Endogenous Criteria for Success*

René Kirkegaard†

July 2022

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. 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. However, given the second-best action, the criteria for success are as stringent as possible. 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, a regulator dictates what it takes to pass a driver's license test, but leaves it to a certified middleman at a test centre to determine if the applicant meets those criteria. Likewise, 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. Finally, in up-or-out systems, the organization simply decides whether to retain the employee or not.

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 "mi-

crofoundation" 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 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 small or very large require substantial bonuses to be incentive compatible. The budget constraint rules out such thresholds.

The second version departs from the traditional approach by focusing squarely

[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.

on the "implementability constraint" by studying environments where the FOA is not valid. While the FOA simplifies the incentive compatibility problem, it requires strong assumptions. It is even less desirable in the current environment, where the probability of success is linked to the underlying performance technology. Without the FOA, there are combinations of actions and thresholds that simply cannot be implemented because no incentive compatible contract exists. Fortunately, it turns out to be 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 properties is identified and the characteristics of the resulting feasible set are described.

In both versions, the optimal threshold is higher than the threshold that optimally implements the first-best action. This is due to a fundamental feature of the model that more stringent criteria for success tend to lower incentive costs. The principal therefore has an incentive to distort the action in such a way that higher thresholds become feasible. However, the two versions of the model differ in how this is achieved. In the first version, the second-best action is lower than the first-best action but this is reversed in the second model under certain regularity assumptions. In either case, the feasibility constraint binds at the second-best solution. The two versions give different conclusions because the feasible sets have dramatically different shapes. This is a stark illustration of the critical role of the FOA in the existing literature.

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, even when allowing for both budget constraints and implementability constraints. 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.

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, Li and Yang (2020) simply assume that the principal wishes to induce the highest possible action. Thus, they do not study how the first-best and second-best 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 in their setting. 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.

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). The FOA is generally not valid in the present paper either. In fact, the problem 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.

The contracting problem in Bond and Gomes (2009) results in two kinds of inefficiencies: A distortion in the total amount of effort, and a distortion in how this total amount is distributed among tasks. They 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; the second-best effort must be no higher than the first-best. Once again, no such assumption is made in this paper. As for the second distortion, the agent in Bond and Gomes (2009) concentrates too much of the total effort on a small number of tasks. In the model presented here, the second dimension of inefficiency comes from the fact that the criteria for success are distorted. As alluded to earlier, when the criterion for success is intrinsically important to the principal, she often makes the criterion more stringent than she would ideally like, since this turns out to make it cheaper to induce effort.

In some ways, the problem in the current paper is simpler than in Li and Yang (2020) and in Bond and Gomes since only partitions into successes and failures are allowed and there is only one task. On the other hand, the implementability problem is tackled in more generality and distortions from the first-best are treated more carefully as it is not assumed that distortions in effort can only be downwards.

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, the support is the same interval for all $a \in (\underline{a}, \overline{a}]$, and the density $f(x|a) = F_x(x|a)$ is strictly positive on this interval, $[\underline{x}, \overline{x}]$, which may be bounded or unbounded.[2] 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 special 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 $a \in [\underline{a}, \overline{a}]$ in case (i) and for all $a \in (\underline{a}, \overline{a}]$ in case (ii). Throughout, it is assumed that $F_a(x|a) < 0$ for all $x \in (\underline{x}, \overline{x})$ and all $a \in [\underline{a}, \overline{a}]$ or $a \in (\underline{a}, \overline{a}]$ in case (i) and (ii), respectively. This assumption implies that 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 assumed to be risk neutral and protected by limited liability. This assumption makes it possible to succinctly characterize implementation costs, but it is not important for the more fundamental discussion of which contracts are incentive compatible.[3] 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

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

[3] 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.

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 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 contracts

To understand the agent's incentives, fix a threshold $t \in (\underline{x}, \overline{x})$ and a bonus b that is paid out if the outcome is a success. Given the participation constraint is slack, it is optimal for the principal to pay nothing if the outcome is a failure. Then, the agent's expected utility from action a is $b\,(1 - F(t|a)) - a$ since the probability of a success is $1 - F(t|a)$. Now assume that the principal is aiming 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. In other words, it must equal

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

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

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

It is useful to think of a contract as a triplet $(a, t, B(a,t))$, specifying a recommended action, 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 contract can be summarized entirely by the pair (a,t). Here (a,t) can 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. In other words, there is more to the incentive compatibility problem. Thus, keep in mind that $W(a,t)$ and $U(a,t)$ are valid only as long as (a,t) is incentive compatible. A general treatment of the incentive compatibility problem is postponed to Section 4.

