $O(\sqrt{V_\gamma})$, so $\kappa \ge \alpha/(2\mu_b)$, and that part costs at most $4\bar{e}\,\varkappa\,V_\gamma/v_b$ in the display above, with $v_b = \alpha\sigma/\mu_b^2$. Let

$$K_H := \max\Big\{8\,\frac{\bar{e}}{e_G}\,\bar{w}\phi(0)\,\bar{\theta}^2,\; 2K''\Big\},$$

with e_G the effort bound of Claim 20 and K'' the constant of the next paragraph. Since $v_b \ge \alpha\sigma/\bar{\theta}^2$, the condition $\sigma \ge \sigma_H(\alpha)$ gives $v_b \ge 8(\bar{e}/e_G)\varkappa$ on the retained range, so $\partial_\gamma \Pi \ge \frac{1}{2}\tilde{e}\,V_\gamma > 0$ for every target $\tilde{e} \ge e_G$. So $\mathcal{V}$ is nondecreasing on $[\gamma_L, \bar{\gamma}]$.

(ii) above $\bar{\gamma}$. Offer the benchmark's optimal contract. Subtracting the two agents' first-order conditions as in Step 2 of the proof of Lemma 4, the agent's effort differs from the benchmark's target by at most $|T|/(\alpha - C_1 - 2C_*) \le 2|T|/\alpha$, and her payoff at a fixed contract is Lipschitz in his effort with a constant K' of the primitives. When $T \ge 0$ she loses nothing relative to the benchmark. When $T < 0$ she loses at most $2K'|T|/\alpha$, and $|T| \le \omega \sup_{\{\beta>0\}}\beta \le 2\omega\bar{e}\,\bar{\theta}\,\varkappa V_\gamma$, because the retention atom is positive and, as above,

$\beta \le 2\bar{e}\,\mu_b \varkappa V_b$ where it is positive. Above $\bar{\gamma}$ the prior itself would be retained, $\Pi_2(g_\theta) \ge \pi_0 - O(V_\gamma^{3/2})$, so by the decomposition in Step 5 of the proof of Lemma 6,

$$\mathcal{V}(\gamma) - 2\pi_0 \;\ge\; OV(\gamma) - D_1(\gamma) - O(V_\gamma^{3/2}) - \frac{4K'\omega\,\bar{e}\,\bar{\theta}\,\varkappa\,V_\gamma}{\alpha}.$$

Step 6 of that proof gives $OV \ge cV_\gamma$ with $c = c_0 s/(2\sqrt{2\pi})$, where $c_0 \ge e_2$ by the proof of Lemma 3. With $\omega/s = 1/\sigma$ and $\varkappa \le \bar{w}\phi(0)/\sigma$, the last term is at most $(K''/(\alpha\sigma^2))\,cV_\gamma$ with

$$K'' := 8\sqrt{2\pi}\,K'\,\bar{\theta}\,\bar{w}\phi(0)\,\frac{\bar{e}}{c_0},$$

a constant of the primitives, since $\bar{e}$ and c_0 both scale with $1/\alpha$. For $\sigma \ge \sigma_H(\alpha)$ this is at most $\frac{1}{2}cV_\gamma$, so the right side is positive for ς large enough, and with the previous paragraph, the model's principal hires exactly on $[\gamma_L, +\infty)$.

(iii) At γ_L, $T > 0$ at every conjecture, so the model's principal implements the benchmark's target at a lower bill, which gives $\mathcal{V}^B(\gamma_L) < \mathcal{V}(\gamma_L) = 2\pi_0$. By (i) the benchmark principal hires nowhere below γ_L^B, and $\gamma_L^B > \gamma_L$. □

Part (E): from the sign of T to the comparison of efforts

How T moves with the prior. Two claims record how T changes as the prior improves, at a fixed target and through the equilibrium's target. They are used in Lemma 9 and in the proof of Proposition 3.

Claim 22 (The direct effect of the prior)**.** *At a zero of T, the derivative in the prior with the conjecture held at its equilibrium value is $T'(\gamma) = -M(\gamma)$, where*

$$M(\gamma) := \frac{\omega}{s^2}\int_{\bar{b}}^{\infty} (b - \bar{b})\,\beta(b)\,L_\gamma(b)\,db \;>\; 0.$$

Proof. By the second form of Claim 19, $T = \omega\big[r(\bar{b})L_\gamma(\bar{b}) - \int_{\bar{b}}^{\infty}\beta L_\gamma\big]$, in which only L_γ depends on the prior. Differentiating under the integral with the score of Claim 14, $\partial_\gamma \log L_\gamma(b) = (b-\gamma)/s^2 - \mu_\gamma$,

$$T' = \omega\Big[r(\bar{b})L_\gamma(\bar{b})\Big(\frac{\bar{b}-\gamma}{s^2} - \mu_\gamma\Big) - \int_{\bar{b}}^{\infty}\beta L_\gamma\Big(\frac{b-\gamma}{s^2} - \mu_\gamma\Big)db\Big].$$

At a zero of T, $r(\bar{b})L_\gamma(\bar{b}) = \int \beta L_\gamma$, so the parts of the score that do not depend on b cancel, and what is left is $T' = -M$.

To see that $M > 0$, recall from Lemma 5 that β is negative below the index b° and positive above it. If $b^\circ > \bar{b}$, then $(b-\bar{b})\beta(b) \ge (b^\circ - \bar{b})\beta(b)$ at every retained b, because on $[\bar{b}, b^\circ)$ both β and $b - b^\circ$ are negative and beyond b° both are positive. Integrating, $\int(b-\bar{b})\beta L_\gamma \ge (b^\circ - \bar{b})\int \beta L_\gamma = (b^\circ - \bar{b})\,r(\bar{b})L_\gamma(\bar{b}) > 0$. If $b^\circ \le \bar{b}$, then $\beta > 0$ on the retained range and the integrand of M is positive. □

Claim 23 (The effect through the equilibrium effort)**.** *Along the equilibrium the conjecture, which equals the equilibrium effort, $\hat{e}_1(\gamma) = e^*(\gamma)$, is differentiable, and there exist constants C_E and N^* such that $|\hat{e}_1'(\gamma)| \le$*

$C_E V_\gamma$ *and* $|\partial_{\hat{e}_1} T| \le N^*$ *at every prior. There exists* $\varsigma_{\text{conj}} \in (0, +\infty)$ *such that for all* $\varsigma \ge \varsigma_{\text{conj}}$*, at every zero of* T*,*

$$|\hat{e}_1'(\gamma)\, \partial_{\hat{e}_1} T| \;\le\; \tfrac{1}{2}\, M(\gamma).$$

