Endogenous Criteria for Success*

René Kirkegaard†

August 2023

Abstract

Economic agents are motivated to undertake costly actions by the prospect of being rewarded for successes and punished for failures. But what determines what a success looks like? This paper endogenizes the criteria for success in an otherwise standard principal-agent model with risk neutrality and limited liability. The set of feasible contracts is constrained by incentive constraints and possibly by a budget constraint. The first-order approach is not required to solve the problem. If the principal manipulates the criteria for success only to lower implementation costs, and depending on which type of constraint is more restrictive, the second-best action may be above or below the first-best action. In a class of problems where the principal's payoff depends directly on the criteria for success, the second-best solution features either more stringent criteria for success or a lower action (or both) than the first-best solution.

JEL Classification Numbers: D82, D86

Keywords: Moral Hazard, Principal-Agent Models.

---

*I thank SSHRC for funding this research.

†Department of Economics and Finance, University of Guelph. Email: rkirkega@uoguelph.ca.

1 Introduction

The following new principal-agent model is proposed and studied. The agent's "performance" is a continuous variable whose distribution is determined by the agent's action. However, the principal is not able to perfectly observe performance. For example, consider a salesman (agent) who has been instructed to sell a product at some price p. The agent's efforts at persuading the customer will make the latter revise his willingness-to-pay for the product. In other words, the agent's "performance" is described by the customer's resulting willingness-to-pay. However, the customer's willingness-to-pay is inside his head and cannot be observed by outsiders. It can only be observed whether he decides to purchase the product or not. In other words, the agent's employer (principal) knows only whether the willingness-to-pay is above or below p. Thus, even though "background performance" is continuous, the observable signal on which remuneration is based is binary. Moreover, the criterion for success is endogenously determined by p, which is after all dictated by the principal. For another example, a firm who is about to market a product newly developed by one of its engineers must decide upon the stringency of product testing prior to launch.

A similar situation may occur even when performance can be observed but the reward structure is restricted. This is often the case when a pass/fail test is taken. The examiner may be able to obtain a fine measure of the examinee's performance, yet much of this information is lost in the coarse marking scheme. The criterion for success – the pass mark or the difficulty of the test – is also often endogenous. For instance, medical boards determine through testing whether the medical school graduate does or does not meet the bar to be awarded a medical license. It is irrelevant if the candidate passes by a wide or a narrow margin. In other words, the medical board decides if the candidate is admitted to the "club", but it is not practical to make his compensation contingent on his precise score on the test.

Performance is one-dimensional in the model and the criterion for success is therefore essentially a performance threshold. If this is exogenous, the model reduces to a standard two-outcome model. However, when it is endogenous, it is entirely possible that the optimal threshold depends on the action that the principal wishes to induce. The standard literature typically does not worry about where the probability of success comes from. One way of thinking about the current model is that a "microfoundation" of sorts is provided, linking the endogenous probability of success to an underlying performance technology. The model thus allows us to ask whether the criterion for success is more or less demanding as varying levels of effort is induced.

Now return to the salesman example. Within this principal-agent relationship, the "first-best" in the absence of moral hazard involves some effort level and a price that is set to maximize monopoly profits (this ignores the effects on parties outside the contractual relationship, such as the consumer). Thus, under asymmetric information, two sources of distortions from the first-best are possible; the employer may decide to distort the action, the price, or both. Thus, the model is richer than the standard two-outcome model, which misses the dependence between one of the principal's choice variables (the price) and the probability of success.

The criterion for success serves a dual purpose. It manipulates implementation costs, and it may also, as in the salesman example, be intrinsically important to the principal. It accomplishes the former by changing the quality of information about the agent's effort. In this sense, the monitoring technology is endogenized. As in Li and Yang (2020), discussed in more detail below, monitoring is tied to, and disciplined by, an underlying performance technology that is outside the principal's control. In contrast, in the more traditional literature, the principal can choose from an ad hoc set of monitoring technologies. Examples include Dye (1986) and Kim (1995).

The agent is assumed to be risk-neutral and protected by limited liability. This makes it possible to characterize implementation costs in a succinct and tractable way. Three versions of the problem are then analyzed. The first two versions assume that the principal does not directly care about the criterion for success, but uses it only to manipulate the cost of incentivizing the agent. Thus, this is a pure monitoring problem. These two versions of the model differ in the nature of the constraint that limits the feasible set of contracts.

The first and shorter version is inspired by the dominant approaches in the existing literature. The first-order approach (FOA) is assumed to be valid but the principal faces a budget constraint.[1] Thresholds that are very large require substantial bonuses to be incentive compatible. The budget constraint rules out such thresholds. However, smaller actions can feasibly be induced with larger thresholds.

The second version represents the heart of the paper and departs from the tradi-

[1]Bounds on payments are analyzed in e.g. Innes (1990), Jewitt, Kadan, and Swinkels (2008), and Poblete and Spulber (2012). However, these papers assume that performance is perfectly observable. A more closely related paper by Bond and Gomes (2009) is discussed below.

tional approach by not relying on the FOA. Instead, the intention is to focus squarely and exhaustively on the "implementability constraint". While the FOA simplifies the incentive compatibility problem, it requires strong assumptions that are not always desirable. Without the FOA, there are combinations of actions and thresholds that simply cannot be implemented because no incentive compatible contract exists.

Crucially, the FOA is not needed to solve the problem. In fact, it is possible to fully characterize the set of incentive compatibility contracts for *any* underlying performance technology. However, the shape of the feasible set depends on the properties of said performance technology. A set of natural regularity properties is identified and the class of performance technologies that satisfy these properties are singled out for more detailed study. The FOA is *not* valid in this class, yet *every* technology within it gives rise to qualitatively similar feasible sets. Specifically, it is feasible to induce larger actions alongside larger thresholds.

Note that the two versions give opposite conclusions regarding which actions allow larger thresholds or more stringent criteria to be used. This is important because it is a fundamental feature of the model that more stringent criteria tend to lower incentive costs. The principal therefore has an incentive to distort the action in such a way that higher thresholds become feasible. In the first version, distorting the action downwards allows more stringent thresholds to be used, but this is reversed in the second version. Thus, the second-best action is lower than the first-best action in the first model but higher than the first-best action in the second model. In either case, the feasibility constraint binds at the second-best solution and the optimal threshold is higher than the threshold that optimally implements the first-best action.

In general, contracting problems are more complicated when the FOA is invalid. Kirkegaard (2017) presents an example with a risk averse agent in which the FOA suggests that the second-best action is close to but below the first-best action. However, neither action is in fact implementable. The nearest implementable action is higher, and for this reason the second-best action turns out to be above the first-best.[2] Thus, it is not necessarily surprising that the second-best action may exceed the first-best action. However, the current paper presents a whole class of distribution functions, disciplined by certain regularity conditions, for which the second-best is higher than the first-best. In this sense, the conclusion is robust within the model.[3] Contrary to

[2] The famous counterexample to the FOA in Mirrlees (1999) exhibits a similar logic, but is slightly harder to interpret since the payoff functions are non-standard.

[3] In comparison, other standard moral hazard models predict that the second-best is weakly

Kirkegaard's (2017) example, the first-best action is implementable but only at low and therefore unprofitable thresholds.

In the third version of the problem, the principal is assumed to intrinsically care about the criterion for success, as in the salesman example. Consequently, the second-best solution is not necessarily on the boundary of the feasible set because the principal may be reluctant to distort the threshold too much. When the first-best is interior and feasible in the second-best problem, the latter features more stringent criteria for success or a lower action (or both) than the first-best solution. In general, both the threshold and the action are distorted away from the first-best. This is typically done in such a way that the threshold is too high compared to what would be socially optimal given the second-best action. In the salesman example, the price then exceeds the monopoly price for the demand curve induced by the agent's action.

The distortions identified so far are consequences of the principal's desire to extract rent from the agent. However, there are settings in which the principal's intrinsic interest in the criterion for success is strong enough to dominate the rent-extraction motive. In such cases, the second-best contract Pareto dominates the contract that implements the first-best action in the way that is optimal to the principal. It is even possible that the second-best contract maximizes social surplus among all incentive compatible contracts, despite limited liability allowing the agent to earn rents.

Various problems with similar flavor have been examined before, but rarely while comparing the first-best and second-best. For instance, other strands of literature have realized that the salesman's incentives are influenced by the price. Thus, Lal (1986) and Mishra and Prasad (2004) debate whether it is better to centralize the pricing decision or delegate it to the agent, contrasting different assumptions about who knows a relevant demand state and what the space of contracts is. However, this literature does not compare first-best and second-best actions or prices.

Li and Yang (2020) examine the design of an optimal monitoring technology in which performance can be partitioned into an exogenously fixed number of categories. This is more general than the current paper, which only allows a partition into success and failure. However, they simply assume that the principal wishes to induce the highest possible action. Thus, they do not study how the first-best and second-best

---

below the first-best action. With risk averse agents, this occurs in the linear-exponential-normal model and in the two-action model. It also occurs in the Demougin and Fluet (1998) model with a risk neutral agent that is discussed in Section 6.

actions differ. Moreover, absent a budget constraint it is always possible to implement the highest action with a sufficiently large bonus. Hence, the implementability problem is minimized. The current paper allows the study of environments consistent with Li and Yang (2020) (with two categories) while allowing the first-best action to be interior. Then, the second-best action may exceed the first-best action, although this depends on how strongly monitoring costs varies with the threshold.

Bond and Gomes (2009) consider an agent who works on a number of independent tasks, each of which either succeeds or fails. The agent's compensation is contingent on the number of successes. The FOA is not valid because the agent may want to deviate to the lowest possible effort profile (shirking on all tasks). Similarly, the problem in the present paper is often precisely that the agent is tempted to deviate to the lowest possible action. Thus, Bond and Gomes (2009) and the current paper share some methodological similarities and are among the few that study the economic consequences of the failure of the FOA. However, Bond and Gomes (2009) assume that the first-best solution involves the agent working as hard as possible on each task. Thus, the only possible distortion in total effort is downwards.

Finally, there is a related literature on what amounts to endogenous criteria for success in settings where performance is fully observable and agents are risk neutral. The set of contracts is richer when performance is observable and contractible, but this does not rule out that optimal contracts take simple forms. Demougin and Fluet (1998) and Oyer (2000) provide different conditions under which threshold contracts are optimal. They assume that the FOA is valid, and their analyses and examples suggest that the second-best action is below the first-best action.

In the current paper, environments are identified in which the FOA is not valid but where threshold contracts are nevertheless optimal even if performance is fully observable. Here, the second-best action coincides with the first-best action when the principal does not intrinsically care about the threshold. Beyond these environments there are settings in which the principal does not benefit from being able to observe failing performances. Examples of such situations include the type of inventory problems studied in Dai and Jerath (2013) and Chu and Lai (2013). The principal incurs sunk costs to stock products for her agent to sell and she can observe how many units are sold if the stock is not depleted. The optimal contract then often entails a bonus that is paid if and only if the stock is exhausted. Thus, threshold contracts remain optimal in many settings even when performance is more finely observable.

2 Model and preliminaries

The principal (she) employs a single agent (he). The agent takes some costly and unobservable action, a, belonging to some compact interval $[\underline{a}, \overline{a}]$. Given the agent's action, his performance is a random variable, X. Let $F(x|a)$ denote the corresponding distribution function, given a. For all $a \in (\underline{a}, \overline{a}]$, assume that there are no mass points and that the support $[\underline{x}, \overline{x}]$, which may be bounded or unbounded, is the same interval for all $a \in (\underline{a}, \overline{a}]$, with density $f(x|a) = F_x(x|a)$ that is strictly positive on $[\underline{x}, \overline{x}]$.[4] At $a = \underline{a}$, the distribution either (i) satisfies these same assumptions or (ii) is degenerate at $\underline{x}$. In the latter case, the agent's performance is guaranteed to be the worst possible if he takes the lowest possible action. This case is included because it is useful in illustrating the workings of the model and because it arises naturally in some parameterized examples. The derivatives F_a, F_{aa}, f_a and their partial and cross partial derivatives are assumed to exist for all $a \in (\underline{a}, \overline{a}]$. Assume that $F_a(x|a) < 0$ for all $x \in (\underline{x}, \overline{x})$ and all $a \in (\underline{a}, \overline{a}]$. Thus, actions are productive in the sense that bad outcomes are less likely the harder the agent works.

Actions are normalized such that the agent's cost function is linear. Thus, the agent incurs a cost of a when he takes action a. Alternatively, think of the agent's action as a decision of what costs to incur. The agent is risk neutral and protected by limited liability. This assumption makes it possible to succinctly characterize implementation costs (i.e. the expected wage), but it is not important for the more fundamental discussion of which contracts are incentive compatible.[5] The minimum wage implied by the limited liability constraint is normalized to zero. The agent's outside option is assumed to be so poor that the participation constraint never binds.

The principal either does not directly observe the agent's performance or his performance is not verifiable. Instead, what is observable and verifiable is whether the agent's performance exceeds a (deterministic) threshold t. Thus, $[\underline{x}, \overline{x}]$ is partitioned into two intervals, $[\underline{x}, t)$ and $[t, \overline{x}]$. The agent fails if his performance falls in the first interval and succeeds otherwise. Note that the threshold t in this way describes the criterion for success. The higher t is, the more stringent is the criterion.

Thus, there are two verifiable outcomes. The novelty is that the criterion for

[4] A subscript indicates the partial derivative with respect to the subscripted variable.

[5] As long as the agent has quasi-linear utility, the bonus from succeeding can be interpreted as being measured in "utils" rather than in monetary terms. The characterization of incentive compatible contracts carries through under this interpretation.

success is endogenized. That is, the principal controls what the threshold is. The stringency of the criterion affects the agent's incentives and therefore the implementation costs that the principal faces.

2.1 Properties of incentive compatible contracts

To understand the agent's incentives, fix a threshold $t \in (\underline{x}, \overline{x})$ and a bonus b that is paid if the outcome is a success. Given the slack participation constraint, the principal pays nothing if the outcome is a failure. The agent's expected utility from action a is $b\,(1 - F(t|a)) - a$. Now assume that the principal aims to induce some specific interior action a while holding fixed the threshold $t \in (\underline{x}, \overline{x})$. Then, b must be calibrated to ensure that the agent's first-order condition is satisfied at the intended action. That is, the bonus is

$$B(a,t) = \frac{-1}{F_a(t|a)}.$$

Hence, if the agent takes the intended action a when offered the bonus $B(a,t)$, his expected wage is

$$W(a,t) = -\frac{1 - F(t|a)}{F_a(t|a)}$$

and his expected utility is

$$U(a,t) = W(a,t) - a.$$

A contract specifies a criterion for success and a bonus if the outcome is a success. However, since $B(a,t)$ is uniquely nailed down by (a,t), the problem can effectively be summarized by the pair (a,t). In the following, (a,t) should therefore be read as: "the principal intends to induce action a by specifying threshold t and committing to the bonus $B(a,t)$." The problem is that the first-order condition is necessary but not always sufficient for the agent's utility to attain a global maximum at the intended action a. Thus, keep in mind that $W(a,t)$ and $U(a,t)$ are valid only as long as (a,t) is globally incentive compatible, or more precisely when

$$a \in \arg\max_{a'} B(a,t)\,(1 - F(t|a')) - a'.$$

A general analysis of the incentive compatibility problem is postponed to Section 4.

