Initiative

Hector Chade[*] Jeroen Swinkels[†]

April 2025

Abstract

In many real-world principal-agent settings, the principal must design incentives to both induce hard work *and* to encourage risky *initiative* instead of safer projects. We provide conditions such that extreme outputs will be rewarded more and middle outputs less than in the classic moral hazard setting, giving an alternative explanation for option-like incentives. We exhibit the structure of optimal contracts when these conditions are not satisfied. Faced by the need to induce initiative, the principal will tend to ask less of the agent if effort is not very important, but ask *more* if effort is important. Effectively, the principal goes big or goes home.

Keywords. Moral Hazard, Project Selection, First-Order Approach, Principal-Agent Problem.
JEL Classification. D82, D86.

[*]Arizona State University, hector.chade@asu.edu
[†]Northwestern University, j-swinkels@northwestern.edu
We are grateful to seminar participants at Northwestern, UCLA, Bocconi, SED-Barcelona, Georgetown, Yale, Boston College, Johns Hopkins, and Washington University, and to Dan Barron, Meghan Busse, Brendan Daley, and Qingmin Liu for helpful comments. Ramtin Salamat and Shusheng Zhong provided excellent research assistance.

1 Introduction

In the classic moral hazard problem, the principal's only problem is to induce the agent to work hard. But, in many real world settings, the agent also chooses on *what* to work. Assume that GM's board has decided on an aggressive transition to electric vehicles. Hence, they want two things from their CEO, Mary Barra. First, as is standard, they want her to work very hard. But, they also want Barra to favor electric over traditional, and not all of her choices of whether to do so are observable. For example, while the board can see the timing of plant transitions from traditional to electric vehicles, a moderate rate of transition could reflect either that Barra was playing it safe, or that she was aggressively pursuing the strategy, but that stochastic market or technological considerations hampered a faster transition. Thus, the same set of rewards that guide her effort choice must also partly guide her degree of *initiative* in pursuing electric. And, to the extent that taking more initiative leads to riskier outcomes, GM needs to be aware that exposing Barra to risk, which is effective at motivating effort, may disincentivize initiative.

Most academics have available "safe" projects that will lead with high probability to publishable output. Our employers (and society), however, may prefer that we take on projects that may turn out to be impossible, but will make a more substantive contribution if successful: the university wants us to *both* show initiative in choosing innovative projects, and then work very hard to make them succeed. And here again, the university can only make a noisy estimate of whether our research agenda is "safe" or ambitious, while we may know quite well. Hence there is a clear tension. Providing poor payoffs in the face of low research output is one very effective way to disincentivize low effort. But "no output" is also the modal outcome for many projects that push the frontiers. Punishing low output thus incentivizes effort, but disincentivizes initiative.

The need to encourage initiative is not just relevant at the top of the firm, or for employees for whom innovation is key. Consider a firm motivating a salesperson. Some clients are highly probable to do some business with the firm, but of limited magnitude. Other clients are more speculative, but have the potential to make large orders. If the type of client pursued is visible to the salesperson but not to the firm, then the firm must use its reward structure both to encourage the pursuit of the right client and to encourage serious effort in doing so. Similar issues arise when an employee negotiating on behalf of the firm is deciding whether to pursue an easy deal or push hard for a better one. Organizations benefit from initiative at all levels.

In this paper, we consider a principal who needs to motivate both effort and initiative. We model a lack of initiative as taking a "safe" action which leads to output which is relatively unlikely to be either very high or very low. In contrast, taking intiative (declining the safe action) places the agent in the classic moral hazard setting (Mirrlees (1975), Holmström (1979)) where effort determines the distribution over outcomes on a risky project. We thus have the need to encourage initiative while maintaining the full richness of the canonical moral hazard setting. As such, our model allows a nuanced understanding of how initiative and effort interact.

We provide a comprehensive analysis of this problem and how it compares to the classic moral-hazard problem. At a high level, there are two main economic insights. First, under reasonable conditions, the optimal contract facing the initiative constraint will cross the contract without the constraint twice, once from above and then once from below. Indeed, in an important class, the need to induce initiative leads to a more convex compensation schemes. Second, there is a tendency for the effort implemented to be pushed away from middle levels with the new constraint. If output is not of very high value, the principal will tend to induce lower effort (or indeed the safer project) facing the need to induce initiative, but if output is of significant value, then the principal will induce higher effort given the extra constraint.

The result that incentives tend to convexify when initiative is added to the model reflects a simple trade-off. When initiative is taken, low outputs become more likely. So, low outputs, while bad news about effort, are good news about initiative. In the face of these mixed messages, the principal does not punish low output as harshly as when initiative is not a consideration. Similarly, medium outputs, while favorable news about effort, are less good news about initiative, and so rewards are lower than before. Finally, high outcomes are good news about *both* effort and initiative, and so are rewarded generously. This suggests a reason why real-world incentive schemes, such as options-based contracts for CEOs and the compensation of tenured academics, seem to be steep in the face of success but flatter in the face of failure. Indeed, if the safe project is sufficiently appealing, then the optimal contract may be non-monotone.[1]

The fact that the need to motivate initiative leads to contracts that punish failure less harshly has precedents both in the literature (see below), and in the popular press. We add significant nuance in two ways. First, we emphasize that the reason why the agent may fail to show initiative is not just because he is afraid of failure, but also because middling outcomes may be too well compensated in the contracts that naturally arise when only moral hazard on effort is considered. The popular wisdom should be amended to state that to encourage initiative, failure should not be punished too harshly, but neither should mediocrity be too comfortable.

Second, we show that there is an important countervailing force to the property that the contract when initiative is a consideration crosses the contract without this consideration twice, first from above, and then from below. Encouraging initiative may indirectly discourage or encourage effort, and the optimal contract must adjust accordingly. We show examples where this overturns the result that low outputs are treated more generously when initiative needs to be encouraged and the result that high outputs are treated more generously.

In the face of this, we exhibit a sufficient condition for the intuitive crossing pattern. The condition is economically interpretable and is satisfied in some very natural settings, but fails

[1] If the agent could destroy output, then there would be an additional monotonicity constraint on compensation, a topic that, for considerations of length, we do not explore in this paper.

in other natural ones. Then, we study a much more general setting. Despite its generality, a remarkable amount of structure emerges. The two contracts now cross at most three times. A range of middle outputs are still treated less generously and a range of higher and lower outputs more generously. Hence if there are only two crossings, they are of the expected pattern. When there is an extra crossing, its location is governed by the interaction of encouraging initiative and encouraging effort. If encouraging initiative discourages effort, then to restore incentives for effort, the principal may end up treating very low outputs more harshly than before. If encouraging initiative encourages effort, then very high outputs may be treated less generously than before.

The result that effort tends to be pushed away from the middle is driven by the fact that in many settings, the cost penalty inherent in the initiative constraint is first increasing and then decreasing in the induced effort. Some intuition for this is that at low efforts, incentives are weak, and so there is not much cost in making sure that middle outcomes are not rewarded too well. But, rewarding middling outputs can be a very effective way to encourage moderate effort, and hence the initiative constraint binds more harshly. Finally, generously rewarding high outputs encourages high effort without also making the safer project attractive. Effectively, low effort levels remove the need to provide strong incentives while high effort levels make it easier for the principal to distinguish whether initiative was taken. But, because the cost of middle efforts rise the most, efforts towards the extremes will be favored in the face of the new constraint. The principal will tend to "go big, or go home" in the face of the need to induce initiative.

An important case is when the agent's utility of income is square-root. All the relevant objects then have closed-form expressions in terms of three basic objects that depend only on the information structure of the problem. The first reflects the informativeness of output about effort, the second the informativeness of output about initiative, and the third the degree to which signals that are good news about effort covary with signals that are good news about initiative. The square-root case is a rich source of examples and insights. For example, it provides a clean comparison of the relevant multipliers, and a closed form expression for the cost penalty and hence effort distortion inherent in the new constraint.

The square-root case turns out to be of much deeper importance. For general utility functions, the equations that implicitly define the optimal contract are intractable. But, for a broad class of utility functions, when the outside option of the agent is large, the cost-minimizing contract for any given effort *converges* to the square-root form.[2] Thus, the insights and intuitions from the square-root case are valid much more generally.

Our detailed results about the form of contracts and effort distortions are unlocked by our use of the first-order approach which relaxes the full incentive constraint on effort to the local

[2]Even in the pure moral hazard case, we meaningfully advance Chade and Swinkels (2020) by fully characterizing the limit contract.

necessary condition. This is valid only if a solution to the relaxed problem exists, and is feasible in the full problem.

To analyze existence, we begin by noting that in the square-root case, a closed form solution exists when the outside option is large enough so that the constraint that payments are non-negative does not bind. We then leverage our convergence result and tools from Kadan, Reny, and Swinkels (2017) to show that a solution to the relaxed problem exists with a large outside option for the same class of utility functions as before. Since optimal contracts can be non-mononote, no previous result justifying the first-order approach applies here. We provide permissive new results for our setting.

Our paper is related to a large literature in economics, finance, and accounting on incentive provision for risk taking and project selection. Indeed, the seminal paper by Grossman and Hart (1983) on the standard principal-agent problem with moral hazard allows for multidimensional actions. Thus, for example, one could think about one dimension as effort and the another one as selecting projects of different risk and return. Indeed they conjecture (see pp.28–29), that in a setting similar to ours low outputs might be rewarded to induce what we refer to as initiative. We make precise these conjectures and explore their implications.[3]

The paper is also closely related to the literature on incentive provision for innovation. Central to this literature is Manso (2011), which analyzes a two-period principal-agent problem where the agent controls a two-armed bandit process, and can choose whether to exert effort on a known arm or explore the other arm. If the agent is risk neutral and exploration is what he calls "radical" then the optimal contract exhibits tolerance for early failure in the sense that the agent's wage for failure in the first period is higher than that for success. It also rewards repeated success (which is evidence of risk taking) more highly than with pure moral hazard. Ederer and Manso (2013) and Azoulay, Zivin, Joshua, and Manso (2011) provide experimental and empirical evidence for the tolerance-for-failure property.[4] Our canonical static principal-agent setting with a risk-averse agent and a continuum of actions and output levels allows a substantially more nuanced examination of how incentives change when initiative is an issue.

Another related paper is Hirshleifer and Suh (1992), who also extend the principal-agent problem with moral hazard to allow for project selection. Their setup allows for a richer set of projects than the binary case we consider. Their most general results are for the case where there

[3] There is also a strand of literature in which the agent costly acquires information about a risky project before deciding between that project and a safe alternative. The seminal paper is Lambert (1986), which shows that without communication the principal distorts project selection, and the distortion can be downward or upwards. Malcomson (2009) analyzes a more general setting and sheds further light on the distortions induced by information acquisition and project selection. Other papers in this literature are Barron and Waddell (2003), which combines theory and estimation of a model with project selection with information acquisition, and Chade and Kovrijnykh (2016), which analyzes a dynamic version and shows that sometimes the principal rewards "bad news." Our setting abstracts from information acquisition before choosing a project, and thus is not closely related to these papers.

[4] Another contribution is Hellmann and Thiele (2011), who analyze optimal contracts to innovate using a multi-task model with moral hazard.

is no risk-return trade-off (projects only differ in their variance) and the distribution of output is normal (an assumption that is technically problematic). When a risk-return trade-off is present, they illustrate via examples that there can be downward distortions in both project selection and effort. Demski and Dye (1999) allows the agent to have private information about the mean and variance of the projects. Under the restriction to compensation schemes that have a quadratic functional form, they find that at the optimal contract the agent underreports the mean of the project chosen. Our setting abstracts from private information, but imposes no restrictions on the set of contracts.[5]

Holmström and Costa (1986) shows that in the presence of career concerns the agent has incentives to take less risk than the principal desires.[6] Under some conditions, the optimal contract protects the agent against low outcomes, thus having an "option-like" shape. We derive a related insight without career concerns.[7]

The organization of the paper is as follows. Section 2 lays out the model. Section 3 presents a simple example to illustrate the two main insights. Section 4 derives the optimality conditions that any solution to the problem must satisfy, and illustrates them with the case in which the agent's utility of income is the square-root function. Section 5 provides a comprehensive analysis of the shape of optimal compensation schemes. Section 6 examines the effort distortions induced by initiative, and illustrates that the distortions can be large. Section 7 derives mild conditions on the agent's utility of income under which solutions converge to the square-root case as the outside option rises. Section 8 discusses existence and when the solution to the relaxed problem is a solution to the full problem. Section 9 concludes. Appendix A contains central omitted proofs and calculations. Appendix B contains the formal development of the existence material. Online Appendix C contains ancillary material.

2 Model

The model is a straightforward extension of the standard principal-agent problem with moral hazard. A principal (she) seeks to hire an agent (he). If the agent accepts, then he makes two choices. First, he faces a choice of projects, where we will term one "safe" and the other "risky," a choice of terminology that we will justify shortly. If he chooses the safe project, which we write as a_s, then effort does not matter, and output is given by a continuous differentiable density f^s

[5]Other papers with project selection and moral hazard are Sung (1995), who analyzes a related problem under linear contracts, and Dittmann, Yu, and Zhang (2017), which calibrates a principal-agent problem and finds empirical support for protecting executives from bad losses and for convex contracts.

[6]For a recent contribution with career concerns, see Laux (2015), who derives a CEO's optimal compensation scheme when pay is restricted to a combination of equity and stock options.

[7]In our setup, the agent is tempted to take less risk than the principal wants. There is also a complementary literature where the opposite is true: contracts are designed to temper the agent's desire to take risk. See, for example, Georgiadis, Barron, and Swinkels (2020) and Biais and Casamatta (1999) and the references therein.

on some interval of the positive reals. If he chooses the risky project, which is what we mean by taking *initiative*, then effort does matter, with $f(\cdot|a)$ being the density on output $x \in [0, \bar{x}]$ given effort level $a \in [0, \bar{a}]$, where $\bar{x}$ and $\bar{a}$ are finite, with $f > 0$ and twice-continuously differentiable.[8] We take f to have the usual structure of the moral hazard problem. In particular, $l(x|a) \equiv \frac{f_a(x|a)}{f(x|a)}$ has the (strict) *monotone likelihood ratio property*, *MLRP*, which is that $l(\cdot|a)$ is strictly increasing for each a. We assume that the support of $f(\cdot|a)$ does not depend on a, and that the support of f^s is a subset of the support of $f(\cdot|a)$.[9] This rules out that certain outcomes are sure evidence that the agent either chose the safe project or chose a non-desired effort level.

To justify our "safe" versus "risky" terminology for the projects, on the support of $f(\cdot|a)$, let $l^s(x|a) \equiv \frac{f^s(x)}{f(x|a)}$ be the likelihood ratio on the safe versus the risky project given effort a and outcome x. We assume that for each a, $l^s(\cdot|a)$ is strictly single peaked, with $l^s(\cdot|a)$ strictly less than one at the extremes of the support of $f(\cdot|a)$. This implies that for each a, $f(\cdot|a) - f^s(\cdot)$ is first strictly positive, then strictly negative, and then again strictly positive. So, when the agent takes initiative, there is less weight on intermediate outcomes and more weight on extreme outcomes than when the agent takes the safe project. To keep things interesting, we assume that for a sufficiently large, $\mathbb{E}[x|a] > \mathbb{E}[x|a_s]$.

The agent's utility is additively separable in income and effort, where an agent with income w who exerts effort a has utility $u(w) - c(a)$. We assume u is strictly increasing, strictly concave and twice differentiable, and that c is increasing, convex and twice differentiable with $c(0) = c_a(0) = 0$. Taking the safe project incurs effort disutility equal to 0.

The principal can see only output, observing neither whether initiative was taken nor the choice of effort. A contract thus specifies a wage for each output x. As is standard, we will work instead with the utility from income that the agent receives, letting $v(x)$ be the utility from income following output x. Let $\varphi = u^{-1}$ give the cost to the principal of inducing any given utility, so that the principal's outlay at outcome x is $\varphi(v(x))$.

Conditional on initiative, the principal values the effort of the agent according to some increasing concave function B. An example we use below is $B(a) = \alpha + \beta\mathbb{E}[x|a]$, so that β is the market price of output, and α reflects the fixed costs or benefits to the principal of employing the agent. The net payoff to the principal when effort is a and the contract is v is $B(a) - \mathbb{E}[\varphi(v(x))|a]$. We also let $B(a_s)$ be the value the principal places on the safe action a_s, where once again, $B(a_s) = \alpha + \beta\mathbb{E}[x|a_s]$ will be a common example. As usual, we analyze the principal's problem in two steps: first minimizing the cost of inducing a given action, and then using the resulting cost function to find the profit-maximizing action.

Note that the safe project can be induced by paying $\bar{u}$ at all outcomes, and hence costs $\varphi(\bar{u})$. Turning to the interesting case, fix a, and consider the problem of inducing the agent to take

[8] Where convenient in examples, we relax various of these assumptions.

[9] When we intend a relationship to be strict, we say so. Throughout, we discard limits of integration and arguments of functions where they are obvious. The symbol $x =_s y$ means that x and y have strictly the same sign.

initiative and then choose effort level a. The cost minimization problem is

$$\min_v \int \varphi(v(x))f(x|a)dx \tag{$\mathcal{P}^{Full}$}$$
$$s.t. \int v(x)f(x|a)dx - \bar{u} - c(a) \geq 0,$$
$$a \in \arg\max_{a'} \int v(x)f(x|a')dx - c(a'), \text{ and}$$
$$\int v(x)f(x|a) - c(a) - \int v(x)f^s(x)dx \geq 0.$$

The first constraint is the participation constraint that the agent prefers to accept the contract than to take his outside option. The second is the incentive-compatibility constraint that conditional on taking initiative, the agent prefers action a to any other action. These two constraints are the usual ones in the standard principal-agent problem with moral hazard. The final constraint reflects that the agent is better off to take initiative than the safe project.

For much of our analysis, we make two simplifications to this program. For convenience, we assume that IR binds at the optimum. This is automatic if u is unbounded below, and in cases like $u(w) = \sqrt{w}$ if the outside option is sufficiently large.[10] More substantively, we only check the first-order condition on the agent's effort choice rather than the full set of incentive constraints. Doing so gives us a tight characterization of optimal contracts. Later we provide conditions under which the first-order approach (FOA) is valid in our setting.[11]

We thus consider the relaxed problem

$$\min_v \int \varphi(v(x))f(x|a)dx \tag{$\mathcal{P}^{PS}$}$$
$$s.t. \int v(x)f(x|a)dx - \bar{u} - c(a) = 0, \tag{IR}$$
$$\int v(x)f_a(x|a)dx - c_a(a) = 0, \text{ and} \tag{IC}$$
$$\bar{u} - \int v(x)f^s(x)dx \geq 0, \tag{PS}$$

where the participation constraint IR is now an equality, the incentive-compatibility constraint IC is relaxed to local optimality, and the initiative (project-selection) constraint PS is simplified using IR. Let $C^{PS}(a)$ be the value of this program. If one discards the constraint PS, one has the standard relaxed moral hazard problem (Holmström (1979), Mirrlees (1975)). Let $\mathcal{P}^{MH}$ be this problem, with value $C^{MH}(a)$.

[10] If u is unbounded below and IR is slack, then removing a small constant from v leaves the incentive and project-selection constraints satisfied and saves the principal money.

[11] We also address existence of an optimum.

We consider two settings. In the first, which with some abuse of notation we refer to as *PS*, initiative is unobservable. The principal chooses $a \in [0, \bar{a}]$ to maximize $B(\cdot) - C^{PS}(\cdot)$, and then induces initiative if and only if $B(a) - C^{PS}(a) \geq B(a_s) - \varphi(\bar{u})$. In the second, which we refer to as *MH*, initiative is observable and contractible: the principal can either insist on a_s or forbid it. Hence the principal solves the same problem but with C^{MH} playing the role of C^{PS}.

3 A Simple Example

Before diving into the formal analysis, let us see the main economic forces at play in a simplified example. We focus on two main economic impacts of the need to motivate initiative. First, for any given effort, high and low outputs are rewarded more generously, but middle outputs less generously. Second, effort choices will often be distorted away from "middle" effort levels in *PS* compared to the observable initiative benchmark *MH*, either towards the safe project or towards a higher one. The principal goes big or goes home.

Example 1 *Let $u(w) = \sqrt{2w}$. There are four actions a_1, a_2, a_3, and a_s and three outputs, x_1, x_2, and x_3. The agent plays it safe with a_s or exerts initiative with a_i, $i = 1, 2, 3$. The action a_s yields x_2 with probability one. If the agent exerts initiative, the probability distribution of output is as follows:*

	x_1	x_2	x_3
a_1	3/4	1/6	1/12
a_2	1/3	1/3	1/3
a_3	0	0	1

The monotone likelihood ratio property holds across a_1, a_2, and a_3, but a_s is not ranked. The middle output x_2 becomes more likely as one moves from a_1 to a_2 but less likely as one moves from a_2 to a_3. Thus, mediocre performance is a positive signal that the agent exerted medium versus low effort, but a negative signal that the agent exerted high versus medium effort.[12] The disutility of effort is a_i for $i = 1, 2, 3$, and 0 for a_s. We take $a_1 = 0$, $a_2 = 1$, and vary a_3. Similarly, we take $x_1 = 0$, $x_2 = 1$, and vary x_3. The agent's reservation utility is $\bar{u} = 1$.

As described in Section 2, in both *MH* and *PS*, a_1, a_2, and a_3 are unobservable. In *MH* the principal faces a pure moral hazard problem over a_1, a_2, and a_3, but can simply require or forbid the agent to take a_s. In *PS* the principal also cannot observe whether the agent took action a_s. We begin with the optimal contracts that implement each action in each informational setting. Either a_s or a_1 is optimally implemented in either *MH* or *PS* by setting utility to $\bar{u}$ at all outputs. Implementing a_3 similarly involves setting utility to 0 at x_1 or x_2 and to $\bar{u} + a_3$ at x_3.

[12] This example is easily modified so that a_3 sometimes generates a worse outcome than a_s, consistent with our interpretation of a_s as the agent playing it safe.

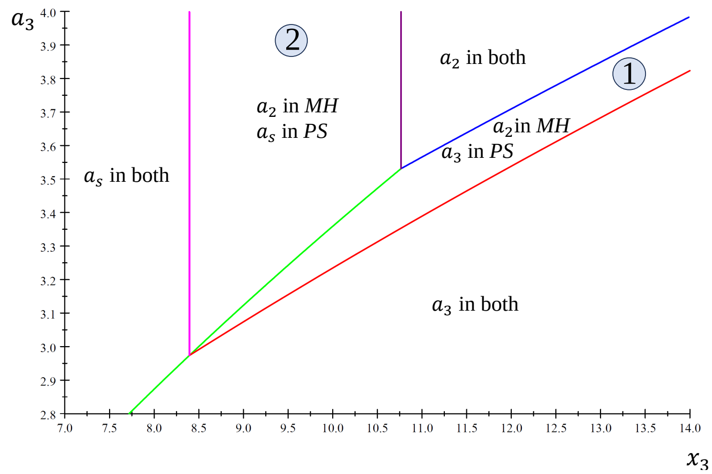

Figure 1: ***Distortions in Effort.*** The figure depicts the regions in which the different actions are optimal under problem *PS* and under problem *MH*. In Region 1, effort is distorted upwards from a_2 to a_3. In Region 2, effort is distorted from a_2 to a_s.

Let us turn to a_2 in *MH*, and focus on values of a_3 where it is the deviation to a_1 that binds rather than the deviation to a_3. The optimal contract is (see Online Appendix C.1 for details)

$$v_1^{MH} \cong 0.42,\ v_2^{MH} \cong 2.63, \text{ and } v_3^{MH} \cong 2.95,$$

where v_i^{MH} is the utility of income following outcome x_i. This contract has cost 2.63.

Now, consider implementing a_2 in *PS*, and continue to focus on values of a_3 where only the downward deviation binds. Unlike in *MH*, no more than $\bar{u} = 1$ can be given at x_2, otherwise the agent will switch to a_s. Instead, rewards above $\bar{u}$ must be concentrated solely on x_3, which is good news about both effort and initiative. Because these rewards occur less often, they must in utility terms be larger, and because the agent is risk averse, this is more expensive to the principal. Indeed, the optimal contract is now

$$v_1^{PS} \cong 0.63,\ v_2^{PS} = 1, \text{ and } v_3^{PS} \cong 4.38$$

at a cost of 3.42. Compared to *MH*, the optimal way to induce initiative and effort in *PS* involves *lower payments at middle outcomes, and higher payments at low and high outcomes.* A major topic of this paper is to understand when this pattern emerges.

Because a_2 becomes more expensive to implement while the other actions do not, there will be a tendency to switch away from a_2 in *PS* compared to *MH*. Figure 1 compares the optimal effort levels implemented in these problems as a function of x_3 and a_3.[13] In many cases, the principal

[13]The principal is indifferent between a_s and a_2 in *MH* along the pink line and between a_2 and a_3 along the red

throws her hands up and now implements a_s despite its lower gross returns. But, interestingly, in other cases, the principal replaces a_2 by the *higher* effort a_3.

The cost of a_2 rose in PS because the frequently arising signal x_2 is good news about effort but bad news about initiative, and so the signal is "conflicted." No such conflict arises following a_3. Another goal in what follows is to understand when higher efforts lead to less conflicted information, and hence an impetus towards implementing higher effort levels in PS.

4 Solving the Optimization Problems

We now analyze the general model. Our task is to understand the conditions under which the two main economic insights illustrated by the example are robust. Problem $\mathcal{P}^{Full}$ is general but does not allow us to say much about either optimal compensation or the resulting cost to the principal. Thus we move to $\mathcal{P}^{MH}$ and $\mathcal{P}^{PS}$, where the first-order approach allows a tractable analysis.

Let $\lambda \geq 0$, μ, and $\eta \geq 0$ be the Lagrange multipliers associated with the participation, incentive, and initiative constraints in $\mathcal{P}^{PS}$. Then the solution is pinned down by

$$\varphi'(v(\cdot)) = \lambda + \mu l(\cdot|a) - \eta l^s(\cdot|a), \tag{1}$$

for almost all x, which differs in structure from the optimality condition in $\mathcal{P}^{MH}$ by the presence of the last term.[14] Below, we tackle whether solutions to these problems exist and are feasible in the full problem, but for now we assume both are true.

Denote the solution to $\mathcal{P}^{PS}$ by $v^{PS}(\cdot, a, \bar{u})$, with multipliers $\lambda^{PS}(a, \bar{u})$, $\mu^{PS}(a, \bar{u})$, and $\eta^{PS}(a, \bar{u})$, and the value of the problem by $C^{PS}(a, \bar{u})$. The corresponding solution and value in $\mathcal{P}^{MH}$ are v^{MH}, λ^{MH}, μ^{MH}, and C^{MH}. If v^{MH} satisfies constraint PS, then it solves $\mathcal{P}^{PS}$, and $\eta^{PS}(a, \bar{u}) = 0$.

4.1 The Square-Root Utility Case

When the agent's utility for income is $u(w) = \sqrt{2w}$, then v^{MH} and v^{PS} and the associated multipliers have particularly transparent and tractable forms. This will allow a more nuanced examination of the crossing properties of v^{MH} and v^{PS}, and will be a continuing source of examples and insight as we move forward. This case is also foundational for our understanding of the case with a large outside option in Section 7.

Under square-root utility, the constraints become linear in the multipliers, which simplifies

line. She is indifferent between a_s and a_3 along the green line in either problem. She is indifferent between a_s and a_2 in PS along the purple line and between a_2 and a_3 along the blue line.

[14] The result is exactly what one would expect from Lagrangian methods (for example, from a careful application of Theorem 1 and problem 7 in Luenberger (1969) Chapter 8), but for completeness, we provide an elementary proof in Online Appendix C.2.

the problem. It is well-known that for given a the multipliers characterizing v^{MH} are

$$\lambda^{MH} = \bar{u} + c \text{ and } \mu^{MH} = \frac{c_a}{I^a}, \tag{2}$$

where $I^a \equiv \int l^2 f$ is the Fisher Information of x about a. To understand v^{PS}, we need two further information-theoretic objects. The first is $\sigma \equiv \int l l^s f$, the covariance of l^s and l. The second is $I^s \equiv \int (l^s)^2 f$, the information in x about whether a_s was chosen or a.