Proof. The bound on $\hat{e}_1'$ is Step 1 below. For N^*, write $T = \omega[r(\bar{b})L_\gamma(\bar{b}) - \int \beta L_\gamma]$ as in Claim 19. Its retention term is at most the retention scale t_R of the proof of Lemma 4, of order $\bar{w}\bar{\Phi}(v_1)\mu_0/\sigma$, and its wage term is smaller, of order V_γ. Each derivative in the conjecture, of $\omega = \hat{e}_1/\sigma^2$, of the density L_γ or of the retention cutoff, multiplies by at most a constant times $(1 + x^2)/\hat{e}_1$, with x the standardized distance of the retention cutoff from the mean of the index, and $L_\gamma(\bar{b})(1 + x^2)$ is bounded by a constant times the peak of L_γ. So $|\partial_{\hat{e}_1} T| \le N^*$ with N^* a constant times $T/\hat{e}_1$, that is, a constant times $\alpha\bar{\Phi}(v_1)/\phi(v_1)$ by the value of $\hat{e}_1$ at the known complementarity. These two bounds do not give the last one, because M can itself be of order V_γ. The proof therefore bounds M from below and $\partial_{\hat{e}_1} T$ from above at a zero of T, in terms of the size of the wage force, which then cancels in the ratio. Write

$$A := \int_{\bar{b}}^{\infty} |\beta(b)|\, L_\gamma(b)\, db$$

for the size of the wage force, and $s = \hat{e}_1/\sigma$, from Claim 14, for the standard deviation of the index given the state.

Step 1: the conjecture moves slowly, $|\hat{e}_1'| \le C_E V_\gamma$**.** The equilibrium conjecture is the agent's fixed point at the principal's optimal cutoff, $\hat{e}_1(\gamma) = e(\kappa^*(\gamma); \gamma)$, so

$$\hat{e}_1' = \partial_\gamma e + \partial_\kappa e \cdot \frac{d\kappa^*}{d\gamma}.$$

At a fixed cutoff, $e(\kappa; \gamma)$ solves $F(e, \gamma; \kappa) = 0$, and $\partial_\gamma e = -\partial_\gamma F/\partial_e F$, where $\partial_e F$ is bounded below by the concavity slack and $\partial_\gamma F$ is a tilt covariance of the prior, $O(V_\gamma)$ by the tilt identities of Part (A). The principal's cutoff solves $\partial_\kappa \widetilde{\Pi} = 0$, so $d\kappa^*/d\gamma = -\partial^2_{\kappa\gamma}\widetilde{\Pi}/\partial^2_\kappa \widetilde{\Pi}$, with the denominator bounded away from zero by Claim 11 and the numerator again a tilt covariance, $O(V_\gamma)$. And $\partial_\kappa e$ is bounded by the implicit function theorem at the unique consistent effort (Lemma 4).

Step 2: M **is not small at a zero,** $M \ge c\, sA/\hat{e}_1$ **with** $A \ge cV_\gamma^2$**.** Two facts about the wage effect on the index distribution are used. First, the posterior mean moves by only $O(V_\gamma s)$ across the distribution: by the tilt identities of Part (A) it moves at rate $V_b \le 1/\varsigma$ per unit of index, and the index has standard deviation $s(1 + O(V_\gamma))$ by Claim 14. Second, β crosses zero once, at b°, with a slope bounded away from zero: by Lemma 5 it changes sign there, and by its proof $\partial_b \beta$ is at least a positive multiple of V_b^2 there, so on the index distribution $\beta(b) = k(b)(b - b^\circ)$ with k between two positive constants of the same order. In particular $A := \int |\beta| L_\gamma \ge cV_\gamma^2$ at a zero of T. Conditional on θ the index is Gaussian with standard deviation s, so at least half of its mass lies at distance at least $s/4$ from b°, and there $|\beta| \ge k_{\min}\, s/4$, with $k_{\min}$ a positive multiple of V_γ^2.

Now $M = \frac{\omega}{s^2} \int (b - \bar{b})\beta L_\gamma$ with $\omega/s^2 = 1/\hat{e}_1$. If $\bar{b} \le b^\circ$, then $M \ge \frac{\omega}{s^2} \int |b - b^\circ|\, |\beta| L_\gamma$, since β has the sign

of $b - b^\circ$ and $\int \beta L_\gamma = r(\bar{b})L_\gamma(\bar{b}) \geq 0$, and by the second fact and the Cauchy–Schwarz inequality,

$$\int |b - b^\circ|\,|\beta| L_\gamma \;\geq\; \frac{k_{\min}}{k_{\max}}\,\mathbb{E}|b - b^\circ| \int |\beta| L_\gamma \;\geq\; c\,sA.$$

If $\bar{b} > b^\circ$, then $\beta \geq 0$ on the retained range, and the zero condition $r(\bar{b})L_\gamma(\bar{b}) = \int \beta L_\gamma$, with $\beta = O(V_\gamma)$ and $r(\bar{b})$ of order one, places the retention cutoff below the mean of the index for ς large, since a cutoff above the mean would make the retention term exceed the wage term by a factor of order $1/V_\gamma$, so the mean distance above it is at least $c\,s$ and again $M \geq c\,sA/\hat{e}_1$.

Step 3: $\partial_{\hat{e}_1} T$ is not large at a zero, $|\partial_{\hat{e}_1} T| \leq KA\log(e/A)/\sigma^2$. In $T = \omega[r(\bar{b})L_\gamma(\bar{b}) - \int \beta L_\gamma]$, the derivative of $\omega = \hat{e}_1/\sigma^2$ multiplies the bracket, which vanishes at a zero. The retention term $S := r(\bar{b})L_\gamma(\bar{b})$ equals $\int \beta L_\gamma \leq A$ at a zero, and its relative rate of change in the conjecture, $\partial_{\hat{e}_1} S/S$, is at most a constant times $(1 + x^2)/\hat{e}_1$, with x the standardized distance of the retention cutoff from the mean of the index. Since $L_\gamma(\bar{b}) \leq S/r(\bar{b})$, $x^2 = O(\log(e/A))$. The wage term moves through β and through L_γ. The first moves the zero b° at rate $\partial_{\hat{e}_1} b^\circ = O(\hat{e}_1/\sigma^2)$, which by the second fact of Step 2 changes $\int |\beta| L_\gamma$ by at most $k_{\max}|\partial_{\hat{e}_1} b^\circ|$, a bounded multiple of $A\hat{e}_1/(\sigma^2 s)$ since $A \geq c\,k_{\min}\,s$. The second has relative rate of change $\partial_{\hat{e}_1} L_\gamma/L_\gamma$ at most a constant times $(1 + y_1^2/\sigma^2)/\hat{e}_1$, whose average against $|\beta|L_\gamma$ is a bounded multiple of $A/\hat{e}_1$. Multiplying by ω gives the bound.