The cheapest way to induce action $\underline{a}$ is to offer a zero bonus, regardless of the threshold. Thus, let $B(\underline{a},t) = W(\underline{a},t) = 0$ and $U(\underline{a},t) = -\underline{a}$. Note that implementation costs are generally discontinuous at $\underline{a}$ (an exception is considered in Section 4.3.1). Similarly, let $B(\overline{a}, t)$ denote the lowest bonus that can be used to induce action $\overline{a}$ with threshold $t \in (\underline{x}, \overline{x})$, and let $W(\overline{a}, t)$ and $U(\overline{a}, t)$ denote the resulting expected wage and expected utility, respectively. Incentive compatibility may necessitate a higher bonus than what the first-order condition suggests. In other words, using the first-order condition gives *lower bounds* on $B(\overline{a}, t)$, $W(\overline{a}, t)$, and $U(\overline{a}, t)$.

Holding fixed the threshold, a standard argument proves that a higher bonus must be offered to induce a higher action. Thus, when the agent is induced to work harder, he benefits not only from a higher bonus but also from a higher probability that he passes the fixed threshold. This double benefit increases his expected wage and more than compensates for the fact that he also incurs higher effort costs. The next proposition records and proves these properties. Proofs are in Appendix A.

Proposition 1 *Fix an interior threshold $t \in (\underline{x}, \overline{x})$ and assume that actions a and a' are implementable, with $a' > a$. Then, $B(a', t) \geq B(a, t)$, $U(a', t) \geq U(a, t)$, and $W(a', t) > W(a, t)$.*

Next, hold $a \in (\underline{a}, \overline{a})$ fixed and consider how the contract depends on the threshold that is used to induce the action. For many of the following results, the monotone likelihood-ratio property (MLRP) is imposed. In fact, for expositional simplicity a strict version of the MLRP is used. Under the (strict) MLRP, the likelihood-ratio $\frac{f_a(x|a)}{f(x|a)}$ is (strictly) increasing in x. An equivalent definition is as follows.

Definition (MLRP): The monotone likelihood-ratio property is satisfied if $f(x|a)$ is strictly log-supermodular in x and a, or

$$\frac{\partial^2 \ln f(x|a)}{\partial x \partial a} > 0.$$

There is a potential trade-off when increasing the threshold. First, the higher the threshold is, the less likely it is that the bonus is paid out. On the other hand, it is possible that the bonus must be increased as well. The MLRP implies that the first effect dominates because the bonus increases relatively slowly. Thus, $W(a, t)$ is strictly decreasing in t. Since the threshold does not directly impact the cost of effort, $U(a, t)$ is strictly decreasing in t as well.[6]

[6] By incentive compatibility and limited liability, $U(a, t) \geq B(a, t)(1 - F(t|\underline{a})) - \underline{a} \geq -\underline{a}$. Thus, utility is bounded below and the participation constraint is slack if the outside option is bad enough.

Proposition 2 *Assume that the MLRP is satisfied. Fix an interior action* $a \in (\underline{a}, \overline{a})$ *and assume that it can be induced with thresholds* t *and* t', *with* $t' > t$. *Then,* $W(a, t') < W(a, t)$ *and* $U(a, t') < U(a, t)$.

2.2 The principal's problem

The principal is assumed to be risk neutral. The cost $W(a, t)$ of implementing an incentive compatible (a, t) pair depends on the threshold. Thus, there is generally an incentive to manipulate the criterion for success in order to manipulate implementation costs. However, the principal may also take a more direct interest in the threshold t. The expected benefit to the principal of (a, t) is $\pi(a, t)$. Her objective is therefore to maximize $\pi(a, t) - W(a, t)$ over incentive compatible (a, t).

The benefit function may depend directly on the criterion for success. The leading example is $\pi(a, t) = (t - c)(1 - F(t|a))$. Here, the principal hires a salesman (agent) to sell a product at price t to a single customer. If successful, the principal incurs a cost c of supplying the product. As mentioned in the introduction, the agent's performance is the willingness-to-pay that he is able to instill in the customer. Given an action a and a price t, the probability of a success is $1 - F(t|a)$, and $\pi(a, t)$ thus describes expected profits. Note that $\pi(a, t)$ is non-monotonic in t in this example. The threshold t will be said to be "intrinsically important" to the principal whenever $\pi(a, t)$ depends on t. This encompasses situations in which it is harder or more costly for the principal to detect if performance exceeds some thresholds rather than others.

There are also environments in which the principal does not care directly about the criterion for success. With some abuse of notation, the benefit function is written more succinctly as $\pi(a)$ in those cases. The obvious example is $\pi(a) = \mathbb{E}[X|a]$. Here, the agent's performance can be interpreted as his productivity, which the principal cares about. However, at the point in time at which the agent must be paid, it can only be verified whether the performance exceeded a pre-set threshold or not.

Social surplus is the difference between the benefits and the effort costs, or

$$S(a, t) = \pi(a, t) - a.$$

The first-best benchmark entails maximizing $S(a, t)$.[7] Thus, any first-best solution

[7] This definition ignores any impact on third parties. For instance, the customer in the salesman example is impacted by (a, t), but this is disregarded in $S(a, t)$.

consists of a pair (a^{FB}, t^{FB}). The point here is that t is directly important for welfare when it is intrinsically important to the principal. Hence, any distortion of the threshold away from t^{FB} is important for welfare reasons. In contrast, t does not matter for social surplus when the benefit function takes the form $\pi(a)$. Stated differently, there is no unique first-best threshold in such cases.

The principal's second-best problem is to maximize

$$V(a,t) = \pi(a,t) - W(a,t).$$

It is often more useful to think of the principal as the "residual claimant," since she claims what is left of social surplus after the risk neutral agent has received his share, or

$$V(a,t) = S(a,t) - U(a,t).$$

Propositions 1 and 2 already reveal important information about $W(a,t)$ and $U(a,t)$. First, fix the threshold t and assume that there is a unique action that maximizes $S(a,t)$. This action identifies the first-best action in a world where the threshold is exogenous. If this action is feasible in the second-best problem, then $V(a,t)$ is maximized at an action that is no higher. The reason is that at higher actions, social surplus is strictly lower and the agent is weakly better off, hence leaving less surplus for the principal. Thus, a standard model with exogenous thresholds predicts that the second-best action is no higher than the first-best action.

Second, $W(a,t)$ is strictly decreasing in the threshold. Thus, when the criterion for success is not intrinsically important, the principal will aim to increase the threshold as much as possible in order to decrease implementation costs. Hence, an existence problem *may* arise because a threshold of $\overline{x}$ is not incentive compatible – the agent never succeeds and will therefore pick action $\underline{a}$ in response.

The following sections consider three different versions of the model in which the existence problem does not arise. First, Section 3 assumes that the principal faces a budget constraint or wage cap. The budget constraint automatically makes it impossible to implement thresholds close to $\overline{x}$.[8]

Second, Section 4 focuses on the incentive compatibility problem by discarding the strong assumption that validates the FOA. Under more realistic assumptions on the

[8] Section 6.2 discusses other ways in which similar existence problems have been addressed in the literature, including discretizing the signal space or assuming the MLRP is violated.

distribution function, contracts with large thresholds are not incentive compatible in the first place. Sections 3 and 4 assume that the principal is not intrinsically interested in the criteria for success.

Third, imagine that the threshold is intrinsically important to the principal and t^{FB} is interior. If the threshold is important enough, then distorting it too much is so undesirable that it does not justify the accompanying decrease in implementation costs. Section 5.1 examines such settings. Section 5.2 assumes that the threshold is intrinsically important but that one of the constraints from Sections 3 or 4 bind.

3 Budget constraints

Assume that the principal faces a budget constraint. That is, she can offer a bonus of at most $\overline{b}$, where $\overline{b} > 0$ is bounded above. Assume also that the FOA is valid. This holds if $F(t|\cdot)$ is globally convex in a, since the agent's problem is then concave. Thus, Rogerson's (1985) Convexity of Distribution Function Condition (CDFC) is imposed.

Definition (CDFC): The Convexity of Distribution Function Condition is satisfied if $F_{aa}(x|a) \geq 0$ for all $x \in [\underline{x}, \overline{x}]$ and all $a \in (\underline{a}, \overline{a}]$.

The CDFC is often criticized. It is used here not because it is a desirable assumption but instead to focus squarely on budget constraints. The next section does the opposite, by ignoring budget constraints but relaxing the CDFC.

Importantly, thresholds near $\overline{x}$ violate the budget constraint. The reason is that failure is almost guaranteed when t is close to $\overline{x}$. Thus, it is harder for the agent to manipulate the chance of success; formally, $F_a(t|a) \to 0$ as $t \to \overline{x}$. To maintain incentives, the bonus would need to grow without bound as $t \to \overline{x}$.

Proposition 1 implies that larger actions necessitate larger bonuses, given t. Hence, the larger the action is, the fewer thresholds meet the budget constraint. The next result describes important qualitative features of the frontier of the feasible set.

Lemma 1 *Assume that the MLRP and CDFC hold. For any given $\overline{b}$, the set of actions that can be implemented is an interval of the form $[\underline{a}, \overline{a}^B]$, where $\overline{a}^B \leq \overline{a}$. The highest possible threshold, $\overline{t}^B(a)$, that can be used to implement a is weakly decreasing on $[\underline{a}, \overline{a}^B]$. That is, weakly higher thresholds are feasible the smaller the action is.*

Turning to optimal contracts, assume that the benefit function $\pi(a)$ is independent of the criterion for success. Thus, the threshold is manipulated with the sole purpose of lowering implementation costs. For expositional convenience, assume that a^{FB} is unique and interior. It is trivial that the second-best action, a^{SB}, is below a^{FB} if the latter cannot be implemented. Thus, the case where $a^{FB} \leq \overline{a}^B$ is more interesting.

From Proposition 2, the cheapest feasible way to induce action $a > \underline{a}$ is to equate the threshold to $\overline{t}^B(a)$. It has already been observed that for an exogenously fixed threshold, the second-best action is no higher than the first-best action. Endogenizing the threshold, Lemma 1 implies that the principal has an additional incentive to lower the action because it enables her to use higher and therefore even cheaper thresholds. Thus, the second-best action can be no higher than the first-best action. As long as $a^{SB} > \underline{a}$, this in turn means that the second-best threshold must be larger than the threshold that optimally implements a^{FB}, or $t^{SB} = \overline{t}^B(a^{SB}) \geq \overline{t}^B(a^{FB})$. In this sense, the criterion for success is stringent. Indeed, since a^{SB} is small and t^{SB} is large, there is a smaller probability that the agent is successful.

Proposition 3 *Assume that the MLRP and CDFC hold. Assume that the principal's benefit function, $\pi(a)$, depends only on a and that there is a unique and interior first-best action, a^{FB}. For any given $\overline{b}$, any second-best action is no greater than the first-best action, $a^{SB} \leq a^{FB}$. If $a^{FB} \leq \overline{a}^B$ and $a^{SB} > \underline{a}$, then $t^{SB} = \overline{t}^B(a^{SB}) \geq \overline{t}^B(a^{FB})$.*

Example 1: Assume that $F(x|a) = \left(\frac{x}{16}\right)^a$, $x \in [0, 16]$, $a \in [1, 4]$. This distribution satisfies the MLRP and the CDFC and is inspired by an example in Rogerson (1985). It is assumed that $\overline{b} = B(3, 13) = 8.98$. Assuming that $\pi(a) = E[X|a] = \frac{16a}{1+a}$, the first-best action is $a^{FB} = 3$. In the second-best problem, $\overline{t}^B(a^{FB}) = 13$ is the most profitable feasible threshold that induces a^{FB}. The resulting contract yields profit of 7.84. In comparison, action $\underline{a} = 1$ can be induced at zero cost, yielding profit of 8. However, the second-best (obtained numerically) is at $(a^{SB}, t^{SB}) = (2.22, 13.66)$, which yields expected profit of 8.37. ▲

4 Implementability constraints

This section studies the problem without imposing the CDFC. Thus, the agent's first-order condition need not be sufficient for incentive compatibility. Consequently, not all (a, t) are generally implementable. To focus on the implementability problem, the principal is assumed not to be budget constrained.

4.1 Characterization of the feasible set

It is possible to completely characterize the set of incentive compatible (a,t).[9] To this end, fix $t \in (\underline{x}, \overline{x})$ and think of it as a parameter. Then, for a fixed bonus b, the curvature of the agent's expected utility depends only on the curvature of $1 - F(t|\cdot)$ with respect to a. Hence, the problem is locally concave in a if F is locally convex in a. Clearly, the CDFC ensures that the agent's problem is globally concave. Without the CDFC, the idea is to "concavify" $1 - F(t|\cdot)$, or equivalently to "convexify" $F(t|\cdot)$.

Thus, starting from the function $F(t|\cdot)$, construct the convex hull (as a function of a), and denote this $F^C(t|\cdot)$. The convex hull is the largest convex function that lies on or below $F(t|\cdot)$. Thus, $F^C(t|a) \leq F(t|a)$ for all $a \in [\underline{a}, \overline{a}]$. For any t, let

$$A^C(t) = \{a \in [\underline{a}, \overline{a}] | F^C(t|a) = F(t|a)\}$$

denote the set of actions for which $F(t|a)$ coincides with $F^C(t|a)$. The end-points of the domain are always in $A^C(t)$, or $\underline{a}, \overline{a} \in A^C(t)$.

For any $t \in (\underline{x}, \overline{x})$, it holds that (a,t) is incentive compatible if and only if $a \in A^C(t)$.[10] Thresholds of $\underline{x}$ or $\overline{x}$ can be used to induce only $\underline{a}$ since $F(\underline{x}|a) = 0$ and $F(\overline{x}|a) = 1$ are independent of a. These results are summarized in the next statement.