As mentioned, incentive compatibility pins down $B(a,t)$, $W(a,t)$, and $U(a,t)$, as

long as $a \in (\underline{a}, \overline{a})$. This is not the case when $\underline{a}$ or $\overline{a}$ is induced. The interesting question in those cases is what the cheapest way to induce the action with threshold t is. For action $\underline{a}$, it is optimal and incentive compatible 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 4.2). 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. A lower bonus would certainly cause the agent to deviate to a lower action. 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. Otherwise, an incentive compatibility constraint is violated. This conclusion does not require a full characterization of the feasible set.

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. All proofs are either in the Appendix or integrated into the main text.

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, move along the other dimension. That is, 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 $\frac{f_a(x|a)}{f(t|a)}$ 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.

Proposition 2 *Assume 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 a feasible contract (a,t) depends on the threshold. Thus, even holding fixed the action, 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 the feasible set of contracts.

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 contracting environments in which the principal does not care directly about the criterion for success. With some abuse of notation, the benefit function will be 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 and the principal cares about his expected productivity. 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.

It is important to distinguish between the case where the threshold is intrinsically important to the principal and the case where it is not. In the latter case, the only role of the threshold is to manipulate implementation costs, whereas it serves a dual purpose when it is intrinsically important.

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 $S(a,t)$.[4] Thus, any first-best solution 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 this case.

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

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

The difference between the first-best and the second-best problem is that there is no moral hazard problem in the former. However, both problems assume the same uncertainty regarding the agent's performance. In other words, the benefit function is the same in either case. In the salesman example, the customer's true willingness-to-pay is hidden to the principal even in the first-best problem. Perfect price discrimination is therefore not possible in the first-best problem. Thus, all distortions in the second-best problem ultimately traces back to the moral hazard problem.

Propositions 1 and 2 already reveal important information about $W(a,t)$ and

[4] 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 $S(a,t)$.

$U(a,t)$ along the a and t dimension, respectively. First, fix the threshold t and assume that there is a unique action that maximizes $S(a,t)$. This would identify the first-best action in a world where the threshold is exogenous. If this 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. However, Section 4 demonstrates that this is no longer necessarily true when actions and thresholds are determined jointly.

Second, holding the action fixed, $W(a,t)$ is strictly decreasing in the threshold (although attention must be restricted to incentive compatible thresholds). 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. This may give rise to an existence problem because a threshold of $\overline{x}$ is not incentive compatible – the agent never succeeds and will therefore pick action $\underline{a}$ in response.

There are at least three ways to deal with the existence problem. First, the problem disappears under realistic restrictions on the set of permissible contracts. For instance, a budget constraint or a wage cap makes it impossible to implement thresholds close to $\overline{x}$. Such contracting environments are examined in Section 3.[5]

Second, under realistic assumptions on the distribution function, contracts with large thresholds are not incentive compatible in the first place. This possibility is explored in Section 4. In comparison, the dominant approach in the existing principal-agent literature downplays the incentive compatibility problem by assuming that the first-order approach is valid. The incentive compatibility problem is taken more seriously in Section 4, where it takes center stage. The conclusions of Sections 3 and 4 are sometimes diametrically opposed. Both sections focus on the case in which the principal is not intrinsically interested in the criteria for success.

Third, assume the threshold is intrinsically important to the principal and t^{FB} is interior. If she cares sufficiently much about the threshold, then distorting it too much is so undesirable that it does not compensate for the accompanying decrease in implementation costs. This possibility is considered in Section 5.

[5] A wage cap can also be thought of as a crude way to model a risk averse agent, specifically one whose utility is constant at wages above a certain level. Thus, risk aversion can also solve the existence problem.

3 Budget constraints

This section assumes that the principal faces a budget constraint. Thus, she can offer a bonus of at most $\overline{b}$, where $\overline{b}$ is bounded. To understand the implications of this restriction, note that 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 negative and then positive as t increases.

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 $f_a(t_0(a)|a) = 0$. Here, the agent is paid if and only if the likelihood-ratio is positive, i.e. whenever $x \geq t_0(a)$ or $f_a(x|a) \geq 0$. This is where it is easiest to incentivize the agent, because he is now paid for all the performance levels that a marginal increase in his effort makes more likely and never for those that are made less likely. However, as Proposition 2 shows, expected wage costs decrease if the threshold exceeds $t_0(a)$ because the bonus, albeit higher, is then paid out less often.