Conclusion. Dividing,

$$\frac{|\hat{e}_1'\,\partial_{\hat{e}_1} T|}{M} \;\leq\; C_E V_\gamma \cdot \frac{KA\log(e/A)}{\sigma^2} \cdot \frac{\hat{e}_1}{c\,sA} \;=\; \frac{C_E K}{c}\,\frac{V_\gamma \log(e/A)}{\sigma},$$

and since $A \geq cV_\gamma^2$, $\log(e/A) \leq \log(e/(cV_\gamma^2))$. So the right side is at most $(C_E K/c)\,V_\gamma \log(e/(cV_\gamma^2))/\sigma$, which goes to zero as $V_\gamma \leq 1/\varsigma$ does, and ς_{conj} is the level at which it is below $\frac{1}{2}$. □

Along the equilibrium. The sign of the corrected manipulation term determines the sign of $e^* - e^B$ (Lemma 9).

The comparison of efforts. The benchmark principal buys the whole incentive with the wage, and the model's principal buys the incentive net of T. The next lemma turns the sign of T, corrected by a term that Step 1 of the assembly bounds, into the sign of $e^* - e^B$. Fix γ. Let $e^B(\gamma)$ and $z^B(\gamma)$ be the benchmark's equilibrium target and cutoff, $C(I;\tilde{e})$ the least-cost wage bill delivering marginal incentive I at target $\tilde{e}$, and $\mathcal{K}_1$ the band of Claim 20, an interval since the required wage incentive is increasing in the target by (U1), which contains every equilibrium target of the model and the benchmark's. The corrected manipulation term is

$$\Lambda(\gamma;\tilde{e}) := T(\gamma;\tilde{e}) + \lambda(\tilde{e})\,\partial_{\tilde{e}} T(\gamma;\tilde{e}),$$

where $\lambda := C_I/(c''C_{II} + C_{I\tilde{e}}) > 0$, with the derivatives of C taken at the target and the incentive it requires, and $\Lambda(\gamma) := \Lambda(\gamma; e^B(\gamma))$.

Lemma 9 (Comparison of efforts)**.**

(i) *For every equilibrium effort e^* of the model at γ, $e^* > e^B(\gamma)$ if $\Lambda(\gamma;e^*) > 0$ and $e^* < e^B(\gamma)$ if $\Lambda(\gamma;e^*) < 0$.*

(ii) *If her problem is strictly concave in the target at γ, then* $\operatorname{sgn}(e^*(\gamma) - e^B(\gamma)) = \operatorname{sgn}\Lambda(\gamma)$.

(iii) *In case (i) of Theorem 3, for $\sigma \geq \sigma^*$ there exists $\varsigma_{\text{pass}}(\alpha, \sigma) \in (0, +\infty)$ such that for all $\varsigma \geq \varsigma_{\text{pass}}$, every zero of Λ is a strict down-crossing.*

Proof. Both principals maximize $\tilde{e}\, m_\gamma - C(\cdot; \tilde{e}) + \Psi_1(\tilde{e})$ with the same continuation Ψ_1. The model's principal needs incentive $c'(\tilde{e}) - T(\tilde{e})$ from the wage where the benchmark's needs $c'(\tilde{e})$, so $\Pi^S = \Pi^B + \Delta$ with

$$\Delta(\tilde{e}) := C(c'(\tilde{e}); \tilde{e}) - C(c'(\tilde{e}) - T(\gamma; \tilde{e}); \tilde{e}),$$

the saving the continuation brings on the wage bill. Differentiating, with the mean value theorem applied to C_I and $C_{\tilde{e}}$ on an interval of length T,

$$\Delta'(\tilde{e}) = \left[c'' C_{II}(\xi) + C_{I\tilde{e}}(\xi')\right] T + C_I\, \partial_{\tilde{e}} T = \left[c'' C_{II}(\xi) + C_{I\tilde{e}}(\xi')\right] \Lambda(\gamma; \tilde{e}).$$

On the threshold family $C_I = 1/\kappa > 0$ by the pricing identity of Claim 9 and $C_{II} = 1/(\kappa^3 \bar{w} \mathcal{D}(\kappa)) > 0$, the score density $\mathcal{D}$ being positive, so $C_I / C_{II} = \kappa^2 \bar{w} \mathcal{D}(\kappa)$ is bounded, uniformly in the prior, by a constant c_λ. For $c'' \geq 2 \sup|C_{I\tilde{e}}| / \inf C_{II}$, one of the levels collected in σ^*, the bracket is at least $c'' C_{II}/2 > 0$, so Δ' has the sign of Λ and $\lambda \leq 2c_\lambda / c''$.

(i) By Lemma 4 the offered contract determines the agent's effort, so her problem is the choice of a target, and by Claim 20 every equilibrium target lies in $\mathcal{K}_1$ and maximizes Π^S there. It is interior, since Π^S is increasing below $\mathcal{K}_1$ and decreasing beyond $e^\dagger$ by Step 2 of the proof of that claim, so $\Pi^{S\prime}(e^*) = 0$ and

$$\Pi^{B\prime}(e^*) = -\Delta'(e^*),$$

which has the sign opposite to $\Lambda(\gamma; e^*)$. By Step 6 of the proof of Proposition 2, Π^B is strictly concave on $\mathcal{K}_1$ with its maximizer e^B in the interior, so $\Pi^{B\prime}(e^*) < 0$ exactly when $e^* > e^B$ and $\Pi^{B\prime}(e^*) > 0$ exactly when $e^* < e^B$, a single-crossing argument in the spirit of Milgrom and Shannon (1994).

(ii) At e^B, $\Pi^{B\prime}(e^B) = 0$, so $\Pi^{S\prime}(e^B) = \Delta'(e^B)$ has the sign of $\Lambda(\gamma; e^B) = \Lambda(\gamma)$. Strict concavity of Π^S in the target converts the sign of the derivative at e^B into the sign of $e^* - e^B$.

(iii) At a fixed target, Λ has the same form as T, an atom at the retention cutoff net of an integral over the retained range, with reweighted coefficients. The argument of Claim 22 then applies to it, which keeps the correction $\lambda\, \partial_{\tilde{e}} T$ together with T rather than bounding its derivative in the prior separately.