When PS binds, v^{PS} is characterized by

$$\lambda^{PS} = \lambda^{MH} + \eta^{PS},\ \mu^{PS} = \mu^{MH} + \frac{\eta^{PS}\sigma}{I^a}, \text{ and } \eta^{PS} = \frac{cI^a + c_a\sigma}{(I^s - 1)I^a - \sigma^2}, \tag{3}$$

where $(I^s - 1)I^a - \sigma^2 > 0$ because it is a particular variance (see Lemmas 2–3 in Appendix A.1).[15] The form of η^{PS} has intuitive content. The numerator is proportional to the amount by which constraint PS is violated at v^{MH}. The denominator measures how easily one can adjust incentives independently of the attractiveness of the safe action.[16] Unambiguously, $\lambda^{PS} > \lambda^{MH}$ when PS bites and thus η^{PS} is strictly positive. The sign of $\mu^{PS} - \mu^{MH}$ is the *same* as the sign of σ, which is a primitive. For some intuition about this result, note that when one adds the term $-\eta^{PS} l^s$ to v^{MH} then outputs where l^s is high are reduced compared to outputs where l^s is low. If $\sigma > 0$, then this lowers incentives for effort, and so μ^{PS} must rise to reestablish *IC*. Conversely if $\sigma < 0$ then μ^{PS} must fall to reestablish *IC*. Appendix A.1 shows conditions for σ negative, a case that will be of special interest.

5 Comparing Compensation Schemes

Let us now turn to the relative shapes of v^{MH} and v^{PS}. Say that v^{PS} is *higher-lower-higher* (*HLH*) if for given a, $v^{PS} - v^{MH}$ crosses zero exactly twice, and is first strictly positive, then strictly negative, and then strictly positive. In our leading example, v^{PS} is *HLH*. Thus, v^{PS} is more lenient towards low outputs, less tolerant of mediocre outputs and more rewarding of excellent outputs than is v^{MH}. Given that the safe action creates outputs that are concentrated towards the middle, this seems the intuitive result of needing to encourage initiative while retaining incentives for effort. Indeed, the first part of this pattern, that when the principal wants the agent to engage in a risky project they must be tolerant of failure is long accepted in the field, with a leading reference being Manso (2011), and a large literature following.

In this section, we do three things. First, we present an economically natural condition on the statistical structure of the problem under which *HLH* in fact holds. Second, we show why

[15]Note that $\bar{u}$ does not enter into μ^{PS} or η^{PS}, and enters additively into λ^{PS}, and so since l^s and l are bounded, the contract v^{PS} is positive for all a for $\bar{u}$ sufficiently large, and similarly for v^{MH}.

[16]Formally, the denominator measures how far the equations that pin down the multipliers are from being colinear.

some such condition is needed for such a result, provide some simple examples where *HLH* fails, and provide intuition for the countervailing force that has been ignored by the literature to date. Finally, we explore a substantially more general class of information structures. In this class we show that there are at most three crossing. If there are two, then *HLH* holds. When a third crossing appears, then depending on the primitives, we have that v^{PS} is either *LHLH* or *HLHL* (in the obvious notation) and so punishes either very low outputs or very high outputs compared to v^{MH}. In the first case, we have invalidated the "tolerance for failure" result that is common in the literature and seems so intuitive. In the second case, we invalidate the equally intuitive "exceptional rewards for exceptional performance" result.

We begin our analysis with a preliminary result about the crossing properties of v^{PS} and v^{MH}.

Lemma 1 (At Least Two Crossings) *For each a and $\bar{u}$ where $\eta > 0$, v^{PS} and v^{MH} cross at least twice.*

The proof is in Appendix A.2, but the idea is very simple. If the contracts do not cross at all, then the higher one provides strictly more utility to the agent than the lower one, contradicting that both satisfy *IR* with equality. And, if they cross only once, then the one that crosses from below provides strictly stronger incentives for effort, contradicting that they both satisfy *IC*.

5.1 Optimality of *HLH* Contracts

The intuition that v^{PS} is *HLH* is in fact correct in many settings. The following theorem shows that one sufficient condition is that if one rescales output such that $l_x(\cdot|a) = 1$, then l^s is strictly concave. This is automatic if l^s is concave and l is convex. But, in the more usual case where l is concave, it is a statement that l^s is more concave than l. See Appendix A.2 for details, an alternative formulation, and a class of examples.

Theorem 1 (Primitives for *HLH*) *Fix $a \neq a_s$ where $\eta > 0$, and assume that $l^s(l^{-1}(\cdot|a)|a)$ is strictly concave. Then v^{PS} is HLH.*

The structure that low outputs are punished less harshly than without project selection, middle outputs are rewarded less generously, and high outputs are rewarded even more generously, resonates with real-world phenomena (see Manso (2011) for a related discussion). Harkening back to the examples in the introduction, CEOs often have generous severance packages, options that are worth little under mediocre firm performance, and what is often thought of as excessive compensation when the firm thrives. The generous severance package in particular is not what the standard moral hazard problem would predict. Nor under reasonable assumptions on the structure of the likelihood ratio would one expect such extreme rewards for success. But, it is this pattern of compensation that is most effective when the CEO needs to be motivated to both work hard *and* pursue strategies that have considerable upside potential but might fail spectacularly. Similarly, the compensation of tenured academics involves considerable downside protection and large rewards for exceptional impact.

To see the proof, observe that l is a strictly increasing function of x and so $v^{PS} - v^{MH}$ has the same sign as $\varphi'\left(v^{PS}\right) - \varphi'\left(v^{MH}\right)$ and hence when $l(x|a) = \tau$ the same sign as

$$D(\tau) \equiv \frac{\lambda^{PS} - \lambda^{MH} + (\mu^{PS} - \mu^{MH})\tau}{\eta} - l^s(l^{-1}(\tau)).$$

This is strictly convex under the premise that $l^s(l^{-1})$ is strictly concave, and so can only cross zero twice, first from above and then from below. But then by Lemma 1, v^{PS} is *HLH*.

When utility is square-root we can considerably sharpen this result. In this case v and φ' coincide, and so under the premise of Theorem 1, $v^{PS} - v^{MH}$ is *strictly convex*. Thus v^{PS} is equal to v^{MH} plus a convex function. This convexification can be very strong; Appendix A.2 shows a well-behaved class of examples in which v^{PS} is *higher* at low outputs than at middle outputs.

5.2 Beyond Two Crossings

The argument proving Theorem 1 suggests that if $l^s(l^{-1})$ changes from concave to convex multiple times, then D can cross zero multiple times as well. And, there are many natural examples where the rescaled l^s is not concave. Concavity fails whenever f^s has less than full support, which is entirely plausible, as the whole point of playing it safe is to avoid bad outcomes at the cost of also giving up on good ones. Concavity also fails if the rescaled l^s looks like a truncated normal distribution, or is decreasing and convex on its support. Because of this, it is very easy to build examples where *HLH* fails. Here are two.

Example 2 (Punishing Failure) *For given* $\tau \in (0, 1)$*, let* $f = 1-\tau-a$ *on* $[0, 1/8]$*,* $f = 1+\tau-a$ *on* $(1/8, 1/4]$*,* $f = 1 + a/3$ *on* $(1/4, 1]$*, and* $f^s = 8$ *on* $[3/8, 1/2]$*, with* $a \in (0, 1-\tau)$ *and* $u = \sqrt{2w}$*. This is the limit of examples in which* f *is continuous,* l *is strictly increasing, and the rescaled* l^s *is strictly concave on its support.*[17] *Appendix A.2 verifies that* $v^{PS} - v^{MH}$ *is LHLH.*

Example 2 is particularly troubling, because it contradicts the received wisdom that encouraging risk involves being gentler in the face of failure. Here, very low outputs (those below 1/8) are punished more harshly in *PS* than in *MH.* The core of this example is that because l^s is strictly positive only where l is strictly positive, encouraging initiative by setting η strictly positive *discourages* effort. Because of this, incentives at places where they do not encourage the safe project must be adjusted to become stronger via a larger μ, and in this example the effect is strong enough at outcomes below 1/8 so as to violate *HLH.*

[17] See Online Appendix C.3 for details.

Example 3 (Punishing Success) *Let $f(x|a) = e^{-x/a}/a$ for $x \in [0,\infty)$ and $a \in [0,\infty)$. If chooses a_s, then output is distributed according to $f^s(x) = e^{-(x-1)}$ on $[1,\infty)$. Let $u = \sqrt{2w}$. Then, as Appendix A.2 verifies, for all relevant effort levels, $v^{PS} - v^{MH}$ is HLHL, and so very high outputs are less generously rewarded than in v^{MH}.*

Example 3 contradicts the intuition that encouraging risk involves especially high rewards in the case of spectacular success. Here, when the principal encourages initiative by setting η positive, she also strengthens the agent's incentives to take effort. To restore *IC*, μ falls, and the the principal reduces compensation at high outputs.

It is tempting at this point to conclude that there is no clear relationship between v^{PS} and v^{HM}. But, while the proof of Theorem 1 provides a recipe book for building examples with any number of crossings, the situation is in fact much more hopeful. In what follows, we will exhibit mild primitives under which (*i*) there are at most three crossings, (*ii*) when there are three crossings, whether *LHLH* or *HLHL* holds depends in an intuitive way on whether addressing the project selection constraint makes satisfying *IC* harder or easier, and (*iii*) when there are two crossings, *HLH* continues to hold.

Say that l^s is *semibellshaped* (*SBS*) if when output is rescaled so that l is linear, l^s never changes from concave to convex before its peak, never changes from convex to concave after its peak and is never linear on the support of f^s. Formally, fix and suppress a, let $[x_\ell, x_h]$ be the support of f^s and let $\tilde{x}$ be the maximizer of l^s. Then, l^s is *SBS* if there is $x_\ell \leq x_1 \leq \tilde{x} \leq x_2 \leq x_h$ such that $l^s(l^{-1}(\cdot))$ is strictly concave on $[l(x_1), l(x_2)]$, is otherwise convex, and is strictly convex on $[l(x_\ell), l(x_h)] \backslash [l(x_1), l(x_2)]$. See Figure 2 for examples and a counterexample, and recall that Appendix A.2 provides an alternative formulation.

Theorem 2 (*SBS* Implies At Most Three Crossings) *If l^s is SBS, then $v^{PS} - v^{MH}$ changes sign at most three times. If there are three crossings, then v^{PS} is LHLH if $\mu^{PS} > \mu^{MH}$ and HLHL if $\mu^{PS} < \mu^{MH}$.*

When $\mu^{PS} > \mu^{MH}$ then addressing project selection makes it harder at the margin to provide incentives for effort. At an intuitive level, this will be true if outputs that are likely under the safe project become more likely as effort is increased. But, when μ is raised, rewards at low outcomes are pushed down, and, as Example 2 shows, this effect can be strong enough to cause very low outputs to be punished relative to v^{MH}. But, v^{PS} is *HLH* after this. Thus, the traditional wisdom of tolerating failure is overturned, but in a disciplined manner. Similarly, when $\mu^{PS} < \mu^{MH}$, then addressing project selection relaxes *IC*. To restore *IC*, some very high outcomes may be rewarded less generously than before. As mentioned, in the square-root case, these two cases are pinned down by the sign of the covariance σ.

Our intuition is that for academics, the sort of work that results from playing it safe is also quite common when one takes initiative and works hard but happens to have limited success.

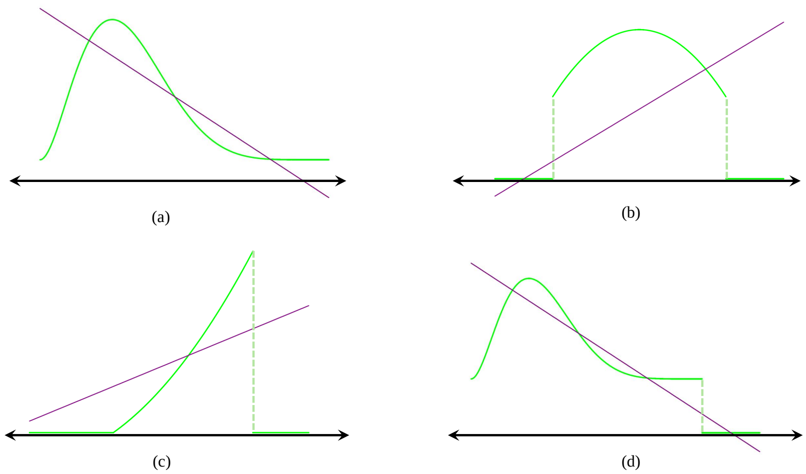

Figure 2: **SBS *Examples and a Counterexample.*** Some examples of $l^s(l^{-1})$, in green, and of $\left((\lambda^{PS} - \lambda^{MH} + (\mu^{PS} - \mu^{MH})\tau\right)/\eta$, in purple, as functions of τ. In (a), f^s has full support and so *SBS* is satisfied despite $f^s(x_\ell) > 0$. The green line is first convex, then concave until past the peak, and then convex again. Where the purple line is above (below) the green line, $v^{PS} - v^{HM}$ is positive (negative). In this example, the purple line slopes down ($\mu^{PS} - \mu^{MH} < 0$), and v^{PS} is *HLHL*. Example (b) satisfies *SBS,* being convex up to the first jump, strictly concave between the jump points, and convex from the second jump point on. Because the purple line slopes up ($\mu^{PS} - \mu^{MH} > 0$), the pattern is *LHLH*. In (c) *SBS* is also satisfied, but the purple line happens to be high enough that the pattern is *HLH*. Example (d) violates *SBS* (there is no way to choose the requisite x_2), and the purple line shows an example where the pattern is *HLHLHL*.

Hence, at an intuitive level, encouraging initiative, which involves lower rewards for middling publications, discourages effort. So, if the reward structure of academics is not *HLH*, it will be *LHLH*, and truly miserable output will be punished. An example is summer money that is contingent on presenting a plausible research agenda, where the inability to do this basic task corresponds to a very low output.

The proof in Appendix A.2 establishes what is evident from the figure. When l^s is *SBS*, no configuration of the purple line can cross the green line more than thrice, where if the purple line is upward sloping and there are three crossings, then as in panel (b), the ordering is *LHLH*, while if the purple line is downward sloping as in panel (a), then the ordering is *HLHL*.

While the theorem allows for as many as three crossings, the case of two crossings remains possible. The next result shows that under a mild condition *HLH* holds whenever this is so. The condition says that while $l^s(l^{-1})$ need not be convex, it does lie above the average of its endpoints. This is satisfied trivially when f^s is zero at its endpoints.[18]

Proposition 1 (*SBS* Plus Two Crossings Implies *HLH*) *Assume that* $l^s(l^{-1}(\cdot))$ *lies above*

[18]It fails if $l^s(l^{-1})$ is strictly positive at one endpoint, but has both slope and value zero at its other endpoint.

the line $\underline{l}^s$ connecting $(l(0), l^s(0))$ to $(l(\bar{x}), l^s(\bar{x}))$, and somewhere strictly. If SBS holds but $v^{PS} - v^{MH}$ nonetheless crosses zero only twice, then v^{PS} is HLH.

The idea is that under the premise if $v^{PS} - v^{MH}$ is negative at both ends, then it is negative everywhere, violating Lemma 1. But then, under *SBS*, if there are only two crossings, *HLH* must hold. An open question of economic interest is to understand primitives distinguishing the two and three crossing cases.

6 Effort and Initiative: Distortions

Besides the comparison of the shapes of the compensation schemes, we would also like to shed light on the effort distortions that can be traced to the need to induce initiative. We stress that signing distortions in effort is notoriously difficult in problems with moral hazard.

To see how the need the initiative problem interacts with the importance of effort to the principal, consider a setting where the benefit of effort to the principal is indexed by $\tau \in [0, \infty)$. In particular, let $B(a, \tau) = \alpha(\tau) + \beta(\tau)\mathbb{E}[x|a]$, where α is increasing in τ and β is strictly increasing in τ, with $\beta(0) = 0$ and $\lim_{\tau\to\infty} \beta(\tau) = \infty$. We will compare the optimal actions for each τ in problems *MH* and *PS*. Let $a^{MH}(\tau)$ and $a^{PS}(\tau)$ be the optimal efforts to induce, conditional on not inducing a_s, in problems *MH* and *PS* respectively.[19]

Define $\Delta(a) \equiv C^{PS}(a) - C^{MH}(a)$ as the cost penalty that is imposed from the extra constraint *PS*. In the square-root utility case (see Appendix A.3)

$$\Delta = \frac{1}{2} \frac{\left(c + c_a \frac{\sigma}{I^a}\right)^2}{I^s - 1 - \frac{\sigma^2}{I^a}}, \tag{4}$$

which depends on the information theoretic objects of the problem and the disutility of effort and its derivative. For some intuition, recall from the discussion of η^{PS} that the expression in the numerator reflects the amount by which *PS* is violated by v^{MH}, and the denominator reflects the amount by which the compensation scheme must be distorted from v^{MH} to reestablish *PS*.[20]

Now, note that in our discrete example in Section 3, Δ is single-peaked, first increasing at low effort levels, and then decreasing. In examples, we consistently arrive at a Δ which is strictly single-peaked over the relevant range of effort levels. Here is one such example.

Example 4 (Cost Penalty in Example 3) *Let $c = a^2$. Then, Appendix A.3 shows that for*

[19] Because B is strictly supermodular, a^{MH} and a^{PS} are single-valued almost everywhere, so we will treat them as functions, breaking ties in favor of, for example, the largest optimal action for given τ.

[20] Online Appendix C.4 shows that Δ decreases in I^s and increases in σ. When σ is positive, Δ decreases in I^a while if σ is negative, we have conflicting forces.

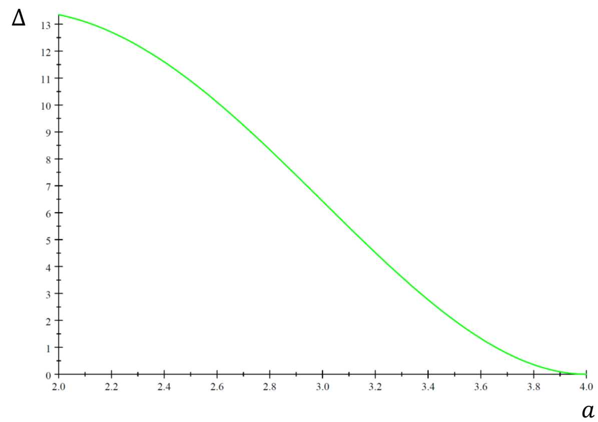

Figure 3: ***Cost Penalty.*** The cost penalty is decreasing in a for all relevant effort levels.

$a \leq 2$, action a_s dominates a, while for $a \geq 4$, $\Delta = 0$. In between, (4) reduces to

$$\Delta = \frac{(a(4-a))^2}{\frac{a^2}{2a-1}e^{\frac{1}{a}} - 1 - \frac{(2-a)^2}{a^2}},$$

which (see Figure 3) is strictly decreasing in a and hence strictly single-peaked.[21]

Since single-peakedness of Δ is our common finding in examples, it is worth exploring what happens to effort under *PS* versus *MH* when Δ is strictly single-peaked.[22] Our next theorem answers this question. Recall that τ indexes the value of effort to the principal. Note that for any τ where the principal induces a_s in *MH* or where $\Delta(a^{MH}(\tau)) = 0$, she trivially induces the same effort in *PS*, since her prefered alternative remains available at the same cost, while the costs to implement other efforts are at least weakly higher.

Theorem 3 (Effort Distortions) *Assume that Δ is strictly single-peaked where it is strictly positive, and that C^{MH} and C^{PS} are differentiable where $\Delta > 0$. Then, there is $\hat{\tau}$ such that for all τ, $a^{PS}(\tau) - a^{MH}(\tau)$ has the same sign as $\tau - \hat{\tau}$ and strictly so if $\Delta(a^{MH}(\tau)) > 0$ and $a^{MH}(\tau)$ is interior.*

Figure 4 provides intuition when the marginal cost functions are strictly increasing (the proof in Appendix A.3 does not rely on this property). The theorem captures in a precise way what we mean by "go big or go home." When effort is not very important to the principal, she responds to the initiative problem by either lowering the amount of effort she asks of the agent or simply

[21]In this example, the magnitude of Δ is quite small and effort is either distorted upwards or to a_s. Appendix A.3 provides a (carefully constructed) example where Δ is large and effort distortions are large in both directions.

[22]Primitives to guarantee single-peakedness of Δ would be desirable, but are complicated, because even in the square-root case, it is hard to disentangle the behavior of the information-theoretic objects.

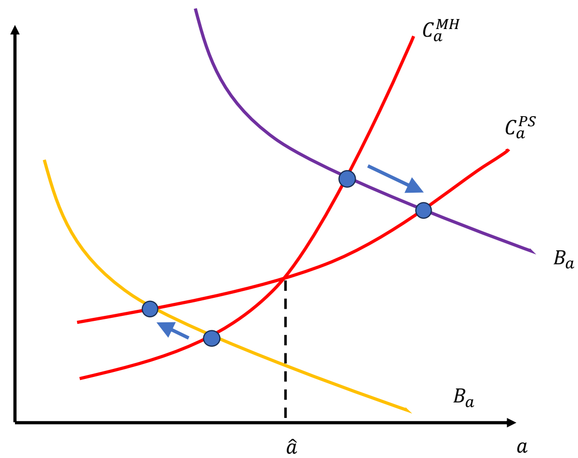

Figure 4: ***Distortions.*** When Δ is single peaked, then the marginal cost of effort C_a^{MH} crosses C_a^{PS} from below. As τ is increased, the marginal benefit curve B_a moves to the northeast. The point $\hat{\tau}$ is determined when B_a goes through the intersection of C_a^{MH} and C_a^{PS}. For higher τ, optimal effort is higher in PS than MH (go big) while for lower τ, optimal effort is lower in PS than MH (go home).

switching the agent from taking initiative to the safer project. But, when effort and initiative are important to the principal, she responds to the project selection problem by continuing to induce initiative but *increasing* the effort that is asked of the agent.

It is a common observation that in a variety of settings including investment banking, consultancy, law firms, and academia, success comes to those who exercise initiative, work at an extreme level, and are lucky. The extreme effort has been explained in a variety of ways including, for example, career concerns. The theorem provides a complementary explanation: by asking extreme effort of the agent, the principal finds it easier to distinguish whether initiative is being taken, which eases the impact of the project selection constraint.

Even when Δ is not single-peaked, we can take the more modest step of asking whether at high levels of effort constraint PS ceases to bind. To see why this is useful, consider a case where constraint PS binds at low effort levels but not at high ones. If so, there must be a region where the marginal cost of inducing effort is lower with constraint PS than without it.[23] This provides an impetus in the direction of going big for some range of τ, that is, of the principal optimally choosing higher effort in PS than in MH. In Online Appendix C.5, we derive two sets of conditions under which PS indeed ceases to bind for high effort levels, and provide a set of natural examples that have appeared in the moral-hazard literature.

[23]That is, if $a^L < a^H$ satisfy $C^{PS}(a^L) > C^{MH}(a^L)$ but $C^{PS}(a^H) = C^{MH}(a^H)$, then $C_a^{PS} < C_a^{MH}$ over some interval between a^L and a^H.

7 High Stakes

Outside of the square-root utility case, the equations describing the multipliers are forbiddingly complex. Despite this, *everything* we learned in the square-root case generalizes to a large class of utility functions when the agent's reservation utility is sufficiently large. In particular, the intuition based on the information-theoretic objects highlighted above extends to this larger class.

Formally, we build on Chade and Swinkels (2020) (henceforth CS) and show that in a class of utility functions the optimal contracts, and hence the behavior of costs, converge in a strong sense to those in the square-root case as $\bar{u}$ grows large. Of course, for this exercise to be relevant, the principal has to want to employ the agent when $\bar{u}$ is large. Thus, while we focus on the cost-minimization problem of implementing each level of effort (a problem that is parametrized by $\bar{u}$), in the background we are considering a sequence of economies where $\bar{u}$ grows, but so does the benefit of effort B to the principal. Hence, the stakes are high, in that both the agent has a good outside option and the principal places large value on his services.

Let $A = -u''/u'$ be the coefficient of absolute risk aversion, and let $P = -u'''/u''$ be the coefficient of absolute prudence. As in CS we will make the following assumption.

Assumption 1 *As $w \to \infty$, $u \to \infty$, $u' \to 0$, $A/u' \to 0$, and $(3A - P)/u' \to 0$.*

As CS show, equivalent to this assumption is that φ has domain with least upper bound ∞, and that as utility goes to ∞, $\varphi' \to \infty$, $\varphi''/\varphi' \to 0$, and $\varphi'''/\varphi'' \to 0$. These assumptions hold with appropriate parameter restrictions for the $HARA$ utility functions, but fail for $u(w) = \log w$, since $\varphi'''/\varphi'' = 1$ for all levels of utility.

Let $v^{SR}(\cdot, a, \bar{u})$ be the optimal contract implementing effort a with outside option $\bar{u}$ with square-root utility.[24] Our next theorem establishes that under Assumption 1, $v^{PS}(\cdot, a, \bar{u})$ and $v^{SR}(\cdot, a, \bar{u})$ become arbitrarily close both in level and slope as $\bar{u}$ grows.[25] To this end, let

$$d(a, \bar{u}) \equiv \sup_x \left| v^{PS}(x, a, \bar{u}) - v^{SR}(x, a, \bar{u}) \right|, \text{ and } d_x(a, \bar{u}) \equiv \sup_x \left| v_x^{PS}(x, a, \bar{u}) - v_x^{SR}(x, a, \bar{u}) \right|$$

be the maximum differences between $v^{PS}(\cdot, a, \bar{u})$ and $v^{SR}(\cdot, a, \bar{u})$ in value and slope.

Theorem 4 (Convergence of Compensation Schemes) *Under Assumption 1, for each $\varepsilon > 0$, there is $\bar{u}^* < \infty$ such that for all a and $\bar{u} > \bar{u}^*$, $d(a, \bar{u}) \leq \varepsilon$ and $d_x(a, \bar{u}) \leq \varepsilon$.*

There are two moving parts to the proof. First, regardless of $\bar{u}$, the optimal compensation scheme stays within a fixed band around $\bar{u}$. Second, given that $\varphi'''/\varphi'' \to 0$, it follows that φ''

[24] That is, v^{SR} is defined by (1), and depending on whether or not constraint PS binds at the solution to $\mathcal{P}^{MH}$, by the multipliers given in (2) or (3).

[25] This is a useful extension of what is shown in CS, who show that *ratios* of multipliers converge, but do not show the limiting form of the contract.

becomes essentially constant over the relevant range of utilities as $\bar{u}$ grows. But, in the square-root case φ'' is a constant and so the two optimization problems become increasingly similar. See Appendix A.4 for details.[26]

8 Existence and the Validity of *FOA*

Two issues that we have not addressed so far are whether the relaxed problem $\mathcal{P}^{PS}$ has a solution, and whether its solution also solves $\mathcal{P}^{Full}$ (that is, whether the first-order approach is valid). In this section, we first discuss some results on feasibility for general utility functions. Then, we turn to the square-root case, where existence is trivial, and where the explicit solution allows us to produce quite general conditions for the validity of *FOA*. Finally, building on Section 7, we show existence and feasibility for a large class of utility functions when the stakes are high.

In some settings, we will show that the solution to $\mathcal{P}^{PS}$ is a solution to $\mathcal{P}^{Full}$ for some but not all actions. To see that this is of value, note that C^{PS} is a lower bound on the true cost of implementation at *all* effort levels. Hence if B is such that $B(\cdot) - C^{PS}(\cdot, \bar{u})$ is maximized at an effort level where feasibility holds, then the same effort remains optimal facing the true cost function and the economics of the situation are indeed driven by the solution to $\mathcal{P}^{PS}$.