Lemma 2 *The feasible set of implementable or incentive compatible* (a,t) *is*

$$\mathcal{I} = \{(a,t) \in [\underline{a}, \overline{a}] \times (\underline{x}, \overline{x}) \, | a \in A^C(t)\} \cup \{(\underline{a}, \underline{x}), (\underline{a}, \overline{x})\}.$$

The "implementability constraint" from now on refers to the condition that the principal must necessarily select a (a,t) pair that belongs to $\mathcal{I}$.

Holding fixed the threshold t, the set $A^C(t)$ traces out all the actions in $\mathcal{I}$. Moving along the other dimension, let $T^C(a)$ denote the set of thresholds for which a can be implemented, $T^C(a) = \{t \in [\underline{x}, \overline{x}] | (a,t) \in \mathcal{I}\}$. Thus, $T^C(a)$ describes the set of thresholds that can be used to incentivize the action a.

[9] The argument leading to Lemma 2 borrows from Kirkegaard (2017). Although he considers a more traditional contracting setting, at the technical level his model is closely related to a problem with binary outcomes as featured in the present model.

[10] For $a \in \{\underline{a}, \overline{a}\}$, this statement is taken to mean that there are bonuses that make the contract incentive compatible. The action $\underline{a}$ can be implemented with any threshold and a fixed wage contract (no bonus). Similarly, the action $\overline{a}$ can be implemented with any interior threshold $t \in (\underline{x}, \overline{x})$ by picking a bonus that is so large that the agent's utility is globally increasing in the action.

Examples that illustrate Lemma 2 can be found in Sections 4.2, 4.3, 5.1, 5.2, and in Appendices B and C. For now, note that $A^C(t) = [\underline{a}, \overline{a}]$ for all $t \in (\underline{x}, \overline{x})$ if and only if the CDFC is satisfied. Thus, *any* departure from the CDFC implies that there are some interior (a, t) that are not implementable.

4.2 The shape of the feasible set

Lemma 2 describes how to derive the feasible set for *any* distribution function. However, the shape of the feasible set depends on the specific properties of the latter. This subsection asks what some reasonable or natural properties are, and what such properties imply for the shape of the feasible set. The next subsection utilizes these results to solve the principal's problem.

To begin, it is helpful to introduce a very simple way to relax the CDFC.

DEFINITION (CAT): *Concavity at the top* is satisfied if, for any $a \in (\underline{a}, \overline{a})$, there exists some $x' \in (\underline{x}, \overline{x})$ such that $F_{aa}(x|a) < 0$ for all $x \in (x', \overline{x})$.

Concavity at the top (CAT) rules out the CDFC and implies that no interior action can be implemented with very high thresholds.

Chade and Swinkels (2020) introduce a *no-upward-crossing* condition, which can be stated as the requirement that $F_{aa}(\cdot|a) - \tau F_a(\cdot|a)$ never crosses 0 from below on $(\underline{x}, \overline{x})$, for any $\tau \in \mathbb{R}$ and any a. An equivalent statement is that $-F_a(\cdot|\cdot)$ is log-supermodular in a and x, or that $\frac{F_{aa}(\cdot|a)}{F_a(\cdot|a)}$ is increasing. Modifying Chade and Swinkels' (2020) terminology slightly, the abbreviation NUC_x will be used for no-upward-crossing with respect to x.

DEFINITION (NUC_x): The *no-upward-crossing* condition (with respect to x) is satisfied if $-F_a(x|a)$ is log-supermodular in a and x.

Given NUC_x, $F_{aa}(\cdot|a)$ is first-positive-then-negative as x increases. Thus, if NUC_x is satisfied but F is not globally convex in a for *any* x then CAT is automatic.

Chade and Swinkels (2020) provide sufficient conditions for NUC_x. Even though they provide counterexamples, they argue that NUC_x is a relatively weak condition.[11] They mention the location families as a special example, such that $F(x|a)$ and $f(x|a)$ can be written as $Q(x-a)$ and $q(x-a)$, respectively. Here, it holds that $-F_a(x|a) = q(x-a) = f(x|a)$. Thus, in this case, the MLRP and NUC_x are the same condition.

[11] In their leading counterexample, $F_{aa}(\cdot|a)$ is first-negative-then-positive as x increases. Hence, CAT is violated and the existence issue arises again.

NUC_x implies that F_{aa} is more likely to be negative the higher the threshold is, suggesting that the set of implementable actions shrinks as t increases. This is correct, but the proof is more involved since the convex hull of $F(t|\cdot)$ must be examined.

Proposition 4 *Assume NUC_x holds. If $t', t \in (\underline{x}, \overline{x})$ and $t' > t$ then $A^C(t') \subseteq A^C(t)$. That is, fewer interior actions can be implemented the higher the threshold is.*

Next, move along the other dimension. Thus, fix a target action and ask which thresholds can work to implement that particular action.

Proposition 5 *Assume NUC_x and CAT are satisfied. Then, for any $a \in (\underline{a}, \overline{a})$, the set $T^C(a)$ is empty or it is an interval of the form $(\underline{x}, \overline{t}^C(a)]$, where $\overline{t}^C(a) < \overline{x}$. Thus, thresholds close to $\overline{x}$ cannot be used to implement a.*

For $a = \overline{a}$, let $\overline{t}^C(\overline{a})$ denote the highest threshold such that the bonus derived from the FOA is incentive compatible. Thresholds above $\overline{t}^C(\overline{a})$ can still be used to induce action $\overline{a}$, but the bonus must be made higher than what is suggested by the FOA.

Introducing a new definition, say that F satisfies *no-downward-crossing* with respect to a, abbreviated NDC_a, if $F_{aa}(x|\cdot)$ never crosses 0 from above on $(\underline{a}, \overline{a})$, for any $x \in [\underline{x}, \overline{x}]$. This allows $F_{aa}(x|\cdot)$ to be first-negative-and-then-positive as a increases. NDC_a is a natural counterpart of NUC_x, which considers the effects of increasing x.

Definition (NDC_a): The *no-downward-crossing* condition (with respect to a) is satisfied if $F_{aa}(x|\cdot)$ never crosses 0 from above on $(\underline{a}, \overline{a})$, for any $x \in [\underline{x}, \overline{x}]$.

Equivalently, NDC_a says that $-F_a$ is unimodal in a. A sufficient condition is that $-F_a$ is log-concave in a. In the location families mentioned before, log-concavity holds under the standard assumption that the density, q, is log-concave.

To understand NUC_x, recall that $-F_a$ is the marginal increase in the probability of succeeding, $1 - F(t|a)$, when a increases. Thus, $\frac{\partial \ln(-F_a)}{\partial a}$ measures how the return to extra effort depends on how large effort was to begin with. In turn, $\frac{\partial^2 \ln(-F_a)}{\partial x \partial a} \geq 0$ implies that the marginal return to extra effort is more sensitive to what the starting action is the larger the threshold is. Chade and Swinkels (2020) explain NUC_x in the context of a jogger trying to run a certain distance in a pre-specified amount of time. The NUC_x disciplines how the marginal increase in the probability of success from additional exercise for a committed jogger (starting at a large a) as compared to a sedentary person (starting at a low a) changes with the threshold (distance).

NDC_a instead says that for a fixed threshold, the probability of success, $1-F(t|a)$, is first-convex-then-concave in effort. For the sedentary person, a bit of additional exercise is not going to improve the chance that he will be able to run the full distance in the allotted time very much. However, as the amount of exercise ramps up, the chance of succeeding increases rapidly, until a point is reached where success is all but guaranteed and the marginal return to further exercise diminishes. Thus, the "learning curve" is s-shaped.

NDC_a implies that the set of feasible actions has a particularly simple structure.

Proposition 6 *Assume NDC_a holds. Then, for any $t \in (\underline{x}, \overline{x})$, the set of implementable actions takes either the form (i) $A^C(t) = \{\underline{a}, \overline{a}\}$, (ii) $A^C(t) = [\underline{a}, \overline{a}]$, or (iii) $A^C(t) = \{\underline{a}\} \cup [\underline{a}^C(t), \overline{a}]$, where $\underline{a}^C(t) \in (\underline{a}, \overline{a})$.*

Thus, if some interior action is implementable, then all higher actions are implementable as well. In case (i), define $\underline{a}^C(t) = \overline{a}$ and in case (ii) define $\underline{a}^C(t) = \underline{a}$. Then, in all three cases, the set $A^C(t)$ can be written in the form $\{\underline{a}\} \cup [\underline{a}^C(t), \overline{a}]$.

It is useful to observe that NDC_a implies that if action a is implemented with a threshold of $\overline{t}^C(a)$ then the agent is exactly *indifferent* between a and the lowest possible action, $\underline{a}$.[12] This property is analytically convenient because the "anchor" $\underline{a}$ is independent of the action that is to be implemented, which in turn makes it easier to make inferences about the agent's utility along the boundary of the feasible set.

The next result combines the previous assumptions and in a way describes the "nicest" shape of the feasible set that can be expected without the CDFC.

Corollary 1 *Assume that CAT, NUC_x, and NDC_a all hold. Then, $\underline{a}^C(t)$ is weakly increasing in t on $(\underline{x}, \overline{x})$. Equivalently, $\overline{t}^C(a)$ is weakly increasing in a on $(\underline{a}, \overline{a})$.*

Intuitively, for any $(a, \overline{t}^C(a))$ pair, the temptation is as mentioned to shirk as much as possible, i.e. deviate to $\underline{a}$. If a is small then the probability of success is small, other things equal. Hence, the temptation to give up is strong. To keep this apathy in check, it is necessary that the threshold $\overline{t}^C(a)$ is lower the lower a is.

Contrasting Lemma 1 and Corollary 1, note that $\overline{t}^B(a)$ and $\overline{t}^C(a)$ move in opposite directions. Thus, the shape of the feasible sets are markedly different.

[12] Intuitively, the problem is that low thresholds would cause a deviation to $\underline{a}$ (recalling that utility is first convex-then-concave under NDC_a). At a threshold of $\overline{t}^C(a)$, this incentive constraint exactly binds. At higher threshold, the constraint is slack.

The remainder of the section assumes that $F(x|a)$ is regular in the following sense.

Definition (Regularity): $F(x|a)$ is *regular* if MLRP, CAT, NUC_x, NDC_a all hold and that for any $a \in (\underline{a}, \overline{a})$, there exists a threshold $\overline{t}^C(a) \in (\underline{x}, \overline{x})$ such that a can be implemented if and only if the threshold is no larger than $\overline{t}^C(a)$.[13]

Example 2: Assume that $F(x|a)$ is the Kumaraswamy distribution,

$$F(x|a) = 1 - (1 - x^a)^\beta \, , \, x \in [0, 1]$$

where $\beta > 0$ is a shape parameter and $\underline{a} \geq 0$. It is easy to verify that the MLRP and the NUC_x hold. Likewise, for any $x \in (0, 1)$, $F_{aa}(x|a)$ has the same sign as $1 - \beta x^a$. Thus, the CDFC is satisfied if $\beta \in (0, 1]$. Indeed, note that $\beta = 1$ reproduces the distribution in Example 1 (with a normalized support), for which the CDFC holds.

For $\beta > 1$, CAT holds since $1 - \beta x^a < 0$ when x is sufficiently close to one. For similar reasons, NDC_a holds as well. If $\underline{a} > 0$, then $F_{aa}(x|a)$ is strictly positive for all $a \in [\underline{a}, \overline{a}]$ when x is sufficiently small. Any such threshold can then be used to implement any action. Combined with CAT, there thus exists a threshold $\overline{t}^C(a) \in (0, 1)$ such that $a \in (\underline{a}, \overline{a})$ can be implemented if and only if the threshold is no larger than $\overline{t}^C(a)$. Hence, regularity is satisfied. The last part of the argument does not hold if $\underline{a} = 0$. However, in this case, $F(x|\underline{a})$ is degenerate. It then turns out to be straightforward to solve for $\overline{t}^C(a)$ and verify directly that $\overline{t}^C(a) \in (0, 1)$ for all $a \in (\underline{a}, \overline{a}]$. By the indifference condition just mentioned, $U(a, \overline{t}^C(a)) = 0$ since action $\underline{a} = 0$ has no chance of earning a bonus. This in turn implies that $\overline{t}^C(a) = z^{\frac{1}{a}}$, where $z \in (0, 1)$ solves $1 = z(1 - \beta \ln z)$. In this example, the probability of success is $(1 - z)^\beta$, and thus constant, along the boundary of the feasible set. ▲

4.3 The second-best solution

Consider again the case in which the principal takes no direct interest in the threshold t. As before, assume that there is a unique and interior first-best action, a^{FB}. The distribution $F(x|a)$ is assumed to be regular. Since wage costs are decreasing in t and π is independent of t, any interior solution to the second-best problem must be on the boundary of the feasible set, i.e. be of the form $(a, \overline{t}^C(a))$.

[13] CAT already ensures that $\overline{t}^C(a) < \overline{x}$. Thus, what is assumed in addition is that $T^C(a)$ is not empty. Hence, all a are implementable with some threshold. This rules out that $F(x|a)$ is globally concave in a for all x.

4.3.1 The best-case scenario

The best-case scenario for the principal is that $F(x|\underline{a})$ is degenerate, or $F(x|\underline{a}) = 1$ for all $x \in [\underline{x}, \overline{x}]$. In words, the agent's performance is guaranteed to be the worst possible if his action is $\underline{a}$.[14] Therefore, for any $t \in (\underline{x}, \overline{x})$, there is no chance that the agent earns the bonus with action $\underline{a}$. Hence, such a deviation is less desirable and therefore easier to prevent. In particular, the indifference condition mentioned after Proposition 6 becomes

$$U(a, \overline{t}^C(a)) = -\underline{a}.$$

Thus, in this special case, the agent is indifferent between all $(a, \overline{t}^C(a))$, $a \in (\underline{a}, \overline{a}]$. Stated differently, the agent appropriates a constant amount of rent and the rest goes to the principal. Thus, the first-best action solves the second-best problem.

Proposition 7 *Assume that $F(x|a)$ is regular and that $F(x|\underline{a})$ is degenerate. Assume that the principal's benefit function, $\pi(a)$, depends only on a and that there is a unique and interior first-best action, a^{FB}. Then the second-best action coincides with the first-best action, $a^{SB} = a^{FB}$.*

The first-best is implemented even though the limited liability constraint prevents the principal from "selling the firm" and extracting all rent in the usual way. However, the logic is similar, since implementing the first-best maximizes social surplus and extracts as much rent as possible from the agent.[15]

Remark 1 *For any interior $(a, \overline{t}^C(a))$, the agent appropriates the smallest possible amount of rent that the limited liability constraint permits. In other words, regardless of the principal's information, it is impossible to extract more rent from the agent. Thus, even if the agent's performance, x, is perfectly observable and verifiable, the principal can do no better than using a threshold contract. In particular, there is no incentive for the principal to incur costs to observe x more finely than what is required for the threshold contract.*

[14] In the literature, such an assumption is sometimes implicit in the model. For instance, Innes (1990) assumes that performance is non-negative but equal to zero in expectation when the lowest possible action is taken. Thus, performance must be zero with probability one, given action $\underline{a}$.