The form of Λ at a fixed target. Fix $\tilde{e}$ and write $T = \omega B$ with $B := r(\bar{b}) L_\gamma(\bar{b}) - \int_{\bar{b}}^{\infty} \beta L_\gamma$ (Claim 19). Here r, β and $\bar{b}$ depend on the target but not on the prior, and $\partial_{\tilde{e}} \omega = \omega / \tilde{e}$. Differentiating B in the target and collecting terms,

$$\Lambda(\gamma; \tilde{e}) = T + \lambda\, \partial_{\tilde{e}} T = \omega \Big[R_\gamma\, L_\gamma(\bar{b}) - \int_{\bar{b}}^{\infty} \hat{\beta}_\gamma\, L_\gamma \Big],$$

with a reweighted retention atom and a reweighted wage force,

$$R_\gamma := \Big(1 + \frac{\lambda}{\tilde{e}}\Big) r + \lambda [\partial_{\tilde{e}} r(\bar{b}) + \beta(\bar{b})\, \partial_{\tilde{e}} \bar{b}] + \lambda\, r [\partial_{\tilde{e}} \log L_\gamma(\bar{b}) + \partial_b \log L_\gamma(\bar{b})\, \partial_{\tilde{e}} \bar{b}],$$
$$\hat{\beta}_\gamma := \Big(1 + \frac{\lambda}{\tilde{e}}\Big) \beta + \lambda\, \partial_{\tilde{e}} \beta + \lambda \beta\, \partial_{\tilde{e}} \log L_\gamma.$$

Here $\partial_{\tilde{e}} r(\bar{b})$ is the total derivative of the atom's coefficient along the moving retention cutoff, and $\beta(\bar{b})\,\partial_{\tilde{e}}\bar{b}$ is the boundary term of the integral. By the proof of Claim 14, $\log L_\gamma(b) = -(b-\gamma)^2/(2s^2) + \log \tilde{Z}(b) - \log Z(\gamma)$, so the only factors that depend on the prior are

$$\partial_\gamma \partial_{\tilde{e}} \log L_\gamma(b) = -\frac{2}{\tilde{e}\, s^2}(b-\gamma) - Z_2(\gamma),$$

with $Z_2 := \partial_\gamma \partial_{\tilde{e}} \log Z$, and $\partial_\gamma \partial_b \log L_\gamma = 1/s^2$. The weight $1 + \lambda\, \partial_{\tilde{e}} \log L_\gamma$ on β is positive, since $\partial_{\tilde{e}} \log L_\gamma$ is $y_1^2/(\sigma^2 \tilde{e}) \geq 0$ up to terms of order $1/\tilde{e}$, and $\lambda\, \partial_{\tilde{e}}\beta$ moves the zero of β by $O(\lambda\tilde{e}/\sigma^2)$, small next to the standard deviation s of the index. So $\hat{\beta}_\gamma$ has the sign structure of β: it crosses zero once, at $\hat{b}^\circ$, with a nonzero slope, negative below and positive above.

The slope at a fixed target. At a zero of $\Lambda(\cdot;\tilde{e})$, $R_\gamma L_\gamma(\bar{b}) = \int \hat{\beta}_\gamma L_\gamma$, and exactly as in Claim 22, using $\partial_\gamma \log L_\gamma = (b-\gamma)/s^2 - \mu_\gamma$ from Claim 14,

$$\partial_\gamma \Lambda(\gamma;\tilde{e}) = -\widetilde{M} + W,$$

where

$$\widetilde{M} := \frac{\omega}{s^2}\int_{\bar{b}}^{\infty} (b-\bar{b})\,\hat{\beta}_\gamma L_\gamma$$

is the slope for the reweighted objects and

$$W := \omega\Big[(\partial_\gamma R_\gamma) L_\gamma(\bar{b}) - \int_{\bar{b}}^{\infty} (\partial_\gamma \hat{\beta}_\gamma) L_\gamma\Big]$$

collects the derivatives of the weights. The argument of Claim 23, applied to $\hat{\beta}_\gamma$, gives $\widetilde{M} \geq c\, s\hat{A}/\tilde{e}$ with $\hat{A} := \int |\hat{\beta}_\gamma| L_\gamma$. With $S := r(\bar{b}) L_\gamma(\bar{b})$, the displayed derivatives give

$$W = \frac{2\omega\lambda}{\tilde{e}\, s^2}\Big[\int_{\bar{b}}^{\infty} (b-\gamma)\beta L_\gamma - (\bar{b}-\gamma) S\Big] - \omega\lambda Z_2\Big[S - \int_{\bar{b}}^{\infty} \beta L_\gamma\Big] + \frac{\omega\lambda}{s^2}\,\partial_{\tilde{e}}\bar{b}\; S.$$

At a zero of Λ, $S - \int \beta L_\gamma$ is of order $\lambda/\tilde{e}$ relative to S, and the first bracket is $\int (b-\bar{b})\beta L_\gamma$ up to that order. This holds whether the retention cutoff sits near the mean of the index, where $\bar{b} - \gamma$ is moderate, or deep in its lower tail, where S is negligible and $\int (b-\gamma)\beta L_\gamma = \int (b-\hat{b}^\circ)\beta L_\gamma$ up to a term of the same order. And $\partial_{\tilde{e}}\bar{b}/s^2 = O(\bar{\theta}/\tilde{e})$. Hence

$$|W| \;\leq\; \frac{2\lambda}{\tilde{e}}\,\widetilde{M}\Big(1 + K\frac{\lambda}{\tilde{e}} + K\frac{\tilde{e}\,\bar{\theta}}{\sigma}\Big).$$

The sign. By Step 1 of the assembly, $\lambda/\tilde{e} \leq 4v_1^2(1+\epsilon)$. Every zero of Λ lies at a prior whose index mass is beyond the zero of the wage effect on balance, where $v_1 \leq v^\circ(1 + O(\sqrt{V_\gamma}))$ with $v^\circ < 1$ the standardized cutoff at the minimizer of the cutoff (Claim 18), and $v^\circ = O(\sigma^{-1/2})$ (Step 4 of the assembly), so $|W| \leq \widetilde{M}/2$ for $\sigma \geq \sigma^*$, and $\partial_\gamma \Lambda(\gamma;\tilde{e}) \leq -\widetilde{M}/2$.