8.1 General Utility Functions

For general utility functions, there are several instances in which we can justify the validity of replacing all the incentive constraints for effort by *IC*. First, if $f(x|\cdot)$ is linear, then no matter the structure of the contract, the agent's expected utility from income is linear in effort, and so satisfying the first-order condition implies satisfying global incentive compatibility (recall that c is convex). This provides a tractable and economically relevant class of examples. Indeed, Example 2 is one such. Second, in some settings, one can show that the solution to the relaxed problem is increasing in x, in which case off-the-shelf conditions such as the convexity of the distribution function condition (*CDFC*) establish the validity of the first-order approach. As an example of this approach recall from Section 6 that in many settings, *PS* ceases to bind at some effort a^0, and so v^{PS} is monotone at a^0. Hence, if l^s is continuous with bounded slope, v^{PS} will continue to be monotone for an interval to the left of a^0.[27] Finally, in examples such as the exponential setting in Example 3, it is easy to numerically check feasibility by brute force.

[26] In Online Appendix C.7 we also show that for $\bar{u}$ large, both C^{MH} and C^{PS} are convex, and so solutions to the principal's first-order conditions on the choice of effort characterize optima.

[27] If *CDFC* holds strictly, then at the lowest a at which v^{PS} is monotone, the agent's payoffs are *strictly* concave in effort, and so they remain concave for a further interval to the left of this point. A similar point applies to the conditions of Jewitt (1988).

8.2 Square-Root Utility

In the square-root utility case, existence of a solution to $\mathcal{P}^{PS}$ is trivial. Let us turn to the validity of *FOA*. If the principal uses the contract $v^{SR}(\cdot, a, \bar{u})$, then the utility of the agent who takes action a' is

$$V(a') \equiv \mu^{PS} \int l f(x|a') - \eta^{PS} \int l^s f(x|a') - c(a'),$$

where μ^{PS}, η^{PS}, l and l^s are evaluated at a. We would like to show that V is quasi-concave with peak at a. Indeed, to facilitate our high-stakes analysis in the next section, we will ask that in addition, V is strictly concave on a neighborhood of a.

Our main approach is to look for conditions on the information structure of the problem under which *each* of the three terms in V is concave, and one term strictly so. Recall that *CDFC* is the condition that F_{aa} is positive. Say that F satisfies *CDFC** if for each a', $F_{aa}(\cdot|a')$ is single-peaked and strictly positive except at its endpoints. Examples satisfying *CDFC* commonly satisfy *CDFC** (see Online Appendix C.8).

Under *CDFC**, $\int l f(x|\cdot)$ is strictly concave. Assume also that a is such that $\mu^{SR} > 0$.[28] Then the first term in V is strictly concave, while $-c$ is concave. So, V will have the required concavity if $-\int l^s f(x|\cdot)$ is concave, or equivalently, $\int l^s f_{aa}(x|a') \geq 0$ for all a'. This is not immediate since l^s is non-monotone. But, note that f_{aa} is positive before F_{aa} reaches its peak. Thus, $\int l^s f_{aa}$ will be positive as desired if f^s has "enough" of its mass before the peak of F_{aa}. Lemma 9 in Appendix A.5 gives a number of conditions formalizing "enough." Starting from any F satisfying *CDFC**, Lemma (9) allows easy construction of densities f^s such that *FOA* is valid.

Of course, for V to be quasi-concave, it need not be that all three terms are concave. For example, since $-c$ is concave, it is enough that the sum of the first two terms is concave. In Appendix A.5, we explore this approach, and show that if c_a/c is large enough, then this is indeed true. Thus, if $c = a^\beta/\beta$, then *FOA* is valid when β is large enough. Similarly, one can simply take c_{aa} large enough to make V strictly concave at any critical point. Each of these exercises imply that c_a and c, which appear in the multipliers, are large for any given $a > 0$, and so such an exercise would be most relevant in a setting where as c_a and c got large, so did the benefit to the firm of effort via B.

8.3 High Stakes

Consider the setting of Section 7. Theorem 6 in Appendix B establishes that for u satisfying Assumption 1, a solution to two relaxed problems $\mathcal{P}^{PS}$ and $\mathcal{P}^{MH}$ exist when $\bar{u}$ is sufficiently large. The proof is novel, but builds on Kadan, Reny, and Swinkels (2017).

Now, let us turn to the validity of *FOA*. We have shown that for any given a, under a

[28] This is equivalent to $(I^s - 1)\, c_a + c\sigma \geq 0$ and so is an assumption on primitives.

variety of primitives $\int v^{SR}(x, a, \bar{u}) f(x|\cdot)$ is quasi-concave with peak at a and is strictly concave on a neighborhood of a. The following theorem closes the loop and shows that when this holds, $v^{PS}(\cdot, a, \bar{u})$ also implements a for large enough $\bar{u}$.

Theorem 5 (*FOA*: High Stakes) *Fix a and assume that $\int v^{SR}(x, a, \bar{u}) f(x|\cdot)$ is quasi-concave with peak at a and is strictly concave on a neighborhood of a. Then under Assumption 1, for $\bar{u}$ large enough, $v^{PS}(\cdot, a, \bar{u})$ is feasible and hence optimal.*

The proof is in Appendix A.5. The idea of the proof is that under the premise, the payoffs to the agent facing v^{SR} are strictly concave near a, and strictly negative for a' further away from a. But then, since $v^{PS} - v^{SR}$ converges to 0 uniformly, the same two properties are true for v^{PS}, and thus a is the unique best response to v^{PS}.

9 Conclusion

In many settings, the principal's problem is not just to get the agent to work hard, but also to work on the right things. We explore a setting which differs from the classic moral hazard problem only in that the agent can "play it safe" by choosing a project that avoids extreme outcomes. This provides a simple model that has nuanced roles for both initiative and effort and allows a detailed comparison of the settings in which initiative is or is not observable.

The need to induce initiative has significant economic implications. Two main insights arise. First, under a simple condition on likelihood ratios, contracts will tend to be "more convex" when initiative must be induced: low outcomes are punished less harshly, middle outcomes are rewarded less generously, and high outcomes are rewards more generously than without the extra constraint. But, while the condition on likelihood ratios is simple and satisfied in many examples, there are also sensible examples where it fails. When it does, the conventional wisdom that failure should be treated leniently when initiative is important can be overturned, as can the intuition that success should be treated more generously. We identify the economic force driving these departures, and then, under a more permissive yet intuitive condition, pin down the relative behavior of the compensation schemes.

Second, the addition of the new constraint often adds a single-peaked function to the cost of implementing effort. When this is true, there is a sharp prediction for the effort the principal will induce compared to what she would do in the classic moral-hazard problem. If the principal has relatively low value for effort, she will *lower* induced effort. But, when the principal values output highly, she will *raise* induced effort. At an intuitive level, asking more effort of the agent creates a larger probability of outcomes that are good news about both effort and initiative, and this relaxes the problem of the principal in rewarding both things simultaneously.

For the case of square root utility, we provide explicit expressions for the relevant objects. They are driven by information-theoretic objects related to the Fisher information, but generalized to this setting. For a large class of utility functions, moreover, the solution in the square-root case also drives the solution when the outside option of the agent is substantial. Finally, in this setting, we provide a novel proof that the relaxed problem using the first-order approach indeed has a solution, and provide primitives under which the first-order approach is valid.

Our results speak to several current issues of organizational design. For example, it suggests that decision-making authority over initiative might be usefully separated from decision-making over effort. Indeed, consider Ford Motor Company's recent reorganization separating the electric vehicle initiative from the internal combustion arm of the firm. One way of rationalizing this decision is that it allows Ford to create very strong incentives for effort on issues like cost control and quality in the well-understood internal combustion area, while creating incentives for initiative in the much more fluid electric vehicle space.

As a second example, consider a firm that wishes to create an environment in which individuals who need work-life balance can thrive. If career concerns are the issue, then the firm can attempt to mitigate the problem by policies such as forbidding email exchanges outside of normal working hours and mandating minimum vacation periods, which are indeed increasingly common. But, if the issue is distinguishing initiative from playing it safe, then firms need to think hard about improving their ability to detect initiative without inducing extreme effort levels.

At a technical level, we take some useful steps towards understanding moral-hazard problems in which the agent has more actions available than a one-dimensional choice of effort. We expect that with square-root utility, information-theoretic objects analogous to those we exploit will continue to play a large role, and that the link between the square-root case and a much larger set of utility functions as the outside option grows large will persist.

Regarding future research, each of the above examples of organizational design calls for further modeling, as in each case the organizational response involves changes in the information structure. It would also be interesting to better understand when the effects of the need to induce initiative are large, and when they are small. Moreover, in our model the agent has no private information about the distribution over outcomes given the various actions. But, a CEO knows a lot about the challenges and opportunities facing her firm, faculty know if they have a great but risky idea available or are going through a less creative period, and a salesperson knows a lot about the likely outcome of aggressively pursuing a more favorable deal with a given customer. Exploring the interaction of the forces we have identified here with this private information seems of first-order interest. Finally, another topic for future research is to understand how the need to motivate initiative affects dynamic interactions between a principal and an agent.

References

Azoulay, P., G. Zivin, S. Joshua, and G. Manso (2011): "Incentives and Creativity: Evidence from the Academic Life Sciences," *The RAND Journal of Economics*, 42, 527–554.

Barron, J., and G. Waddell (2003): "Executive Rank, Pay and Project Selection," *Journal of Financial Economics*, 67, 305–349.

Beesack, P. R. (1957): "A Note on an Integral Inequality," *Proceedings of the Americal Mathematical Society*, 8(5), 875–879.

Bertsekas, D., and S. Shreve (1978): *Stochastic Optimal Control: The Discrete-Time Case.* Athena Scientific.

Biais, B., and C. Casamatta (1999): "Optimal Leverage and Aggregate Investment," *Journal of Finance*, 54, 1291–1323.

Chade, H., and N. Kovrijnykh (2016): "Delegated Information Acquisition with Moral Hazard," *Journal of Economic Theory*, 162, 55–92.

Chade, H., and J. Swinkels (2020): "The Moral Hazard Problem with High Stakes," *Journal of Economic Theory*, 187.

Demski, J., and R. Dye (1999): "Risk, Return, and Moral Hazard," *Journal of Accounting Research*, 37, 27–55.

Dittmann, I., K. Yu, and D. Zhang (2017): "How Important are Risk-Taking Incentives in Executive Compensation," *Review of Finance*, 5, 1805–1846.

Ederer, F., and G. Manso (2013): "Is Pay for Performance Detrimental to Innovation?," *Management Science*, 59, 1496–1513.

Georgiadis, G., D. Barron, and J. Swinkels (2020): "Optimal Contracts with a Risk-Taking Agent," *Theoretical Economics*, 15, 715–761.

Grossman, S., and O. Hart (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

Hellmann, T., and V. Thiele (2011): "Incentives and Innovation: A Multitasking Approach," *American Economic Journal: Microeconomics*, 3, 78–128.

Hirshleifer, D., and Y. Suh (1992): "Risk, Managerial Effort, and Project Choice," *Journal of Financial Intermediation*, 2, 308–345.

Holmström, B. (1979): "Moral Hazard and Observability," *Bell Journal of Economics*, 10, 74–91.

Holmström, B., and J. Costa (1986): "Managerial Incentives and Capital Management," *Quarterly Journal of Economics*, 101, 835–860.

Jewitt, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, pp. 1177–1190.

Kadan, O., P. Reny, and J. Swinkels (2017): "Existence of Optimal Mechanisms in Principal-Agent Problems," *Econometrica*, 85(3), 769–823.

Lambert, R. (1986): "Executive Effort and Selection of Risky Projects," *The RAND Journal of Economics*, pp. 77–88.

Laux, V. (2015): "Executive Pay, Innovation, and Risk-Taking," *Journal of Economics & Management Strategy*, 24, 275–305.

LiCalzi, M., and S. Spaeter (2003): "Distributions for the First-Order Approach to Principal-Agent Problems," *Economic Theory*, 21, 167–173.

Luenberger, D. (1969): *Optimization by Vector Space Methods*. John Wiley & Sons.

Manso, G. (2011): "Motivating Innovation," *The Journal of Finance*, 66, 1823–1860.

Mirrlees, J. (1975): "On Moral Hazard and the Theory of Unobservable Behavior," Nuffield College.

Sung, J. (1995): "Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems," *The RAND Journal of Economics*, pp. 720–743.

A Appendix A: Properties of the Optimal Contract

A.1 Proofs for Section 4

Details for Square-Root Utility Case. We start with the following preliminary lemma.

Lemma 2 (Sign of $I^s - 1 - \frac{\sigma^2}{I^a}$) *The expression $I^s - 1 - \frac{\sigma^2}{I^a}$ is strictly positive for all a.*

Proof Define $\zeta(x,a) \equiv 1 + \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)} - \frac{f^s(x)}{f(x|a)}$, noting that $\int \zeta f = 0$. Since $-\frac{f^s(\cdot)}{f(\cdot|a)}$ is strictly quasi-convex, with interior minimum at some $\tilde{x}$ for each a, while $\frac{f_a(\cdot|a)}{f(\cdot|a)}$ is strictly monotone, it follows that regardless of the sign of $\frac{\sigma}{I^a}$, $\zeta(\cdot,a)$ is either strictly increasing to the right of $\tilde{x}$ or strictly decreasing to the left of $\tilde{x}$, and so is not everywhere zero. Hence, $\int \zeta^2(x,a)f(x|a)dx > 0$. But, using that $(a+b+c)^2 = a^2+b^2+c^2+2ab+2ac+2bc$,

$$\begin{aligned}
\int \zeta^2(x,a)f(x|a)dx &= \int \left(1 + \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)} - \frac{f^s(x)}{f(x|a)}\right)^2 f(x|a)dx \\
&= \int \left(1 + \left(\frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)}\right)^2 + \left(\frac{f^s(x)}{f(x|a)}\right)^2\right) f(x|a)dx \\
&+ 2\int \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)} f(x|a)dx + 2\int \left(-\frac{f^s(x)}{f(x|a)}\right) f(x|a)dx \\
&- 2\int \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)}\frac{f^s(x)}{f(x|a)} f(x|a)dx \\
&= 1 + \frac{\sigma^2}{I^a} + I^s + 0 - 2 - 2\frac{\sigma^2}{I^a} \\
&= I^s - 1 - \frac{\sigma^2}{I^a},
\end{aligned}$$

and we are done. □

Lemma 3 (Solution Square Root Utility) *Let $u(w) = \sqrt{2w}$. Assume the constraint that $v \geq 0$ does not bind. If $c(a)I^a + c_a(a)\sigma \leq 0$, then the solution to the pure moral hazard problem $\mathcal{P}^{MH}$ solves $\mathcal{P}^{PS}$, and the multipliers are $\lambda^{MH} = \bar{u} + c(a)$ and $\mu^{MH} = \frac{c_a(a)}{I^a}$, while if $c(a)I^a + c_a(a)\sigma \geq 0$, then PS binds, and the multipliers are*

$$\lambda^{PS} = \lambda^{MH} + \eta^{PS},\ \mu^{PS} = \mu^{MH} + \frac{\eta^{PS}\sigma}{I^a},\ \textit{and}\ \eta^{PS} = \frac{c(a)I^a + c_a(a)\sigma}{(I^s-1)I^a - \sigma^2}.$$

Proof Note that $\varphi(\hat{u}) = \hat{u}^2/2$, and so $\varphi'(\hat{u}) = \hat{u}$. Thus, we can replace

$$v(x) = \varphi'(v(x)) = \lambda + \mu l(x|a) - \eta l^s(x|a)$$

in the constraints to arrive, in the case where all three constraints bind, but the constraint that $v \geq 0$ does not, at the system of equations

$$\int (\lambda + \mu l(x|a) - \eta l^s(x|a)) f(x|a)dx = \bar{u} + c(a)$$
$$\int (\lambda + \mu l(x|a) - \eta l^s(x|a)) f_a(x|a)dx = c_a(a)$$
$$\int (\lambda + \mu l(x|a) - \eta l^s(x|a)) f^s(x)dx = \bar{u}.$$

This can then be rewritten as

$$\lambda - \eta = \bar{u} + c(a)$$
$$\mu I^a - \eta\sigma = c_a(a)$$
$$\lambda + \mu\sigma - \eta I^s = \bar{u}$$

to which it can easily be verified the solution is as claimed, where by Lemma 2, $\eta^{PS} =_s c(a)I^a + c_a(a)\sigma$. The multipliers for $\mathcal{P}^{MH}$ are derived similarly. Finally, note that the value to the agent of taking the safe action facing v^{MH} is

$$\bar{u} + c(a) + \frac{c_a(a)}{I^a}\int l(x|a)f^s(x)dx = \bar{u} + c(a) + \frac{c_a(a)}{I^a}\sigma(a),$$

and so if $c(a)I^a + c_a(a)\sigma \leq 0$ then v^{MH} solves $\mathcal{P}^{PS}$. □

Lemma 4 (Negative σ: Sufficient Conditions) *If $l(\cdot|a)$ is convex then sufficient for $\sigma(a) < 0$ is that $\mathbb{E}[x|a] > \mathbb{E}[x|a_s]$. If $l(\cdot|a)$ is concave then sufficient for $\sigma(a) < 0$ is that $\mathbb{E}[x|a_s] < \hat{x}(a)$.*

Proof Consider first the case that l is convex. Note that since l^s is single peaked, $F - F^s$ is first positive and then negative, and let $\grave{x}$ be such that $F - F^s$ is positive to the left of $\grave{x}$ and negative to the right of $\grave{x}$. Then,

$$\begin{aligned}\sigma(a) &= \int l(x|a)f^s(x|a)dx = \int l(x|a)\left(f^s(x|a) - f(x|a)\right)dx \\ &= \int l_x(x|a)\left(F(x|a) - F^s(x|a)\right)dx \\ &\leq l_x(\grave{x}|a)\int \left(F(x|a) - F^s(x|a)\right)dx \\ &= l_x(\grave{x}|a)\left(\mathbb{E}[x|a_s] - \mathbb{E}[x|a]\right) < 0,\end{aligned} \tag{5}$$

where the second equality uses that $\int lf = \int f_a = 0$, and the third integrates by parts. The inequality uses that convexity of l and the sign pattern of $F - F^s$ together imply that $l_x(\grave{x}|a) -$

$l_x(x|a) =_s F(x|a) - F^s(x|a)$.

Now assume that $l(\cdot|a)$ is concave. Then by Jensen's inequality,

$$\sigma(a) = \int l(x|a) f^s(x) dx \leq l(\mathbb{E}[x|a_s]|a), \tag{6}$$

and so sufficient for $\sigma < 0$ is that $l(\mathbb{E}[x|a_s]|a) < 0$, or equivalently, $\mathbb{E}[x|a_s] < \hat{x}(a)$. □

A.2 Proofs for Section 5

At Least Two Crossings. We now prove that $v^{PS} - v^{MH}$ crosses at least twice.

Proof of Lemma 1 Since both contracts satisfy *IR*, v^{PS} and v^{MH} must cross at least once. Assume they cross exactly once, where, for example, v^{MH} crosses v^{PS} from below. Then, since by *IR* , $\int \left(v^{MH}(x) - v^{PS}(x)\right) f(x|a)dx = 0$, and since $\frac{f_a}{f}$ is increasing, it follows from an inequality in Beesack (1957) that

$$0 < \int \left(v^{MH}(x) - v^{PS}(x)\right) f(x|a) \frac{f_a(x|a)}{f(x|a)} dx = \int v^{MH}(x) f_a(x|a) dx - \int v^{PS}(x) f_a(x|a) dx$$

which is inconsistent with *IC* being satisfied for both v^{PS} and v^{MH}. We conclude that v^{PS} and v^{MH} cross at least twice. □

Rescaling Output and the Function $l^s(l^{-1}(\cdot|a))$. Consider the function $l^s(l^{-1}(\cdot|a))$ which has domain $[l(0|a), l(\bar{x}|a)]$. This is the function that arises when one rescales output such that $l(\cdot|a)$ is the identity. Let us first establish that this is strictly concave if and only if l^s_x/l_x is strictly decreasing. This follows since

$$(l^s(l^{-1}(\tau|a)))_\tau = \frac{l^s_x(l^{-1}(\tau|a)|a)}{l_x(l^{-1}(\tau|a)|a)}, \text{ and thus } (l^s(l^{-1}(\tau|a)))_{\tau\tau} = \left(\frac{l^s_x}{l_x}\right)_x l^{-1}_\tau(\tau|a) =_s \left(\frac{l^s_x}{l_x}\right)_x.$$

Similarly, l^s is semibellshaped if and only if $l^s(l^{-1}(\cdot|a))$ does not shift from strictly concave to strictly convex anywhere before the peak of l^s nor from strictly convex to strictly concave anywhere beyond the peak. Thus, *SBS* holds if (1) l^s_x/l_x is strictly quasiconcave on $[x_\ell, \tilde{x}]$ and is strictly quasiconvex on $[\tilde{x}, x_h]$ and (2) if $x_\ell > 0$ and $f^s(x_\ell) > 0$ then l^s_x/l_x is strictly decreasing on $[x_\ell, \tilde{x}]$ while if $x_h < \bar{x}$ and $f^s(x_h) > 0$ then l^s_x/l_x is strictly decreasing on $[\tilde{x}, x_h]$.

For an example, let f^s be uniform on $[0, \bar{x}]$. Then one can show that

$$\left(\frac{\left(\frac{f^s}{f}\right)_x}{\left(\frac{f_a}{f}\right)_x}\right)_x =_s f_x f_{axx} - f_{ax} f_{xx}.$$

Thus, necessary and sufficient for l^s_x/l_x to be strictly decreasing is that $f_x f_{axx} - f_{ax} f_{xx} < 0$.

Equivalently, $|f_x|$ is log-submodular. In the spanning case where $f = (1-a)f_l + af_h$,

$$f_x f_{axx} - f_{ax} f_{xx} = f_{lx} f_{hxx} - f_{hx} f_{lxx},$$

and so sufficient is that f_l is strictly convex and strictly decreasing and f_h is strictly convex and strictly increasing.

If f^s has less than full support, then under the same conditions, l^s is semibellshaped because it is convex on its support. In general l^s will be semi-bellshaped if $|f_x|$ does not change from log-submodular to log-supermodular before the minimum of f, or from log-supermodular to log-submodular after the minimum of f.

A Non-Monotone v^{PS}. We asserted in main text that v^{PS} can be decreasing for low outputs. To see this, note that when l^s is differentiable, since $l^s_x(0) > 0$, a sufficient condition for $v_x(0) < 0$ is $\mu^{PS} < 0$. But, substituting from (3) and simplifying, $\mu^{PS} =_s c_a(a)(I^s - 1) + c(a)\sigma$. So for example, let $f(x|a) = (1-a)f_\ell(x) + af_h(x)$, where $\frac{f_h}{f_\ell}$ is increasing, and let $c(a) = a^2$, noting that since f is linear in a, there is no issue about the validity of the first-order approach. One can show that μ^{PS} is negative at $a = 1$ if and only if

$$\int (2f^s - f_\ell - f_h) \frac{f^s}{f_h} dx < 0.$$

Thus, consider $f^s = 6x(1-x)$, $f_h = bx^{b-1}$, $f_\ell = dx^{d-1}$ on $[0,1]$. Note that for $b > d$, f^s is single-peaked, while f is single-troughed, and so our condition that f^s crosses f first from below and then from above is satisfied. It is easily checked numerically that $\mu^{PS}(1) < 0$ for $b \in [2, 2.2]$, and $d \in [.2, 5]$, and hence $\mu^{PS} < 0$ for a sufficiently close to 1.

Details for Example 2. Note first that

$$\sigma = \int_{3/8}^{1/2} \frac{8}{1+\frac{a}{3}} \frac{1}{3} dx = \frac{1}{a+3}$$

while

$$I^a = \int_0^{1/8} \frac{(-1)^2}{1-\tau-a} dx + \int_{1/8}^{1/4} \frac{(-1)^2}{1+\tau-a} dx + \int_{1/4}^{1} \frac{\left(\frac{1}{3}\right)^2}{1+a/3} dx = \frac{1}{4\,(a+3)} \frac{\tau^2+4a-4}{-a^2+2a+\tau^2-1},$$

and so

$$\frac{\sigma}{I^a} = \frac{\frac{1}{a+3}}{\frac{1}{4(a+3)} \frac{\tau^2+4a-4}{-a^2+2a+\tau^2-1}} = \frac{4\left(1+a^2-2a-\tau^2\right)}{4-\tau^2-4a},$$

where when $a < 1-\tau$, both top and bottom are positive.

But, on $[0,1/8]$, $l=-1/(1-\tau-a)$, and $l^s=0$, and so

$$v^{PS}-v^{MH}=_s 1+\frac{\sigma}{I^a}l-l^s=1+\frac{4\left(1+a^2-2a-\tau^2\right)}{4-\tau^2-4a}\frac{-1}{1-\tau-a}=-\tau\frac{\tau+4}{4-\tau^2-4a}<0.$$

Similarly, on $(1/8,1/4]$

$$v^{PS}-v^{MH}=_s 1+\frac{4\left(1+a^2-2a-\tau^2\right)}{4-\tau^2-4a}\frac{-1}{1+\tau-a}=\tau\frac{4-\tau}{4-\tau^2-4a}>0,$$

on $(1/4,3/8)$ and $(1/2,1]$

$$v^{PS}-v^{MH}=_s 1+\frac{4\left(1+a^2-2a-\tau^2\right)}{4-\tau^2-4a}\frac{\frac{1}{3}}{1+\frac{a}{3}}>0,$$

while on $[3/8,1/2]$

$$\begin{aligned}
v^{PS}-v^{MH} &=_s 1+\frac{4\left(1+a^2-2a-\tau^2\right)}{4-\tau^2-4a}\frac{\frac{1}{3}}{1+\frac{a}{3}}-\frac{8}{1+\frac{a}{3}}\\
&=_s (17-a)\,\tau^2-80(1-a)=80a+17\tau^2-a\tau^2-80\\
&\leq (17-(1-\tau))\,\tau^2-80(1-(1-\tau))\\
&=\tau\left(\tau^2+16\tau-80\right)<0.
\end{aligned}$$

Details for Example 3. The Fisher information I^a is $1/a^2$. To verify this, note that $f_a(x|a)=-e^{-\frac{x}{a}}\left(a-x\right)/a^3$ and so

$$l(x|a)=\frac{-\frac{1}{a^3}e^{-\frac{1}{a}x}\left(a-x\right)}{\frac{1}{a}e^{-\frac{x}{a}}}=\frac{1}{a^2}(x-a)$$

But then,

$$I^a=\int\left(\frac{1}{a^2}(x-a)\right)^2\frac{1}{a}e^{-\frac{x}{a}}dx=\frac{1}{a^2}.$$

Next, note that

$$\sigma=\int\frac{f^s(x|a)}{f(x|a)}f_a(x|a)dx=\int_1^\infty\frac{e^{-(x-1)}}{\frac{1}{a}e^{-x/a}}\left(-\frac{1}{a^3}e^{-\frac{1}{a}x}\left(a-x\right)\right)dx=-\frac{1}{a^2}e^{1-x}\left(x-a+1\right)|_1^\infty=\frac{2-a}{a^2}.$$

Finally, since $f^s=e^{-(x-1)}$ on $[1,\infty)$ we have

$$I^s=\int_1^\infty\frac{\left(e^{-(x-1)}\right)^2}{\frac{1}{a}e^{-x/a}}dx=-a^2\frac{\exp\left(\frac{1}{a}\left(2a+x-2ax\right)\right)}{2a-1}\Bigg|_1^\infty,$$

which is infinite for $a \in (0, \frac{1}{2})$, while for $a > \frac{1}{2}$ it is equal to

$$I^s = \frac{a^2}{2a-1} e^{\frac{1}{a}}.$$

Next, let us derive the crossing behavior of v^{MH} and v^{PS}. We have that

$$v^{PS} - v^{MH} =_s \frac{v^{PS} - v^{MH}}{\eta} = 1 + \frac{\sigma}{I^a} l(x|a) - l^s(x|a) = 1 + (a-2)\frac{1}{a^2}(a-x) - l^s(x|a)$$

where $l^s(x|a) = 0$ for $x < 1$, and $l^s(x|a) = ae^{-(x-1)+x/a}$ for $x \geq 1$. For $x \in [0,1)$, the last expression is clearly positive for $a \geq 2$. For $x \geq 1$, it is routine to establish that for $a \in (2,4)$ the last expression is strictly concave in x, strictly negative at $x = 1$, strictly positive at $x = 5$ and strictly negative for x large enough. Hence v^{PS} is *HLHL*.