Conversely, more extreme thresholds, whether they are high or low, require higher bonuses. The reason is that $F(t|a)$ is close to 0 or 1 when t is close to $\underline{x}$ or $\overline{x}$, respectively. Thus, it is harder for the agent to manipulate the chance of success. To entice him to work harder, the bonus must therefore be substantial. In fact, since $F_a(t|a) \to 0$ as $t \to \overline{x}$ or $t \to \underline{x}$, the bonus needs to grow without bound as more and more extreme thresholds are used. In contrast, when t takes an intermediate value, $F(t|a)$ is easier to manipulate through a, and so a smaller bonus is required.

Thus, the budget constraint implies that only thresholds in an intermediate range can be used to incentivize the agent. However, the set of thresholds that work depends on the action that the principal intends to induce. To focus on this issue, it is assumed, in line with most of the existing literature, that the first-order approach is valid and that all interior (a,t) are implementable. This is satisfied as long as $F(t|\cdot)$ is globally convex in a, since this implies that the agent's problem is 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 assump-

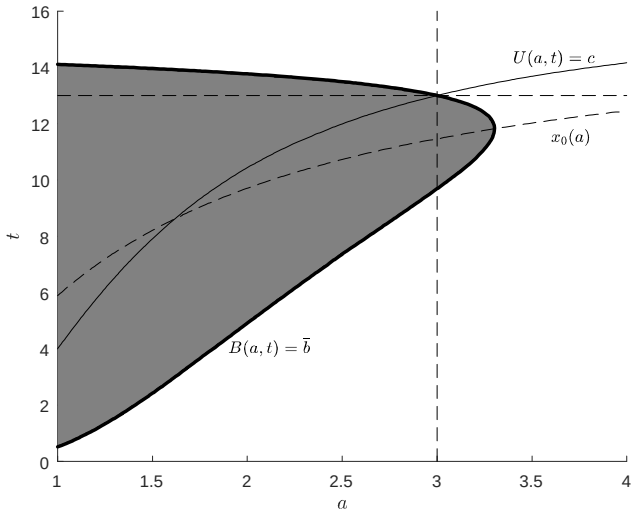

Figure 1: The feasible set given a budget constraint.

tion but instead to focus squarely on budget constraints. The next section does the opposite, by ignoring budget constraints but relaxing the CDFC. The CDFC implies that the cheapest way to induce action $\overline{a}$ with threshold t is to use the bonus that satisfies the first-order condition. Thus, Propositions 1 and 2 also hold when $a = \overline{a}$.

To visualize the problem, start by sketching the iso-bonus curve where $B(a,t) = \overline{b}$ in (a,t) space. Fix some action a and assume that $B(a, t_0(a)) < \overline{b}$, meaning that there exists threshold that implements a while satisfying the budget constraint. Given that $B(a,t)$ is u-shaped in t, there is one threshold below $t_0(a)$ and one threshold above it for which $B(a,t) = \overline{b}$. Next, it follows from Proposition 1 that $B(a,t)$ is weakly increasing in a. Thus, holding fixed the threshold, if $B(a,t)$ satisfies the budget constraint, then so do all smaller actions. The slope of the iso-bonus line is $-\frac{B_a(t|a)}{B_t(t|a)}$ or $-\frac{F_{aa}(t|a)}{f_a(t|a)}$. Thus, the iso-bonus curve slopes downward if $t > t_0(a)$ and upwards if $t > t_0(a)$. Figure 1 illustrates (the details are in Example 1 at the end of this section).

Note that the iso-bonus curve has two "prongs." Any (a,t) that is inside the area between the prongs satisfies the budget constraint and is therefore feasible. In contrast, Propositions 1 and 2 imply that iso-wage curves, $W(a,t) = c$, and indifference curves, $U(a,t) = c$, slope upwards when the CDFC and MLRP are satisfied. Expected wages and expected utility are higher *below* the respective curves.

Assume that the benefit function $\pi(a)$ does not depend on the criterion for success.

Thus, in the second-best problem, the threshold is manipulated with the sole purpose of lowering implementation costs. Assume that a^{FB} is unique and interior.