Moving the target with the prior. Along the family,

$$\frac{d\Lambda}{d\gamma} = \partial_\gamma \Lambda(\gamma;\tilde{e})\Big|_{\tilde{e}=e^B} + \frac{de^B}{d\gamma}\,\partial_{\tilde{e}}\Lambda + \partial_\gamma \lambda\; \partial_{\tilde{e}} T.$$

By Proposition 2, e^B is C^1 with $|de^B/d\gamma| \le C_E V_\gamma$: as in Claim 23, the benchmark effort at a fixed cutoff and the benchmark's optimal cutoff both move with the prior through tilt covariances, over concavity moduli, and likewise $|\partial_\gamma \lambda| \le C_E V_\gamma$. The bound of Claim 23 on $\partial_{\tilde e} T$, taken one derivative further, gives $|\partial_{\tilde e}\Lambda| \le K\hat{A}\log^2(e/\hat{A})/\sigma^2$, each derivative in the target adding at most a factor $(1+x^2)/\tilde e$. Relative to $\widetilde{M}$, the last two terms are therefore of order $V_\gamma \log^2(1/V_\gamma)/\sigma$, below $\frac{1}{4}$ for ς large, and $d\Lambda/d\gamma \le -\widetilde{M}/4 < 0$ at every zero of Λ. □

Part (F): assembling the proof

Proof of Theorem 3. The four steps combine Parts (B)–(D). Step 1 shows that the correction in Lemma 9(i) never flips the sign of T where the sign is used. Steps 2 and 3 treat the two cases, and Step 4 removes the concavity that Step 2 uses.

Step 1: the correction. Let $\Lambda := T + \lambda\,\partial_{\tilde e} T$ be the corrected comparison of Lemma 9, whose sign is the sign of $e^* - e^B$.

The slope of T in the target. In Claim 19 the target enters T through the coefficient $\omega = \tilde e/\sigma^2$, through the density L of the realized index, whose mean $\gamma + \iota\mu_\gamma$, with $\iota := \tilde e^2/\sigma^2$, and scale $s = \tilde e/\sigma$ move with $\tilde e$, and through the modified base, which changes $r(\bar b)$ and β by relative amounts of order $\tilde e\theta_{\max}^2/\sigma^2$. Since $\omega L_\gamma(b) = \phi(v)/\sigma$ at standardized distance v from the expected index,

$$\partial_{\tilde e}\log(\omega L) = \frac{v^2}{\tilde e} + \frac{2\mu_\gamma v}{\sigma} \;\ge\; -\frac{\mu_\gamma^2 \tilde e}{\sigma^2},$$

so the $\tilde e$-derivative of each term of T is at least $-C_T$ times that term, with $C_T := \theta_{\max}^2 \bar e\,(1 + c_{\text{base}})/\sigma^2$ and c_{base} the constant bounding the modified base's relative effect. The bound is one-sided, since $v^2/\tilde e$ is of order $1/\tilde e$. At optimistic priors in case (i) an upper bound holds as well: there T is the wage integral up to the retention atom, which is exponentially small because the retention cutoff lies many standard deviations below the mean of the index (Lemma 7), and $\tilde e\,\partial_{\tilde e}$ of the integral reweights it by v^2, so $|\tilde e\,\partial_{\tilde e} T| \le c_2|T|$ with c_2 the second moment of the standardized index under the weighting by the wage effect, bounded since the wage effect is nearly constant over the index there.

The size of λ. By the proof of Lemma 9, $\lambda \le 2c_\lambda/c''$ for $\sigma \ge \sigma^*$, so at every admissible curvature

$$\lambda C_T \;\le\; \frac{2c_\lambda\,\theta_{\max}^2 \bar e\,(1 + c_{\text{base}})}{c''\,\sigma^2} \;<\; \tfrac{1}{3}.$$

At optimistic priors in case (i), with $\tilde e \ge I_1/(2c'')$ on $\mathcal{K}_1$, $|\lambda\,\partial_{\tilde e} T| \le 4c_\lambda c_2|T|/I_1$, and $c_\lambda/I_1 = \kappa^2\bar w\mathcal{D}(\kappa)/I_1 = v_1^2(1 \pm \epsilon)$ with $v_1 \le \bar\alpha\sigma/\mu_0^2 = O(\sigma^{-1/2})$ by Claim 16, so $|\lambda\,\partial_{\tilde e} T| \le |T|/3$ there for $\sigma \ge \sigma^*$.

Conclusion. In case (ii), where both terms of T are nonnegative, $\Lambda \ge \frac{2}{3}T$ at every prior. In case (i), $\Lambda \ge \frac{2}{3}T$ at γ_L, where the atom dominates and the lower bound applies, and $\Lambda \le \frac{2}{3}T$ at optimistic priors by the upper bound.

Step 2: case (i) where her problem is concave. By Lemma 6, $T(\gamma_L) > 0$, and by Lemma 7, $T < 0$ at optimistic priors, so by Step 1, $\Lambda(\gamma_L) > 0$ and $\Lambda < 0$ at optimistic priors. Λ is continuous in γ, so it has a zero, and by Lemma 9(iii) every zero is a strict down-crossing. Hence Λ has exactly one zero γ^*, and by

Lemma 9(ii) $e^* > e^B$ on (γ_L, γ^*) and $e^* < e^B$ beyond.

Step 3: case (ii). For $\alpha \geq \bar{\alpha}$, Lemma 5 gives $\beta \leq 0$ in the family, so both terms of T are nonnegative and $T > 0$ at every prior and every conjecture in the range. By Step 1, $\Lambda(\gamma; \tilde{e}) \geq \frac{2}{3}T > 0$ at every prior and target, so by Lemma 9(i) every equilibrium over-experiments.

Step 4: case (i) where her problem is not concave. Step 2 used the concavity of her problem, through Lemma 9(ii) and (iii), and concavity can fail only near initiation, where it is not needed. Where the expected index lies below the zero b° of the wage effect (Lemma 5) by more than $c\sqrt{V_\gamma}$, both terms of T are positive up to contributions that are exponentially small in the standardized distance to the cutoff, so $\Lambda(\gamma; \cdot) \geq \frac{2}{3}T > 0$ on $\mathcal{K}_1$ by Step 1, and every equilibrium over-experiments by Lemma 9(i). At optimistic priors, $\Lambda(\gamma; \cdot) < 0$ on $\mathcal{K}_1$ by Lemma 7 and Step 1, and every equilibrium under-experiments. At the remaining priors Claim 21 gives either $T > 0$ on $\mathcal{K}_1$, and again every equilibrium over-experiments, or strict concavity, where the equilibrium is unique, moves continuously with the prior, and by Lemma 10 every zero of T along it is a down-crossing. At the boundary between the two sets $T > 0$ along the equilibrium, since the retention term is at least twice the wage term there. So T has a single down-crossing γ^*, and every equilibrium over-experiments on (γ_L, γ^*) and under-experiments on $(\gamma^*, +\infty)$, the endpoint γ_L excluded only because the principal is indifferent there between hiring and not. □

The levels of Theorem 3.