Finally, let us calculate the value to the agent of deviating to effort t facing $v^{PS}(\cdot; a)$, the contract that implements effort a in the relaxed problem. The utility gain from the deviation is

$$\begin{aligned} -\bar{u} - t^2 + \int v^{PS}(x;a) f(x|t)dx &= -\bar{u} - t^2 + \int \left(\bar{u} + c(a) + \frac{c_a(a)}{I^a} l(x|a) + \eta \left(1 + \frac{\sigma}{I^a} l(x|a) - l^s(x|a) \right) \right) f(x|t)dx \\ &= c(a) + \eta - t^2 + \left(\frac{c_a(a)}{I^a} + \eta \frac{\sigma}{I^a} \right) \int l(x|a) f(x|t)dx - \eta \int l^s(x|a) f(x|t)dx \\ &= a^2 + \eta - t^2 + \left(2a + \eta \frac{2-a}{a^2} \right) \int (x-a) \frac{1}{t} e^{-\frac{x}{t}} dx - \eta \frac{a}{t} \int_1^\infty e^{-(x-1)+\frac{x}{a}-\frac{x}{t}} dx \end{aligned}$$

But,

$$\int (x-a) \frac{1}{t} e^{-\frac{x}{t}} dx = -e^{-\frac{1}{t}x} (t - a + x) \Big|_0^\infty = t - a,$$

and

$$\int_1^\infty e^{-(x-1)+\frac{x}{a}-\frac{x}{t}} dx = -\frac{e^{1-x\left(1-\frac{1}{a}+\frac{1}{t}\right)}}{1-\frac{1}{a}+\frac{1}{t}} \Bigg|_1^\infty = \frac{e^{\frac{1}{a}-\frac{1}{t}}}{1-\frac{1}{a}+\frac{1}{t}}$$

where since $a \geq 2$, $1 - \frac{1}{a} + \frac{1}{t}$ is strictly positive. The gain to the deviation is thus

$$a^2 + \eta - t^2 + \left(2a + \eta \frac{2-a}{a^2} \right) (t-a) - \eta \frac{a}{t} \frac{e^{\frac{1}{a}-\frac{1}{t}}}{1-\frac{1}{a}+\frac{1}{t}} = -(a-t)^2 + \eta \left(1 + \frac{2-a}{a^2}(t-a) - \frac{a}{t} \frac{e^{\frac{1}{a}-\frac{1}{t}}}{1-\frac{1}{a}+\frac{1}{t}} \right).$$

Note finally that

$$\eta = \frac{cI^a + c_a \sigma}{(I^s - 1) I^a - \sigma^2} = \frac{a^2 \frac{1}{a^2} + 2a \frac{2-a}{a^2}}{\left(\frac{a^2}{2a-1} e^{\frac{1}{a}} - 1 \right) \frac{1}{a^2} - \left(\frac{2-a}{a^2} \right)^2} = \frac{a(4-a)}{\frac{a^2}{2a-1} e^{\frac{1}{a}} - 1 - \left(\frac{2-a}{a} \right)^2}.$$

In Online Appendix C.6, we show that if the principal chooses to induce initiative, she will induce at least an effort of 2.8 or above (the intuition is that because the safe project can be induced by

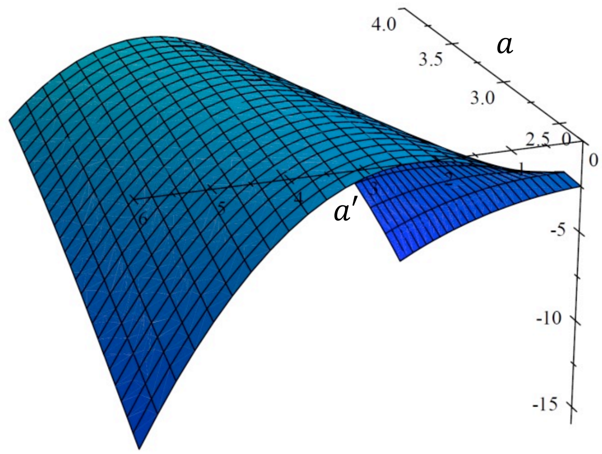

Figure 5: ***Agent Optimality.*** The figure depicts the agent's expected payoff from deviating from a to a'. It shows that this payoff is negative.

paying the outside option, inducing an effort a little above 2 makes sense only if β is very large. But then, a higher effort is better still). Figure 5 plots for $a \in [2.8, 4]$, the value to the agent of deviating to any given action a' when faced with the contract solving $\mathcal{P}^{PS}$ for that a. It is clear that the agent has no profitable deviation.

SBS and at Most Three Crossings. We now prove the results regarding the structure of compensation under the semi-bellshaped condition.

Proof of Theorem 2 Fix and suppress a and let $\underline{\tau} \equiv l(0)$, $\bar{\tau} \equiv l(\bar{x})$, $\tau_\ell = l(x_\ell)$, $\tau_h = l(x_h)$, and $\tilde{\tau} = l(\tilde{x})$, where recall that $[x_\ell, x_h]$ is the support of f^s and $\tilde{x}$ be the maximizer of l^s.

Consider first the case $\mu^{PS} - \mu^{MH} = 0$. Then, it cannot be that $\lambda^{PS} - \lambda^{MH} \leq 0$, since then D is always negative, violating Lemma 1. Thus, any crossing of D with zero occurs where $l^s(l^{-1}) > 0$ and so on $[\tau_\ell, \tau_h]$ an interval over which $D(\tau)$ is strictly quasiconvex given $\mu^{PS} - \mu^{MH} = 0$ and thus crosses zero at most two times. By Lemma 1 it follows that v^{PS} is *HLH*.

Now, let us turn to the case $\mu^{PS} - \mu^{MH} > 0$. If D changes sign three or more times, then over some interval D must have sign pattern $-/+/-$. Take the rightmost region $J = [\tau', \tau''] \subseteq (\underline{\tau}, \bar{\tau})$ over which D is positive, and where D changes sign at τ' and τ''. We will show that D is strictly negative on $[\underline{\tau}, \tau')$. But then, the pattern $-/+/-$ occurs over at most one interval, and if it does, then there can be one more region where D is positive to the right of τ'', but this region must include $\bar{\tau}$. No further crossings of D are possible. Hence, the only sign pattern consistent with more than two crossings results in v^{PS} being *LHLH*.

Assume first that $\tau' < \tau_\ell$. Then since D is continuous at τ', $D(\tau') = 0$ and D is strictly increasing on $[\underline{\tau}, \tau_\ell)$ and it it immediate that D is strictly negative on $[\underline{\tau}, \tau')$. Next, assume $\tau' = \tau_\ell$. Then, if l^s jumps up at τ_ℓ then D is strictly positive on an interval to the right of

τ', contradicting that D changes sign at τ'. Thus τ' is again a continuity point of D, and so $D(\tau') = 0$ and D is strictly increasing on $[\underline{\tau}, \tau_\ell]$ and thus D is strictly negative on $[\underline{\tau}, \tau')$. Finally, assume $\tau' > \tau_\ell$. If $\tau'' \geq \tilde{\tau}$, then we have a contradiction, since $D \geq 0$ on (τ', τ'') and D is strictly increasing on $[\tilde{\tau}, \bar{\tau}]$ (using that l^s is decreasing and $\mu^{PS} - \mu^{MH} > 0$) and hence D is strictly positive on $[\tau'', \bar{\tau}]$ contradicting that D changes sign at τ''. Hence, we have $\tau_\ell < \tau' < \tau'' < \tilde{\tau}$ and so D is continuous at τ' and τ'' and so is equal to zero at each. It follows that l^s is strictly convex at τ'. To see this, note that if $l^s(l^{-1})$ is concave at τ' then by *SBS* it is strictly concave on $(\tau', \tau'']$. But then since $D(\tau') = D(\tau'') = 0$, D is strictly negative on $[\tau', \tau'']$, a contradiction. Thus, $D(\tau') = 0$, $D_x(\tau') \geq 0$, D is strictly concave on (τ_ℓ, τ') and concave on $[\underline{\tau}, \tau']$ and so $D < 0$ on $[\underline{\tau}, \tau')$, and done.

A similar argument establishes that if $\mu^{PS} - \mu^{MH} < 0$, and if $J = [\tau', \tau'']$ is interior to $[\underline{\tau}, \bar{\tau}]$, with D positive on J and changing signs at τ' and τ'', then D is strictly negative everywhere to the right of τ'' and so the only sign pattern consistent with more than two crossings is *HLHL*. In particular, if $\tau'' \geq \tau_h$ then since D is strictly decreasing to the right of τ_h, it is strictly negative to the right of τ'', while if $\tau'' < \tau_h$ then one argues symmetrically to above to show that $\tilde{\tau} < \tau' < \tau'' < \tau_h$ and so D must be concave and hence strictly negative from τ'' onwards. □

Proof of Proposition 1 Note that $\lambda^{PS} - \lambda^{MH} + (\mu^{PS} - \mu^{MH})\tau - \eta \underline{l}^s(\tau)$ is linear in τ, and hence if negative at both $\underline{\tau}$ and $\bar{\tau}$, is negative everywhere. But then, using the premise, $v^{PS} - v^{MH}$ is everywhere negative, violating Lemma 1. Thus, v^{PS} is strictly above v^{MH} at at least one of 0 and $\bar{x}$. But then, if there are only two crossings, *HLH* holds. □

A.3 Proofs for Section 6

Derivation of Δ in Square-Root Case. Note that $v^{PS} = v^{MH} + \eta(1 + (\sigma/I^a)l(x|a) - l^s(x|a))$. Thus,

$$\begin{aligned}
C^{PS}(a, \bar{u}) &= \frac{1}{2}\int \left(v^{PS}(x)\right)^2 f(x|a)dx \\
&= \frac{1}{2}\int \left(v^{MH} + \eta\left(1 + \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)} - \frac{f^s(x)}{f(x|a)}\right)\right)^2 f(x|a)dx \\
&= \frac{1}{2}\int \left(v^{MH}\right)^2 f(x|a)dx + \frac{\eta}{2}\int v^{MH}\left(1 + \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)} - \frac{f^s(x)}{f(x|a)}\right) f(x|a)dx \\
&+ \frac{\eta^2}{2}\int \left(1 + \frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)} - \frac{f^s(x)}{f(x|a)}\right)^2 f(x|a)dx,
\end{aligned}$$

where we note that $\frac{1}{2}\int \left(v^{MH}\right)^2 f(x|a)dx = C^{MH}(a)$. Consider the second term, and note that

$$\int \left(1 + \frac{\sigma}{I^a}\frac{f_a}{f} - \frac{f^s}{f}\right) f dx = \int f dx + \frac{\sigma}{I^a}\int f_a dx - \int f^s dx = 0.$$

Hence,

$$\int v^{MH}\left(1+\frac{\sigma}{I^a}\frac{f_a}{f}-\frac{f^s}{f}\right)dx=\int\left(\lambda^{MH}+\mu^{MH}\frac{f_a}{f}\right)\left(1+\frac{\sigma}{I^a}\frac{f_a}{f}-\frac{f^s}{f}\right)fdx$$
$$=\mu^{MH}\int\frac{f_a}{f}\left(1+\frac{\sigma}{I^a}\frac{f_a}{f}-\frac{f^s}{f}\right)fdx$$
$$=\mu^{MH}\left(\int f_a dx+\frac{\sigma}{I^a}\int\frac{f_a^2}{f}dx-\int\frac{f_a}{f}\frac{f^s}{f}f\right)dx$$
$$=\mu^{MH}\left(0+\frac{\sigma}{I^a}I^a-\sigma\right)$$
$$=0$$

and so we have

$$\Delta=\frac{\eta^2}{2}\int\left(1+\frac{\sigma}{I^a}\frac{f_a(x|a)}{f(x|a)}-\frac{f^s(x)}{f(x|a)}\right)^2 f(x|a)dx.$$

But then, by Lemma 2,

$$\Delta=\frac{\eta^2}{2I^a}((I^s-1)I^a-\sigma^2).$$

Recalling that

$$\eta=\frac{cI^a+c_a\sigma}{(I^s-1)I^a-\sigma^2}$$

we have, after taking the cancellation, that

$$\Delta=\frac{1}{2I^a}\frac{\left(c+c_a\frac{\sigma}{I^a}\right)^2}{(I^s-1)-\frac{\sigma^2}{I^a}},$$

where since by assumption PS binds, we have $cI^a+c_a\sigma>0$, and we are done.

To sketch a different proof, note that if one replaces the *rhs* of PS by $\bar{u}+\tau$, and solves for the multipliers, then for $\tau\in[0,c+\frac{c_a}{I^a}\sigma]$,

$$\eta(\tau)=\frac{\sigma c_a+(c-\tau)\,I^a}{(I^s-1)I^a-\sigma^2},$$

while for higher τ, v^{MH} is feasible and hence optimal. But then, since $\eta(\tau)$ is the shadow value of tightening PS,

$$\Delta=\int_0^{c+c_a\frac{\sigma}{I^a}}\eta(\tau)d\tau$$

which is easily shown to agree with our previous expression.

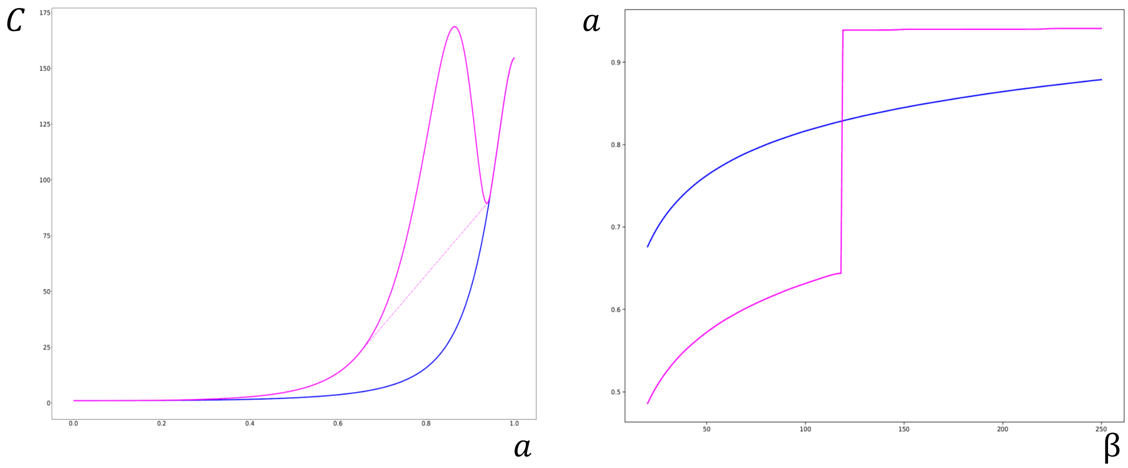

Figure 6: ***Costs and Optimal Efforts.*** The left panel depicts that the difference between C^{PS} and C^{MH} is single-peaked and large. The right panel shows that optimal effort in *PS* is first substantially below and then substantially above that in *MH*.

Details for Example 4. Using the expressions from Example 3, we obtain that for $a \geq \frac{1}{2}$,

$$\Delta = \frac{(c(a)I^a + c_a(a)\sigma)^2}{-\sigma^2 + (I^s - 1)I^a}\frac{1}{I^a} = \frac{\left(a^2 \frac{1}{a^2} + 2a\left(\frac{2-a}{a^2}\right)\right)^2}{-\left(\frac{2-a}{a^2}\right)^2 + \left(\frac{a^2}{2a-1}e^{\frac{1}{a}} - 1\right)\frac{1}{a^2}}\frac{1}{\frac{1}{a^2}} = \frac{(a(4-a))^2}{\frac{a^2}{2a-1}e^{\frac{1}{a}} - 1 - \frac{(2-a)^2}{a^2}},$$

which is equivalent to the expression stated in main text.

An Example Where Effort Distortions are Large in Both Directions. Consider the following (carefully constructed) example. There are four outputs, $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, and $x_4 = 10$. Effort lies in $[0, 1]$ with probabilities of output given $a \in [0, 1]$ given by $p_1 = (1/4)(1-a)$, $p_2 = (1/4)(1-a)$, $p_3 = 0.35(1+a)$, and $p_4 = 0.15(1+a)$, while under a_s, $p_2^s = 0.05$ and $p_3^s = 0.95$.[29] Since probabilities are linear in a, the first-order approach is valid. The disutility of effort a is $c(a) = (1/(1.15 - a)) - (1/1.15) - (1/(1.15)^2)a$, utility of income is $u(w) = \log w$, and $\bar{u} = 0$. In Figure 6, the left panel shows C^{MH} in magenta and C^{PS} in blue. The difference between them is single-peaked and PS ceases to bind for a close to one. The right panel shows the optimal efforts as a function of β.[30] The jump in a^{PS} occurs where $\beta\mathbb{E}_a[x|a]$ equals the slope of the dotted line in the left panel.[31] This generates an extreme example of Theorem 3.

A perhaps surprising feature of the example is that C^{PS} is not monotone. The crux is that when a is in the relevant range, the principal finds it very attractive to reduce v_2, where p^s/p is

[29] The example can be modified to make *MLRP* strict.

[30] It is easily verified that $\mathbb{E}[x|0] > \mathbb{E}[x|a_s]$. Hence, since a flat contract that pays the outside option induces a_s and $a = 0$ in either $\mathcal{P}^{MH}$ and $\mathcal{P}^{PS}$, it follows that for any $\beta > 0$, the principal prefers implementing $a = 0$ to a_s in either *MH* or *PS*. *A fortiori*, she is better off to implement the optimal effort than a_s.

[31] The jump can be made arbitrarily large by lowering p_2^s, or by raising p_3 while lowering p_4.

large, to discourage a_s. This has the side effect of providing excessive incentives for effort, and so to restore IC, payments at x_1 have to be *larger* than payments at x_4. Increasing a then lowers costs both because it becomes easier to distinguish a_s from a and because the distortion between payments at x_1 and x_4 becomes smaller.

Effort Distortions. Let $\hat{a}$ be the action at which Δ reaches its maximum, and let $\hat{\tau}$ be such that for $\tau < \hat{\tau}$ we have $a^{PS}(\tau) \leq \hat{a}$ and for $\tau > \hat{\tau}$, $a^{PS} \geq \hat{a}$, noting that $B - C^{PS}$ is strictly supermodular in (a, τ), and so such a $\hat{\tau}$ exists.

Proof of Theorem 3 Note that for any τ,

$$B(a^{PS}(\tau), \tau) - C^{PS}(a^{PS}(\tau)) \geq B(a^{MH}(\tau), \tau) - C^{PS}(a^{MH}(\tau))$$
$$B(a^{MH}(\tau), \tau) - C^{MH}(a^{MH}(\tau)) \geq B(a^{PS}(\tau), \tau) - C^{MH}(a^{PS}(\tau))$$

and so, adding the two inequalities and manipulating,

$$C^{PS}(a^{MH}(\tau)) - C^{MH}(a^{MH}(\tau)) \geq C^{PS}(a^{PS}(\tau)) - C^{MH}(a^{PS}(\tau))$$

or

$$\Delta(a^{MH}(\tau)) \geq \Delta(a^{PS}(\tau)). \tag{7}$$

Assume that $\tau > \hat{\tau}$ so that $a^{PS}(\tau) \geq \hat{a}$. If $\Delta(a^{MH}(\tau)) = 0$, then $a^{PS}(\tau) = a^{MH}(\tau)$, and we are done. So, assume $\Delta(a^{MH}(\tau)) > 0$ and towards a contradiction, that $a^{MH}(\tau) > a^{PS}(\tau)$. Then, Δ is strictly decreasing between $a^{PS}(\tau)$ and $a^{MH}(\tau)$, contradicting (7). Thus, $a^{PS}(\tau) \geq a^{MH}(\tau)$.

Finally, assume $a^{MH}(\tau)$ is interior, and towards a contradiction, assume that $a^{PS}(\tau) = a^{MH}(\tau)$. Then $B_a(a^{MH}(\tau), \tau) - C_a^{PS}(a^{MH}(\tau)) = B_a(a^{MH}(\tau), \tau) - C_a^{MH}(a^{MH}(\tau)) = 0$, and so $\Delta_a(a^{MH}(\tau)) = 0$. But then, $a^{PS}(\tau) = a^{MH}(\tau) = \hat{a}$ where by assumption $\hat{a}$ is interior. Let $\tau' = (\hat{\tau} + \tau)/2$. Since B is strictly supermodular, $a^{PS}(\tau') \leq \hat{a}$ and since $B_a(\hat{a}, \tau') - C_a^{PS}(\hat{a}) < 0$, in fact $a^{PS}(\tau') < \hat{a}$, contradicting the definition of $\hat{\tau}$, and we are done. □

A.4 Proofs for Section 7

The proof of Theorem 4 will follow from several technical lemmas, which will also allow us to derive some additional properties of the problem when $\bar{u}$ is sufficiently large. Some of the proofs of these lemmas are in Online Appendix C.7.

Let us first derive the equations that define the multipliers. Recall that $\varphi = u^{-1}$. We have the following expressions for λ, μ, and η.

Lemma 5 (Multipliers) *Where PS binds, the multipliers λ, μ, and η are implicitly defined by*

$$\lambda = \int \varphi'(v^{PS}(x,a,\bar{u}))f(x|a)dx + \eta,$$
$$\mu = \frac{\int \varphi'(v^{PS}(x,a,\bar{u}))f_a(x|a)dx}{I^a} + \frac{\eta\sigma}{I^a}, \text{ and}$$
$$\eta = \frac{\int \varphi'(v^{PS}(x,a,\bar{u}))\left[I^a\left(1-l^s(x|a)\right)+\sigma l(x|a)\right]f(x|a)dx}{I^a\left(I^s-1\right)-\sigma^2}.$$

For a given contract v, define

$$W(v) = \max_x v(x) - \min_x v(x),$$

as the maximum amount by which v differs at its highest and lowest points, where W is mnemonic for "wiggle." The following lemma shows that if v^{PS} has bounded wiggle, then as $\bar{u}$ diverges, the multipliers λ, μ, and η take on very simple forms. The predicate $W(v^{PS}(\cdot,a,\bar{u})) < J$ will automatically hold for some $J < \infty$ when PS is satisfied at v^{MH} as shown in CS Lemma 3. The reason for this at an intuitive level is that v^{MH} is monotone, and a monotone contract that rises by more than a certain amount will provide excessively strong incentives, violating IC . But, because PS contracts may cease to be monotone, and because of the complexities that η adds, we will have to work harder to bound W. We do so below.

Lemma 6 (Limit Multipliers) *Let Assumption 1 hold, let $0 < J < \infty$, and let $\varepsilon > 0$. Then, there is $\bar{u}^* < \infty$ such that for all $\bar{u} > \bar{u}^*$, and for all a, if $W(v^{PS}(\cdot,a,\bar{u})) < J$, and if PS binds, then*

$$\left|\frac{\lambda^{PS}}{\varphi'(\bar{u}+c(a))} - 1\right| < \varepsilon, \left|\frac{\mu^{PS}}{\varphi''(\bar{u}+c(a))} - \frac{(I^s-1)c_a(a)+\sigma c(a)}{(I^s-1)I^a-\sigma^2}\right| < \varepsilon, \left|\frac{\eta^{PS}}{\varphi''(\bar{u}+c(a))} - \frac{c(a)I^a+c_a(a)\sigma}{(I^s-1)I^a-\sigma^2}\right| < \varepsilon.$$

If PS does not bind, so that $v^{PS} = v^{MH}$, then $\eta^{PS} = 0$, and

$$\left|\frac{\lambda^{MH}}{\varphi'(\bar{u}+c(a))} - 1\right| \le \varepsilon, \text{ and} \left|\frac{\mu^{MH}}{\varphi''(\bar{u}+c(a))} - \frac{c_a(a)}{I^a}\right| < \varepsilon.$$

Note that where $c(a)I^a + c_a(a)\sigma = 0$, we have $c(a) = -c_a(a)\frac{\sigma}{I^a}$. But then,

$$\frac{(I^s-1)c_a(a)+\sigma c(a)}{(I^s-1)I^a-\sigma^2} = \frac{c_a(a)}{I^a},$$

and so the two versions of v^{SR} agree, and thus v^{SR} is continuous. Note also that since $(I^s-1)I^a - \sigma^2 > 0$ and $I^s - 1 > 0$ all the limiting multipliers are positive, with μ strictly positive. Hence, since for $x > \mathbb{E}[x|a_s]$ sufficiently large, $-l^s(\cdot|a)$ is strictly increasing , while $l(\cdot|a)$ is everywhere strictly increasing, v^{SR} is not constant except when $a = 0$.

Let $J^{SR} \equiv \max_a W\left(v^{SR}(\cdot, a, \bar{u})\right)$ be the maximum wiggle that v^{SR} takes on as one varies a. This is finite, since $\bar{u}$ cancels out, and the remaining expression of a and x is continuous over a compact set. It is also strictly positive, since v^{SR} is not constant when $a > 0$.

Now, let us consider v^{PS}. We will show that in a very strong sense, $v^{PS}(\cdot, a, \bar{u})$ behaves in the limit like $v^{SR}(\cdot, a, \bar{u})$. Recall the definition of $d(a, \bar{u})$ and $d_x(a, \bar{u})$ given in Section 7.

We begin by showing that where $c(a)I^a + c_a(a)\sigma < 0$, PS ceases to bind for large $\bar{u}$, and the contract converges to one that is simply $v^{SR}(\cdot, a)$, which in this case is the standard contract in the square-root case with pure moral hazard.

Lemma 7 (*PS* Not Binding) *Let Assumption 1 hold, and let $c(a)I^a + c_a(a)\sigma < 0$. Then, for all $\varepsilon > 0$, there is $\bar{u}^* < \infty$ such that for all $\bar{u} > \bar{u}^*$,*

$$v^{PS}(\cdot, a, \bar{u}) = v^{MH}(\cdot, a, \bar{u}),\ d_x(a, \bar{u}) < \varepsilon \text{ and } d(a, \bar{u}) < \varepsilon.$$

If $c(a)I^a + c_a(a)\sigma > 0$ then for large $\bar{u}$, PS fails at $v^{MH}(\cdot, a, \bar{u})$.

Proof Choose a where $c(a)I^a + c_a(a)\sigma < 0$, and consider first $v^{MH}(\cdot, a, \bar{u})$. Consider any $\bar{u} > \bar{u}^*$, and let ρ be the function defined by $\varphi'(\rho(\tau)) = \tau$. Since $v^{MH}(x, a, \bar{u}) = \rho(\lambda + \mu l(x|a))$,

$$v_x^{MH}(x, a, \bar{u}) = \rho'\left(\lambda + \mu l(x|a)\right)\mu l_x(x|a) > 0.$$

But, since $\varphi'(\rho(\tau)) = \tau$, we have $\varphi''(\rho(\tau))\rho'(\tau) = 1$, and so

$$\rho'\left(\lambda + \mu l(x|a)\right) = \frac{1}{\varphi''(v^{MH}(x, a, \bar{u}))}.$$

Substituting and then multiplying and dividing by $\varphi''(\bar{u} + c)$, we obtain

$$v_x^{MH}(x, a, \bar{u}) = \frac{\varphi''(\bar{u} + c)}{\varphi''(v^{MH}(x, a, \bar{u}))} \frac{\mu}{\varphi''(\bar{u} + c)} l_x(x|a).$$

But, by CS, Lemma 3, there is some $J < \infty$ such that for all $\bar{u}$ sufficiently large, $v^{MH}(x, a, \bar{u}) - \bar{u} - c(a) < J$ for all x and a. It follows from CS Lemma 1 that

$$\frac{\varphi''(\bar{u} + c)}{\varphi''(v^{MH}(x, a, \bar{u}))} \to 1$$

uniformly in x and a. Also by CS, Proposition 1,

$$\frac{\mu}{\varphi''(\bar{u} + c(a))} \to \frac{c_a(a)}{I^a}$$

uniformly in a, and so it follows that

$$v_x^{MH}(x,a,\bar{u}) - \frac{c_a(a)}{I^a} l_x(x|a) \to 0$$

uniformly in x and a, establishing that for $\bar{u}$ sufficiently large and for all a, $d_x(a,\bar{u}) < \varepsilon$. Thus, recalling that $\hat{x}(a)$ is the point where $l(x|a) = 0$,

$$v^{MH}(x,a,\bar{u}) - v^{MH}(\hat{x}(a),a,\bar{u}) \to \frac{c_a(a)}{I^a} l(x|a) \tag{8}$$

uniformly in x.