[15] Gürtler and Kräkel (2010) consider a framework where shirking can be detected and litigated. Depending on how rents are shared or destroyed under litigation, the first-best action may be implemented under the threat of litigation. A key consideration is to prevent a deviation to $\underline{a}$.

4.3.2 The general case

Next, remove the assumption that $F(x|\underline{a})$ is degenerate. A deviation to $\underline{a}$ now carries with it a strictly positive probability that the agent earns the bonus. Since the bonus depends on $(a, \bar{t}^C(a))$, the agent's utility therefore also depends on $(a, \bar{t}^C(a))$. Thus, the agent's utility is no longer constant along the boundary of the feasible set. The principal therefore has an incentive to distort the action away from the first-best, since this allows her to manipulate the rent that has to be portioned off to the agent.

Propositions 1 and 2 imply a trade-off: An increase in a benefits the agent, while the associated increase in $\bar{t}^C(a)$ makes him worse off. The latter effect can be shown to dominate, meaning that the agent is worse off the higher a is in the pair $(a, \bar{t}^C(a))$. Thus, there are only two possibilities. First, the principal may induce an action no lower than the first-best action, with the aim of lowering the agent's rent. Second, since implementation costs are discontinuous at $a = \underline{a}$, it cannot be ruled out that inducing $\underline{a}$ is preferable to the principal. Thus, either $a^{SB} = \underline{a}$ or $a^{SB} \geq a^{FB}$.

Proposition 8 *Assume that $F(x|a)$ is regular. Assume that the principal's benefit function, $\pi(a)$, depends only on a and that there is a unique and interior first-best action, a^{FB}. Then the second-best action is either $\underline{a}$ or it is no smaller than the first-best action, or $a^{SB} \geq a^{FB}$. In the latter case, the second-best threshold is no smaller than the threshold that optimally implements a^{FB} subject to feasibility, or $t^{SB} = \bar{t}^C(a^{SB}) \geq \bar{t}^C(a^{FB})$.*

Example 3: Assume that $F(x|a) = 1 - e^{-\frac{x}{4\sqrt{1+a}}}$, $x \geq 0$, $a \in [0, 6]$. Here, the agent's performance is exponentially distributed with mean $E[X|a] = 4\sqrt{1+a}$ and the distribution is regular. Assuming that $\pi(a) = E[X|a]$, the first-best action is $a^{FB} = 3$. In the second-best problem, $\bar{t}^C(a^{FB}) = 19.46$ is the most profitable threshold that can induce a^{FB}. The resulting contract yields profit of 4.71 whereas inducing $\underline{a} = 0$ yields profit of 4. It can be verified that inducing action $\bar{a} = 6$ yields profit of at most 4.58, depending on the threshold that is used. The second-best (obtained numerically) is at $(a^{SB}, t^{SB}) = (3.54, 20.39)$, which yields expected profit of 4.73. ▲

Comparing Propositions 3 and 8 reveals that it matters whether the budget constraint or the "implementability constraint" is binding. The reason is that the feasible set are shaped so differently, since $\bar{t}^B(a)$ and $\bar{t}^C(a)$ move in opposite directions. On the other hand, Propositions 3 and 8 agree on the conclusion that the second-best threshold is typically larger than the threshold that would optimally implement the first-best action subject to feasibility. This distortion is due to the fact that higher thresholds are cheaper to implement. To be able to use higher thresholds, however, the action must be distorted downwards under budget constraints but upwards when it is the implementability constraint that binds.

Examples 2–6 provide examples of regular distribution functions. Appendix C contains an example that shows that some natural distribution functions are not regular. However, the problem can also be solved in such cases.

5 Intrinsically important criteria for success

This section considers benefit functions $\pi(a,t)$ that depend both on the action and the criterion for success. The analysis is broken into two parts, each examining a different class of environments.

5.1 Benefits versus implementation costs

This subsection assumes that the first-best solution (a^{FB}, t^{FB}) is unique and interior. It is also assumed that (a^{FB}, t^{FB}) can feasibly be implemented in the second-best problem. First, this requires that the budget is large enough, or $\bar{b} \geq B(a^{FB}, t^{FB})$. Second, (a^{FB}, t^{FB}) must be incentive compatible, or $(a^{FB}, t^{FB}) \in \mathcal{I}$. This is automatic if the CDFC is satisfied. More generally, however, the complication is that while Section 4 describes the shape of $\mathcal{I}$ for any $F(x|a)$, the first-best solution depends not only on $F(x|a)$ but also on $\pi(a,t)$. The next result confirms that $(a^{FB}, t^{FB}) \in \mathcal{I}$ for two specifications of $\pi(a,t)$ that are closely related to the salesman problem.

Proposition 9 *Let $v(x)$ be a strictly increasing and differentiable function defined on $[\underline{x}, \overline{x}]$. Assume also that there exists some $c \in (\underline{x}, \overline{x})$ such that $v(c) = 0$. Assume that either*

1. *$\pi(a,t) = v(t)(1 - F(t|a))$, or*

2. *$\pi(a,t) = \int_t^{\overline{x}} v(x) f(x|a) dx$ and that F is regular.*

In either case, any first-best solution (a^{FB}, t^{FB}) is in $\mathcal{I}$.

The first case fits the salesman example when $v(t) = t - c$, and where c represents the cost of production. The second case is relevant to up-or-out employment contracts where the action (a) is the agent's effort during the trial period to build up job-specific human capital (x). Human capital accumulation is stochastic and determines the agent's productivity $(v(x))$ if he remains in the organization after the trial period. In this case, c can be interpreted as the minimum level of competency that is required for the agent's continued employment to be productive to the principal. This interpretation also partially fits the pass/fail and licensing examples in the introduction. However, in such cases the principal often controls the bonus at best in an indirect way. Section 6.1 discusses the possibility of exogenous bonuses.

The first case in Proposition 9 does not rule out the CDFC. Thus, in the absence of a budget constraint, the existence problem may seemingly raise its head. However, since $\pi(a,t) \rightarrow 0$ as $t \rightarrow \overline{x}$, large thresholds are unlikely to be optimal. Hence, the properties of $\pi(a,t)$ may on their own be enough to ensure the existence of a solution to the second-best problem. Thus, in the next result, it is assumed only that a solution to the second-best problem exists, but it is left unspecified whether this is due to a budget constraint, the implementation constraint, or the properties of $\pi(a,t)$. Indeed, the result is driven only by the interaction between a and t as benefits and implementation costs are traded off.

Any departure from (a^{FB}, t^{FB}) strictly lowers social surplus. Likewise, under the MLRP, any incentive compatible contract that weakly increases a and/or weakly decreases t makes the agent at least weakly better off, by Propositions 1 and 2. Thus, such a contract leaves strictly less surplus to the principal than she would get from inducing the first-best. In other words, the second-best cannot have both a larger a and a smaller t than the first-best.

Corollary 2 *Assume the MLRP holds. Assume that the first-best solution* (a^{FB}, t^{FB}) *is unique, interior, and feasible in the second-best problem. Assume that a second-best solution* (a^{SB}, t^{SB}) *exists and that it is different from the first-best.*[16] *Then, the second-best features either a strictly higher threshold than the first-best* $(t^{SB} > t^{FB})$*, a strictly lower action* $(a^{SB} < a^{FB})$*, or both.*

It is important to realize that both the action and the criterion for success are distorted. Hence, the moral hazard problem implies a welfare loss along both dimen-

[16] The first-best and second-best may coincide in special cases, such as when $\pi(a,x)$ is discontinuous at (a^{FB}, x^{FB}).

sions. The next example is a version of the salesman example and shows that *both* the action and the threshold may be distorted downwards.

EXAMPLE 4: Assume that $\pi(a,t) = t(1 - F(t|a))$ and that $F(x|a) = 1 - e^{-\frac{x}{a^\beta}}$, $x \in [0,\infty)$, $a \in [0,1]$, and $\beta \in (0,1)$. Thus, the agent's performance is exponentially distributed with mean $h(a) = a^\beta$. Note that F is regular and degenerate at $a = \underline{a} = 0$. These functional-form assumptions make it possible to solve the first-best and second-best problems analytically. The details are in Appendix B, which in fact outlines a solution procedure for any concave $h(a)$ function for which $h(0) = 0$ and F is regular.

In the first-best problem,

$$a^{FB} = \left(\beta e^{-1}\right)^{\frac{1}{1-\beta}} \text{ and } t^{FB} = \left(a^{FB}\right)^\beta = \left(\beta e^{-1}\right)^{\frac{\beta}{1-\beta}}.$$

Thus, t^{FB} equals the mean performance, in equilibrium. Likewise, $1 - F(t^{FB}|a^{FB}) = e^{-1}$ regardless of β. Thus, the probability that the agent succeeds is always the same.

In the second-best problem, the boundary of the feasible set is described by $\bar{t}^C(a) = \frac{1}{\beta}a^\beta$, thus confirming that the first-best solution is feasible in the second-best problem. The second-best solution depends qualitatively on the size of β. If β is below $\frac{\sqrt{5}-1}{2} = 0.618$ then the solution is in the interior of the feasible set. For higher β values, the solution is on the boundary of the feasible set. In particular,

$$a^{SB} = \begin{cases} \left((1+\beta)^2 \beta^2 e^{-(1+\beta)}\right)^{\frac{1}{1-\beta}} & \text{if } \beta \in (0, \frac{\sqrt{5}-1}{2}] \\ e^{-\frac{1}{\beta(1-\beta)}} & \text{if } \beta \in (\frac{\sqrt{5}-1}{2}, 1) \end{cases}$$

and

$$t^{SB} = \begin{cases} (1+\beta)\left(a^{SB}\right)^\beta & \text{if } \beta \in (0, \frac{\sqrt{5}-1}{2}] \\ \frac{1}{\beta}\left(a^{SB}\right)^\beta & \text{if } \beta \in (\frac{\sqrt{5}-1}{2}, 1) \end{cases}.$$

Note that $1 + \beta \leq \frac{1}{\beta}$ if and only if $\beta \leq \frac{\sqrt{5}-1}{2}$.

It can be verified that $a^{SB} < a^{FB}$ for all $\beta \in (0,1)$. On the other hand, $t^{SB} < t^{FB}$ if $\beta < 0.492$ and $t^{SB} > t^{FB}$ if $\beta > 0.492$. Thus, the threshold is distorted below the first-best if β is small. It is also when β is small that the second-best solution is in the interior of the feasible set and therefore that $t^{SB} < \bar{t}^C(a^{SB})$. The reason is that the agent is more productive the smaller β is. Then, it is relatively more important for the principal to manipulate $\pi(a,t)$ than $W(a,t)$.

The probability that the agent succeeds is $1 - F(t^{SB}|a^{SB}) = \min\{e^{-(1+\beta)}, e^{-\frac{1}{\beta}}\}$. This is u-shaped in β and minimized at $\beta = \frac{\sqrt{5}-1}{2}$. Since $1 - F(t^{SB}|a^{SB}) < e^{-1}$, the agent succeeds less often in the second-best than in the first-best despite the fact that the threshold may be lower in the second-best. ▲

Example 4 demonstrates the possibility that $t^{SB} < t^{FB}$. Nevertheless, it is a fairly general conclusion that the second-best threshold is higher than the socially optimal threshold that implements a^{SB} (rather than a^{FB}). Thus, the threshold is too stringent for the action that is actually taken in equilibrium.

To be more precise, assume that for any action a, there is a unique and interior threshold that maximizes social surplus. Let $\widehat{t}(a) = \arg\max_t (\pi(a,t) - a)$ denote the threshold in question. Assume that $(a, \widehat{t}(a))$ is feasible in the second-best problem for any a. Then, $t^{SB} \geq \widehat{t}(a^{SB})$ as shown next. An implication is that the price in the salesman example is set above the monopoly price for the demand curve described by $1 - F(\cdot|a^{SB})$. For instance, in Example 4 it holds that $\widehat{t}(a) = a^{\beta} \leq \frac{1}{\beta}a^{\beta} = \overline{t}^C(a)$, implying that $(a, \widehat{t}(a))$ is feasible. Indeed, note that $\widehat{t}(a^{SB}) = \left(a^{SB}\right)^{\beta} < \min\{(1+\beta)\left(a^{SB}\right)^{\beta}, \frac{1}{\beta}\left(a^{SB}\right)^{\beta}\} = t^{SB}$ as claimed.

Corollary 3 *Assume that the MLRP holds. Assume that* $\widehat{t}(a) = \arg\max_t (\pi(a,t) - a)$ *is unique and interior for all* a *and that* $(a, \widehat{t}(a))$ *is feasible in the second-best problem for any* a*. Then,* $t^{SB} \geq \widehat{t}(a^{SB})$*.*

Example 5: The central argument in the proof of Corollary 3 relies only on the feasibility of $(a^{SB}, \widehat{t}(a^{SB}))$ in the second-best problem, but stating the condition that way is somewhat more obscure since a^{SB} is endogenous. The second specification in Proposition 9 illustrates the issue. Here $\widehat{t}(a) = c$ for all a. While $(a^{FB}, \widehat{t}(a^{FB})) = (a^{FB}, t^{FB})$ was shown to be feasible, it is not a given that $(a, \widehat{t}(a))$ is feasible for all a since $\overline{t}^C(a) < c$ is possible when a is small. Appendix B considers in detail an example in which $F(x|a) = 1 - e^{-\frac{x}{a^{\beta}}}$, $x \in [0, \infty)$, $a \in [0,1]$, and $\beta \in (0,1)$ as in Example 4, and where $\pi(a,t)$ is as in the second specification in Proposition 9, but with $v(x) = x - c$. It is shown that the second-best action is never in the range where $\overline{t}^C(a) < c$. Hence, $t^{SB} \geq c = \widehat{t}(a^{SB})$. ▲

5.2 Intrinsically important success probabilities

The distortions identified so far are driven by the principal's attempt to extract rent from the agent. However, there are situations in which the principal finds it optimal to concede rent to the agent compared to what is earned if the first-best action is implemented. To illustrate, assume in the following that the principal's benefit function takes the form

$$\pi(a,t) = \nu(a) + \kappa(F(t|a)), \tag{1}$$

where ν and κ are continuous and bounded functions. Here, $\nu(a)$ is some direct benefit deriving from the agent's action, such as $\nu(a) = \mathbb{E}[X|a]$. In contrast, $\kappa(F(t|a))$ describes an additional benefit that depends only on the probability that the agent fails or succeeds. If κ is constant, then the principal is not intrinsically interested in the criteria for success and the model reduces to the one studied in Sections 3 and 4.