The proof above imposes finitely many requirements of the form "σ large enough" or "ς large enough", each uniform in the prior, and defines σ^* and ς^* as the largest of their levels. Every constant in them is a supremum or infimum over the set of parameter values the posteriors take, replaced by a bound in primitives in the direction that keeps the requirement sufficient. That set is described by $e, \hat{e}_1 \leq \bar{e}$, retained posterior means $\mu_b \in [\underline{\mu}, \bar{\theta}]$, standardized cutoffs in the interval spanned by the known-complementarity cutoff $\alpha\sigma/\mu^2$ for $\mu \in [\underline{\mu}, \bar{\theta}]$, $\varkappa \leq (\bar{w}/\sigma)\phi(0)$, posterior precision at least ς, and $\mu_- \leq C_0 e^{-\varsigma\underline{\mu}^2/2}$. The noise level is

$$\sigma^*(\alpha) := \max\left\{\sigma^{\mathrm{sc}},\ \sigma^B,\ \sigma^U,\ \sigma_B(\alpha),\ \sigma_\alpha,\ \sigma_T(\alpha),\ \sigma_H(\alpha),\ \frac{\bar{w}\theta_{\max}\bar{\nu}K'^2}{2\underline{C}_I}\right\},$$

which may depend on α and on the fixed primitives $\bar{w}$, Θ and π_0, the last through $\underline{\mu}$, the lowest retained mean. Superscripts mark levels stated in results of the text, and subscripts levels used only here. So the levels are of two kinds. $\sigma_B(\alpha)$, σ_α, $\sigma_T(\alpha)$ and $\sigma_H(\alpha)$ have the form $(K/\alpha)^{1/2}$, that is, they are the requirement $\alpha \geq K\sigma^{-2}$ with $K \in \{K_B, K_U, K_T, K_H\}$ a constant of the primitives. The others do not depend on α. With $\underline{K}$ the largest of the four constants and σ_0 the largest of the α-independent levels,

$$\sigma^*(\alpha) \ \leq\ \max\{\sigma_0, (\underline{K}/\alpha)^{1/2}\} \qquad \text{on } \alpha < \bar{\alpha},$$

The precision level is

$$\varsigma^*(\alpha, \sigma) := \max\{\varsigma^R, \varsigma^B, \varsigma^U, \varsigma_\alpha, \varsigma_{\mathrm{exp}}, \varsigma_{\mathrm{peak}}, \varsigma_{\mathrm{conc}}, \varsigma_{\mathrm{init}}, \varsigma_H, \varsigma_{\mathrm{conj}}, \varsigma_{\mathrm{pass}}\},$$

which also depends on the fixed primitives $\bar{w}$, Θ and π_0.

The levels, what each enforces, and where it is established:

Noise levels		
$\sigma^{\mathrm{sc}} = \bar{w}/2$	retention is a threshold rule, and $v_\mu < 1$ at the minimizer of the cutoff	Lemma 3, Claim 18
σ^B, σ^U	the levels of the benchmark and of the unique implemented effort	Proposition 2 and Lemma 4
$\sigma_B(\alpha) = (K_B/\alpha)^{1/2}$	$\alpha > \alpha^B = C_1$, since $C_1 \le K_B\sigma^{-2}$	Step 2 of the proof of Corollary 2
$\sigma_\alpha = (K_U/\alpha)^{1/2}$	$\sup_\gamma \alpha^U(\gamma) < \alpha$ for $\varsigma \ge \varsigma_\alpha$, since $\sup_\gamma \alpha^U(\gamma) \le K_U\sigma^{-2}$ there, so existence and a unique effort at every prior at which she hires	the same step
$\sigma_T(\alpha) = (K_T/\alpha)^{1/2}$	$\lambda C_T < \frac{1}{3}$, where λC_T is at most a constant of the primitives times $(\alpha\sigma^2)^{-2}$	Step 1 of the assembly
$\sigma_H(\alpha) = (K_H/\alpha)^{1/2}$	a better prior's direct gain to the principal dominates the ratchet's erosion of the continuation incentive	Lemma 8
$\bar{w}\theta_{\max}\bar{\nu}K'^2/(2\underline{C}_I)$	the level of the adoption application	proof of Proposition 3
Precision levels		
ς^R, ς^B, ς^U	the levels of the threshold lemma, the benchmark and the unique implemented effort	Lemma 3, Proposition 2 and Lemma 4
ς_α	the concentration corrections to α^U are at most its leading term	Step 3 of the proof of Corollary 2
$\varsigma_{\exp}$	$:= \max\{8\bar{e}^2/\sigma^2, 16/\underline{\mu}^2\}$, so that every denominator in the concentrated-posterior expansion is within a factor two of its point-mass value	Claim 17
$\varsigma_{\mathrm{peak}}$	$:= \max\{\varsigma_{\mathrm{turn}}, \varsigma_{\mathrm{away}}\}$ with $\varsigma_{\mathrm{turn}} := \sup_{\lvert R\rvert\le\delta}(2K_2/z^{*\prime\prime})^2$ and $\varsigma_{\mathrm{away}} := \sup_{\lvert R\rvert\ge\delta}\max\{(4\lvert A\rvert C_3/c_\delta)^2, 4K_1/c_\delta\}$, $c_\delta := \inf_{\lvert R\rvert\ge\delta}\lvert z^{*\prime}\rvert$: near the minimizer of the cutoff its curvature, and away from it its slope, dominate the concentration correction	Lemma 5
$\varsigma_{\mathrm{conc}}$	her reduced problem is concave wherever the retention term is small	Claim 21
$\varsigma_{\mathrm{init}}$	the two conditions of the pessimistic-prior lemma hold	Step 6 of the proof of Lemma 6
ς_H	hiring is a threshold in the prior	Lemma 8
$\varsigma_{\mathrm{conj}}$	$(C_E K/c)\,V\log(e/(cV^2))/\sigma < 1$ with $V = 1/\varsigma$	Claim 23
$\varsigma_{\mathrm{pass}}$	every zero of Λ is a strict down-crossing	Lemma 9(iii)

Each supremum is over the set described above and is finite.

Proof of Corollary 2. The proof shows that $\bar{\alpha}$ falls like $\sigma^{-3/2}$ as the noise grows while every lower bound on the curvature collected in $\sigma^*(\alpha)$ falls like σ^{-2}, and then checks regime (P2) on the resulting interval.

Step 1: the boundary.

By Claim 16, $\bar{\alpha}$ exists, is unique, and equals $\bar{\theta}^2\sqrt{\bar{w}\phi(0)}\,\sigma^{-3/2}(1+O(\sigma^{-1}))$.

Step 2: the thresholds.