Now, from IR,

$$\int v^{MH}(x,a,\bar{u}) f(x|a)dx - \bar{u} - c(a) = 0,$$

and so, adding and subtracting $v^{MH}(\hat{x}(a),a,\bar{u})$ and rearranging,

$$v^{MH}(\hat{x}(a),a,\bar{u}) - \bar{u} - c(a) + \int \left(v^{MH}(x) - v^{MH}(\hat{x}(a))\right) f(x|a)dx = 0$$

But, by (8),

$$\int \left(v^{MH}(x,a,\bar{u}) - v^{MH}(\hat{x}(a),a,\bar{u})\right) f(x|a)dx \to \frac{c_a(a)}{I^a} \int l(x|a) f(x|a)dx = 0$$

and hence

$$v^{MH}(\hat{x}(a),a,\bar{u}) - \bar{u} - c(a) \to 0.$$

It follows that

$$v^{MH}(x,a,\bar{u}) - \left(\bar{u} + c(a) + \frac{c_a(a)}{I^a} l(x|a)\right) \to 0,$$

uniformly in x and a, and so since $v^{SR}(\cdot,a) = \bar{u} + c(a) + \frac{c_a(a)}{I^a} l(\cdot|a)$ where $c(a)I^a + c_a(a)\sigma < 0$, we have shown that for all $\bar{u}$ sufficiently large and for all a, $d(a,\bar{u}) < \varepsilon$, establishing the first claim.

To establish the remaining claims, note that the value of taking a_s over $\bar{u}$ facing v^{MH} is

$$\begin{aligned}\int v^{MH}(x,a,\bar{u}) f^s(x)dx - \bar{u} &= \int \left(v^{MH}(x,a,\bar{u}) - \bar{u}\right) f^s(x)dx \\ &\to \int \left(c(a) + \frac{c_a(a)}{I^a} l(x|a)\right) f^s(x)dx \\ &= c(a) + \frac{c_a(a)}{I^a}\sigma,\end{aligned}$$

and so if $c(a)I^a + c_a(a)\sigma < 0$ then for high $\bar{u}$, PS does not bind at $v^{MH}(\cdot,a,\bar{u})$, while if $c(a)I^a + c_a(a)\sigma > 0$ then for high $\bar{u}$, $v^{MH}(\cdot,a,\bar{u})$ fails PS . □

Our next lemma shows that as $\bar{u}$ grows, for each a, one of two things happens. Either $v^{PS}(\cdot, a, \bar{u})$ and $v^{SR}(\cdot, a, \bar{u})$ grow arbitrarily close to each other, or they stay a large distance apart. Intermediate outcomes do not occur.

Lemma 8 (Distance between v^{PS} and v^{SR}) *Let Assumption 1 hold. Then, for each $\varepsilon \in (0, J^{SR}/2)$, there is a threshold $\bar{u}^* < \infty$ such that for all $\bar{u} > \bar{u}^*$, and for all a, either $d(a, \bar{u}) \leq \varepsilon$ and $d_x(a, \bar{u}) \leq \varepsilon$ or $d(a, \bar{u}) \geq J^{SR}$.*

Proof Note first that where $c(a)I^a + c_a(a)\sigma < 0$, then by Lemma 7, we are always in the first case for large enough $\bar{u}$. Consider $c(a)I^a + c_a(a)\sigma > 0$, and assume that the second case fails, so that $d(a, \bar{u}) < 3J^{SR}$, and note that since for large enough $\bar{u}$, PS binds, we have that $v^{PS}(x, a, \bar{u}) = \rho(\lambda + \mu l(x|a) - \eta l^s(x|a))$, and thus

$$\begin{aligned} v_x^{PS}(x, a, \bar{u}) &= \rho' \left((\lambda + \mu l(x|a) - \eta l^s(x|a)) \right) (\mu l_x(x|a) - \eta l_x^s(x|a)) \\ &= \frac{1}{\varphi''(v^{PS}(x, a, \bar{u}))} (\mu l_x(x|a) - \eta l_x^s(x|a)) \end{aligned}$$

and so, multiplying and dividing by $\varphi''(\bar{u} + c(a))$, we have

$$v_x^{PS}(x, a, \bar{u}) = \frac{\varphi''(\bar{u} + c(a))}{\varphi''(v^{PS}(x, a, \bar{u}))} \left(\frac{\mu}{\varphi''(\bar{u} + c(a))} l_x(x|a) - \frac{\eta}{\varphi''(\bar{u} + c(a))} l_x^s(x|a) \right).$$

But, since $d(a, \bar{u}) < J^{SR}$, it follows that $W(v(\cdot, a, \bar{u})) < J^{SR} + 2J$ and since by *IR* at some point $v(x, a, \bar{u}) = \bar{u} + c(a)$, we have as in the proof of Lemma 6 applied to $J = J^{SR} + 2J$ that

$$\frac{\varphi''(\bar{u} + c(a))}{\varphi''(v^{PS}(x, a, \bar{u}))} \to 1,$$

and by Lemma 6

$$\begin{aligned} \frac{\mu}{\varphi''(\bar{u} + c(a))} &\to \frac{(I^s - 1)c_a(a) + \sigma c(a)}{(I^s - 1)I^a - \sigma^2}, \text{ and} \\ \frac{\eta}{\varphi''(\bar{u} + c(a))} &\to \frac{c(a)I^a + c_a(a)\sigma}{(I^s - 1)I^a - \sigma^2}, \end{aligned}$$

and so

$$v_x^{PS}(x, a, \bar{u}) \to \frac{(I^s - 1)c_a(a) + \sigma c(a)}{(I^s - 1)I^a - \sigma^2} l_x(x|a) - \frac{c(a)I^a + c_a(a)\sigma}{(I^s - 1)I^a - \sigma^2} l_x^s(x|a) = v_x^{SR}(x, a, \bar{u})$$

uniformly in x. But then, since each of $v^{PS}(\cdot, a, \bar{u})$ and $v^{SR}(\cdot, a, \bar{u})$ satisfy *IR*, it follows that for $\bar{u}$ sufficiently large, $d(a, \bar{u}) < \varepsilon$ and $d_x(a, \bar{u}) < \varepsilon$, as claimed. □

Proof of Theorem 4 Choose $\bar{u}^*$ such that for $\bar{u} > \bar{u}^*$ the conclusion of Lemma 8 holds and thus, for each a, either $d(a,\bar{u}) \le \varepsilon$, or $d(a,\bar{u}) \ge J^{SR}$.

Now, note that to implement effort 0, a contract that is flat at $\bar{u}$ is optimal, and so $d(0,\bar{u}) = 0$. But, $d(\cdot,\bar{u})$ is continuous, and so, since $d(0,\bar{u}) \le \varepsilon$ and since $d(a,\bar{u})$ can never lie in $\left(\varepsilon, J^{SR}\right)$ it follows that $d(a,\bar{u}) \le \varepsilon$ everywhere, and we are done. □

A.5 Proofs for Section 8

FOA AND SQUARE ROOT UTILITY. Let $\tilde{x}$ be the smallest point at which the peak of $F_{aa}(\cdot|a')$ occurs as a' varies.

Lemma 9 (*FOA*: Square-Root Utility) *Assume CDFC*. Let $\mu^{SR}(a) > 0$. Then, $\int v^{SR}(x,a,\bar{u})f(x|\cdot)$ is quasi-concave with peak at a and is strictly concave on a neighborhood of a whenever $\int l^s(x|a)f_{aa}(x|a') \ge 0$. This holds under any of the following conditions:*

(i) f^s has support contained in $[0,\tilde{x}]$,

(ii) $l^s(0|a) \ge l^s(\tilde{x}|a)$,

(iii) for each a', $\int_0^{\tilde{x}} l^s(x|a)\frac{f_{aa}(x|a')}{\int_0^{\tilde{x}} f_{aa}(s|a')ds}dx \ge l^s(\tilde{x}|a)$, or

(iv) for each a', if $F_{aa}(x|a') = F_{aa}(z|a')$ with $z > x$, then $f^s(x) \ge f^s(z)$.

Recalling that l^s is single peaked, conditions (i)–(iii) are successively more general, where (iii) says that l^s at $\tilde{x}$ is no bigger than a particular weighted average of l^s on $[0,\tilde{x}]$. Condition (iv) is weaker than (i) but otherwise unranked. Each condition captures a sense in which f^s is larger before the peak in F_{aa} than after.

Proof of Lemma 9 Let us first prove sufficiency of (iii). Fix a and a'. Since $F_{aa}(\cdot|a') > 0$ for all interior x, it follows that $\int l(x|a)f_{aa}(x|a')dx < 0$. Thus, since $c_{aa} \ge 0$ it suffices to show that $\int l^s(x|a)f_{aa}(x|a')dx \ge 0$. But,

$$\begin{aligned}\int l^s(x|a)f_{aa}(x|a')dx &= \int_0^{\tilde{x}} l^s(x|a)f_{aa}(x|a')dx + \int_{\tilde{x}} l^s(x|a)f_{aa}(x|a')dx \\ &\ge \int_0^{\tilde{x}} l^s(\tilde{x}|a)f_{aa}(x|a')dx + \int_{\tilde{x}} l^s(x|a)f_{aa}(x|a')dx \\ &= \int \chi(x)f_{aa}(x|a')dx,\end{aligned}$$

where $\chi(x)$ equals $l^s(\tilde{x}|a)$ on $[0,\tilde{x}]$ and $l^s(x|a)$ on $[\tilde{x},1]$. But then, since $l^s(\cdot|a)$ is quasi-concave and $\tilde{x}$ is beyond the peak of $l^s(\cdot|a)$, χ is decreasing. And since $\int f_{aa}(x|a')dx = 0$ and f_{aa} is first positive and then negative, Beesack's inequality (Beesack (1957)) yields $\int \chi(x)f_{aa}(x|a')dx \ge 0$.

Clearly (i) implies (iii). To see that (ii) implies (iii) note that since l^s is quasi-concave $l^s(0|a) \ge l^s(\tilde{x}|a)$, if follows that $l^s(x|a) \ge l^s(\tilde{x}|a)$ for all $x \in [0,\tilde{x}]$ and (iii) follows.

Finally, let us turn to (iv). Note that $\omega_x(x,a') = \frac{f_{aa}(x|a')}{f_{aa}(\omega(x,a)|a)}$, and consider

$$W(x|a') \equiv \int_0^x f^s(s) f_{aa}(s|a') + \int_{\omega(x|a)}^{\bar{x}} f^s(s) f_{aa}(s|a') ds.$$

We have $W(0|a') = 0$, and for each $x \in [0, \hat{x}(a')]$, that

$$\begin{aligned} W_x(x|a') &= f^s(x) f_{aa}(x|a') - \omega_x(x|a) f^s(\omega(x|a)) f_{aa}(\omega(x|a)|a') \\ &= f^s(x) f_{aa}(x|a') - \frac{f_{aa}(x|a')}{f_{aa}(\omega(x,a)|a)} f^s(\omega(x|a)) f_{aa}(\omega(x|a)|a') \\ &= \left(f^s(x) - f^s(\omega(x|a))\right) f_{aa}(x|a') \\ &\geq 0, \end{aligned}$$

using that $x < \hat{x}(a)$, and hence $f_{aa}(x|a) > 0$. Thus, $\int f^s(s) f_{aa}(s|a') = W(\hat{x}(a')|a') \geq 0$. $\square$

An Alternative Approach to FOA Concavity of $-\int l^s(x|a) f(x|\cdot)$ is far from necessary. For example, since c is convex, it is enough that

$$\mu^{PS} \int l f_{aa}(x|a') - \eta^{PS} \int l^s f_{aa}(x|a') \leq 0,$$

which can be rewritten as

$$\int l^s f_{aa}(x|a') \geq \theta \int l f_{aa}(x|a'),$$

where

$$\theta \equiv \frac{\mu^{PS}}{\eta^{PS}} = \frac{(I^s - 1)\frac{c_a}{c} + \sigma}{I^a + \frac{c_a}{c}\sigma}.$$

Note that θ is increasing in $\frac{c_a}{c}$, and that if $\sigma \leq 0$, then θ diverges as $\frac{c_a}{c} \to \frac{I^a}{-\sigma}$. But then, under any conditions such that $\int l f_{aa}(x|a') < 0$, we will have the needed concavity as long as $\frac{c_a}{c}$ is large enough. One needs to exercise some care here in constructing examples, since if $\frac{c_a}{c}$ is too large, then $I^a + \frac{c_a}{c}\sigma < 0$, at which point η is zero.

This approach can also be used to provide conditions under which the solution to $\mathcal{P}^{PS}$ is increasing, allowing the use of standard conditions for the validity of *FOA*. Since it is natural for contracts in our setting to violate monotonicity, we do not pursue this further.

Proof of Theorem 5 Since $\int v^{SR}(x,a) f_{aa}(x|\hat{a}) - c_{aa}(\hat{a}) < 0$ at $\hat{a} = a$, and is continuous in $\hat{a}$, there is a neighborhood $(a-\delta, a+\delta)$ and $\tau > 0$ such that $\int v^{SR}(x,a) f_{aa}(x|\hat{a}) - c_{aa}(\hat{a}) < -\tau < 0$ on the neighborhood. Thus, in particular, for any $a' \in [a, a+\delta]$, since

$$\int v^{SR}(x,a) f_a(x|a) - c_a(a) = 0$$

and since $\int v^{SR}(x,a) f_{aa}(x|\hat{a}) - c_{aa}(\hat{a}) < -\tau$, it follows that

$$\int v^{SR}(x,a) f_a(x|a') - c_a(a') < -(a'-a)\tau$$

and so

$$\int v^{SR}(x,a) f(x|a) - c(a) - \left(\int v^{SR}(x,a) f(x|a+\delta) - c(a+\delta)\right) > \int_a^{a+\delta} (a'-a)\tau da' = \tau\frac{\delta^2}{2}$$

and thus, since $\int v^{SR}(x,a) f_a(x|a') - c_a(a') < 0$ for $a' > a$, a fortiori,

$$\int v^{SR}(x,a) f(x|a) - c(a) - \left(\int v^{SR}(x,a) f(x|\hat{a}) - c(\hat{a})\right) > \tau\frac{\delta^2}{2}$$

for all $\hat{a} \geq a+\delta$, and similarly for $\hat{a} \leq a-\delta$.

But then, since $\left|v^{PS}(x,a',\bar{u}) - v^{SR}(x,a',\bar{u})\right| \to 0$ uniformly in x and a', it follows that

$$\int v^{PS}(x,a',\bar{u}) f(x|a') - c(a') \to \int v^{SR}(x,a',\bar{u}) f(x|a') - c(a')$$

uniformly in a' as $\bar{u}$ grows, and so for $\bar{u}$ large enough, any action outside of $(a-\delta, a+\delta)$ is dominated by a facing $v^{PS}(\cdot,a,\bar{u})$. And, for $\bar{u}$ large enough, $\int v^{PS}(x,a,\bar{u}) f_{aa}(x|\hat{a}) - c_{aa}(\hat{a}) < -\frac{\tau}{2} < 0$ on $(a-\delta, a+\delta)$ and so, since $\int v^{PS}(x,a,\bar{u}) f_a(x|a) - c_a(a) = 0$, by construction, it follows that the unique best response to $v^{PS}(\cdot,a,\bar{u})$ is a, and we are done. □

B Appendix B: Existence and Continuity

Our results hinge on $\mathcal{P}^{PS}$ having a solution, and hence on the relevant multipliers existing, and on those multipliers being continuous. This cannot be true with full generality, because there are well-known counterexamples to existence already in the pure moral-hazard problem. But, when we restrict attention to utility functions satisfying Assumption 1, then existence indeed follows for a sufficiently large outside option.

We will prove existence of a solution to $\mathcal{P}^{PS}$ with continuous multipliers. The proof for $\mathcal{P}^{MH}$ is a simplified version of the same proof. Consider the problem $\hat{\mathcal{P}}^{PS}(a,\bar{u})$ which is $\mathcal{P}^{PS}$ augmented by a bounded payment constraint that $v(x) \in [0, 2\bar{u}]$ for all x. Throughout this section, we will impose Assumption 1.

While the space of functions v is ill-behaved, the space of distributions on rewards cross signals is not. So, let us first move to mechanisms that allow for a randomized reward following any given signal. A mechanism is thus defined by a transition probability, that is, a measurable function $\kappa : [0,1] \to \Delta[0,\infty)$, with the interpretation that following signal $x \in [0,1]$, the agent receives rewards according to $\kappa(\cdot|x)$. A special case is that $\kappa(\cdot|x)$ is Dirac at some particular value, a case which will turn out to be central to us.

Following a small twist to an idea of Kadan, Reny, and Swinkels (2017), for given κ, let π be the measure on $\Delta([0,\infty)\times[0,1])$ that arises if one first takes x uniform $[0,1]$, and then draws r according to $\kappa(\cdot|x)$. Let $\mathcal{M}$ be the set of probability measures on $\Delta([0,\infty)\times[0,1])$ with marginal onto signals equal to the uniform distribution. Note also that by Corollary 7.27.2 in Bertsekas and Shreve (1978), every measure $\pi \in \mathcal{M}$ is associated with a transition probability that is defined uniquely up to sets of x of Lebesgue measure zero.

We will thus move our search for an optimal mechanism to the space $\mathcal{M}$. To do so, note that, letting g be the density that is 1 on $[0,1]$, the utility of the agent facing κ of action a is

$$\int\left(\int r d\kappa(r|x)\right) f(x|a)dx = \int\int r\frac{f(x|a)}{g(x)}d\kappa(r|x)g(x)dx = \int rf(x|a)d\pi(x,r),$$

and so we can can rewrite all of the constraints in terms of π, and similarly for incentives and the utility of the outside option. We will take the distance d^P between any two distributions as given by the Levy-Prokhorov metric. This induces the topology of weak convergence.

We will use the following construction repeatedly. Let $\omega : [0,\infty)\times[0,1] \to [0,\infty)$ be measurable, and satisfy that $\omega(r,x) - r < \tau$ for all r and x. Start from a measure π, and let $\tilde{\pi}$ be constructed by first drawing (r,x) according to π, and then replacing r by $\omega(r,x)$. Then, $d^p(\pi,\tilde{\pi}) \leq \tau$. To see this, for any Borel set $\mathcal{A}$ of $[0,\infty)\times[0,1]$, let $\mathcal{A}^\varepsilon$ be the set of all points within ε of some point in $\mathcal{A}$. Then, $\tilde{\pi}(\mathcal{A}) \leq \pi(\mathcal{A}^\tau)$ since for the final realization to be in $\mathcal{A}$, the initial realization must be within of τ of $\mathcal{A}$, and similarly, $\pi(\mathcal{A}) \leq \tilde{\pi}(\mathcal{A}^\tau)$ since any point in $\mathcal{A}$ ends up somewhere in $\mathcal{A}^\tau$.

Lemma 10 (Distributional Mechanism) *Fix $\bar{u}^* > 2J^{SR}$. Then, for all $(a,\bar{u}) \in [0,\bar{a}]\times[\bar{u}^*,\infty)$, an optimal distributional mechanism $\hat{\pi}(\cdot,a,\bar{u})$ exists, is unique, and is continuous in (a,u).*

Proof We will apply Berge's theorem. Let

$$\Theta(a,\bar{u}) = \left\{ \pi\in\mathcal{M} \;\middle|\; \begin{array}{l} \int rf(x|a)d\pi(x,r) = \bar{u}+c(a) \\ \int rf_a(x|a)d\pi(x,r) = c_a(a) \\ \int rf^s(x|a)d\pi(x,r) \leq \bar{u} \\ \pi\left([0,2\bar{u}]\times[0,1]\right) = 1 \end{array} \right\}.$$

That is, $\pi \in \Theta(a,\bar{u})$ satisfies *IR* , *IC* , and *PS* , it never gives utility less than 0 or more than $2\bar{u}$, and it has the right marginal on signals. Let $\pi^{SR}(\cdot,a,\bar{u})$ be the distribution associated with $v^{SR}(\cdot,a,\bar{u})$, and note that since $\bar{u}^* \geq 2J^{SR}$, $\pi^{SR}(\cdot,a,\bar{u}) \in \Theta(a,\bar{u})$, and so Θ is non-empty. Let $(a^k,\bar{u}^k) \to (a',\bar{u}')$, and let $\pi^k \in \Theta(a^k,\bar{u}^k)$. Then, since for k large, $\pi\left([0,4\bar{u}']\times[0,1]\right) = 1$, π^k is a sequence of measures on a compact space, and so there is a subsequence along which π^k

converges to some limit π'. But, all the integrals defining Θ are of bounded continuous functions on $[0, 4\bar{u}') \times [0, 1]$, and so since π^k converges to π' in the weak topology, it follows that $\pi' \in \Theta(a', \bar{u}')$. Hence, Θ is upper hemi-continuous and compact valued.

Next, let us show that Θ is lower hemicontinuous. Fix $(a', \bar{u}')$, $\pi' \in \Theta(a', \bar{u}')$, a sequence $(a^k, \bar{u}^k) \to (a', \bar{u}')$, and $\varepsilon > 0$. Let us show that for $\hat{k}$ sufficiently large and for each $k > \hat{k}$, there is $\pi^k \in \Theta(a^k, \bar{u}^k)$ such that $d^P(\pi^k, \pi') < 2\varepsilon$. This is enough, as one can then construct a subsequence along which $\pi^k \to \pi'$.

We begin by modifying π' so that it never pays near 0 or $2\bar{u}'$. Draw (r, x) according to π', then replace r by

$$(1 - \varepsilon')r + \varepsilon' v^{SR}(x, a', \bar{u}'),$$

where $\varepsilon' \in (0, \varepsilon)$ is chosen so that the resultant measure, call it π'', satisfies $d^p(\pi', \pi'') \leq \varepsilon$. Now

$$\begin{aligned} \int r f(x|a) d\pi'' &= \int \left((1 - \varepsilon')r + \varepsilon' v^{SR}(x, a', \bar{u}')\right) f(x|a) d\pi' \\ &= (1 - \varepsilon') \int r f(x|a) d\pi' + \varepsilon' \int v^{SR}(x, a', \bar{u}') f(x|a) dx, \end{aligned}$$

and similarly for $\int r f_a(x|a) d\pi''$ and $\int r f^s(x) d\pi''$. Thus, $\pi'' \in \Theta(a', \bar{u}')$. Note also that since $v^{SR}(x, a', \bar{u}') > \bar{u}^* - J^{SR} > 1$, π'' never pays less than ε', and similarly never more than $2\bar{u}' - \varepsilon'$.

Now, pick $x^\ell < x^m < x^h$ where $l^s(x^\ell|a') = l^s(x^h|a')$. Using Lemma 14, choose $\gamma > 0$ small enough that for all a within γ of a'

$$\det \underbrace{\begin{bmatrix} \int_{x^\ell-\gamma}^{x^\ell+\gamma} f(x|a)dx & \int_{x^m-\gamma}^{x^m+\gamma} f(x|a)dx & \int_{x^h-\gamma}^{x^h+\gamma} f(x|a)dx \\ \int_{x^\ell-\gamma}^{x^\ell+\gamma} f_a(x|a)dx & \int_{x^m-\gamma}^{x^m+\gamma} f_a(x|a)dx & \int_{x^h-\gamma}^{x^h+\gamma} f_a(x|a)dx \\ \int_{x^\ell-\gamma}^{x^\ell+\gamma} f^s(x)dx & \int_{x^m-\gamma}^{x^m+\gamma} f^s(x)dx & \int_{x^h-\gamma}^{x^h+\gamma} f^s(x)dx \end{bmatrix}}_{Y(a)} < 0.$$

But, to construct a distributional mechanism satisfying IR , IC , and PS at $(a, \bar{u})$, we can solve

$$Y(a) \begin{bmatrix} \psi^\ell(a, u) \\ \psi^m(a, u) \\ \psi^h(a, u) \end{bmatrix} = \begin{bmatrix} \bar{u} + c(a) - (1 - \varepsilon) \int r f(x|a) d\pi' - \varepsilon \int v^*(x, a', \bar{u}') f(x|a) dx \\ c_a(a) - (1 - \varepsilon) \int r f_a(x|a) d\pi' - \varepsilon \int v^*(x, a', \bar{u}') f_a(x|a) dx \\ \bar{u} - \bar{u}' \end{bmatrix}$$

and take $\tilde{\pi}(\cdot, a, \bar{u})$ as the measure that results when one draws (r, x) according to π'' and then modifies any (r, x) with $x \in x^d$ by adding ψ^d to r.

Now, the column on the righthand side is arbitrarily close to 0 for $(a, \bar{u})$ close to $(a', \bar{u}')$, and so the determinant of the matrix formed by replacing a column of $Y(a)$ with this column is arbitrarily small, while as $a \to a'$, $\det Y(a) \to \det Y(a') > 0$. But then, by Cramer's rule $(\psi^\ell(a, u), \psi^m(a, u), \psi^h(a, u)) \to 0$. Thus, in particular, for $(a, \bar{u})$ sufficiently close to $(a', \bar{u}')$,

$|\psi^d(a,\bar{u})| < \frac{\varepsilon'}{2}$, and so $\tilde{\pi}(\cdot,a,\bar{u})$ places no weight on payments below 0 or above $2\bar{u}$. Thus $\tilde{\pi}(\cdot,a,\bar{u}) \in \Theta(a,\bar{u})$ and $d^p(\tilde{\pi}(\cdot,a,\bar{u}),\pi'') < \varepsilon$ so that $d^p(\tilde{\pi}(\cdot,a,\bar{u}),\pi') < 2\varepsilon$, and we are done.

Since Θ is non-empty, compact valued, and continuous, and since $\int \varphi(r)f(x|a)d\pi$ is continuous in π, we can apply Berge's theorem to conclude that an optimum exists and that the set of optima is upper hemicontinuous in $(a,\bar{u})$.

Let π' be optimal for $(a',\bar{u}')$, and let κ' be a transition probability for π'. We claim that κ' is degenerate at almost all x. To see this, note that φ is strictly convex, and thus

$$\varphi\left(\int r d\kappa'(r|x)dx\right) < \int \varphi\left(r\right) d\kappa'(r|x)dx,$$

unless κ' is degenerate. Thus, taking $v'(x) = \int r d\kappa'(r|x)dx$ for each x, and noting that replacing the agent's lottery over utilities at each outcome by its expectation does not affect incentives, we have that v' is optimal for $(a',\bar{u}')$. Next, assume there is a second optimum π'' at $(a',\bar{u}')$ with corresponding $v'' \neq v'$. Then the contract that provides utilty $\frac{1}{2}v'(x) + \frac{1}{2}v''(x)$ at each x is also feasible, and by strict convexity of φ, cheaper still. Thus, the optimal solution is unique, where we can let $\hat{v}(\cdot,a,\bar{u})$ be the optimal contract, and $\hat{\pi}(\cdot,a,\bar{u})$ the associated distributional contract. Finally, since $\hat{\pi}$ is unique, it follows that the optimum correspondence, which we already know from Berge's theorem to be upper hemicontinuous, is in fact continuous. □

Our next tasks are to show that $\hat{v}$ is characterized by multipliers, and that these multipliers move continuously in $(a,\bar{u})$. We begin with the analog to Proposition 2 for the case of $\hat{\mathcal{P}}^{PS}$.