If $B(a^{FB}, t_0(a^{FB})) > \bar{b}$ then a^{FB}, or any higher action, is not implementable in the second-best problem. Then, the second-best action must be below the first-best action, or $a^{SB} < a^{FB}$. Thus, assume instead that $B(a^{FB}, t_0(a^{FB})) \leq \bar{b}$. This is the case in Figure 1, where $a^{FB} = 3$. In the second-best problem and for any a, it is optimal for the principal to increase the threshold as much as is feasible. Hence, the optimal threshold is to be found on the upper prong or the downwards-sloping part of the iso-bonus line. Now imagine inducing a^{FB} with the highest possible threshold. At this point, the weakly downward-sloping iso-bonus curve is intersected from below by a weakly increasing indifference curve. Thus, any feasible contract that involves a higher action is on or below the indifference curve, which means that it has both strictly lower social surplus and gives the agent weakly higher utility. This cannot be optimal for the principal. Thus, the second-best action can be no higher than the first-best, or $a^{SB} \leq a^{FB}$.[6] As long as $a^{SB} > \underline{a}$, this means that the threshold must be larger than the threshold that would optimally implement a^{FB}, subject to the budget constraint. Thus, 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 and that the principal is not intrinsically interested in the criterion for success. Assume the first-best action a^{FB} is unique and interior. Assume that the principal faces a budget constraint, $\bar{b} < \infty$. Then, any second-best action is no greater than the first-best action, $a^{SB} \leq a^{FB}$. If $\bar{b}$ allows a^{FB} to be implemented and $a^{SB} > \underline{a}$, then the second-best threshold is no smaller than the threshold that optimally implements a^{FB} subject to feasibility.*

EXAMPLE 1: Figure 1 illustrates Proposition 3 when $F(x|a) = \left(\frac{x}{16}\right)^a$, $x \in [0, 16]$, $a \in [1, 4]$. This distribution satisfies the MLRP and the CDFC and is inspired by an example in Rogerson (1985). It is assumed that $\bar{b} = B(3, 13) = 8.98$. Assuming that $\pi(a) = E[X|a] = \frac{16a}{1+a}$, the first-best action is $a^{FB} = 3$. In the second-best problem, the contract $(a, t) = (3, 13)$ is the best feasible contract that induces a^{FB}. This contract yields profit of 7.84. In comparison, action $\underline{a} = 1$ can be induced at

[6] If $\pi(a)$ is differentiable and $F_{aa}(x|a) > 0$ for all $x \in (\underline{x}, \overline{x})$, then $a^{SB} < a^{FB}$. Intuitively, a small distortion away from a^{FB} has no first-order effect on social surplus but it has a first-order effect on the agent's expected utility.

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. ▲

Demougin and Fluet (2001) consider a model with a *finite* number of signals and risk neutral parties. The MLRP and the CDFC hold and there is a limited liability constraint. Interestingly, in such a setting there is no loss of generality in restricting the compensation structure to be binary even if performance is perfectly observable. A bonus is paid if and only if the very best signal is realized, which has a strictly positive probability of occurring in their finite-signal model. Note that there is no existence problem.

The principal in Demougin and Fluet (2001) can invest in different monitoring technologies, each leading to a different relationship between actions and the probability of realizing the highest signal. They make the point that the optimal monitoring system may depend on the action that is implemented. Something similar occurs in the present model, where the optimal threshold depends on the action.

4 Implementability constraints

This section studies the problem without imposing the CDFC. Thus, some (a, t) may not be incentive compatibility. This places new restrictions on the principal's problem. To focus on these implementability constraints, 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 contracts.[7] 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, or $F_{aa} \geq 0$. Now, 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

[7] The argument leading to Lemma 1 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.

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 $F^C(t|a)$. With some abuse of terminology, say that a is "on the convex hull" of $F(t|\cdot)$ if $a \in A^C(t)$. The end-points of the domain are always on the convex hull, 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 $a \in A^C(t)$.[8] 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. Finally, 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 1 *The set of implementable* (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 contract that belongs to $\mathcal{I}$. Recall that the first-order condition pins down $B(a, t)$, $W(a, t)$, and $U(a, t)$ for any interior (a, t) that is in $\mathcal{I}$.

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.

Examples that illustrate Lemma 1 can be found in Sections 4.3, 5.1, and in the Appendix. 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.

[8] 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 $x \in (\underline{x}, \overline{x})$ by picking a bonus that is so large that the agent's utility is globally increasing in the action.