The curvature levels α^B and $\alpha^U(\gamma)$ are each of order σ^{-2}, with constants of the primitives, because every requirement collected in them is of one of three types: a lower bound on the curvature that falls like σ^{-2}, a condition on σ alone, or a condition on ς alone. The level α^{FOA} is of the first type, with constants built from $\bar{w}$, $\theta_{\max}$, the Gaussian moments $\nu_1 = O(\sigma^{-1})$ and $\nu_2 = O(\sigma^{-2})$ and the derivative bounds of Claim 4, none of which depends on the prior beyond its support. In the condition (U) of Lemma 4, (U1) and (U3) are of the first type, with constants that depend on ν_1 only through c_w, which is bounded below where ν_1 is bounded, (U2) adds a condition on σ alone, and (U4) is a condition on ς alone, all as evaluated in Step 4 of the proof of that proposition. The bound (S) is a condition on σ alone, and the concavity condition of Claim 21 is a level on ς, which enters ς^*.

Hence $\sup_\gamma \alpha^U(\gamma) \le K_c\,\sigma^{-2} + \eta(\sigma,\varsigma)$ for a constant K_c of the primitives, where $\eta(\sigma,\varsigma) = O(\varsigma^{-1/2}\sigma^{-2})$ collects the $O(\sqrt{V})$ corrections at a concentrated prior, and $\alpha^B = C_1 = 4\phi(1)\bar{w}\theta_{\max}^2/\sigma^2 \le K_B\sigma^{-2}$ (Step 1 of the proof of Proposition 2).

Step 3: the comparison.

By Step 2, for $\varsigma \ge \varsigma_\alpha$, the level at which $\eta \le K_c\sigma^{-2}$, $\sup_\gamma \alpha^U(\gamma) \le K_U\sigma^{-2}$ with $K_U := 2K_c$, and $\alpha^B \le K_B\sigma^{-2}$: these are the constants of the levels σ_α and $\sigma_B(\alpha)$ in $\sigma^*(\alpha)$, and with them $\sigma^*(\alpha) \le \max\{\sigma_0, (\underline{K}/\alpha)^{1/2}\}$ on $\alpha < \bar{\alpha}$, as shown after the definition of $\sigma^*(\alpha)$. So for $\sigma \ge \sigma_0$, every α with $\underline{K}/\sigma^2 < \alpha < \bar{\alpha}(\sigma)$ satisfies $\sigma \ge \sigma^*(\alpha)$. By Step 1, $\bar{\alpha}(\sigma) \ge \overline{K}\sigma^{-3/2}$ for $\sigma \ge \sigma_1$, so the interval is nonempty for $\sigma \ge \max\{\sigma_0, \sigma_1, (\underline{K}/\overline{K})^2\}$.

Step 4: regime (P2).

At a known complementarity $\theta > 0$ the second-period profit is $\pi(\alpha\sigma/\theta^2)$, with $\pi(v) := \bar{w}[\phi(v)/v - \bar{\Phi}(v)]$ strictly decreasing from $+\infty$ to 0, and it is zero at $\theta \le 0$. Since $\underline{\theta} \le 0$, $\pi^- = 0 < \pi_0$. Let v_A solve $\pi(v_A) = \pi_0$, so that $\pi_0 < \pi^+$ if and only if $\alpha\sigma < v_A\bar{\theta}^2$. On the interval, $\alpha\sigma < \sigma\bar{\alpha}(\sigma) = \bar{\theta}^2 v^\circ(\sigma)$, where $v^\circ(\sigma)$ solves $v^2(1-\varkappa) = \varkappa$ with $\varkappa = \bar{w}\phi(v)/\sigma$ (proof of Claim 16), and $v^\circ(\sigma)$ decreases to zero as σ grows, since $\varkappa$ decreases in σ at every v. So $\pi_0 < \pi^+$ on the whole interval once $v^\circ(\sigma) \le v_A$, that is, for $\sigma \ge \sigma_A$. With $\sigma_{\min} := \max\{\sigma_0, \sigma_1, (\underline{K}/\overline{K})^2, \sigma_A\}$, for every $\sigma \ge \sigma_{\min}$ the interval is nonempty and every curvature in it satisfies $\sigma \ge \sigma^*(\alpha)$, (P2) and $\alpha < \bar{\alpha}(\sigma)$, which proves the corollary. The constants σ_0, σ_1, σ_A, $\underline{K}$ and $\overline{K}$ depend only on $\bar{w}$, Θ and π_0, since every level they come from does.

When $\underline{\theta} > 0$, instead, $\pi^- = \pi(\alpha\sigma/\underline{\theta}^2)$ is below π_0 if and only if $\alpha\sigma > v_A\underline{\theta}^2$. On the interval $\alpha\sigma > \underline{K}/\sigma$, so (P2) holds on the whole interval for $\sigma_{\min} \le \sigma < \underline{K}/(v_A\underline{\theta}^2)$, a nonempty range when $v_A\underline{\theta}^2\sigma_{\min} < \underline{K}$. This is the statement after the corollary for a productive worst complementarity. □

Proof of Proposition 3 (Technology adoption)

The comparison of efforts evaluates the manipulation term at the benchmark's target, the adoption decision at the equilibrium's. The two switch at nearby priors.

Lemma 10 (The manipulation term crosses zero only downward)**.** *There exists* $\varsigma_{\text{conj}} \in (0, +\infty)$ *such that for all* $\varsigma \geq \varsigma_{\text{conj}}$*, at every zero of* T *along the equilibrium,* $dT/d\gamma < 0$*.*

Proof. In case (i), every zero of T lies at a prior whose mean is at least $\mu^\circ - O(\sqrt{V_\gamma})$: at a more pessimistic prior every retained posterior has mean below the minimizer of the cutoff, so $\beta < 0$ on the retained range by Lemma 5, and both terms of T in Claim 19 are positive. At a zero of T alternative (b) of Claim 21 holds, so the equilibrium is unique there. So the equilibrium is well defined near every zero. Along it the conjecture is correct, $\hat{e}_1(\gamma) = e^*(\gamma)$, and the equilibrium effort moves with the prior, so

$$\frac{dT}{d\gamma} = T'(\gamma) + \hat{e}_1'(\gamma)\, \partial_{\hat{e}_1} T,$$

where T' is the derivative in the prior with the conjecture held at its equilibrium value $e^*(\gamma)$. Claim 22 shows that $T' = -M$ at a zero, with $M > 0$. Claim 23 shows that the effect through the movement of the equilibrium effort is at most $\frac{1}{2}M$ in absolute value. Together, $dT/d\gamma \leq -\frac{1}{2}M < 0$. □

Claim 24 (The two switches are close)**.** *In case (i), the zero* γ^{**} *of* $T(\gamma; e^*(\gamma))$ *is unique, and* $|\gamma^* - \gamma^{**}| \leq \delta^* := \sup|\lambda\, \partial_{\tilde{e}} T| / \inf|dT/d\gamma|$*, the supremum and infimum over a neighborhood of* γ^{**}*.*