Lemma 11 (Characterization of $\hat{v}$) *Fix $\bar{u}^* > 2J^{SR}$. Then, for each $(a,\bar{u})$ with $\bar{u} \geq \bar{u}^*$, $v(\cdot)$ solves $\hat{\mathcal{P}}^{PS}$ if and only if it is feasible and there is (λ,μ,η) with $\eta \geq 0$, and $\eta\left(\bar{u} - \int v(x)f^s(x)dx\right) = 0$ such that*

$$\begin{aligned}\varphi'(v(\cdot)) &= \lambda + \mu l(\cdot|a) - \eta l^s(\cdot|a) \text{ if } \varphi'(0) < \lambda + \mu l(\cdot|a) - \eta l^s(\cdot|a) < \varphi'(2\bar{u}), \\ v(x) &= 0 \text{ if } \lambda + \mu l(\cdot|a) - \eta l^s(\cdot|a) \leq 0, \text{ and} \\ v(x) &= 2\bar{u} \text{ if } \lambda + \mu l(\cdot|a) - \eta l^s(\cdot|a) \geq \varphi'(2\bar{u}).\end{aligned} \tag{9}$$

If $\eta = 0$, then $v = v^{MH}$. If $v(x) \in (0,2\bar{u})$ for all x, then $\lambda > 0$.

Proof Sufficiency is exactly as in the proof of Proposition 2 (see Online Appendix C.2) with small additions to deal with the cases where $v(x) \in \{0,2\bar{u}\}$, where a perturbation is only feasible in one direction. The proof of the existence of multipliers follows from a variation of the necessity part of Proposition 2, where we add the condition $v(x) < 2\bar{u}$ to merit inclusion in X^- and the condition $v(x) > 0$ to merit inclusion in X^+. As in the proof of Proposition 2, $\eta \geq 0$, $\eta(\bar{u} - \int v(x)f^s(x)dx) = 0$, and if $\eta = 0$, then $\hat{v}^{PS} = \hat{v}^{MH}$. Finally, if $v(x) \in (0,2\bar{u})$ for all x, then exactly as before, $\lambda > 0$. □

Next we show that an optimal contract only pays at the boundaries with small probability.

Lemma 12 (Payments at Boundaries) *Fix $\tau \in (0, \frac{1}{2})$. Then, there is $\bar{u}^*$ such that for all $\bar{u} > \bar{u}^*$ and for all a,*

$$\int_{\{x|\hat{v}(x,a,\bar{u})\in\{0,2\bar{u}\}\}} f(x|a)dx < 2\tau.$$

Proof Choose $\bar{u}^*$ large enough such that for $\bar{u} > \bar{u}^*$,

$$\frac{\tau}{1-\tau}\bar{u} > J^{SR}.$$

Fix $\bar{u} > \bar{u}^*$, and a, and assume that $\hat{v}(\cdot, a, \bar{u})$ pays 0 with probability $\tau' \geq \tau$. Let ζ be the average utility given when it is not 0. The distribution of utilities under $\hat{v}(\cdot, a, \bar{u})$, which may not be constant when it is more than 0, is thus a mean preserving spread of the distribution which pays 0 with probability τ' and ζ with probability $1 - \tau'$.

Now, by *IR*, $(1 - \tau')\zeta = \bar{u} + c(a)$, and so

$$\zeta = \frac{\bar{u} + c(a)}{1 - \tau'} > \frac{\bar{u}}{1 - \tau'} = \bar{u} + \frac{\tau'}{1 - \tau'}\bar{u} \geq \bar{u} + \frac{\tau}{1 - \tau}\bar{u} > \bar{u} + J^{SR}.$$

But then, $\hat{v}$ gives utilities that are a mean preserving spread of those given by v^{SR}. Since φ is strictly convex, $v^{SR}(\cdot, a, \bar{u})$, which implements a, is strictly less expensive than $\hat{v}(\cdot, a, \bar{u})$, and so $\hat{v}(\cdot, a, \bar{u})$ is not optimal, a contradiction. Similarly, $\hat{v}(\cdot, a, \bar{u})$ pays $2\bar{u}$ less than τ of the time. □

With this, we can prove that the multipliers move continuously in $(a, \bar{u})$. Let

$$\tau^* = \frac{1}{2} \min_a \min\{F(\hat{x}^s(a)|a), 1 - F(\hat{x}^s(a)|a)\}.$$

Note that $\tau^* > 0$, since the functions involved are continuous, and since we have assumed that $\hat{x}^s$ is everywhere interior. For each $(a, \bar{u})$ with $\bar{u} > \bar{u}^*$, let $\hat{\lambda}(a, \bar{u})$, $\hat{\mu}(a, \bar{u})$, and $\hat{\eta}(a, \bar{u})$ be the multipliers associated with $\hat{v}(\cdot, a, \bar{u})$.

Lemma 13 (Continuity of Multipliers) *Fix $\bar{u}^* \geq 2J^{SR}$ and large enough that Lemma 12 applies for $\tau = \tau^*$. Then, $\hat{\lambda}$, $\hat{\mu}$, and $\hat{\eta}$ are continuous at all $(a, \bar{u})$ with $\bar{u} > \bar{u}^*$.*

Proof Let $(a^k, \bar{u}^k) \to (a', \bar{u}')$ where $\bar{u}' > \bar{u}^*$. Then, by Lemma 10, $\pi(\cdot, a^k, \bar{u}^k)$ converges to $\pi(\cdot, a', \bar{u}')$. To prove that $(\hat{\lambda}(a^k, \bar{u}^k), \hat{\mu}(a^k, \bar{u}^k), \hat{\eta}(a^k, \bar{u}^k))$ converges to $(\hat{\lambda}(a', \bar{u}'), \hat{\mu}(a', \bar{u}'), \hat{\eta}(a', \bar{u}'))$, note first that if either or both of $\hat{\mu}(a^k, \bar{u}^k)$ or $\hat{\eta}(a^k, \bar{u}^k)$ diverge, then $\hat{\lambda}(a^k, \bar{u}^k) + \hat{\mu}(a^k, \bar{u}^k)l(x|a^k) - \hat{\eta}(a^k, \bar{u}^k)l^s(x|a^k)$ becomes arbitrarily steep to the right of $\hat{x}^s$ if $\hat{\mu}(a^k, \bar{u}^k) \geq 0$, and arbitrarily steep to the left of $\hat{x}^s$ if $\hat{\mu}(a^k, \bar{u}^k) \leq 0$, and so for k large, $\hat{v}(\cdot, a^k, \bar{u}^k)$ is interior only on an arbitrarily short interval of one of $[0, \hat{x}^s]$ or $[\hat{x}^s, 1]$, which is inconsistent with Lemma 12. But, since $\hat{\mu}(a^k, \bar{u}^k)$ and $\hat{\eta}(a^k, \bar{u}^k)$ are bounded, *IR* implies that $\hat{\lambda}(a^k, \bar{u}^k)$ is bounded as well. Thus, along a subsequence if needed, $(\hat{\lambda}(a^k,\bar{u}),\hat{\mu}(a^k,\bar{u}),\hat{\eta}(a^k,\bar{u}))$ converges to some $(\hat{\lambda}',\hat{\mu}',\hat{\eta}')$. But then, by the sufficiency part of Lemma 11, the contract characterized by $(\hat{\lambda}',\hat{\mu}',\hat{\eta}')$ is optimal in $\hat{\mathcal{P}}^{PS}(a',\bar{u}')$. But then, optima are unique, it must be that $(\hat{\lambda}',\hat{\mu}',\hat{\eta}') = (\hat{\lambda}(a',\bar{u}'),\hat{\mu}(a',\bar{u}),\hat{\eta}(a',\bar{u}))$, and we are done. □

We are finally in a position to prove existence of a continuous solution to $\mathcal{P}^{PS}$.

Theorem 6 (Existence) *Let Assumption 1 hold. Then, there is $\bar{u}^* < \infty$ such that for all $(a,\bar{u})$ with $\bar{u} > \bar{u}^*$, solutions to both $\mathcal{P}^{PS}$ and $\mathcal{P}^{MH}$ exist. The multipliers characterizing these solutions are continuous in $(a,\bar{u})$ where $\bar{u} > \bar{u}^*$.*

Proof We will prove the existence of an optimal solution to $\mathcal{P}^{PS}$ and continuity of multipliers that characterize the solution. The proof for $\mathcal{P}^{MH}$ is similar. Recall that $|v^{SR} - \bar{u}| < J^{SR}$, and so we can thus choose $\bar{u}^*$ large enough that for all $\bar{u} > \bar{u}^*$, $v^{SR}(x,a,\bar{u}) \in [2J^{SR}, 2\bar{u} - 2J^{SR}]$ for all a and x. And, by Lemma 8 for any given $\varepsilon \in (0, \frac{J^{SR}}{2})$, there is $\bar{u}^*$ large enough such that for all $\bar{u} > \bar{u}^*$, either $d(a,\bar{u}) < \varepsilon$ or $d(a,\bar{u}) > J^{SR}$.

Let

$$\hat{d}(a,\bar{u}) \equiv \max_x \left|\hat{v}^{PS}(x,a,\bar{u}) - v^{SR}(x,a,\bar{u})\right|.$$

Consider any a where $\hat{d}(a,\bar{u}) < J^{SR}$. Then, it follows that $\hat{v}^{PS}(x,a,\bar{u}) \in (0, 2\bar{u})$ for all x, and so the multipliers associated with $\hat{v}^{PS}(x,a,\bar{u})$ also characterize an optimum of $\mathcal{P}^{PS}$ which hence exists, and so $\hat{v}^{PS}(\cdot,a,\bar{u}) = v^{PS}(\cdot,a,\bar{u})$ and thus $\hat{d}(a,\bar{u}) = d(a,\bar{u})$. Thus, by definition of $\bar{u}^*$, $\hat{d}(a,\bar{u}) < \varepsilon$. Finally, note that $\hat{d}(0,\bar{u}) = 0$, since the optimal solution in $\hat{\mathcal{P}}^{PS}(0,\bar{u})$ is to pay $\bar{u}$ at all outcomes which is what v^{SR} also specifies. But then, since $\hat{d}$ is continuous, and is never in the interval (ε, J), $\hat{d}(a,\bar{u}) < \varepsilon$ for all a. But then, for all a, $\hat{v}^{PS}(\cdot,a,\bar{u})$ solves the sufficient conditions for optimality in $\mathcal{P}^{PS}(a,\bar{u})$, and hence $v^{PS}(\cdot,a,\bar{u})$ exists and is equal to $\hat{v}^{PS}(\cdot,a,\bar{u})$, and so by Lemma 13 is defined by continuous multipliers. □

C Online Appendix

C.1 Details for Example 1

Recall that the signal technology is given by

	x_1	x_2	x_3
a_1	$\frac{3}{4}$	$\frac{1}{6}$	$\frac{1}{12}$
a_2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
a_3	0	0	1

It is clear that in both *MH*, and *PS*, a_1 and a_s can be implemented by offering $\bar{u}$ at all outcomes for a cost of $\frac{1}{2}$, while for $a_3 \leq 5$, a_3 can be implemented by offering utility 0 at x_1 and x_2 and

$\bar{u} + a_3$ at x_3 for a cost of $\frac{1}{2}(\bar{u} + a_3)^2$.[32]

Let us turn to a_2. The minimization problem the principal faces in *MH* is

$$\begin{aligned}
&\min_{v_1, v_2, v_3} \left(\frac{1}{3}\frac{v_1^2}{2} + \frac{1}{3}\frac{v_2^2}{2} + \frac{1}{3}\frac{v_3^2}{2} \right) \\
s.t.\ &\frac{1}{3}v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3 - 1 \geq \bar{u} \\
&\frac{1}{3}v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3 - 1 \geq \frac{3}{4}v_1 + \frac{1}{6}v_2 + \frac{1}{12}v_3 \\
&\frac{1}{3}v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3 - 1 \geq v_3 - a_3
\end{aligned}$$

where the first constraint is the participation constraint (IR), the second the constraint (IC_1) that the agent does not want to deviate to a_1, and the third the constraint (IC_3) that the agent does not want to deviate to a_3. Let λ, μ_1, and μ_3 be the Lagrange multipliers of these constraints. Then, the relevant first-order conditions are

$$\begin{aligned}
\frac{1}{3}v_1 - \lambda\frac{1}{3} - \mu_1\left(\frac{1}{3} - \frac{3}{4}\right) - \mu_3\left(\frac{1}{3}\right) &= 0, \\
\frac{1}{3}v_2 - \lambda\frac{1}{3} - \mu_1\left(\frac{1}{3} - \frac{1}{6}\right) - \mu_3\left(\frac{1}{3}\right) &= 0, \text{ and} \\
\frac{1}{3}v_3 - \lambda\frac{1}{3} - \mu_1\left(\frac{1}{3} - \frac{1}{12}\right) - \mu_3\left(\frac{1}{3} - 1\right) &= 0.
\end{aligned}$$

Let us look at case where IR and IC_1 bind and IC_3 is slack so that $\mu_3 = 0$, and then check when the solution to the relaxed problem in fact satisfies IC_3. We then have 5 equations in 5 unknowns, vis the three just displayed along with IR and IC_1 as equalities. The solution to this system is

$$\lambda = 2, \mu_1 = \frac{24}{19}, v_1 = \frac{8}{19}, v_2 = \frac{50}{19}, \text{ and } v_3 = \frac{56}{19}.$$

For IC_3 to be slack, we need $\bar{u} > v_3 - a_3$, or $a_3 > \frac{37}{19}$.

For *PS*, we have the additional constraint $v_2 \leq \bar{u}$ to which we adjoin the Lagrange multiplier η. Taking the first-order conditions and focusing on the case where IC_3 is slack, so $\mu_3 = 0$, we

[32] Higher values of a_3 can be implemented when $\bar{u}$ is higher.

have the 6 equations in 6 unknowns

$$\frac{1}{3}v_1 - \lambda\frac{1}{3} - \mu_1\left(\frac{1}{3} - \frac{3}{4}\right) = 0$$
$$\frac{1}{3}v_2 - \lambda\frac{1}{3} - \mu_1\left(\frac{1}{3} - \frac{1}{6}\right) + \eta = 0$$
$$\frac{1}{3}v_3 - \lambda\frac{1}{3} - \mu_1\left(\frac{1}{3} - \frac{1}{12}\right) = 0$$
$$v_2 = 1$$
$$\frac{1}{3}v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3 - 1 = 1$$
$$\frac{1}{3}v_1 + \frac{1}{3}v_2 + \frac{1}{3}v_3 - 1 = \frac{3}{4}v_1 + \frac{1}{6}v_2 + \frac{1}{12}v_3,$$

the solution to which is

$$\lambda = \frac{95}{32}, \eta = \frac{31}{32}, \mu_1 = \frac{15}{8}, v_1 = \frac{5}{8}, v_2 = 1, \text{ and } v_3 = \frac{35}{8}.$$

For IC_3 to be slack, we need $v_3 - a_3 < \bar{u}$, or $a_3 > \frac{27}{8}$.

We thus have

$$C^{MH}(a_2) = \left(\frac{1}{3}\frac{\left(v_1^{MH}\right)^2}{2} + \frac{1}{3}\frac{\left(v_2^{MH}\right)^2}{2} + \frac{1}{3}\frac{\left(v_3^{MH}\right)^2}{2}\right) = \frac{50}{19}$$

and similarly, $C^{PS}(a_2) = \frac{219}{64}$. Let B_i and B_s be the gross returns to the principal of the various actions. To generate Figure 1, we note that $a_2 \succ a_s$ under *MH* if $B_2 - C^{MH}(a_2) \geq B_s - C^{MH}(a_s)$, or $\frac{1}{3} + \frac{1}{3}x_3 - \frac{50}{19} \geq 1 - \frac{1}{2}$, from which we have $x_3 \geq \frac{319}{38} \cong 8.39$ (the pink line). Similarly, $a_2 \succ a_s$ under *PS* if $\frac{1}{3} + \frac{1}{3}x_3 - \frac{219}{64} \geq 1 - \frac{1}{2}$, or $x_3 \geq \frac{689}{64} \cong 10.77$ (the purple line). Next, $a_2 \succ a_3$ under *MH* if $B_2 - C^{MH}(a_2) \geq B_3 - C^{MH}(a_3)$ or $\frac{1}{3} + \frac{1}{3}x_3 - \frac{50}{19} \geq x_3 - \frac{1}{2}(1 + a_3)^2$, from which $a_3 \geq \sqrt{\frac{4}{3}x_3 + \frac{262}{57}} - 1$ (the red line), and $a_2 \succ a_3$ under *PS* if $B_2 - C^{PS}(a_2) \geq B_3 - C(a_3)$, or $\frac{1}{3} + \frac{1}{3}x_3 - \frac{219}{64} \geq x_3 - \frac{1}{2}(1 + a_3)^2$, from which $a_3 \geq \sqrt{\frac{4}{3}x_3 + \frac{593}{96}} - 1$ (the blue line). Finally, a^3 is preferred to a_s if $B_3 - C(a_3) \geq B_s - C(a_s)$, or $x_3 - \frac{1}{2}(1 + a_3)^2 \geq 1 - \frac{1}{2}$, from which $a_3 \leq \sqrt{2x_3 - 1} - 1$ (the green line). Figure 1 is generated by graphing each of the most binding equation for each x_3. It can be checked that at all relevant a_3 for each of the *MH* and *PS* cases (that is, along the red and blue segments displayed in the figure), a_3 is large enough that the omitted constraint IC_3 does not bind. For example, for x_3 above 10.77, effort is always above $\frac{27}{8}$, and so the omitted constraint is satisfied. Below 10.77, the green line is below the blue line, and so the binding constraint is driven by switching from a_3 to a_s. The fact that when a_3 is this small, a_2 may be more expensive to implement than the given calculation is then irrelevant as we simply have that an already ruled out choice is even less attractive than it seemed.

C.2 Derivation of Optimality Conditions

Say that v is feasible if it satisfies IR , IC , and PS.

Proposition 2 (Optimality Condition) *Fix a and $\bar{u}$. Then, $v(\cdot)$ solves $\mathcal{P}^{PS}$ if and only if it is feasible and there is (λ, μ, η) with $\lambda \geq 0$, $\eta \geq 0$, and $\eta \left(\bar{u} - \int v(x) f^s(x) dx\right) = 0$ such that*

$$\varphi'(v(\cdot)) = \lambda + \mu l(\cdot|a) - \eta l^s(\cdot|a). \tag{10}$$

If such a v and (λ, μ, η) exists, then it is unique. If $\eta = 0$, then $v = v^{MH}$.

While this is reasonably obvious given standard convex optimization techniques, we provide a self-contained proof. The proof uses the following lemma:

Lemma 14 (Determinant) *Fix any a', and any triple $x^\ell < x^m < x^h$ where $l^s(x^\ell|a') = l^s(x^h|a')$. Then there is $\gamma > 0$ such that if we take*

$$Q(a) \equiv \begin{bmatrix} \int_{x^\ell-\gamma}^{x^\ell+\gamma} f(x|a)dx & \int_{x^m-\gamma}^{x^m+\gamma} f(x|a)dx & \int_{x^h-\gamma}^{x^h+\gamma} f(x|a)dx \\ \int_{x^\ell-\gamma}^{x^\ell+\gamma} f_a(x|a)dx & \int_{x^m-\gamma}^{x^m+\gamma} f_a(x|a)dx & \int_{x^h-\gamma}^{x^h+\gamma} f_a(x|a)dx \\ \int_{x^\ell-\gamma}^{x^\ell+\gamma} f^s(x)dx & \int_{x^m-\gamma}^{x^m+\gamma} f^s(x)dx & \int_{x^h-\gamma}^{x^h+\gamma} f^s(x)dx \end{bmatrix},$$

then $\det Q(a) < 0$ *for all* $a \in [a' - \gamma, a' + \gamma]$.

Proof We have that $\det Q(a) =_s \det \frac{Q(a)}{2\gamma}$, were we note that when γ is small, $\frac{Q(a)}{2\gamma}$ is term by term as close as is desired to

$$R = \begin{bmatrix} f(x^\ell|a') & f(x^m|a') & f(x^h|a') \\ f_a(x^\ell|a') & f_a(x^m|a') & f_a(x^h|a') \\ f^s(x^\ell) & f^s(x^m) & f^s(x^h) \end{bmatrix}.$$

But,

$$\begin{aligned}
\det R &= f(x^\ell|a')\left(f_{a'}(x^m|a')f^s(x^h|a') - f^s(x^m|a')f_{a'}(x^h|a')\right)\\
&- f(x^m|a')\left(f_{a'}(x^\ell|a')f^s(x^h|a') - f^s(x^\ell|a')f_{a'}(x^h|a')\right)\\
&+ f(x^h|a')\left(f_{a'}(x^\ell|a')f^s(x^m|a') - f^s(x^\ell|a')f_{a'}(x^m|a')\right)\\
&=_s l(x^m|a')l^s(x^h|a') - l^s(x^m|a')l(x^h|a') - \left(l(x^\ell|a')l^s(x^h|a') - l^s(x^\ell|a')l(x^h|a')\right)\\
&+ l(x^\ell|a')l^s(x^m|a') - l^s(x^\ell|a')l(x^m|a')\\
&= -l^s(x^m|a')l(x^h|a') - l(x^\ell|a')l^s(x^\ell|a') + l^s(x^\ell|a')l(x^h|a') + l(x^\ell|a')l^s(x^m|a')\\
&= -\left(l^s(x^m|a') - l^s(x^\ell|a')\right)\left(l(x^h|a') - l(x^\ell|a')\right)\\
&< 0,
\end{aligned}$$

where at the second line, we divided by $f(x^\ell|a')f(x^m|a')f(x^h|a')$, the third line uses that $l^s(x^\ell|a') = l^s(x^h|a')$, and the inequality follows since l^s is strictly single peaked, and l is strictly increasing. Thus, since the determinant is continuous in the entries of the matrix, we are done. □

Proof of Proposition 2 To see sufficiency, let $\tilde{v}$ be any other feasible contract that differs from v on a positive $f(\cdot|a)$-measure set of outcomes, and define

$$\Psi(\delta) \equiv \int \varphi((1-\delta)v(x) + \delta\tilde{v}(x))f(x|a)dx.$$

Assume that $\tilde{v}$ has costs lower than v, so that

$$\Psi(1) = \int \varphi(\tilde{v}(x))f(x|a)dx \leq \int \varphi(v(x))f(x|a)dx = \Psi(0).$$

Since u is strictly concave, φ is strictly convex, and thus since $\tilde{v}$ differs from v on a positive $f(\cdot|a)$-measure set of outcomes, Ψ is strictly convex as well. Thus, $\Psi_\delta(0) < 0$. But,

$$\begin{aligned}
\Psi_\delta(0) &= \int \varphi'(v(x))(\tilde{v}(x) - v(x))f(x|a)dx\\
&= \lambda \int (\tilde{v}(x) - v(x))f(x|a)dx + \mu \int (\tilde{v}(x) - v(x))f_a(x|a)dx - \eta \int (\tilde{v}(x) - v(x))f^s(x)dx\\
&= -\eta \int (\tilde{v}(x) - v(x))f^s(x)dx,
\end{aligned}$$

where the second equality follows since $\varphi'(v(x)) = \lambda + \mu l(x|a) - \eta l^s(x|a)$, and the third equality follows since IR and IC hold as equalities for both v and $\tilde{v}$. If $\eta = 0$, then we have $\Psi_\delta(0) = 0$, contradicting $\Psi_\delta(0) < 0$. So, assume $\eta > 0$. Then, $\int v(x)f^s(x)dx = \bar{u}$, and since $\tilde{v}$ satisfies PS , $\int \tilde{v}(x)f^s(x)dx \leq \bar{u}$, and so $\int(\tilde{v}(x) - v(x))f^s(x)dx \leq 0$. But then, $\Psi_\delta(0) \geq 0$, again contradicting that $\Psi_\delta(0) < 0$. It follows that v is the unique solution to $\mathcal{P}^{PS}$.

To see necessity, fix any a, and any triple $x^\ell < x^m < x^h$ where $l^s(x^\ell|a) = l^s(x^h|a)$. Appealing to Lemma 14, for $d \in (\ell, m, h)$ choose $\gamma > 0$ and then define $I^d = [x^d - \gamma, x^d + \gamma]$. Consider the effect of changing v by adding ψ^d on I^d for each of $d \in (\ell, m, h)$. The effect on $\int vf$, $\int vf_a$, and $\int vf^s$ can be seen to be the top, middle and bottom elements of

$$Q\left[\psi^\ell, \psi^m, \psi^h\right]^T,$$

where here $a = a'$ is constant, and so we suppress the argument of Q.

Since Q has non-zero determinant, to vary $\int vf$ at rate one while holding fixed $\int vf_a$ and $\int vf^s$, one can vary $(\psi^\ell, \psi^m, \psi^h)$ at rate

$$\left[\psi^\ell_{IR}, \psi^m_{IR}, \psi^h_{IR}\right] \equiv Q^{-1}\left[1, 0, 0\right]^T,$$

the marginal cost of which is

$$\lambda \equiv \sum_{d \in \{\ell, m, h\}} \psi^d_{IR} \int_{I^d} \varphi'(v(x)) f(x|a) dx.$$

Similarly, if we define

$$\left[\psi^\ell_{IC}, \psi^m_{IC}, \psi^h_{IC}\right] \equiv Q^{-1}\left[0, 1, 0\right]^T,$$

then one can vary $\int vf_a$ while holding $\int vf$ and $\int vf^s$ constant at cost

$$\mu \equiv \sum_{d \in \{\ell, m, h\}} \psi^d_{IC} \int_{I^d} \varphi'(v(x)) f(x|a) dx$$

and if we define

$$\left[\psi^\ell_{PS}, \psi^m_{PS}, \psi^h_{PS}\right] \equiv Q^{-1}\left[0, 0, -1\right]^T,$$

then one can reduce $\int vf^s$ while holding $\int vf$ and $\int vf_a$ constant at cost

$$\eta \equiv \sum_{d \in \{\ell, m, h\}} \psi^d_{PS} \int_{I^d} \varphi'(v(x)) f(x|a) dx.$$

Of course one can take linear combinations of these perturbations.

Let

$$X^- = \left\{x | \varphi'\left(v(x)\right) < \lambda + \mu l(x|a) - \eta l^s(x|a)\right\},$$

and assume $F(X^-|a) > 0$. Increase v at rate one on X^-, and undo the effect by the perturbations. The direct rate of change of costs is $\int_{X^-} \varphi'\left(v(x)\right) f(x|a) dx$ while the benefit of undoing the changes using our three perturbations is

$$\lambda \int_{X^-} f(x|a)dx + \mu \int_{X^-} f_a(x|a)dx - \eta \int_{X^-} f^s(x|a)dx$$

and so the net benefit to the principal of this perturbation is

$$\int_{X^-} \left(\varphi'\left(v(x)\right) - \lambda + \mu l(x|a) - \eta l^s(x|a)\right) f(x|a)dx > 0.$$

This contradicts that $v(\cdot)$ is optimal. Similarly, if we define

$$X^+ = \left\{x|\varphi'\left(v(x) > \lambda + \mu l(x|a) - \eta l^s(x|a)\right\},$$

then reducing payoffs on X^+ at rate one and undoing the effect via the perturbations is strictly profitable unless $F(X^+|a) = 0$. It follows that on an $F(\cdot|a)$-measure one set of x, (10) holds for the λ, μ, and η we derived.