The formulation in (1) is inspired by Li and Yang's (2020) monitoring problem. In their setting, monitoring is costly and modelled by partitioning $[\underline{x}, \overline{x}]$ into a number of performance categories. They assume that the principal wishes to induce the highest possible action. Li and Yang (2020) devote particular attention to the special case in which monitoring costs depend only on the number of categories. If the cost of additional categories is high enough, then two categories (success and failure) are optimal.[17] The results in Sections 3 and 4 then apply directly. Unlike Li and Yang (2020) these results allow a comparison of first-best and second-best actions.

However, Li and Yang (2020) also allow for monitoring cost functions that depend on the probability of success. This produces a benefit function as in (1). In fact, they assume that the naming of the categories do not matter for monitoring costs, which means that the cost is symmetric in the probability of success/failure. Moreover, $\kappa(F(t|a))$ is maximized when $F(t|a) = 0$ and when $F(t|a) = 1$, because in either case there is effectively only one performance category. Thus, there is a first-best solution with $t^{FB} = \underline{x}$ and another with $t^{FB} = \overline{x}$ (assuming the support of X is bounded). This is in contrast to the previous subsection, where t^{FB} is assumed to be unique and interior. In Li and Yang (2020), it is reasonable to assume that κ is u-shaped and minimized at $F(t|a) = \frac{1}{2}$, i.e. when the two categories are equally "large."

There are other applications of benefit functions of the form in (1). For instance, imagine that it is costly to process the payment of the bonus. Then, κ is increasing in its argument, because the larger the probability is that the agent fails, the less likely the principal is to have to incur the cost. Conversely, κ is decreasing if a failure means that the principal will have to incur fixed costs of restarting a research endeavour or

[17] Note that in Section 4.3.1 there is zero return to having more than two categories (Remark 1).

incur search costs to replace the agent. Finally, consider some professional body that controls the admission of candidates or apprentices into a "club" (e.g. a guild or other professional organization). This body may have in mind an ideal size, or pass-rate. In this case, t^{FB} is interior and κ is hump-shaped.

The first-best problem is to maximize

$$S(a,t) = \nu(a) - a + \kappa(F(t|a)).$$

Hence, whatever action is chosen, t must be calibrated to ensure that $F(t|a)$ achieves a value that maximizes $\kappa(F(t|a))$, if such a t value exists. Since $\sup_t \kappa(F(t|a))$ is independent of a, define the first-best action as $a^{FB} \in \arg\max_a (\nu(a) - a)$.

The second-best problem is to maximize

$$S(a,t) - U(a,t) = [\nu(a) - a - U(a,t)] + \kappa(F(t|a)).$$

Thus, there are two forces at play. The bracketed terms take the same form as the principal's problem in Sections 3 and 4. This describes the trade-off between the action and implementation costs. The last term captures the added consideration that comes from the principal's direct interest in the interaction between a and t.

To continue, assume that $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t in the interior of $\mathcal{I}$. Since $U(a,t)$ is strictly decreasing in t in the interior of $\mathcal{I}$, this assumption holds if κ is either increasing or if it is not too sensitive to changes in the probability of failure. Then, if a^{SB} is interior, the accompanying threshold must be the highest feasible threshold. Now, the results in Sections 3 and 4 reveal which direction along the boundary of the feasible set to move in for $\nu(a) - a - U(a,t)$ to increase. Recall that what motivates the principal to travel along the boundary is that doing so can lower the agent's rent, at the cost of distorting the action away from a^{FB}. However, moving in said direction may or may not increase $\kappa(F(t|a))$. If the two effects agree or if the first effect dominates, then the conclusions from Sections 3 and 4 obviously stand. Otherwise, it is optimal to move in the opposite direction to what is suggested in Sections 3 and 4. In such cases, the distortion away from a^{FB} increases the agent's rent and is therefore mutually beneficial to both parties.

Corollary 4 *Assume that $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t in the interior of $\mathcal{I}$. Assume that a^{FB} is unique and interior. Then, if a^{SB} is interior, it compares* *to* a^{FB} *as follows:*

1. *If the MLRP and CDFC hold and the principal faces a budget constraint, then either (i)* $a^{SB} \leq a^{FB}$ *or (ii) or* $(a^{SB}, \bar{t}^{B}(a^{SB}))$ *is a Pareto improvement over* $(a^{FB}, \bar{t}^{B}(a^{FB}))$*.*

2. *If* $F(x|a)$ *is regular then either (i)* $a^{SB} \geq a^{FB}$ *or (ii) or* $(a^{SB}, \bar{t}^{C}(a^{SB}))$ *is a Pareto improvement over* $(a^{FB}, \bar{t}^{C}(a^{FB}))$*. If* $F(x|\underline{a})$ *is degenerate then any interior second-best action maximizes* $S(a, \bar{t}^{C}(a))$*.*[18]

In comparison, if t is not intrinsically important to the principal then the agent is always worse off in the second-best when $a^{SB} \neq a^{FB}$ compared to the counterfactual situation where the principal induces a^{FB} in the most profitable manner. After all, the only reason for the principal to distort the action in such cases it to extract more rent from the agent.

Finally, note that when κ is increasing and $t^{FB} = \overline{x}$, it is necessarily the case that $t^{SB} < t^{FB}$ when a^{SB} is interior. Hence, it is possible that $a^{SB} > a^{FB}$ and $t^{SB} < t^{FB}$ at the same time. This is contrary to the conclusion that is obtained in Corollary 2, where it is assumed that the first-best is feasible in the second-best problem.[19]

The next example considers an environment consistent with Li and Yang (2020).

Example 6: Assume that $F(x|a)$ is a version of the Lomax distribution, with

$$F(x|a) = 1 - \left(1 + \frac{x}{a}\right)^{-2},\ x \in [0, \infty)$$

and let $\underline{a} = 1$. This distribution is regular and $\bar{t}^{C}(a) = \frac{1}{2}\left(a + \sqrt{a\,(a+8)}\right)$. It can be verified that $F(\bar{t}^{C}(a)|a)$ is decreasing in a and that it is always greater than $\frac{3}{4}$.

Consistent with Li and Yang's (2020) problem, assume that $\nu(a) = 4\sqrt{a}$ and that $\kappa(F(t|a)) = \varepsilon\left(F(t|a) - \frac{1}{2}\right)^2$, $\varepsilon \geq 0$. The first-best action is $a^{FB} = 4$. Turning to the second-best problem, $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t in the interior of $\mathcal{I}$ whenever $\varepsilon \leq 27$. However, $\kappa(F(\bar{t}^{C}(a)|a))$ is decreasing in a, to the principal's

[18] Note that $F(x|\underline{a})$ is degenerate and $F(\bar{t}^{C}(a)|a)$ is constant in (i) Example 2 when $\beta > 1$ and $\underline{a} = 0$ and in (ii) Example 4 when $h(a) = a^{\beta}$, $\beta \in (0,1)$ and $\underline{a} = 0$. Thus, $\kappa(F(t|a))$ does not change along the boundary of $\mathcal{I}$. In such cases, a^{SB} and a^{FB} coincide and the welfare cost of moral hazard comes only from the fact that t^{SB} is distorted away from t^{FB}.

[19] Corollaries 2 and 4 are not in general mutually exclusive. For instance, both may apply if κ is hump-shaped.

detriment. Thus, the principal has an incentive to lower a, but this competes with the opposing incentive to increase a in order to lower implementation costs. The latter effect dominates if and only if $\varepsilon < \frac{\sqrt{6075}+108}{23} = 8.084$. Thus, if ε is below this cut-off, then the second-best action exceeds the first-best action, or $a^{SB} > a^{FB}$. However, if ε is below the cut-off then $a^{SB} < a^{FB}$ and the agent is made better off. Thus, in this example, the higher $\varepsilon \in [0, 27]$ is – i.e. the more the principal intrinsically cares about the threshold – the less the agent works and the better off he is. ▲

Corollary 4 applies more broadly to benefit functions for which a^{FB} exists and $\pi(a,t) - U(a,t)$ is strictly increasing in t in the interior of $\mathcal{I}$. The functional form in (1) is used for concreteness and for the fact that a^{FB} as defined always exists.

6 Discussion

6.1 Exogenous thresholds or bonuses

Proposition 4 has some relevance to the literature that uses the first-order approach with a continuum of actions but two outcomes on the one hand, and the literature that assumes binary actions and two outcomes on the other hand.

Corollary 5 *Assume NUC_x holds. If $t', t \in (\underline{x}, \overline{x})$ and $t' > t$ then $A^C(t) = [\underline{a}, \overline{a}]$ if $A^C(t') = [\underline{a}, \overline{a}]$. In this case, the first-order approach is valid for any fixed threshold that is below t'. That is, the first-order condition is sufficient for incentive compatibility. Likewise, $A^C(t') = \{\underline{a}, \overline{a}\}$ if $A^C(t) = \{\underline{a}, \overline{a}\}$. In this case, for any fixed threshold that is above t, only $\underline{a}$ and $\overline{a}$ can be implemented and the model reduces to a binary action model with two outcomes.*

Thus, consider an environment with an exogenous threshold. If the threshold is small and it is easy to succeed, then the first-order approach is valid and a continuum of actions can be implemented. In contrast, if it is hard to succeed then only the two extreme actions can be implemented. Hence, there is a link between how stringent the criterion for success is and whether it is appropriate to model the agent as having effectively a continuum of actions or binary actions.

Imagine now instead that the bonus is fixed, but that the principal has flexibility to adjust the criterion for success. If the bonus is fixed at $\overline{b}$, the restriction that $B(a,t) = \overline{b}$ must now be respected and the set of feasible (a,t) shrinks as a result.

In comparison, Section 3 assumed that $B(a,t) \leq \bar{b}$ but showed that $B(a,t) = \bar{b}$ is optimal when the CDFC is satisfied and the principal does not intrinsically care about t. Thus, Proposition 3 is robust and it is still the case that the second-best action is distorted downwards when the bonus is exogenously fixed.

For any given a, there are two thresholds that exactly satisfy $B(a,t) = \bar{b}$ (see the proof of Lemma 1) Proposition 3 is concerned only with the higher of these thresholds, $\bar{t}^B(a)$. The conclusion in Proposition 3 relies on the fact that $U(a, \bar{t}^B(a))$ is increasing in a, so a downwards distortion in a is required to extract more rent from the agent. However, for the smaller threshold, $\underline{t}^B(a)$, it can be verified that $U(a, \underline{t}^B(a))$ is decreasing in a. Now note that if $F(x|a)$ is regular and a^{FB} is small, then $\bar{t}^B(a^{FB})$ may fall outside $T^C(a^{FB})$, in which case the principal is forced to use thresholds $\underline{t}^B(a)$ to induce actions close to a^{FB}. In this case, there is again an incentive to distort the action upwards. Hence, whether the conclusion in Proposition 8 survives depends on the size of a^{FB} compared to $\bar{b}$.

6.2 Observable performance and endogenous criteria

Demougin and Fluet (1998) consider a model with a *finite* number of signals, limited liability, and risk neutral parties. Assuming that the FOA is valid and that performance is perfectly observable, they show that there is no loss of generality in restricting the compensation structure to be binary. A bonus is paid if and only if the "most favorable" signal – the one with the highest likelihood-ratio – is realized. This event has a strictly positive probability of occurring in their finite-signal model. Thus, there is no existence problem.

Under the MLRP, the most favorable signal is the highest signal. Then, it is optimal to award a bonus if and only if the very highest signal is observed, regardless of which action is to be induced.[20] This situation is therefore similar to the case of an exogenous threshold in the present model. The same logic as in Section 2.2 then proves that the second-best action is distorted downwards.

Demougin and Fluet (1998) note that without the MLRP, the most favorable signal is generally not the highest signal. Then, the optimal contract is not monotonic

[20] Demougin and Fluet (2001) build on Demougin and Fluet (1998), while assuming that the MLRP and the CDFC hold. The principal can invest in different monitoring technologies, each leading to a different relationship between actions and the probability of realizing the highest signal. Then, the optimal monitoring system may depend on the action that is implemented.

in the signal. They also make the point that without the MLRP the most favorable signal may depend on the action that is to be induced. This observation is similar to how the criteria of success are sensitive to the action in Sections 3 and 4 of the current paper. However, Demougin and Fluet (1998) do not impose enough structure to say exactly how the most favorable signal depends on the action when the MLRP is violated, or how the second-best action compares to the first-best action.

In a related paper, Oyer (2000) assumes a continuum of observable signals, limited liability, risk neutrality, a slack participation constraint, and the additional constraint that compensation must be non-decreasing in the signal. Assuming the FOA is valid, the uniquely optimal contract (if one exists) consists of a threshold below which the limited liability constraint binds and above which a fixed bonus is paid. The contract therefore has the same structure as in the current paper. Acknowledging the existence problem, Oyer (2000) thus considers environments where the MLRP does not hold, in which case the criterion for success depends on the action that is to be implemented. In his examples, the second-best action is always below the first-best action.

Demougin and Fluet's (1998) and Oyer (2000) assume that the FOA is valid but the MLRP does not necessarily hold. In the current paper, the MLRP is maintained throughout, but the FOA is not assumed to be valid in Section 4. As discussed in Remark 1, threshold contracts are optimal when $F(x|a)$ is regular and $F(x|\underline{a})$ is degenerate, even when performance is fully observable. In such cases, $a^{FB} = a^{SB}$.

6.3 Granular failures

The analysis in Section 5 is under some conditions also still valid if the principal has granular information about failures, meaning that she can observe the agent's exact performance when he fails to meet the threshold. An application is the inventory problem studied by Chu and Lai (2013) and Dai and Jerath (2013). The principal is a retailer who decides upon the quantity of product to stock. Since a higher stock is costlier, the principal is intrinsically interested in its level and she would ideally like to not stock too much. She can observe how much of the stock the salesman or agent manages to sell. Thus, she has exact information if the agent fails to exhaust the stock. However, if the agent depletes the stock, the principal obviously cannot tell how much he would have been able to sell had the stock been higher. Thus, the principal has an incentive to increase the stock above the first-best level in order to gather more information and lower implementation costs. Dai and Jerath (2013) assume that there are just two effort levels. Chu and Lai (2013) assume that there is a continuum of actions and that the FOA is valid. They provide an example in which the second-best action is above the first-best action.