Proof. $T(\cdot; e^*(\cdot))$ is positive at γ_L (Lemma 6) and negative at optimistic priors (Lemma 7), continuous in γ, and by Lemma 10 every zero is a strict down-crossing, so the zero is unique. At γ^*, $e^B(\gamma^*) = e^*(\gamma^*)$ by Lemma 9(ii), so the condition $\Lambda(\gamma^*) = 0$ reads $T(\gamma^*; e^*(\gamma^*)) = -\lambda\, \partial_{\tilde{e}} T$, and $|T(\gamma^*; e^*(\gamma^*))| \leq \sup|\lambda\, \partial_{\tilde{e}} T|$. Since $T(\cdot; e^*(\cdot))$ vanishes at γ^{**}, the mean value theorem between γ^{**} and γ^* gives the bound. □

Proof. Without the technology there is nothing to learn, so the agent's continuation contains no information, $T \equiv 0$, and the equilibrium and benchmark problems coincide: Π^0 is common to the two, so the comparison of adoption reduces to the sign of $\Pi(\gamma) - \Pi^B(\gamma)$, where $\Pi(\gamma) = \Pi^S(e^*(\gamma))$ and $\Pi^B(\gamma) = \Pi^B(e^B(\gamma))$ are the two principals' values at prior γ, the maxima of their reduced objectives Π^S and Π^B over the target. With the technology, the continuation is the same function of first-period effort in both problems, so implementing a given target $\tilde{e}$ at prior γ yields the same continuation value, and the two problems differ only in the wage bill. Write $C(I; \tilde{e})$ for the least-cost wage bill delivering marginal incentive I at target $\tilde{e}$ (Lemma 9). The model's principal needs $c'(\tilde{e}) - T(\gamma; \tilde{e})$ from the wage where the benchmark's needs $c'(\tilde{e})$, and C is increasing in I on the branch on which the cutoff is above the expected signal.

Decomposition. Write $S(\gamma, \tilde{e}) := C(c'(\tilde{e}); \tilde{e}) - C(c'(\tilde{e}) - T(\gamma; \tilde{e}); \tilde{e})$ for the wage saving the continuation incentive provides at target $\tilde{e}$, so that $\Pi^S(\tilde{e}) = \Pi^B(\tilde{e}) + S(\gamma, \tilde{e})$. Then

$$\Pi(\gamma) - \Pi^B(\gamma) = \Pi^S(e^*) - \Pi^B(e^B) = S(\gamma, e^*) - [\Pi^B(e^B) - \Pi^B(e^*)].$$

The bracket is nonnegative, since e^B maximizes Π^B, and second order in the distortion $e^* - e^B$, since the first derivative of Π^B vanishes at e^B. Three constants enter, each a supremum or infimum over the compact ranges of $(\gamma, \tilde{e}, \kappa)$ the two problems visit at the given (σ, ς), and each finite by the bounds already established: $\bar{\nu} := \sup|\partial_{\tilde{e}}^2 \Pi^B|$, finite by the dominations of Lemma 1. $\underline{C}_I := \inf \partial_I C = 1/\sup \kappa > 0$, since $\partial_I C = 1/\kappa$ by the pricing identity of Claim 9 and the cutoffs on the branch are bounded over the compact range of targets,

and $K := \sup|\partial_T e^*|$, the sensitivity of the model's target to T in Lemma 9(ii), bounded by $C_I/\underline{\nu}$ with $\underline{\nu}$ the concavity modulus of Π^S from Claim 11 in the preliminaries of Appendix B. With these, the strong concavity of Π^B gives $\Pi^B(e^B) - \Pi^B(e^*) \le \frac{1}{2}\bar{\nu}(e^* - e^B)^2$ and Lemma 9(ii) gives $|e^* - e^B| \le K|T(\gamma; e^B)|$. Since $|T(\gamma; e^B) - T(\gamma; e^*)| \le N^*|e^B - e^*|$, this yields $|e^* - e^B| \le K'|T(\gamma; e^*)|$ with $K' := K/(1 - KN^*)$, provided $KN^* < 1$. This holds for $\sigma \ge \sigma^*$: K is the coefficient λ of Lemma 9, at most $4\nu_1^2\bar{e}\,(1+\epsilon)$ by Step 1 of the assembly of that theorem's proof, and N^* is at most a constant times $\alpha\bar{\Phi}(\nu_1)/\phi(\nu_1)$ by Claim 23, so with $\bar{e}$ of order $\bar{w}\phi(\nu_1)\mu_0/(\alpha\sigma)$ the product KN^* is at most a constant times $\nu_1^2\bar{\Phi}(\nu_1)\,\bar{w}\bar{\theta}/\sigma$, which vanishes as σ grows since $\nu^2\bar{\Phi}(\nu)$ is bounded. The level at which $KN^* < 1$ is one of those collected in σ^*. The saving is first order: S has the sign of $T(\gamma; e^*)$ and $|S| \ge \underline{C}_I|T(\gamma; e^*)|$.

The sign. Where $T(\gamma; e^*) \le 0$, $S \le 0$ and the bracket is nonnegative, so $\Pi \le \Pi^B$. Where $T(\gamma; e^*) > 0$, and whenever $T < 2\underline{C}_I/(\bar{\nu}K'^2)$,

$$\Pi - \Pi^B \;\ge\; T\big(\underline{C}_I - \tfrac{1}{2}\bar{\nu}K'^2\,T\big) \;>\; 0.$$

The manipulation term is bounded at every prior: by (8) it is the integral of the retained continuation rent against $\partial_e f_1$, the rent is at most $\bar{w}$, and $\int|\partial_e f_1| \le \theta_{\max}\nu_1 = 2\phi(0)\theta_{\max}/\sigma$, so $T \le \bar{w}\theta_{\max}/\sigma$, so the condition holds at every prior once $\sigma > \bar{w}\theta_{\max}\bar{\nu}K'^2/(2\underline{C}_I)$, one of the finitely many levels collected in σ^* in the proof of Theorem 3. Hence $\Pi - \Pi^B$ has the sign of $T(\gamma; e^*(\gamma))$ at every prior.

Locating the threshold. For $\alpha < \bar{\alpha}$, $T(\gamma; e^*(\gamma))$ has a unique zero γ^{**}, positive below and negative above, with $|\gamma^{**} - \gamma^*| \le \delta^*$, by Claim 24. So there is weakly more adoption below γ^{**} and weakly less above it. For $\alpha \ge \bar{\alpha}$, $T > 0$ at every prior (Step 3 there), so there is weakly more adoption throughout. □