Assume that $\eta > 0$. Then, if $\bar{u} - \int v(x)f^s(x)dx > 0$ one can increase $\int vf^s$ at benefit η using $-(\psi^{\ell}_{PS}, \psi^{m}_{PS}, \psi^{h}_{PS})$ to strictly benefit the principal, a contradiction, and so $\bar{u} - \int v(x)f^s(x)dx = 0$. Assume that η is 0. Then, exactly as above, one can show that v solves $\mathcal{P}^{MH}$ (see Proposition 1 of Kadan, Reny, and Swinkels (2017)) so that $v = v^{MH}$. Finally, assume that $\eta < 0$. By the analysis from above with v^{MH} playing the role of $\tilde{v}$, it follows from the optimality of v that

$$0 > \Psi_\delta(0) = -\eta \int (v^{MH}(x) - v(x))f^s(x)dx,$$

and so, since $\eta < 0$, $\int(v^{MH}(x) - v(x))f^s(x)dx < 0$. But then, since v satisfies *PS*, a fortiori v^{HM} satisfies *PS*. And, since l^s is non-monotone, it follows that v^{MH} and v differ on a positive measure set. Thus, since v^{MH} is the unique optimum in $\mathcal{P}^{MH}$, we have that

$$\int \varphi(v(x))f(x|a)dx > \int \varphi(v^{MH}(x))f(x|a)dx,$$

contradicting that v was optimal in $\mathcal{P}^{PS}$. Hence, $\eta \geq 0$.

Finally, note that if we add constant to all utilities, then *IR* is relaxed at rate 1, *IC* is unaffected, and *PS* is tightened at rate one. So, if one does this, and then undoes the effects on *IR* and *PS* using our perturbations, then the net benefit is

$$\int \varphi'(v(x))f(x|a)dx - \lambda + \eta,$$

and so, for this variation not to pay, we must have

$$\lambda = \int \varphi'(v(x)) f(x|a) dx + \eta > 0,$$

and we are done. □

C.3 Details for Footnote 17

Start from the Example 2. Let us begin by making f have strict *MLRP*. To do so, let

$$f^{\varepsilon} = f + \varepsilon \left(\frac{1}{2} - x \right),$$

and note that for each interval, $f_a^{\varepsilon} = f_a$ is constant, while f^{ε} is strictly decreasing, and so $f_a^{\varepsilon}/f^{\varepsilon}$ is strictly increasing. Now, let us make f^{ε} continuous. To do so, let $\delta < 1/16$ (half the radius of the smallest interval over which f was constant in x), and consider the function

$$\alpha(z, \theta) = \frac{1}{2} + \frac{1}{\pi} \arctan \left(\theta z + \frac{1}{\delta - z} - \frac{1}{z + \delta} \right),$$

for $z \in [-\delta, \delta]$, where it is easy to verify that for any given $\theta > 0$, α is a strictly increasing function with $\alpha(-\delta) = 0$, $\alpha(\delta) = 1$, and $\alpha(-z) + \alpha(z) = 1$. It is also easy to verify that as θ diverges, α converges to a step function which is 0 for $z < 0$ and 1 for $z > 0$. For each jump point x_J, and on the interval $(x_J - \delta, x_J + \delta)$, let

$$f^{\varepsilon,\theta} = (1 - \alpha(x - x_J)) f^{\varepsilon}(x_J - \delta) + \alpha(x - x_J) f^{\varepsilon}(x_J + \delta).$$

Let us first verify that $f^{\varepsilon,\theta}$ is a density. To see this, note that (suppressing a)

$$\begin{aligned}
\int_{x_J - \delta}^{x_J + \delta} f^{\varepsilon,\theta}(x) dx &= \int_{-\delta}^{\delta} \left((1 - \alpha(z)) f^{\varepsilon}(x_J - \delta) + \alpha(z) f^{\varepsilon}(x_J + \delta) \right) dz \\
&= f^{\varepsilon}(x_J - \delta) \int_{-\delta}^{\delta} (1 - \alpha(z)) \, dz + f^{\varepsilon}(x_J + \delta) \int_{-\delta}^{\delta} \alpha(z) dz \\
&= \left(f^{-}(x_J) - \varepsilon \left(x_J - \delta - \frac{1}{2} \right) \right) \int_{-\delta}^{\delta} (1 - \alpha(z)) \, dz + \left(f^{+}(x_J) - \varepsilon \left(x_J + \delta - \frac{1}{2} \right) \right) \int_{-\delta}^{\delta} \alpha(z) dz
\end{aligned}$$

But,

$$\begin{aligned}
\int_{-\delta}^{\delta} (1 - \alpha(z)) \, dz &= \int_{-\delta}^{0} (1 - \alpha(z)) \, dz + \int_{0}^{\delta} (1 - \alpha(z)) \, dz = \int_{-\delta}^{0} (\alpha(-z)) \, dz + \int_{0}^{\delta} (1 - \alpha(z)) \, dz \\
&= \int_{0}^{\delta} (\alpha(z)) \, dz + \int_{0}^{\delta} (1 - \alpha(z)) \, dz = \delta
\end{aligned}$$

and similarly, $\int_{-\delta}^{\delta}\alpha(z)dz=\delta$, and so the last expression is equal to

$$\delta\left(f^{-}(x_J)-\varepsilon\left(x_J-\delta-\frac{1}{2}\right)+f^{+}(x_J)-\varepsilon\left(x_J+\delta-\frac{1}{2}\right)\right)=\delta\left(f^{-}(x_J)+f^{+}(x_J)-2\varepsilon\left(x_J-\frac{1}{2}\right)\right).$$

But,

$$\begin{aligned}\int_{x_J-\delta}^{x_J+\delta}f^{\varepsilon}(x)dx&=\int_{-\delta}^{0}f^{\varepsilon}(x_J+z)dz+\int_{0}^{\delta}f^{\varepsilon}(x_J+z)dz\\&=\int_{-\delta}^{0}\left(f(x_J+z)-\varepsilon\left(x_J+z-\frac{1}{2}\right)\right)dz+\int_{0}^{\delta}\left(f(x_J+z)-\varepsilon\left(x_J+z-\frac{1}{2}\right)\right)dz\\&=\delta f^{-}(x_J)-\varepsilon\int_{-\delta}^{0}\left(x_J+z-\frac{1}{2}\right)dz+\delta f^{+}(x_J)-\varepsilon\int_{0}^{\delta}\left(x_J+z-\frac{1}{2}\right)dz\\&=\delta f^{-}(x_J)+\delta f^{+}(x_J)-2\varepsilon\delta\left(x_J-\frac{1}{2}\right)\end{aligned}$$

and so $\int_{x_J-\delta}^{x_J+\delta}f^{\varepsilon,\theta}(x)dx=\int_{x_J-\delta}^{x_J+\delta}f^{\varepsilon}(x)dx$, and $f^{\varepsilon,\theta}$ is a density.

Finally, let us check that $f^{\varepsilon,\theta}$ satisfies *MLRP*. On any given interval $(x_J-\delta,x_J+\delta)$, we have

$$\frac{f_a^{\varepsilon,\theta}}{f^{\varepsilon,\theta}}=\frac{f_a^{\varepsilon}(x_J-\delta|a)+\alpha(x,\theta)\left(f_a^{\varepsilon}(x_J+\delta|a)-f_a^{\varepsilon}(x_J-\delta|a)\right)}{f^{\varepsilon}(x_J-\delta|a)+\alpha(x,\theta)\left(f^{\varepsilon}(x_J+\delta|a)-f^{\varepsilon}(x_J-\delta|a)\right)}$$

and so since $\alpha'>0$,

$$\begin{aligned}\left(\frac{f_a^{\varepsilon,\theta}}{f}\right)_x&=_s\left(f_a^{\varepsilon}(x_J+\delta|a)-f_a^{\varepsilon}(x_J-\delta|a)\right)\left(f^{\varepsilon}(x_J-\delta|a)+\alpha(x,\theta)\left(f^{\varepsilon}(x_J+\delta|a)-f^{\varepsilon}(x_J-\delta|a)\right)\right)\\&\quad-\left(f_a^{\varepsilon}(x_J-\delta|a)+\alpha(x,\theta)\left(f_a^{\varepsilon}(x_J+\delta|a)-f_a^{\varepsilon}(x_J-\delta|a)\right)\right)\left(f^{\varepsilon}(x_J+\delta|a)-f^{\varepsilon}(x_J-\delta|a)\right)\\&=\left(f_a^{\varepsilon}(x_J+\delta|a)-f_a^{\varepsilon}(x_J-\delta|a)\right)f^{\varepsilon}(x_J-\delta|a)-f_a^{\varepsilon}(x_J-\delta|a)\left(f^{\varepsilon}(x_J+\delta|a)-f^{\varepsilon}(x_J-\delta|a)\right)\\&=f_a^{\varepsilon}(x_J+\delta|a)f^{\varepsilon}(x_J-\delta|a)-f_a^{\varepsilon}(x_J-\delta|a)f^{\varepsilon}(x_J+\delta|a)\\&=_s\frac{f_a^{\varepsilon}(x_J+\delta|a)}{f^{\varepsilon}(x_J+\delta|a)}-\frac{f_a^{\varepsilon}(x_J-\delta|a)}{f^{\varepsilon}(x_J-\delta|a)}>0\end{aligned}$$

establishing *MLRP*.

Finally, note that if we perturb f^s to be strictly concave, as for example

$$f^s(x)=\frac{8+\kappa\left(x-\frac{7}{16}\right)^2}{1+\kappa\frac{1}{6144}}$$

then

$$\frac{f^s}{f}=\frac{1}{1+\kappa\frac{1}{6144}}\frac{8+\kappa\left(x-\frac{7}{16}\right)^2}{1+\frac{a}{3}+\varepsilon\left(\frac{1}{2}-x\right)}$$

$$\left(\frac{f^s}{f}\right)_x = \frac{1}{1+\kappa\frac{1}{6144}}\frac{2\kappa\left(x-\frac{7}{16}\right)\left(1+\frac{a}{3}+\varepsilon\left(\frac{1}{2}-x\right)\right)-\left(8+\kappa\left(x-\frac{7}{16}\right)^2\right)(-\varepsilon)}{\left(1+\frac{a}{3}+\varepsilon\left(\frac{1}{2}-x\right)\right)^2}$$

$$\left(\frac{f_a}{f}\right)_x = \frac{\frac{1}{3}(-\varepsilon)}{\left(1+\frac{a}{3}+\varepsilon\left(\frac{1}{2}-x\right)\right)^2}$$

and so

$$\frac{l_x^s}{l_x^{\varepsilon,\theta}} = \frac{1}{1+\kappa\frac{1}{6144}}\frac{2\kappa\left(x-\frac{7}{16}\right)\left(1+\frac{a}{3}+\varepsilon\left(\frac{1}{2}-x\right)\right)-\left(8+\kappa\left(x-\frac{7}{16}\right)^2\right)(-\varepsilon)}{\frac{1}{3}(-\varepsilon)}$$

from which

$$\left(\frac{l_x^s}{l_x^{\varepsilon,\theta}}\right)_x =_s -(6+2a+3\varepsilon-6x\varepsilon)$$

which is negative for $\varepsilon < 2$.

C.4 Details for Footnote 20

In Footnote 20 we asserted a few comparative statics results regarding Δ. Here are the proofs of those assertions. Note first that it is immediate that Δ decreases in I^s, using that $(I^s-1)I^a-\sigma^2 > 0$. This is intuitive since if f^s and f are easier to tell apart, then PS hurts less. Let us consider how Δ changes with I^a. We have

$$\Delta = \frac{1}{2}c^2 I^a \frac{\left(1+\frac{c_a}{c}\frac{\sigma}{I^a}\right)^2}{(I^s-1)I^a-\sigma^2} = \frac{1}{2}\frac{(cI^a+c_a\sigma)^2}{(I^s-1)(I^a)^2-\sigma^2 I^a},$$

and thus

$$\begin{aligned}\Delta_{I^a} &= \frac{1}{2}\frac{2(cI^a+c_a\sigma)c\left((I^s-1)(I^a)^2-\sigma^2 I^a\right)-(cI^a+c_a\sigma)^2\left((I^s-1)2I^a-\sigma^2\right)}{\left((I^s-1)(I^a)^2-\sigma^2 I^a\right)^2}\\ &=_s 2c\left((I^s-1)(I^a)^2-\sigma^2 I^a\right)-(cI^a+c_a\sigma)\left((I^s-1)2I^a-\sigma^2\right)\\ &= -\sigma\left(\sigma(cI^a+c_a\sigma)+2c_a\left((I^s-1)I^a-\sigma^2\right)\right)\end{aligned}$$

where we know that $(I^s-1)I^a-\sigma^2$ is strictly positive from Lemma 2 and $cI^a+c_a\sigma$ is positive since PS binds. Hence, if σ is positive, then Δ_{I^a} is negative, while if σ is negative, we have conflicting forces. This provides one more example where we care about the sign of σ.

Finally,

$$\begin{aligned}\Delta_\sigma &=_s 2\left(c(a)I^a + c_a(a)\sigma\right) c_a(a) \left((I^s-1)I^a - \sigma^2\right) + 2\left(c(a)I^a + c_a(a)\sigma\right)^2 \sigma \\ &= 2a^2 I^a \left(\sigma c_a + cI^a\right)\left(c\sigma - c_a + c_a I^s\right) \\ &=_s c_a(I^s-1) + c\sigma \\ &> \frac{c_a}{I^a}\left(I^a(I^s-1) - \sigma^2\right) > 0.\end{aligned}$$

where the first inequality follows since PS binds and so $c > -\frac{\sigma c_a}{I^a}$, and the second by Lemma 2.

C.5 Conditions for Nonbinding *PS* at Large Effort

In Section 4.1, we showed that if $c(a)I^a + c_a(a)\sigma < 0$ then constraint PS is slack. We now provide two sets of sufficient conditions under which such is the case for large enough values of a. To this end, let $\underline{l}_x(a) \equiv \min_x l_x(x|a)$ and let $\bar{l}_x(a) \equiv \max_x l_x(x|a)$. We have the following result.

Lemma 15 (Non-Binding *PS* for Large Effort) *Constraint PS ceases to bind for large enough values of a if either of the following sets of conditions hold:*

(i) $a \in [0,1]$; c_a/c diverges as a approaches 1; and $\lim_{a\to 1} \sigma(1)/I^a(1) < 0$;

(ii) $a \in [0,\infty)$; either $l(\cdot|a)$ is convex and for sufficiently large a, $\mathbb{E}[x|a] > \mathbb{E}[x|a_s]$, or $l(\cdot|a)$ is concave and there is $\tilde{x} \in [0,1]$ such that for all a sufficiently large, $\hat{x}(a) \geq \tilde{x} > \mathbb{E}[x|a_s]$; there is an $\upsilon > 0$ such that $\frac{\underline{l}_x(a)}{\bar{l}_x(a)} \geq \upsilon$ for all a sufficiently large; and $a\mathbb{E}_a[x|a] \to 0$.

Proof Part (i) follows since under the premises, we have

$$\lim_{a\to 1} \frac{c_a(a)}{c(a)} \frac{\sigma(a)}{I^a(a)} = -\infty,$$

and so is less than -1 for a sufficiently close to 1, which implies that $c(a)I^a + c_a(a)\sigma < 0$. Sufficient for $\lim_{a\to 1} \sigma(1)/I^a(1) < 0$ is that $I^a(1) < \infty$, for which a bounded likelihood ratio is sufficient, and $\sigma(1) < 0$, which says that when the agent works at her maximum possible effort, the covariance between l and l^s is negative.

For some intuition for part (ii), note that c convex implies that $\frac{ac_a(a)}{c(a)} \geq 1$, and thus $c(a)I^a + c_a(a)\sigma < 0$ as long as $\frac{\sigma}{aI^a} < -1$. The proof shows that, under the stated premises, $\frac{\sigma(a)}{aI^a(a)}$ not only is eventually less than -1, but in fact diverges to negative infinity. One version of (ii) deals with the case in which l is convex, and the other with the case in which l is concave and for a large, $\hat{x}(a)$, the point at which $f_a = 0$, is above $\mathbb{E}[x|a_s]$ by a strictly positive amount. In turn, the ratio condition states that as a diverges, $\frac{\underline{l}_x(a)}{\bar{l}_x(a)}$ remains bounded away from zero. Since l has been assumed either concave or convex, this involves a comparison of $l_x(0,a)$ with $l_x(\bar{x},a)$, where $\bar{x}$ is the upper bound of the support of $f(\cdot|a)$, and where we abuse notation if $\bar{x} = \infty$. Finally, we assume that as effort diverges, $a\mathbb{E}_a[x|a] \to 0$. It is easily shown that this holds if $\mathbb{E}[x|a]$ is concave in a and bounded.[33] Of course, $\mathbb{E}[x|a]$ will be concave if $F_{aa} \geq 0$, the convexity of the distribution function condition. It can be shown that $a\mathbb{E}_a[x|a] \to 0$ also holds if $\mathbb{E}[x|a]$ is unbounded but grows more slowly than $\log a$.

To prove part (ii) formally, note that

$$\frac{\sigma(a)}{aI^a(a)} \geq \upsilon \frac{\frac{\sigma(a)}{\underline{l}_x(a)}}{\frac{aI^a(a)}{\bar{l}_x(a)}}.$$

We will show that the numerator of the right hand side is negative for sufficiently large a and bounded away from zero, while the denominator is positive and converges to zero.

Consider the numerator. Assume first that $l(\cdot|a)$ is convex. Then, from (5), for all a such that $\mathbb{E}[x|a] > \mathbb{E}[x|a_s]$, if we let $\grave{x}$ be such that $F - F^s$ is positive to the left of $\grave{x}$ and negative to the right of $\grave{x}$ we have

$$\frac{\sigma(a)}{\underline{l}_x(a)} \leq \frac{l_x(\grave{x}|a)}{\underline{l}_x(a)} \int (F(x|a) - F^s(x|a))\, dx \leq -\left(\mathbb{E}[x|a] - \mathbb{E}[x|a_s]\right).$$

The last expression is decreasing in a, and strictly negative for sufficiently large a. If instead $l(\cdot|a)$ is concave, then using (6),

$$\frac{\sigma(a)}{\underline{l}_x(a)} \leq \frac{l(\mathbb{E}[x|a_s]|a)}{\underline{l}_x(a)} = -\frac{1}{\underline{l}_x(a)} \int_{\mathbb{E}[x|a_s]}^{\hat{x}(a)} l_x(x|a)dx \leq -\left(\hat{x}(a) - \mathbb{E}[x|a_s]\right) \leq -\left(\tilde{x} - \mathbb{E}[x|a_s]\right).$$

Turning to the denominator, we have

$$\begin{aligned}
\frac{aI^a(a)}{\bar{l}_x(a)} &= \frac{a}{\bar{l}_x(a)} \int l(x|a) f_a(x|a)dx = \frac{a}{\bar{l}_x(a)} \int l_x(x|a)(-F_a(x|a))dx \\
&\leq a \int (-F_a(x|a))dx \\
&= a\mathbb{E}_a[x|a],
\end{aligned}$$

where the second inequality is by integration by parts and the inequality uses that $-F_a(x|a) \geq 0$. We are thus done since by assumption $a\mathbb{E}_a[x|a] \to 0$. □

Example 5 (Distributions for which *PS* Ceases to Bind) *In each of the following parameterized families of distributions, constraint PS ceases to bind at high levels of effort for appropriate choice of* $\mathbb{E}[x|a_s]$.

[33]To see this, note that by concavity, $0 \leq a\mathbb{E}_a[x|a] \leq 2\left(\mathbb{E}[x|a] - \mathbb{E}[x|\frac{a}{2}]\right)$, where the rightmost term goes to zero, since both $\mathbb{E}[x|a]$ and $\mathbb{E}[x|\frac{a}{2}]$ converge to the same finite limit.

(1) Let $F(x|a)$ be $1-e^{-\frac{x}{a}}$, let F^s be arbitrary, and c be sufficiently convex that $a\frac{c_a(a)}{c(a)} \geq \theta$ for some $\theta > 1$ (as for example if $c(a) = a^\theta$ for any $\theta > 1$).

(2) Fix $\delta > 0$, and let $F(x|a) = \frac{(x+\delta)^a}{(1+\delta)^a - \delta^a}$ on $[0,1]$, where $\delta > 0$ ensures that l is bounded.

(3) Let $f(x|a) = \frac{1}{a}f^L(x) + \left(1-\frac{1}{a}\right)f^H(x)$ on $[0,1]$, where f_H/f_L is increasing and concave.

(4) As in LiCalzi and Spaeter (2003), let $F(x|a) = x + \frac{x-x^2}{a+1}$ for $x \in [0,1]$ and $a \in [0,\infty)$.

(5) As in LiCalzi and Spaeter (2003), let $F(x|a) = x^k e^{a(x-1)}$ for $x \in [0,1]$ and $a \in [0,\infty)$.[34]

To see this, consider first $f(x|a) = \frac{1}{a}e^{-\frac{x}{a}}$ (as in Example 3) and let f^s be arbitrary. Then, as before $I^a = 1/a^2$, and similarly,

$$\sigma = \frac{1}{a^2}\int f^s(x)(x-a)dx = \frac{1}{a^2}\left(\mathbb{E}_{f^s}(x) - a\right).$$

Thus,

$$\lim_{a\to\infty}\frac{\sigma}{aI^a} = \lim_{a\to\infty}\frac{\mathbb{E}_{f^s}(x) - a}{a} = -1.$$

But then,

$$\lim_{a\to 1}\frac{c_a(a)}{c(a)}\frac{\sigma}{I^a} \leq \theta < -1,$$

and we are done.

Consider now

$$F(x|a) = \frac{(x+\delta)^a}{(1+\delta)^a - \delta^a}$$

on $[0,1]$. Then our conditions are satisfied. To see this, note that

$$f(x|a) = \frac{a\,(x+\delta)^{a-1}}{(1+\delta)^a - \delta^a},$$

and so

$$\log f(x|a) = \log a + (a-1)\log(x+\delta) - \log\left((1+\delta)^a - \delta^a\right).$$

Thus

$$l(x|a) = \frac{1}{a} + \log(x+\delta) - \frac{(1+\delta)^a\log(1+\delta) - \delta^a\log\delta}{(1+\delta)^a - \delta^a},$$

from which $l(\cdot|a)$ is clearly concave, and

$$l_x(x|a) = \frac{1}{x+\delta} \in \left[\frac{1}{1+\delta}, \frac{1}{\delta}\right]$$

[34] LiCalzi and Spaeter (2003) provide two classes of distributions satisfying *MLRP* and the convexity of distribution function condition (*CDFC*) of which this and the previous example are lead examples. For the first class, it is easy to find conditions under which l is convex, and so our results apply generally. Primitives for l to be concave or convex in the second class are forbidding.

and so we can set υ in Lemma 15 (ii) equal to $\frac{\delta}{1+\delta}$. It can be numerically checked that F satisfies *CDFC*. Hence, $\mathbb{E}[x|a]$ is concave in a, and so $a\mathbb{E}_a[x|a] \to 0$. Finally, $\hat{x}(a)$ is defined by

$$\log(x+\delta) = \frac{(1+\delta)^a \log(1+\delta) - \delta^a \log \delta}{(1+\delta)^a - \delta^a} - \frac{1}{a}$$

where the rhs converges to $\log(1+\delta)$, and so $\hat{x}(a)$ converges to 1. Hence, as long as $\mathbb{E}[x|a_s] < 1$, we can take $\tilde{x} \in (\mathbb{E}[x|a_s], 1)$, and satisfy the relevant condition.

Consider next

$$f(x|a) = \frac{1}{a} f^L(x) + \left(1 - \frac{1}{a}\right) f^H(x)$$

where f_H/f_L is increasing and concave, and note that

$$l(x|a) = \frac{1}{a^2} \frac{\frac{f^H(x)}{f^L(x)} - 1}{\frac{1}{a} + \left(1 - \frac{1}{a}\right) \frac{f^H(x)}{f^L(x)}},$$

from which

$$l_x(x|a) = \frac{1}{a^2} \frac{\left(\frac{f^H(x)}{f^L(x)}\right)_x}{\left(\frac{1}{a} + \left(1 - \frac{1}{a}\right) \frac{f^H(x)}{f^L(x)}\right)^2}$$

from which it is clear that l is concave, since then the top is positive and decreasing in x, while the bottom is positive and increasing in x. Note also that

$$\frac{\frac{\left(\frac{f^H(1)}{f^L(1)}\right)_x}{\left(\frac{1}{a} + \left(1 - \frac{1}{a}\right) \frac{f^H(1)}{f^L(1)}\right)^2}}{\frac{\left(\frac{f^H(0)}{f^L(0)}\right)_x}{\left(\frac{1}{a} + \left(1 - \frac{1}{a}\right) \frac{f^H(0)}{f^L(0)}\right)^2}} = \frac{\left(\frac{f^H(1)}{f^L(1)}\right)_x}{\left(\frac{f^H(0)}{f^L(0)}\right)_x} \frac{\left(\frac{1}{a} + \left(1 - \frac{1}{a}\right) \frac{f^H(0)}{f^L(0)}\right)^2}{\left(\frac{1}{a} + \left(1 - \frac{1}{a}\right) \frac{f^H(1)}{f^L(1)}\right)^2}$$

$$\to \frac{\left(\frac{f^H(1)}{f^L(1)}\right)_x}{\left(\frac{f^H(0)}{f^L(0)}\right)_x} \frac{\left(\frac{f^H(0)}{f^L(0)}\right)^2}{\left(\frac{f^H(1)}{f^L(1)}\right)^2}$$

and so we can take the constant υ in Lemma 15 (ii) to be

$$\upsilon = \frac{1}{2} \frac{\left(\frac{f^H(1)}{f^L(1)}\right)_x}{\left(\frac{f^H(0)}{f^L(0)}\right)_x} \frac{\left(\frac{f^H(0)}{f^L(0)}\right)^2}{\left(\frac{f^H(1)}{f^L(1)}\right)^2}$$

Next, note that

$$\mathbb{E}[x|a] = \frac{1}{a} \int x f^L(x) dx + \left(1 - \frac{1}{a}\right) \int x f^H(x) dx$$

which is clearly concave and bounded and so $a\mathbb{E}_a[x|a] \to 0$ as desired. Finally, note from our expression for l that $\hat{x}$ is constant, and occurs where $\frac{f^H}{f^L} = 1$, and so the existence of $\tilde{x}$ follows any time $\mathbb{E}[x|a_s]$ occurs to the left of this point.

Next, let $F(x|a) = x + \frac{x-x^2}{a+1}$, so that

$$l(x|a) = \frac{\frac{2x-1}{(a+1)^2}}{1+\frac{1-2x}{a+1}}$$

and so

$$l_x(x|a) = \frac{\partial}{\partial x}\frac{\frac{2x-1}{(a+1)^2}}{1+\frac{1-2x}{a+1}} = \frac{2}{(a-2x+2)^2}.$$

and

$$l_{xx}(x|a) = \frac{8}{(a-2x+2)^3}$$

and so l is convex. Hence we can take the constant υ in Lemma 15 (ii) equal to

$$\upsilon = \frac{1}{2}\lim_{a\to\infty}\frac{\frac{2}{(a+2)^2}}{\frac{2}{a^2}} = \frac{1}{2}.$$

Also, clearly $F_{aa} > 0$, and so $\mathbb{E}[x|a]$ is concave in a and, having finite support, is bounded. Thus $a\mathbb{E}_a[x|a] \to 0$. Finally, $\hat{x} = \frac{1}{2}$, and so $\tilde{x}$ exists as long as $\mathbb{E}[x|a_s] < \frac{1}{2}$.

Next, let $F(x|a) = x^k e^{a(x-1)}$ so that $f(x|a) = x^{k-1}e^{a(x-1)}(k+xa)$. Then,

$$\log f(x|a) = (k-1)\log x + a(x-1) + \log(k+ax)$$

and hence

$$l(x|a) = x - 1 + \frac{x}{k+ax}$$

from which

$$l_x(x|a) = 1 + \frac{k}{(k+ax)^2}$$

which is decreasing in x. Thus, we can set υ in Lemma 15 (ii) equal to

$$\upsilon = \frac{1}{2}\lim_{a\to\infty}\frac{1+\frac{k}{(k+a)^2}}{1+\frac{k}{(k)^2}} = \frac{1}{2}\frac{k}{1+k}.$$

Next,

$$F_{aa}(x|a) = \left(x^k e^{a(x-1)}\right)_{aa} = x^k e^{a(x-1)}(x-1)^2 > 0$$

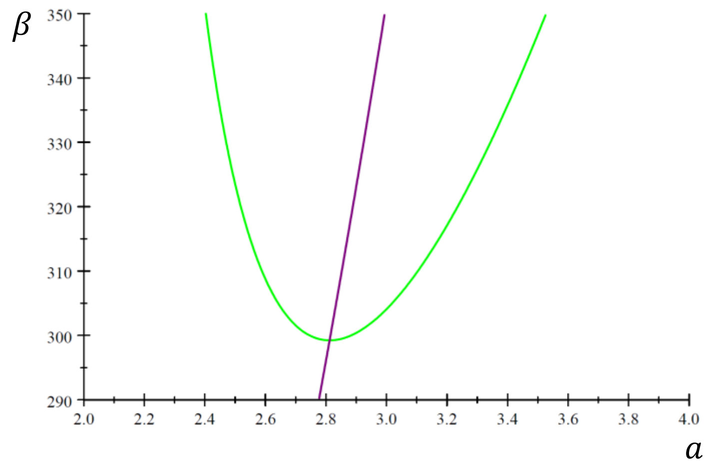

Figure 7: ***Initiative vs. Safe.*** On the green curve, the principal is indifferent between effort a and the safe project. On the purple curve, she has chosen effort optimally.