If the agent's first-order condition is sufficient for incentive compatibility, then the MLRP implies that the optimal contract is a do-or-die contract: A bonus is paid only if the entire stock is sold, whereas the minimum wage is paid if the agent fails (see Appendix D for details). In other words, the granular data about failures is not used. If the cost of stocking the product is sufficiently high, the optimal stock in both the first-best and second-best problem must be relatively small and, when the distribution is regular, the first-order condition is then sufficient for incentive compatibility. Hence, in this case the analysis is valid in the inventory problem as well.[21] If $F(x|\underline{a})$ is degenerate, then it is not even necessary for the cost of stocking the product to be high. After all, if the principal wishes to maintain a stock $t > \bar{t}^C(a^{SB})$, there is no cheaper way to incentivize the agent than to pay him a bonus if and only if sales exceeds $\bar{t}^C(a^{SB})$. In other words, the quota that triggers a bonus is below the stock level. This can also occur in Chu and Lai (2013), but only when the participation constraint is binding and prevents the quota from being raised.

These observations are also pertinent to the salesman example. Return to Example 4, where $F(x|\underline{a})$ is degenerate and $t^{SB} > \widehat{t}(a^{SB})$. Imagine that the principal is now able to infer the willingness-to-pay of the customer if the agent failed to make the sale, for instance because she undertakes a (presumably costly) survey or interview. If the action a^{SB} and price t^{SB} are such that $t^{SB} < \bar{t}^C(a^{SB})$, then, by the argument in the previous paragraph, the granular information is of no use in lowering implementation costs. However, the granular information may potentially be useful in writing a contract that is incentive compatibility even at prices $t^{SB} \geq \bar{t}^C(a^{SB})$. Nevertheless, using the logic in Remark 1, the implementation cost of any incentive compatible contract is at least a^{SB}, but this is exactly the cost that is incurred from using the price $\bar{t}^C(a^{SB})$ and ignoring the granular information failures. Thus, there can be no cost-saving and since t has moved even further away from $\widehat{t}(a^{SB})$, the principal's profit π also decreases. In sum, the principal has no use for granular information about failures and should not incur costs to collect it.

[21] The FOA can be used to bound implementation costs. If the solution that is obtained in this manner is in $\mathcal{I}$, then it is the correct second-best solution.

7 Conclusion

This paper endogenizes the criteria for success that form the basis of the agent's compensation. These criteria are disciplined by the agent's underlying performance technology and possibly a budget constraint. Depending on whether the budget constraint or implementability constraints are more restrictive, the second-best action may be distorted upwards or downwards compared to the first-best, but it is typically the case that the criteria for success are in some sense too stringent. For instance, in situations such as the salesman example, the terms that are offered to the customer are distorted in order to make it cheaper to incentivize the agent, who acts as a middleman. Charging a price that is above the monopoly price makes it more difficult for the agent to sell the product and gives him stronger incentives to try harder. Thus, the distortion that comes from the moral hazard problem spills over into the market.

The FOA is not needed and many of the central results come from tackling the implementability problem in more generality. A key observation is that the second-best solution is often on the boundary of the feasible set, the shape of which depends on the performance technology. Thus, the solution is sensitive to the properties of the performance technology in a way that is absent when the FOA approach is valid.

A class of distributions with certain regularity properties are identified in which the second-best action is distorted upwards when the principal does not intrinsically care about the threshold. Central to this result is the fact that a "non-local" incentive constraint is binding in the second-best solution. Specifically, the agent is indifferent between taking the intended action and the very lowest action. A low action cannot be incentivized with a high threshold, since the bonus is then unlikely to be realized and the agent might as well shirk completely. Thus, to be able to maintain high standards – which tends to be cheaper for the principal – the agent must be incentivized to work harder, thereby explaining the upwards distortion in the action. This logic relies on the FOA being invalid and therefore demonstrates the need for further research into contracting without the FOA.

References

Athey, S., 2002, "Monotone Comparative Statics under Uncertainty," *The Quarterly Journal of Economics*, 117: 187–223

Bond, P. and A. Gomes, 2009, "Multitask principal–agent problems: Optimal contracts, fragility, and effort misallocation," *Journal of Economic Theory*, 144: 175–211.

Chade, H. and J.M. Swinkels, 2020, "The no-upward-crossing condition, comparative statics, and the moral-hazard problem," *Theoretical Economics*, 15: 445–476.

Chu, L. Y. and G. Lai, 2013, "Salesforce Contracting Under Demand Censorship," *Manufacturing & Service Operations Management*, 15: 320–334

Dai, T. and K. Jerath, 2013, "Salesforce Compensation with Inventory Considerations," *Management Science*, 59(11):2490–2501.

Demougin, D. and C. Fluet, 2001, "Monitoring versus incentives," *European Economic Review*, 45: 1741–1764.

Demougin, D. and C. Fluet, 1998, "Mechanism Sufficient Statistic in the Risk-Neutral Agency Problem," *Journal of Institutional and Theoretical Economics*, 154: 622–639.

Dye, R.A., 1986, "Optimal monitoring policies in agencies," *RAND Journal of Economics*, 17: 339–350.

Gürtler, O. and M. Kräkel, 2010, "Double-Sided Moral Hazard, Efficiency Wages, and Litigation," *The Journal of Law, Economics, and Organization*, 26: 337–364.

Innes, R.D., 1990, "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory*, 52: 45–67.

Jewitt, I., O. Kadan and J.M. Swinkels, 2008, "Moral hazard with bounded payments," *Journal of Economic Theory*, 143: 59–82.

Kim, S.K., 1995, "Efficiency of an Information System in an Agency Model," *Econometrica*, 63: 89–102.

Kirkegaard, R., 2017, "Moral Hazard and the Spanning Condition without the First-Order Approach", *Games and Economic Behavior*, 102: 373–387.

Lal, R., 1986, "Delegating Pricing Responsibility to the Salesforce," *Marketing Science*, 5: 159–168.

Li, A. and M. Yang, 2020, "Optimal incentive contract with endogenous monitoring technology," *Theoretical Economics*, 15: 1135–1173.

Mirrlees, J.A., 1999, "The Theory of Moral Hazard and Unobservable Behavior: Part I," *Review of Economic Studies*, 66, 3–21.

Mishra, B.K. and A. Prasad, 2004, "Centralized Pricing Versus Delegating Pricing to the Salesforce Under Information Asymmetry," *Marketing Science*, 23: 21–27.

Oyer, P., 2000, A Theory of Sales Quotas with Limited Liability and Rent Sharing, *Journal of Labor Economics*, 18: 405–426.

Poblete, J. and D. Spulber, 2012, "The form of incentive contracts: agency with moral hazard, risk neutrality, and limited liability," *RAND Journal of Economics*, 43: 215–234.

Rogerson, W.P., 1985, "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 53 (6): 1357–1367.

Appendix A: Omitted proofs

Proof of Proposition 1. Fix a threshold t. The statement is trivial if $t \in \{\underline{x}, \overline{x}\}$ since only $\underline{a}$ can be implemented in that case. Thus, assume that $t \in (\underline{x}, \overline{x})$ and assume that there are at least two implementable actions. Using a standard argument, compare two implementable actions, a and a', with $a' > a$. Incentive compatibility requires that

$$B(a', t)\,(1 - F(t|a')) - a' \geq B(a', t)\,(1 - F(t|a)) - a$$

if a' is induced and

$$B(a, t)\,(1 - F(t|a)) - a \geq B(a, t)\,(1 - F(t|a')) - a'$$

if a is induced. Combining the two yields

$$B(a', t)\,(F(t|a) - F(t|a')) \geq a' - a \geq B(a, t)\,(F(t|a) - F(t|a'))\,.$$

Since $F(t|a) - F(t|a') > 0$, it follows that $B(a', t) \geq B(a, t)$. Then, $1 - F(t|a') > 1 - F(t|a)$ implies that $W(a', t) > W(a, t)$. Finally, it follows from incentive compatibility and $B(a', t) \geq B(a, t)$ that

$$\begin{aligned} U(a', t) &= B(a', t)\,(1 - F(t|a')) - a' \\ &\geq B(a', t)\,(1 - F(t|a)) - a \\ &\geq B(a, t)\,(1 - F(t|a)) - a \\ &= U(a, t). \end{aligned}$$

This completes the proof. ■

Proof of Proposition 2. Since log-supermodularity survives integration, the MLRP implies that the distribution function $F(x|a)$ as well as the survival function $1 - F(x|a)$ are also strictly log-supermodular; see e.g. Athey (2002, Lemma 3). This observation is relevant because $W(a, t)$ can be written

$$W(a, t) = \left(\frac{\partial \ln (1 - F(t|a))}{\partial a} \right)^{-1}.$$

It follows immediately from simple differentiation that $W(a,t)$ is strictly decreasing in t. Thus, $U(a,t) = W(a,t) - a$ is strictly decreasing in t as well. ■

Proof of Lemma 1. The CDFC ensures that any (a,t) pair is incentive compatible (given the bonus is $B(a,t)$). At $\overline{a}$, the bonus $B(\overline{a},t)$ is the lowest possible bonus that is incentive compatible. Higher bonuses are also incentive compatible, but unnecessarily costly. Thus, Propositions 1 and 2 also hold when $a = \overline{a}$. Since the first-order approach is valid under the CDFC, the feasible set of implementable (a,t) pairs is restricted only by the budget constraint that $B(a,t) \leq \overline{b}$.

By Proposition 1, $B(a,t)$ is weakly increasing in a. Hence, if (a,t) satisfies the budget constraint then any (a',t) with $a' < a$ also satisfies the constraint. This proves that the set of actions that can be implemented is an interval of the form $[\underline{a}, \overline{a}^B]$.

The rest of the lemma is trivial if $\overline{a}^B = \underline{a}$, so assume in the remainder that the budget is high enough that $\overline{a}^B > \underline{a}$. As mentioned in the text, since $F_a(t|a) \to 0$ as $t \to \overline{x}$, thresholds that are close to $\overline{x}$ violate the budget constraint.[22] Thus, for any $a \in (\underline{a}, \overline{a}^B]$, there exists some $\overline{t}^B(a) \in (\underline{x}, \overline{x})$ for which $B(a, \overline{t}^B(a)) = \overline{b}$ and $B(a,t) > \overline{b}$ for all $t > \overline{t}^B(a)$. Then, Proposition 1 also implies that $\overline{t}^B(a)$ is weakly decreasing in a. The reason is that if $a' < a$ and $t > \overline{t}^B(a')$ then $\overline{b} < B(a',t) \leq B(a,t)$, meaning that thresholds above $\overline{t}^B(a')$ cannot be used to implement a.[23] ■

Proof of Proposition 3. The proposition is trivial if $a^{FB} > \overline{a}^B$. Thus, assume that $a^{FB} \leq \overline{a}^B$. To implement $a \in (\underline{a}, \overline{a}^B]$, the cheapest threshold is $\overline{t}^B(a)$ and the agent consequently earns utility $U(a, \overline{t}^B(a))$. However, since $\overline{t}^B(a)$ is weakly decreasing in a by Lemma 1, it follows from Propositions 1 and 2 that $U(a, \overline{t}^B(a))$ is weakly increasing in a. Then, starting from a^{FB} and the associated threshold $\overline{t}^B(a^{FB})$, an increase in a and associated decrease in t strictly decreases social surplus and at the same time weakly increases the agent's utility. Hence, the principal is unambiguously worse off. Thus, $a^{SB} > a^{FB}$ cannot be optimal, and it follows that $a^{SB} \leq a^{FB}$.[24] Then, if

[22] Incidentally, the same argument also implies that the budget constraint is violated by thresholds that are close to $\underline{x}$. The MLRP implies that the bonus $B(a,t)$ is u-shaped in t since $B_t(a,t) = \frac{f_a(t|a)}{F_a(t|a)^2}$ is first strictly negative and then strictly positive as t increases. Thus, for a given action a, the bonus $B(a,t)$ is minimized at the threshold $t = t_0(a)$ for which the likelihood-ratio is zero, or $f_a(t_0(a)|a) = 0$. Note that $\overline{a}^B$ satisfies $B(\overline{a}^B, t_0(\overline{a}^B)) = \overline{b}$ if $\overline{a}^B < \overline{a}$.

[23] In the special case where $F(t|a)$ is linear in a, $\overline{t}^B(a)$ is a constant because $B(a,t)$ is independent of a.

[24] If $\pi(a)$ is differentiable and $\overline{t}^B(a)$ is stricly decreasing in a, then $a^{SB} < a^{FB}$. Intuitively, a small distortion away from $(a^{FB}, \overline{t}^B(a^{FB})$ has no first-order effect on social surplus but it has a first-order

$a^{SB} > \underline{a}$ it must hold that $t^{SB} = \bar{t}^B(a^{SB}) \geq \bar{t}^B(a^{FB})$ since $\bar{t}^B(a)$ is weakly decreasing in a. ■

Proof of Lemma 2. Fix some $t \in (\underline{x}, \overline{x})$. Then, (a, t) is incentive compatible if and only if it is true that there is no profitable deviation, or

$$B(a,t)\,(1 - F(t|a)) - a \geq B(a,t)\,(1 - F(t|a')) - a'$$

for all $a' \in [\underline{a}, \overline{a}]$. For any $a \in (\underline{a}, \overline{a})$, the first-order condition dictates that the bonus is $B(a,t) = \frac{-1}{F_a(t|a)}$. Then, keeping in mind that $F_a(t|a) < 0$ when $t \in (\underline{x}, \overline{x})$, a rearrangement of the first condition yields

$$F(t|a) + (a' - a)\,F_a(t|a) \leq F(t|a')$$

for all $a' \in [\underline{a}, \overline{a}]$. Thus, the tangent line to $F(t|\cdot)$ through a must lie always below the function itself. This is the case if and only if $a \in A^C(t)$.

An alternative way to understand the intuition is as follows. First, since $1 - F^C(t|\cdot) \geq 1 - F(t|\cdot)$, the agent's expected utility is at least as high in an imaginary problem where his technology is described by $F^C(t|\cdot)$ rather than $F(t|\cdot)$. Moreover, expected utility is concave in the imaginary problem. Thus, if utility in the imaginary problem is maximized at some $a \in A^C(t)$ then it is maximized at the same action in the real problem.

For $a \in \{\underline{a}, \overline{a}\}$, note first that a zero bonus induces $\underline{a}$ regardless of the threshold. Likewise, for any $t \in (\underline{x}, \overline{x})$, a sufficiently high bonus makes the agent's expected utility globally increasing in a and therefore incentivizes action $\overline{a}$. ■

Proof of Proposition 4. Recall that $\underline{a}$ and $\overline{a}$ are always in $A^C(t)$. Hence, the result is trivial if $A^C(t') = \{\underline{a}, \overline{a}\}$. Thus, assume that there is some action $a^* \in (\underline{a}, \overline{a})$ that belongs to $A^C(t')$. Since $a^* \in A^C(t')$,

$$F(t'|a^*) + (a - a^*)\,F_a(t'|a^*) \leq F(t'|a)$$

or

$$F(t'|a^*) - F(t'|a) + (a - a^*)\,F_a(t'|a^*) \leq 0 \tag{2}$$

effect on the agent's expected utility.

for all $a \in [\underline{a}, \overline{a}]$. By contradiction, assume that $a^* \notin A^C(t)$. Then, there exists some $a \neq a^*$ for which

$$F(t|a^*) - F(t|a) + (a - a^*) F_a(t|a^*) > 0. \tag{3}$$

For any a that satisfies (2) and (3), combining the two inequalities leads to the conclusion that

$$-F_a(t'|a^*) (F(t|a^*) - F(t|a)) > -F_a(t|a^*) (F(t'|a^*) - F(t'|a)),$$

or

$$\int_a^{a^*} (F_a(t|a^*)F_a(t'|z) - F_a(t'|a^*)F_a(t|z)) \, dz > 0. \tag{4}$$

The integrand is weakly negative if and only if

$$\ln(-F_a(t|a^*)) + \ln(-F(t'|z)) \leq \ln(-F_a(t'|a^*)) + \ln(-F_a(t|z)). \tag{5}$$

Now recall that $t' > t$ and assume to begin that $a < a^*$, implying that $z \leq a^*$. Then, since (a^*, t') is the componentwise maximum and (z, t) the componentwise minimum, the supermodularity of $\ln(-F_a)$ implies that the inequality in (5) is satisfied. However, this contradicts (4). If $a > a^*$, then the inequality in (5) is reversed, but this again yields the contradiction to (4) that

$$-\int_{a^*}^{a} (F_a(t|a^*)F_a(t'|z) - F_a(t'|a^*)F_a(t|z)) \, dz < 0.$$

Hence, there can be no a for which (3) holds, given that (2) holds for all actions. Thus $a^* \in A^C(t')$ implies that $a^* \in A^C(t)$, or $A^C(t') \subseteq A^C(t)$. ■

Proof of Proposition 5. Fix an interior action a. Due to CAT, $F_{aa}(\cdot|a)$ must be negative when t is large enough. Such threshold are not incentive compatible and cannot implement a. In combination, NUC_x and CAT imply that $F_{aa}(\cdot|a)$ is either always negative or it is first-positive-then-negative in x.

Recall from Proposition 4 that if $t', t \in (\underline{x}, \overline{x})$ and $t' > t$ then $A^C(t') \subseteq A^C(t)$. Thus, if $a \notin A^C(t)$ then $a \notin A^C(t')$. In words, if some threshold t cannot implement a then no higher threshold works either.

Combining the two observations implies that if $T^C(a)$ is not empty then it must take the form $(\underline{x}, \overline{t}^C(a)]$, where $\overline{t}^C(a) < \overline{x}$. ■

Proof of Proposition 6. If $F_{aa}(t|\cdot) < 0$ for all a then $A^C(t) = \{\underline{a}, \overline{a}\}$. If $F_{aa}(t|\cdot) \geq 0$ for all a then $A^C(t) = [\underline{a}, \overline{a}]$. NDC$_a$ permits only one additional possibility, namely that $F_{aa}(t|\cdot)$ is first-negative-then-positive as a increases. In this case, $A^C(t)$ either consists only of $\{\underline{a}, \overline{a}\}$ or of $\underline{a}$ and an interval that extends to $\overline{a}$. In the latter case $A^C(t) = \{\underline{a}\} \cup [\underline{a}^C(t), \overline{a}]$, where $\underline{a}^C(t) \in (\underline{a}, \overline{a})$. More formally, if $a \in (\underline{a}, \overline{a})$ belongs to $A^C(t)$ then it holds that $F_{aa}(t|a) \geq 0$ and that

$$F(t|a) + (a' - a)\, F_a(t|a) \leq F(t|a')$$

for all $a' \in [\underline{a}, \overline{a}]$. The left hand side is decreasing in a when $F_{aa} \geq 0$. Hence, given NDC$_a$, if $a \in (\underline{a}, \overline{a})$ is in $A^C(t)$ then all higher actions are also in $A^C(t)$. ■

Proof of Corollary 1. This follows from combining the conclusion that $A^C(t)$ shrinks when t increases with the conclusion that $A^C(t) = \{\underline{a}\} \cup [\underline{a}^C(t), \overline{a}]$. ■

Proof of Proposition 7. Recall that $U(a, \overline{t}^C(a)) = -\underline{a}$, or

$$W(a, \overline{t}^C(a)) = a - \underline{a}$$

for $a \in (\underline{a}, \overline{a}]$. Thus, if action $a \in (a, \overline{a}]$ is implemented with threshold $\overline{t}^C(a)$, then the cost of implementation is $a - \underline{a}$. Similarly, $\underline{a}$ can be implemented with a zero bonus. At the other end of the support, for action $\overline{a}$ there is no benefit to making the threshold exceed $\overline{t}^C(\overline{a})$ since the indifference condition must also hold at such thresholds (the binding incentive compatibility constraint is the no-jump constraint to $\underline{a}$). Thus, implementation costs are continuous in a.

In summary, the principal's payoff is

$$\pi(a) - a + \underline{a},$$

which by definition is no greater than $\pi(a^{FB}) - a^{FB} + \underline{a}$. In this form, the problem is easy to solve. Simply induce action a^{FB} by picking the threshold $t = \overline{t}^C(a^{FB})$. Thus, the first-best action is implemented, and it is implemented with the highest feasible threshold. ■

Proof of Proposition 8. The proof begins by showing that $U(a, \overline{t}^C(a))$ is decreasing in a. Any point on the boundary of the feasible set, $(a, \overline{t}^C(a))$, is characterized by the condition that

$$F(t|a) - F(t|\underline{a}) + (\underline{a} - a)\, F_a(t|a) = 0.$$

From Corollary 1, $\bar{t}^C(a)$ is weakly increasing in a on $(\underline{a}, \bar{a})$. In fact, the slope equals

$$\frac{d\bar{t}^C(a)}{da} = -\frac{(\underline{a} - a)\, F_{aa}(t|a)}{f(t|a) - f(t|\underline{a}) + (\underline{a} - a)\, f_a(t|a)},$$

where $F_{aa}(t|a) \geq 0$ is necessary for $(a, \bar{t}^C(a))$ to be incentive compatible and where the denominator is non-negative because

$$\begin{aligned} f(t|a) - f(t|\underline{a}) + (\underline{a} - a)\, f_a(t|a) &= f(t|a) - f(t|\underline{a}) - \frac{F(t|a) - F(t|\underline{a})}{F_a(t|a)} f_a(t|a) \\ &= \int_{\underline{a}}^{a} F_a(t|z) \left(\frac{f_a(t|z)}{F_a(t|z)} - \frac{f_a(t|a)}{F_a(t|a)} \right) dz \end{aligned}$$

is non-negative since $F_a(t|z) < 0$ and NUC_x implies that $\frac{f_a(t|z)}{F_a(t|z)}$ is weakly increasing in z.

Since

$$\begin{aligned} U_a(a,t) &= \frac{F_{aa}(t|a)(1 - F(t|a))}{F_a(t|a)^2} \\ U_t(a,t) &= \frac{f(t|a)F_a(t|a) + f_a(t|a)\,(1 - F(t|a))}{F_a(t|a)^2} \end{aligned}$$

and utilizing $(\underline{a} - a)\, F_a(\bar{t}^C(a)|a) = F(\bar{t}^C(a)|\underline{a}) - F(\bar{t}^C(a)|a)$, it can now be seen that

$$\begin{aligned} \frac{dU(a, \bar{t}^C(a))}{da} &= U_a(a, \bar{t}^C(a)) + U_t(a, \bar{t}^C(a)) \frac{d\bar{t}^C(a)}{da} \\ &= \frac{(1 - F(t|\underline{a}))(1 - F(t|a))}{F_a(t|a)^2} \frac{F_{aa}(t|a)}{f(t|a) - f(t|\underline{a}) + (\underline{a} - a)\, f_a(t|a)} \\ &\quad \times \left[\frac{f(t|a)}{1 - F(t|a)} - \frac{f(t|\underline{a})}{1 - F(t|\underline{a})} \right], \end{aligned}$$

where t is evaluated at $t = \bar{t}^C(a)$. The term in front of the brackets is proportional to the slope of $\bar{t}^C(a)$, which is non-negative. By the MLRP, $1 - F(t|a)$ is log-supermodular in (a, t), implying that $\frac{f(t|a)}{1 - F(t|a)}$ is decreasing in a. Therefore, the term in the brackets is negative. Hence, $U(a, \bar{t}^C(a))$ is weakly decreasing in a.

The rest of the proof mirrors the argument in the proof of Proposition 3. The cheapest way to induce any $a > \underline{a}$ is to use threshold $\bar{t}^C(a)$. The optimal contract that induces a^{FB} gives the agent utility $U(a^{FB}, \bar{t}^C(a^{FB}))$. In comparison, inducing any $a \in (\underline{a}, a^{FB})$ weakly increases the agent's utility and strictly decreases social surplus, thus leaving strictly less surplus to the principal. Hence, the second-best action does not belong to $(\underline{a}, a^{FB})$. The rest of the proposition follows. ■

Proof of Proposition 9. Assume first that $\pi(a,t) = v(t)(1 - F(t|a))$. The first-best threshold, t^{FB}, is then between c and $\overline{x}$. Given t^{FB}, the first-best action, a^{FB}, then solves the problem

$$\max_a v(t^{FB})(1 - F(t^{FB}|a)) - a.$$

However, this is equivalent to maximizing the agent's utility with respect to a, for a fixed threshold t^{FB} and a fixed bonus, $b = v(t^{FB}) > v(c) = 0$. Hence, the optimal action must necessarily belong to $A^C(t^{FB})$. Thus, $(a^{FB}, t^{FB}) \in \mathcal{I}$. Note that this does not require F to be regular.

Assume next that $\pi(a,t) = \int_t^{\overline{x}} v(x) f(x|a)dx$ and that F is regular. Note that the first-best threshold is $t^{FB} = c$. Any first-best action, a^{FB}, solves

$$\max_a \int_c^{\overline{x}} v(x) f(x|a)dx - a$$

or

$$\max_a \int_c^{\overline{x}} v'(x)(1 - F(x|a))\, dx - a.$$

As always, the first-best (a^{FB}, t^{FB}) is implementable in the second-best problem if a^{FB} is at one of the corners. Thus, consider the more interesting case in which a^{FB} is interior. The first- and second-order conditions are

$$\begin{aligned} -\int_c^{\overline{x}} v'(x) F_a(x|a^{FB})dx &= 1 \\ -\int_c^{\overline{x}} v'(x) F_{aa}(x|a^{FB})dx &\leq 0. \end{aligned}$$

Recall that F is regular. By NUC_x, the second-order condition necessitates that

$F_{aa}(c|a^{FB}) \geq 0$. By definition,

$$\int_c^{\overline{x}} v'(x)\left(1 - F(x|a^{FB})\right) dx - a^{FB} \geq \int_c^{\overline{x}} v'(x)\left(1 - F(x|a)\right) dx - a \text{ for all } a \in [\underline{a}, \overline{a}]$$

Using the first-order condition, this can be rewritten as

$$\int_c^{\overline{x}} v'(x)\left(F(x|a) - F(x|a^{FB})\right) dx - \left(a - a^{FB}\right) \int_c^{\overline{x}} v'(x) F_a(x|a^{FB}) dx \geq 0 \text{ for all for all } a \in [\underline{a}, \overline{a}],$$

or

$$\int_c^{\overline{x}} v'(x)\left(F(x|a) - \left(F(x|a^{FB}) + \left(a - a^{FB}\right) F_a(x|a^{FB})\right)\right) dx \geq 0 \text{ for all } a \in [\underline{a}, \overline{a}]. \tag{6}$$

Thus, the tangent line to $F(x|\cdot)$ through a^{FB} is "in expectation" below the function $F(x|\cdot)$ at any possible alternative action. Now, (a^{FB}, t^{FB}) is implementable in the second-best problem if and only if

$$F(c|a') - \left(F(c|a^{FB}) + \left(a' - a^{FB}\right) F_a(c|a^{FB})\right) \geq 0 \text{ for all } a \in [\underline{a}, \overline{a}]. \tag{7}$$

Thus, assume to the contrary that there exist some $a' \neq a^{FB}$ such that

$$F(c|a') - \left(F(c|a^{FB}) + \left(a' - a^{FB}\right) F_a(c|a^{FB})\right) < 0.$$

Since $F_{aa}(c|a^{FB}) \geq 0$ it holds by NDC$_a$ that $F_{aa}(c|a) \geq 0$ for all $a \geq a^{FB}$. This rules out that $a' > a^{FB}$. Since $a' < a^{FB}$, NDC$_a$ further implies that

$$F(c|\underline{a}) - \left(F(c|a^{FB}) + \left(\underline{a} - a^{FB}\right) F_a(c|a^{FB})\right) < 0.$$

In words, (a^{FB}, c) is not implementable because the agent could profitably deviate to $\underline{a}$. Indeed, regularity implies that if the threshold increases from c to some higher level, then the new contract is also not implementable because a deviation to $\underline{a}$ remains profitable. That is,

$$F(x|\underline{a}) - \left(F(x|a^{FB}) + \left(\underline{a} - a^{FB}\right) F_a(x|a^{FB})\right) < 0 \text{ for all } x \in [c, \overline{x}).$$

However, this violates (6). Thus, (7) must hold and it follows that $(a^{FB}, t^{FB}) \in \mathcal{I}$. ■

Proof of Corollary 2. The corollary follows from the argument in the text that $t^{SB} \leq t^{FB}$ and $a^{SB} \geq a^{FB}$ cannot be jointly optimal, as a consequence of Propositions 1 and 2. ■

Proof of Corollary 3. Given action a^{SB}, threshold $t = \widehat{t}(a^{SB})$ dominates any $t < \widehat{t}(a^{SB})$ from the principal's point of view, since the latter has lower social surplus and gives more rent to the agent. Hence, $t^{SB} \geq \widehat{t}(a^{SB})$. ■

Proof of Corollary 4. Assume that the MLRP and CDFC hold and the principal faces a budget constraint. If $a^{SB} > a^{FB}$ then $U(a^{SB}, \bar{t}^B(a^{SB})) \geq U(a^{FB}, \bar{t}^B(a^{FB}))$, by the arguments in the proof of Proposition 3. Hence, the agent is better off. Since $(a^{FB}, \bar{t}^B(a^{FB}))$ is feasible, $(a^{SB}, \bar{t}^B(a^{SB}))$ is in the principal's interest only if it makes her weakly better off. Hence, $(a^{SB}, \bar{t}^B(a^{SB}))$ is a Pareto improvement over $(a^{FB}, \bar{t}^B(a^{FB}))$.