Finally, $\hat{x}(a)$ is the solution to

$$0 = l(\hat{x}(a)|a) = \hat{x}(a) - 1 + \frac{\hat{x}(a)}{k + a\hat{x}(a)},$$

from which $\lim_{a\to\infty} \hat{x}(a) = 1$, and where

$$\hat{x}_a(a) = \frac{\hat{x}(a)}{(k + a\hat{x}(a))^2 + k} > 0,$$

and so any $\mathbb{E}[x|a_s] < 1$ will do.

C.6 A Minimal Effort in the Exponential Example

We have that

$$C^{PS}(a) = \frac{1}{2}\left(\left(12 + a^2\right)^2 + 4a^4\right) + a^4\left(2a - 1\right)\frac{(a-4)^2}{a^4 e^{\frac{1}{a}} - 12a + 10a^2 - 4a^3 + 4},$$

and so for any α and β, the difference between implementing effort a and implementing a_s is

$$\beta a - C^{PS}(a) - \left(2\beta - \frac{12^2}{2}\right).$$

The green line in Figure 7 plots the set of β and a where this expression equals zero, and so the principal is indifferent between initiative and a_s. As can be seen, for β below around 300, the principal is better to implement a_s than any level of effort under initiative. The purple line shows the solution to $\frac{\partial}{\partial a}(\beta a - C^{PS}(a)) = 0$, which gives the optimal effort to implement as a function of β (a graph shows that C^{PS} is convex). Since the objective function is supermodular in β and a, optimal effort increases in β. Thus, for any β where it is worth implementing initiative, it is worth implementing at least an initiative a little above 2.8.

C.7 Omitted Proofs from Appendix A on Section 7

Proof of Lemma 5 We have that

$$\varphi'(v^{PS}(x,a,\bar{u})) = \lambda + \mu l(x|a) - \eta l^s(x|a)$$

and so, multiplying both sides by $f(x|a)$ and integrating yields

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f(x|a)dx = \lambda - \eta.$$

Similarly, multiplying both sides by $f_a(x|a)$ and integrating yields

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f_a(x|a)dx = \mu \int l(x|a)f_a(x|a)dx - \eta \int l^s(x|a)f_a(x|a)dx$$

or

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f_a(x|a)dx = \mu I^a(a) - \eta\sigma,$$

and multiplying both sides by $f^s(x)$ and integrating yields

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f^s(x)dx = \lambda + \mu \int l(x|a)f^s(x|a)dx - \eta \int l^s(x|a)f^s(x)dx$$

or

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f^s(x|a)dx = \lambda + \mu\sigma - \eta I^s.$$

But, from the system of equations

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f(x|a)dx = \lambda - \eta$$
$$\int \varphi'(v^{PS}(x,a,\bar{u}))f_a(x|a)dx = \mu I^a(a) - \eta\sigma$$
$$\int \varphi'(v^{PS}(x,a,\bar{u}))f^s(x|a)dx = \lambda + \mu\sigma - \eta I^s$$

we obtain

$$\int \varphi'(v^{PS}(x,a,\bar{u}))f^s(x|a)dx = \eta + \int \varphi'(v^{PS}(x,a,\bar{u}))f(x|a)dx + \left(\frac{\int \varphi'(v^{PS}(x,a,\bar{u}))f_a(x|a)dx}{I^a} + \frac{\eta\sigma}{I^a}\right)\sigma - \eta I^s$$

and so we arrive with a little manipulation at the claimed expressions. □

We claimed in main text that, as a by product of the large $\bar{u}$ case, we obtain the convexity of C, a difficult property to ensure from primitives. To show this we need a few steps. To begin, note that from the envelope theorem applied to $\mathcal{P}^{PS}$, we have

$$C_a^{PS}(a) = \int \varphi(v(x)) f_a(x|a)dx - \mu \left(\int v(x) f_{aa}(x|a)dx - c_{aa}(a) \right),$$

noting that the term in λ drops out using IC , and that a does not enter into PS . We begin with a key lemma about the derivatives of λ, μ, and η with respect to a.

Lemma 16 (Limit Derivatives of Multipliers) *Each of $\frac{\lambda_a}{\lambda}$, $\frac{\mu_a}{\lambda}$, and $\frac{\eta_a}{\lambda}$ converges to zero in $\bar{u}$, and does so uniformly in a.*

Proof For given a and $\bar{u}$ where PS binds, λ, μ, and η are defined implicitly by

$$\int \rho(\lambda + \mu l - \eta l^s) f = \bar{u} + c$$
$$\int \rho(\lambda + \mu l - \eta l^s) f_a = c_a$$
$$\int \rho(\lambda + \mu l - \eta l^s) f^s = \bar{u},$$

and so differentiating with respect to a yields

$$\int \rho'(\lambda_a + \mu_a l + \mu l_a - \eta_a l^s - \eta l_a^s) f + \int \rho f_a = c_a$$
$$\int \rho'(\lambda_a + \mu_a l + \mu l_a - \eta_a l^s - \eta l_a^s) f_a + \int \rho f_{aa} = c_{aa}$$
$$\int \rho'(\lambda_a + \mu_a l + \mu l_a - \eta_a l^s - \eta l_a^s) f^s = 0,$$

where we use IC to simplify the first equation, and then rearrange so that $(\lambda_a, \mu_a, \eta_a)$ solve

$$\int (\lambda_a + \mu_a l - \eta_a l^s)\rho' f = \int (\eta l_a^s - \mu l_a)\rho' f$$
$$\int (\lambda_a + \mu_a l - \eta_a l^s) l \rho' f = c_{aa} - \int \rho f_{aa} + \int (\eta l_a^s - \mu l_a) l \rho' f$$
$$\int (\lambda_a + \mu_a l - \eta_a l^s) l^s \rho' f = \int (\eta l_a^s - \mu l_a) l^s \rho' f,$$

or equivalently, dividing both sides by $\varphi' \int \rho' f$ (where we take φ' to mean $\varphi'(\bar{u} + c(a))$), and then expressing things in matrix form,

$$\underbrace{\begin{bmatrix} 1 & \int l \frac{\rho' f}{\int \rho' f} & \int l^s \frac{\rho' f}{\int \rho' f} \\ \int l \frac{\rho' f}{\int \rho' f} & \int l^2 \frac{\rho' f}{\int \rho' f} & \int l l^s \frac{\rho' f}{\int \rho' f} \\ \int l^s \frac{\rho' f}{\int \rho' f} & \int l l^s \frac{\rho' f}{\int \rho' f} & \int (l^s)^2 \frac{\rho' f}{\int \rho' f} \end{bmatrix}}_{M} \begin{bmatrix} \frac{\lambda_a}{\varphi'} \\ \frac{\mu_a}{\varphi'} \\ -\frac{\eta_a}{\varphi'} \end{bmatrix} = \begin{bmatrix} \frac{\int (\eta l_a^s - \mu l_a)\rho' f}{\varphi' \int \rho' f} \\ \frac{c_{aa} - \int \rho f_{aa} + \int (\eta l_a^s - \mu l_a) l \rho' f}{\varphi' \int \rho' f} \\ \frac{\int (\eta l_a^s - \mu l_a) l^s \rho' f}{\varphi' \int \rho' f} \end{bmatrix}.$$

Consider first the column vector on the right. Note that

$$\frac{\int (\eta l_a^s - \mu l_a)\rho' f}{\varphi' \int \rho' f} = \int \left(\frac{\eta}{\varphi'} l_a^s - \frac{\mu}{\varphi'} l_a \right) \frac{\rho' f}{\int \rho' f} \to 0,$$

using that $\frac{\eta}{\varphi'} \to 0$ and $\frac{\mu}{\varphi'} \to 0$, and that as CS show (and is intuitive since ρ' converges to a constant over the relevant range) $\frac{\rho' f}{\int \rho' f} \to f$. Similarly,

$$\frac{\int (\eta l_a^s - \mu l_a) l \rho' f}{\varphi' \int \rho' f} \to 0 \text{ and } \frac{\int (\eta l_a^s - \mu l_a) l^s \rho' f}{\varphi' \int \rho' f} \to 0.$$

Also,

$$\frac{-\int \rho f_{aa}}{\varphi' \int \rho' f} = \frac{\int \rho' F_{aa}}{\varphi' \int \rho' f} = \frac{\int \frac{F_{aa}}{f} \rho' f}{\varphi' \int \rho' f} = \frac{\int \frac{F_{aa}}{f} \xi}{\varphi'} \to 0$$

since the top converges to $\int F_{aa}$ which is finite, and the bottom diverges. Finally, since $\varphi'(\rho(\tau)) = \tau$ we have $\varphi''(\rho(\tau))\rho'(\tau) = 1$ and thus

$$\frac{c_{aa}}{\varphi'(\bar{u} + c(a)) \int \rho' f} = \frac{c_{aa}}{\varphi'(\bar{u} + c(a)) \int \frac{1}{\varphi''(\rho)} f} = \frac{c_{aa}}{\int \frac{\varphi'(\bar{u}+c(a))}{\varphi'(\rho)} \frac{\varphi'(\rho)}{\varphi''(\rho)} f} \to 0,$$

since $\frac{\varphi'(\rho)}{\varphi''(\rho)} \to \infty$, and $\frac{\varphi'(\bar{u}+c(a))}{\varphi'(\rho)} \to 1$. Thus, the right side converges to the zero vector.

But, since $\frac{\rho' f}{\int \rho' f} \to f$,

$$M \to M^{\lim} \equiv \begin{vmatrix} 1 & 0 & 1 \\ 0 & I^a & \sigma \\ 1 & \sigma & I^s \end{vmatrix}.$$

The determinant of $M^{\lim}$ is $I^a(I^s - 1) - \sigma^2$ which is strictly positive by Lemma 2. Hence $M^{\lim}$ is invertible, and the unique solution to the system

$$M^{\lim} \begin{bmatrix} \tau^1 \\ \tau^2 \\ \tau^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is $\tau^1 = \tau^2 = \tau^3 = 0$. But then, for $\bar{u}$ large, $|M|$ is also strictly positive, and hence the solution to the system of equations is continuous as $\bar{u}$ diverges. Thus $\frac{\lambda_a}{\varphi'} \to 0$, $\frac{\mu_a}{\varphi'} \to 0$, and $\frac{\eta_a}{\varphi'} \to 0$. $\square$

Proof of Lemma 6 Write $v(x)$ where we more properly mean $v^{PS}(x, a, \bar{u})$. Using Lemma 5, start from

$$\eta = \frac{-I^a \int \varphi'(v(x)) \left[(f^s(x) - f(x|a))\right] dx + \int \varphi'(v(x)) \left[\sigma l(x|a)\right] f(x|a) dx}{I^a \left(I^s - 1\right) - \sigma^2}$$

and integrate by parts and divide by $\varphi''(\bar{u} + c(a))$ to arrive at

$$\frac{\eta}{\varphi''(\bar{u} + c(a))} = \frac{-I^a \int \frac{\varphi''(v(x))}{\varphi''(\bar{u}+c(a))} v_x(x)(F(x|a) - F^s(x|a))dx + \sigma \int \frac{\varphi''(v(x))}{\varphi''(\bar{u}+c(a))} v_x(x)(-F_a(x|a))dx}{I^a \left(I^s - 1\right) - \sigma^2}.$$

But, by IR and continuity of $v(\cdot)$, we must have $v(x) = \bar{u} + c(a)$ for some x. Hence, since $a \in [0, 1]$, $v(x) \in [\bar{u} - J, \bar{u} + c(1) + J]$ for all x. But then, using CS, Lemma 1, $\frac{\varphi''(v(x))}{\varphi''(\bar{u}+c(a))} \to 1$ as $\bar{u}$ diverges, and does so uniformly in a. Thus, uniformly in a,

$$\frac{\eta}{\varphi''(\bar{u} + c(a))} \to \frac{-I^a \int v_x(x)(F(x|a) - F^s(x|a))dx + \sigma \int v_x(x)(-F_a(x|a))dx}{I^a \left(I^s - 1\right) - \sigma^2},$$

where we observe that

$$\int v_x(x)(F(x|a) - F^s(x|a))dx = \int v(x))(f^s(x|a) - f(x|a))dx = \bar{u} - (\bar{u} + c(a)) = -c(a)$$

and $\int v_x(x)(-F_a(x|a))dx = c_a(a)$, and so

$$\frac{\eta}{\varphi''(\bar{u} + c(a))} \to \frac{c(a)I^a + c_a(a)\sigma}{(I^s - 1)I^a - \sigma^2}$$

uniformly in a. As a reality check, for the square root case where $u = \sqrt{2w}$ we have $\varphi'' = 1$, and so this expression agrees with the one derived for that case.

Continuing, we then have that

$$\lim_{\bar{u}\to\infty} \frac{\lambda}{\varphi'(\bar{u} + c(a))} = \lim_{\bar{u}\to\infty} \int \frac{\varphi'(v(x))}{\varphi'(\bar{u} + c(a))} f(x|a)dx - \lim_{\bar{u}\to\infty} \left(\frac{\varphi''(\bar{u} + c(a))}{\varphi'(\bar{u} + c(a))} \frac{c(a)I^a + c_a(a)\sigma}{(I^s - 1)I^a - \sigma^2} \right)$$

and so, since $\frac{\varphi''}{\varphi'} \to 0$ uniformly in a, and $\frac{\varphi'(v(x))}{\varphi'(\bar{u}+c(a))} \to 1$ uniformly in a,

$$\frac{\lambda}{\varphi'(\bar{u} + c(a))} \to 1$$

uniformly in a. This also agrees with the square root case, since in that case, $\varphi'(\bar{u}+c(a)) = \bar{u}+c(a)$.

Finally,

$$\mu = \frac{\int \varphi'(v(x)) f_a(x|a) dx}{I^a} + \frac{\eta\sigma}{I^a} = \frac{\int \varphi''(v(x)) v_x(x) \left(-F_a(x|a)\right) dx}{I^a} + \frac{\eta\sigma}{I^a}$$

and so

$$\frac{\mu}{\varphi''(\bar{u} + c(a))} = \frac{\int \frac{\varphi''(v(x))}{\varphi''(\bar{u}+c(a))} v_x(x) \left(-F_a(x|a)\right) dx}{I^a} + \lim_{\bar{u}\to\infty} \frac{\eta}{\varphi''(\bar{u} + c(a))} \frac{\sigma}{I^a}$$

from which, since $\frac{\varphi''(v(x))}{\varphi''(\bar{u}+c(a))} \to 1$ uniformly in a, using that $\int v_x(x) \left(-F_a(x|a)\right) dx = -c_a(a)$, and using our limiting expression for $\frac{\eta}{\varphi''}$, we have that uniformly in a,

$$\frac{\mu}{\varphi''(\bar{u} + c(a))} \to \frac{c_a(a)}{I^a} + \frac{c(a)I^a + c_a(a)\sigma}{(I^s - 1)I^a - \sigma^2} \frac{\sigma}{I^a} = \frac{(I^s - 1)c_a(a) + \sigma c(a)}{(I^s - 1)I^a - \sigma^2}.$$

which once again agrees with the square root case.[35] The expressions for the multipliers for v^{MH} are proven in CS by similar techniques. □

We are now ready to prove the following result on the first and second derivatives of C as $\bar{u}$ diverges. Since c_{aa} is strictly positive, it will follow from the proposition that C is eventually convex in a for sufficiently large $\bar{u}$.

Proposition 3 (Limits of Derivatives of C) *Let Assumption 1 hold. As $\bar{u}$ diverges, then uniformly in a,*

$$\frac{C_a^{PS}(a)}{\varphi'(\bar{u} + c(a))c_a(a)} \to 1, \text{ and } \frac{C_{aa}^{PS}(a)}{\varphi'(\bar{u} + c(a))c_{aa}(a)} \to 1.$$

Proof Note that

$$\begin{aligned}\frac{C_a(a)}{\varphi'(\bar{u} + c(a))c_a(a)} &= \frac{\int \varphi(v(x)) f_a(x|a) dx - \mu \left(\int v(x) f_{aa}(x|a) dx - c_{aa}(a)\right)}{\varphi'(\bar{u} + c(a))c_a(a)} \\ &= \frac{\int \varphi(v(x)) f_a(x|a) dx}{\varphi'(\bar{u} + c(a))c_a(a)} - \mu \frac{\int v(x) f_{aa}(x|a) dx - c_{aa}(a)}{\varphi'(\bar{u} + c(a))c_a(a)}.\end{aligned}$$

[35] As a reality check, note that

$$I^s = \int l^s(x|a) f^s(x) dx = \int \left(\frac{f^s(x)}{f(x|a)}\right)^2 f(x|a) dx$$

which is convex in the term in parentheses. Hence,

$$\int \left(\frac{f^s(x)}{f(x|a)}\right)^2 f(x|a) dx \geq \left(\int \frac{f^s(x)}{f(x|a)} f(x|a) dx\right)^2 = 1.$$

It follows that μ is positive.

Now,

$$\frac{\int \varphi(v(x)) f_a(x|a)dx}{\varphi'(\bar{u}+c(a))c_a(a)} = \frac{-\int \frac{\varphi'(v(x))}{\varphi'(\bar{u}+c(a))} v_x(x) F_a(x|a)dx}{c_a(a)} \to \frac{-\int v_x(x) F_a(x|a)dx}{c_a(a)} = 1,$$

and so it is enough to show that the second fraction converges to 0. Note that

$$\begin{aligned} 0 &\geq \int v(x) f_{aa}(x|a)dx - c_{aa}(a) \\ &= -\int v_x(x) F_{aa}(x|a)dx - c_{aa}(a) \\ &\geq -\max_{a,x} |F_{aa}(x|a)| \int v_x(x)dx - \max_a c_{aa}(a) \\ &\geq -J \max_{a,x} |F_{aa}(x|a)| - \max_a c_{aa}(a) \end{aligned}$$

using Lemma 8 and so $\left|\int v(x) f_{aa}(x|a)dx - c_{aa}(a)\right|$ is uniformly bounded.

So, consider

$$\begin{aligned} \frac{\mu}{\varphi'(\bar{u}+c(a))c_a(a)} &= \frac{\mu}{\varphi''(\bar{u}+c(a))\frac{(I^s-1)c_a(a)+\sigma c(a)}{(I^s-1)I^a-\sigma^2}} \frac{\varphi''(\bar{u}+c(a))\frac{(I^s-1)c_a(a)+\sigma c(a)}{(I^s-1)I^a-\sigma^2}}{\varphi'(\bar{u}+c(a))c_a(a)} \\ &= \frac{\varphi''(\bar{u}+c(a))}{\varphi'(\bar{u}+c(a))} \frac{\mu}{\varphi''(\bar{u}+c(a))\frac{(I^s-1)c_a(a)+\sigma c(a)}{(I^s-1)I^a-\sigma^2}} \frac{I^s-1+\sigma\frac{c(a)}{c_a(a)}}{(I^s-1)I^a-\sigma^2}. \end{aligned}$$

The first fraction converges to 0 by Assumption 1, while the second converges uniformly to 1 using Lemma 6, and so it is enough that the third fraction has bounded absolute value. But, the denominator of the third fraction is bounded away from zero, since $(I^s-1)I^a-\sigma^2$ is strictly positive everywhere and continuous, I^s is bounded by assumption, and $\frac{c(a)}{c_a(a)} \leq 1$ since c is convex, and we have established the claimed form of C_a^{PS}.

To analyze C_{aa}^{PS}, note from our expression for C_a^{PS}, that it follows that

$$\begin{aligned} C_{aa}^{PS}(a) = &\int \varphi'(v) v_a f_a + \int \varphi(v) f_{aa} - \mu \left(\int v(x) f_{aaa} - c_{aaa}\right) \\ &- \mu \int v_a f_{aa} - \mu_a \left(\int v(x) f_{aa}(x|a)dx - c_{aa}\right), \end{aligned}$$

and we shall be interested in the limiting behavior of

$$\frac{C_{aa}^{PS}}{\varphi'(\bar{u}+c)c_{aa}(a)}.$$

Note first that the bracketed term in the fifth term is finite as argued above, and similarly for the bracketed term in the third term. But then, since,

$$\frac{\mu}{\varphi'(\bar{u}+c)} \to 0, \text{ and } \frac{\mu_a}{\varphi'(\bar{u}+c)} \to 0,$$

we can dispense with the third and fifth terms without loss. Integrate the second term by parts, and make the substitution

$$\varphi'(v) = \lambda + \mu l - \eta l^s$$

to arrive at

$$\begin{aligned} C_{aa}^{PS} &\cong \lambda \int v_a f_a + \mu \int v_a l f_a - \eta \int l^s v_a f_a \\ &+ \lambda \int v_x \left(-F_{aa}\right) + \mu \int v_x l \left(-F_{aa}\right) - \eta \int v_x l^s \left(-F_{aa}\right) - \mu \int v_a f_{aa}. \end{aligned}$$

The term $\mu \int v_x l \left(-F_{aa}\right) \leq \mu J \max_{a,x} |l F_{aa}|$, and so disappears on division by $\varphi'(\bar{u})$, and similarly for $\eta \int v_x l^s \left(-F_{aa}\right)$. But, $\int v f_a = c_a$ is an identity, and so, differentiating,

$$\int v_a f_a = c_{aa} - \int v f_{aa} = c_{aa} + \int v_x F_{aa}.$$

Making this substitution and cancelling the two terms involving $\int v_x F_{aa}$,

$$C_{aa}^{PS} \cong \lambda c_{aa} + \mu \int v_a l f_a - \eta \int l^s v_a f_a - \mu \int v_a f_{aa}.$$

Note next that

$$l_a = \left(\frac{f_a}{f}\right)_a = \frac{f_{aa} f - f_a^2}{f^2}$$

and so

$$f l_a = f_{aa} - \frac{f_a^2}{f} = f_{aa} - l f_a.$$

Substituting this in the second term and then cancelling with the last term,

$$C_{aa}^{PS} \cong \lambda c_{aa} - \mu \int v_a f l_a - \eta \int l^s v_a f_a.$$

Since for large $\bar{u}$ the multiplier λ behaves like $\varphi'(\bar{u}+c)$, we would be done if we can show that

$$\frac{\mu}{\lambda} \int v_a f l_a - \frac{\eta}{\lambda} \int l^s v_a f_a \to 0,$$

for which it is enough that $\frac{\mu}{\lambda} \int v_a f l_a$ and $\frac{\eta}{\lambda} \int l^s v_a f_a$ each go to zero. Consider the first. Expanding

v_a and then multiplying and dividing by $\int \rho' f$ gives that

$$\frac{\mu}{\lambda}\int v_a f l_a = \mu \int \rho' f \int \left(\frac{\lambda_a}{\lambda} + \frac{\mu_a}{\lambda} l - \frac{\eta_a}{\lambda} l^s + \frac{\mu}{\lambda} l_a - \frac{\eta}{\lambda} l_a^s\right) l_a \frac{\rho' f}{\int \rho' f}.$$

But, since $\frac{\lambda_a}{\lambda}$ and its ilk all converge to 0, and since $\frac{\rho' f}{\int \rho' f}$ converges to f, the second integral converges to 0, and so it is enough to show that $\mu \int \rho' f$, or equivalently,

$$\frac{\mu}{\varphi''(\bar{u}+c(a))}\varphi''(\bar{u}+c(a))\int \rho' f$$

is bounded. But we know from Lemma 6 that

$$\left|\frac{\mu}{\varphi''(\bar{u}+c(a))} - \frac{(I^s-1)c_a(a)+\sigma c(a)}{(I^s-1)I^a-\sigma^2}\right| \to 0$$

where the righthand ratio within the absolute value sign is independendent of $\bar{u}$. Hence it is enough to know that $\varphi''(\bar{u}+c(a))\int \rho' f$ is bounded. But,

$$\int \rho' f = \int \frac{1}{\varphi''(v)} f$$

and so we desire to show that

$$\int \frac{\varphi''(\bar{u}+c(a))}{\varphi''(v)} f$$

is bounded. But, since $\max_x v^{SR}(x,a,\bar{u}) - \min_x v^{SR}(x,a,\bar{u})$ is finite and independent of $\bar{u}$ and $d(a,\bar{u}) \to 0$, it follows from Lemma 1 in *CS* that that $\frac{\varphi''(\bar{u}+c(a))}{\varphi''(v)} \to 1$ uniformly in x. □

C.8 Examples with *CDFC**

The following examples, which satisfy *CDFC*, also satisfy the condition of F_{aa} being single peaked and strictly positive at interior outputs, and hence satisfy *CDFC**.

Example 6 *Let $F(x|a) = x + \frac{x-x^2}{a+1}$ for $x \in [0,1]$ and $a \geq 0$. Then, F_{aa} is single-peaked with peak at $x = 1/2$. Let $F(x|a) = x^k e^{a(x-1)}$ for $x \in [0,1]$ and $a \in [0,\infty)$. Then, F_{aa} is single-peaked, with peak at or above $\frac{k}{k+2}$. Finally, let $F(x|a) = x^{a+\beta}$ for $x \in [0,1]$ and $a \geq 0$, where $\beta > 0$. Then, F_{aa} is single-peaked with peak at $e^{-\frac{2}{a+\beta}} \geq e^{-\frac{2}{\beta}}$.*

The case $F(x|a) = x + \frac{x-x^2}{a+1}$ is straightforward. If $F = x^k e^{a(x-1)}$, then $F_{aa} = x^k (x-1)^2 e^{a(x-1)}$. We want to show that this is strictly single-peaked. Since F_{aa} is zero at $x = 0$ and 1, it follows that F_{aa} has an interior critical point. It is enough to show that any such interior critical point is a strict local maximum. But,

$$f_{aa} = x^{k-1}e^{a(x-1)}\,(1-x)\left(-2x + k\,(1-x) - ax^2 + ax\right) =_s k + ax - \frac{2x}{1-x} \equiv j(x,a,k),$$

and so, since $k \geq 0$, where $f_{aa} = 0$, we have $2 \geq a\,(1-x)$. But then, where $f_{aa} = 0$,

$$\begin{aligned} f_{aax} &= ((k+ax)\,(1-x) - 2x)_x = a - k - 2ax - 2 \\ &\leq a - k - 2ax - a\,(1-x) = -k - ax < 0. \end{aligned}$$

Note also that $j(\cdot, a, k)$ is strictly concave, with $j(0,a,k) = k > 0$, and with j tending to $-\infty$ as x tends to one. Hence, $j(\cdot, a, k)$ crosses zero once and is strictly decreasing where it does so. But then, when a is increased, the crossing point moves to the right. Hence the solution x to $j(\cdot, a, k) = 0$ is smallest when a is zero and thus $x = \frac{k}{k+2}$.

Finally, let $F(x|a) = x^{a+\beta}$. Then,

$$f_{aa}(x|a) = x^{a+\beta-1}\,(\ln x)\,(\beta \ln x + a \ln x + 2) =_s -\,(\beta \ln x + a \ln x + 2)\,,$$

where the last object has derivative $-\frac{1}{x}\,(a+\beta) < 0$. Hence, F_{aa} is single-peaked, with peak at $e^{-\frac{2}{a+\beta}} \geq e^{-\frac{2}{\beta}}$.