Assume next that $F(x|a)$ is regular. By the proof of Proposition 8, $U(a^{SB}, \bar{t}^C(a^{SB})) \geq U(a^{FB}, \bar{t}^C(a^{FB}))$ if $a^{SB} < a^{FB}$, and it follows by the same arguments as above that $(a^{SB}, \bar{t}^C(a^{SB}))$ is a Pareto improvement over $(a^{FB}, \bar{t}^C(a^{FB}))$. If $F(x|\underline{a})$ is degenerate, then $U(a, \bar{t}^C(a))$ is constant by the argument leading to Proposition 7. Hence, $\pi(a, \bar{t}^C(a)) - W(a, \bar{t}^C(a))$ is proportional to $S(a, \bar{t}^C(a))$ and the last part of the corollary follows. ■

Proof of Corollary 5. The corollary follows from Proposition 4. ■

Appendix B: Details of Examples 4 and 5

Details of Example 4: Assume that $\pi(a,t) = t(1-F(t|a))$. To begin, consider a more general specification of the distribution function than in the main text. In particular, assume that the agent's performance is exponentially distributed with mean $h(a)$, where $h'(a) > 0$, $h''(a) < 0$, and $h(0) = 0$. Thus, $F(x|a) = 1 - e^{-\frac{x}{h(a)}}$, $x \in [0,\infty)$, and where a belongs to an interval of the form $[0,\overline{a}]$. The fact that $h'(\cdot) > 0$ implies that the MLRP and the NUC$_x$ hold.

Assume that $h(\overline{a})e^{-1} \leq \overline{a}$. Since

$$\max_t \pi(\overline{a},t) = h(\overline{a})e^{-1},$$

the assumption implies that $S(\overline{a},t) < 0$ for all t. Thus, social surplus from $\overline{a}$ is smaller than social surplus from $\underline{a}$. This in turn means that $\overline{a}$ cannot be optimal in the first-best or second-best problems. This corner can therefore be ignored. Assume also that $h'(0) > e$. This implies that the first-best action exceeds $\underline{a}$ (see the next paragraph). These assumptions are satisfied if $h(a) = a^\beta$, $\beta \in (0,1)$ and $\overline{a} > e^{-1}$.

The first-best problem is to maximize $\pi(a,t) - a$ or

$$\max_{a,t} te^{-\frac{t}{h(a)}} - a.$$

The necessary first-order condition for t^{FB} reveals that $\frac{t^{FB}}{h(a^{FB})} = 1$. This implies that the agent succeeds with probability $1 - F(t^{FB}|a^{FB}) = e^{-1}$ regardless of the functional form of h. Utilizing $t^{FB} = h(a^{FB})$ in the first-order condition for a^{FB} yields the conclusion that a^{FB} is uniquely determined by $h'(a^{FB}) = e$, which in turn pins down $t^{FB} = h(a^{FB})$.

Turning to the implementability problem, note that

$$\begin{aligned} F_{aa}(x|a) &= \left[-\frac{d}{da}\left(\frac{h'(a)}{h(a)^2}\right) - x\left(\frac{h'(a)}{h(a)^2}\right)^2\right] xe^{-\frac{x}{h(a)}} \\ &= \left[\frac{d}{da}\left(\frac{h(a)^2}{h'(a)}\right) - x\right]\left(\frac{h'(a)}{h(a)^2}\right)^2 xe^{-\frac{x}{h(a)}}. \end{aligned}$$

The first term inside the brackets is positive, by concavity of h. Thus, $F_{aa}(\cdot|a)$ changes sign as x increases. Hence, the CDFC is not satisfied, but NUC$_x$ and CAT are. NDC$_a$

holds if and only if the first term is increasing in a, or in other words if and only if $\frac{h(a)^2}{h'(a)}$ is convex. It can be verified that this holds true if $h(a) = a^\beta$, $\beta \in (0, 1)$, and also if $h(a) = \ln(1 + a)$ or $h(a) = 1 - e^{-a}$.

Proceeding under the assumption that NDC_a is satisfied, any (a, t) on the boundary of the feasible set is characterized by

$$F(t|0) = F(t|a) + (0 - a)F_a(t|a).$$

Utilizing $F(t|0) = 1$, this can be solved for

$$\bar{t}^C(a) = \frac{h(a)^2}{ah'(a)},$$

which is $\bar{t}^C(a) = \frac{1}{\beta}a^\beta$ if $h(a) = a^\beta$, $\beta \in (0, 1)$.

The principal's second-best problem is to maximize $\pi(a, t) - W(a, t)$, or

$$\max_{a,t} te^{-\frac{t}{h(a)}} - \frac{h(a)^2}{h'(a)}\frac{1}{t},$$

subject to feasibility. It is surprisingly easy to solve the first-order conditions simultaneously if the feasibility constraint is ignored. Each first-order condition can be solved for $e^{-\frac{t}{h(a)}}$. Equating these expressions and simplifying yields an equation that is linear in t but non-linear in a. Thus, it is easy to solve for t, for any given a. Substituting this back into one of the first-order conditions then makes it possible to solve (either numerically or analytically) for a, and with it the accompanying t value. Once a solution has been obtained, it can then be verified whether it satisfies the feasibility constraint. If it does not, then the second-best solution must be on the boundary of the feasible set. In this case, the second-best solutions takes the form $(a^{SB}, \bar{t}^C(a^{SB}))$, where a^{SB} solves

$$\max_a \pi(a, \bar{t}^C(a)) - W(a, \bar{t}^C(a)).$$

Note that the main role of the $h(0) = 0$ assumption is to provide a convenient analytical characterization of $\bar{t}^C(a)$. However, recall that since $F(\cdot|0)$ is degenerate in this case, $W(a, \bar{t}^C(a))$ equals the first-best implementation costs, or $W(a, \bar{t}^C(a)) = a$ (see the discussion leading up to Proposition 7).

Applying the procedure to the example where $h(a) = a^{\beta}$, $\beta \in (0,1)$, yields the analytical solution in the main body of the text. ▲

Details of Example 5: Assume that $F(x|a) = 1 - e^{-\frac{x}{a^{\beta}}}$, $x \in [0, \infty)$, $a \in [0,1]$, and $\beta \in (0,1)$ as in Example 4, and that $\pi(a,t) = \int_t^{\overline{x}} v(x) f(x|a) dx$, with $v(x) = x - c$ for some $c \in (0, \infty)$. From Example 4, $\overline{t}^C(a) = \frac{1}{\beta} a^{\beta}$. Note that $\overline{t}^C(a) < c$ if a is small.

Recall that $\pi(a,t)$ is increasing in t for $t < c$ and that $W(a,t)$ is globally decreasing in t on the feasible set. Thus, given some interior second-best action, a^{SB}, the second-best threshold must be no smaller than c whenever such a threshold is feasible. The only way a smaller threshold can be optimal is when $\overline{t}^C(a) < c$, i.e. when a^{SB} is small. It will now be shown that the second-best action cannot be interior and in this range.

Thus, consider implementing an action for which $\overline{t}^C(a) < c$, or $a < (\beta c)^{\frac{1}{\beta}}$. As mentioned, the optimal threshold is then $t = \overline{t}^C(a)$. Thus, wage costs are $W(a, \overline{t}^C(a)) = a$, while

$$\begin{aligned} \pi(a, \overline{t}^C(a)) &= \int_{\overline{t}^C(a)}^{\infty} (x - c) \frac{1}{a^{\beta}} e^{-\frac{x}{a^{\beta}}} dx \\ &= e^{-\frac{1}{\beta}} \left(\frac{1+\beta}{\beta} a^{\beta} - c \right). \end{aligned}$$

The principal's expected payoff is

$$\pi(a, \overline{t}^C(a)) - W(a, \overline{t}^C(a)) = e^{-\frac{1}{\beta}} \left(\frac{1+\beta}{\beta} a^{\beta} - c \right) - a,$$

which is evidently concave in a. The first derivative is

$$\frac{d\left(\pi(a, \overline{t}^C(a)) - W(a, \overline{t}^C(a))\right)}{da} = e^{-\frac{1}{\beta}} (1+\beta) a^{\beta - 1} - 1,$$

which is positive when a is small. It is increasing in a for all $a < (\beta c)^{\frac{1}{\beta}}$ if c is so small that $c < \frac{1}{\beta} \left(\frac{1}{1+\beta} \right)^{\frac{\beta}{\beta-1}} e^{\frac{1}{\beta-1}} = \overline{c}$. In this case, it cannot be optimal to induce an action $a < (\beta c)^{\frac{1}{\beta}}$, since inducing a marginally higher action leads to higher expected payoff.

Thus, assume that c is large, or $c \geq \bar{c}$. Then, the first-order condition is satisfied at

$$a^* = \left(\frac{e^{\frac{1}{\beta}}}{1+\beta} \right)^{\frac{1}{\beta-1}},$$

at which point expected profit is

$$\begin{aligned} \pi(a^*, \bar{t}^C(a^*)) - W(a^*, \bar{t}^C(a^*)) &= e^{-\frac{1}{\beta}} \left((1-\beta) \frac{1}{\beta} \left(\frac{1}{\beta+1} \right)^{\frac{1}{\beta-1}} e^{\frac{1}{\beta-1}} - c \right) \\ &= e^{-\frac{1}{\beta}} \left((1-\beta) \bar{c} - c \right), \end{aligned}$$

but this is negative for all $c \geq \bar{c}$. Hence, this cannot be part of the second-best solution because inducing $a = 0$ gives zero payoff to the principal. ▲

Appendix C: Non-regular distribution functions

EXAMPLE 7: Assume that $F(x|a)$ is the normal distribution with variance σ^2 and mean $h(a)$, with $h'(a) > 0$ and $h''(a) \leq 0$. An implication of $h'(a) > 0$ is that the MLRP and the NUC_x are satisfied. The sign of $F_{aa}(x|a)$ is determined by the sign of $\gamma(a, x) = \frac{h(a)-x}{\sigma} - \frac{h''(a)}{h'(a)^2}$. The sign depends on x, implying that the CDFC is not satisfied. However, CAT is satisfied. Whether NDC_a is satisfied depends on $h(a)$ and possibly σ. It is sufficient that $\gamma(a, x)$ is increasing in a for the NDC_a to hold. This is the case if $h(a) = k - e^{-a}$ for some $k \in \mathbb{R}$. Note that if $\gamma(a, x)$ is increasing in a for some σ, then this remains the case as σ decreases. Thus, the NDC_a is more likely to hold the less noisy the distribution is.

Next, assume that $\sigma^2 = 1$ and $h(a) = \sqrt{a}$, $a \in [0, 4]$. Then, $\gamma(a, x) = \frac{1}{\sqrt{a}} - x + \sqrt{a}$. For any $x \in \mathbb{R}$, this is minimized where $a = 1$ and it is therefore no smaller than $2 - x$. Consequently, if $t \leq 2$ then $F_{aa}(t|\cdot) \geq 0$ for all $a \in [0, 4]$ and all actions can thus be implemented. However, if $x > 2$ then $F_{aa}(x|\cdot)$ changes sign. For instance, $F_{aa}(2.1|\cdot)$ is zero at $a = 0.5327$ and $a = 1.8773$, and is positive-negative-positive as a increases. The NDC_a does not hold in this case. Indeed, it can be verified that the set of implementable actions with threshold $t = 2.1$ is $A^C(2.1) = [0, 0.2776] \cup [2.6013, 4]$. Thus, there is a "hole" or gap in the set of actions that can be implemented. Proposition 4 thus implies that $\bar{t}^C(a)$ is u-shaped in a. Finally, if $t \geq 2.5$ then $F_{aa}(t|\cdot)$ is first-positive-then-negative. Then, the set of implementable actions consists of $\overline{a}$ and a set of action close to (and including) $\underline{a}$. This is a mirror image of the conclusion in the third part of Proposition 6, which assumes NDC_a.

The $h(a) = \sqrt{a}$ setting is interesting for a couple of reasons. First, $h'(0) = \infty$, meaning that the marginal return in the agent's expected performance to a small increase in his action starting from zero is infinite. This may or may not be realistic. Second, the model is isomorphic to a setting in which $h(a) = a$ but where the agent's cost function is $c(a) = a^2$ rather than a. The latter specification is common in e.g. the literature on rank-order tournaments. ▲

Appendix D: Contracting with granular failures

Fix an interior action a to implement and an interior threshold t. With granular failures, performances below t can be observed. Thus, for performance $x < t$, let $w(x)$ denote the wage that is paid. Let b denote the wage that is paid if the threshold is met. The principal's problem is then to manipulate $w(x)$ and b to minimize

$$W = \int_{\underline{x}}^{t} w(x) f(x|a) dx + b(1 - F(t|a)),$$

subject to incentive compatibility and the limited liability constraint. For incentive compatibility, the agent's necessary first-order condition implies that

$$bF_a(t|a) = \int_{\underline{x}}^{t} w(x) f_a(x|a) dx - 1,$$

which means that

$$\begin{aligned} W &= \int_{\underline{x}}^{t} w(x) f(x|a) dx + \frac{1 - F(t|a)}{F_a(t|a)} \left(\int_{\underline{x}}^{t} w(x) f_a(x|a) dx - 1 \right) \\ &= \int_{\underline{x}}^{t} w(x) \left(1 + \frac{1 - F(t|a)}{F_a(t|a)} \frac{f_a(x|a)}{f(x|a)} \right) f(x|a) dx - \frac{1 - F(t|a)}{F_a(t|a)}. \end{aligned}$$

Since $F_a(t|a) < 0$, the MLRP implies that the sum inside the parenthesis is minimized at $x = t$. Moreover, using the fact that $1 - F(t|a)$ is log-supermodular due to the MLRP, the sum can easily be shown to be positive at $x = t$. Thus, it is positive for any $x < t$. Since the objective is to minimize W, it is therefore optimal to minimize $w(x)$ by equating it to the minimum wage. Thus, the candidate for the optimal contract entails a flat wage for any $x < t$ and a bonus that is paid only if the threshold is met. This is a candidate contract in the usual sense that the agent's first-order condition is necessary but not sufficient for full incentive compatibility.