4.2 The shape of the feasible set

Lemma 1 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. Thus, this assumption solves the existence problem mentioned in Section 2.

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. 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.

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. This 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, this holds under the standard assumption that the density, q, is log-concave.

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 action is how much the jogger exercises. The NUC$_x$ disciplines how the marginal increase in the probability of success from additional exercise for a committed jogger as compared to a sedentary person 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. It is particularly important to note that if the action $\underline{a}^C(t)$ is induced, then the agent is exactly indifferent between the target action and the lowest action, $\underline{a}$. 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 $\{\underline{a}\} \cup [\underline{a}^C(t), \overline{a}]$.

The next result combines the previous regularity assumptions. It can be thought of as describing the "nicest" shape of the feasible set that can be expected once the CDFC no longer holds.

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 $(\underline{a}, \overline{a})$.*

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}. In addition, assume that the distribution function is well-behaved 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 $\overline{t}^C(a)$.[9]

The regularity assumption is interesting in part because it turns out to yield conclusions that are diametrically opposed to those obtained in Section 3.

Note that $(a, \overline{t}^C(a))$ traces out the boundary of the feasible set on $(\underline{a}, \overline{a}) \times (\underline{x}, \overline{x})$. 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, i.e. be of the form $(a, \overline{t}^C(a))$. By Corollary 1, the threshold is more stringent the higher the target action is.

What is particularly important for the upcoming analysis is that NDC$_a$ implies that if action a is implemented with a threshold of exactly $\overline{t}^C(a)$ then the agent is exactly *indifferent* between a and the lowest possible action, $\underline{a}$ (see the discussion following Proposition 6). This 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.

[9] 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.

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. ▲

The best-case scenario for the principal is that $F(x|\underline{a}) = 1$ for all $x \in [\underline{x}, \overline{x}]$. Example 2 provided one example (when $\underline{a} = 0$) and Section 5.1 contains another example. In words, the agent's performance is guaranteed to be the worst possible if his action is $\underline{a}$. Therefore, for any $t \in (\underline{x}, \overline{x})$, there is no chance that the agent earns the bonus with action $\underline{a}$. This is advantageous to the principal, because such a deviation is less desirable and therefore easier to prevent. In particular, the aforementioned indifference condition is now

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

Thus, the agent is indifferent between all $(a, \overline{t}^C(a))$, $a \in (\underline{a}, \overline{a}]$. Stated differently, along the boundary of the feasible set, the agent appropriates a constant amount of rent and the rest goes to the principal. This is reminiscent of a standard principal-agent problem with risk neutral parties and a binding participation constraint. Thus, the first-best action solves the second-best problem.

Proposition 7 *Assume that $F(x|a)$ is regular and that $F(x|\underline{a}) = 1$ for all $x \in [\underline{x}, \overline{x}]$. 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, $a^{SB} = a^{FB}$.*

In this setting, there is no distortion of the optimal action. However, this is the case only because the threshold is endogenous and can be adjusted. If t is exogenous and fixed, the conclusion is much different. For instance, if t is fixed at a level that exceeds the solution to the second-best problem, then only $\underline{a}$ and actions above a^{FB} are feasible. Thus, the agent is either induced to take the lowest possible action or an action that exceeds the first-best. This example illustrates in a rather extreme way the benefit to being able to endogenize the criterion for success.

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 $(a, \overline{t}^C(a))$, the agent's utility therefore also depends on $(a, \overline{t}^C(a))$. Thus, the agent's utility is no longer constant along the boundary of the feasible set. This gives the principal 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. It will now be proven that the action is distorted *upwards*, and that the threshold as a consequence must be "large."

The argument centers on comparing the agent's indifference curve to the boundary of the feasible set, summarized by the function $\overline{t}^C(a)$. As already noted, the indifference curve must slope upwards on the feasible set, but so does $\overline{t}^C(a)$. However, it can be shown that at any point of intersection, the indifference curve is flatter than $\overline{t}^C(a)$. Thus, any indifference curve crosses $\overline{t}^C(a)$ at most once, and if so from above. Figure 2 illustrates this property (see Example 3, below, for details). In comparison, in Proposition 7 there is an indifference curve that coincides everywhere with $\overline{t}^C(a)$.

Consider the indifference curve through the feasible point $(a^{FB}, \overline{t}^C(a^{FB}))$, as in Figure 2. Any feasible point below this curve is better for the agent and must have social surplus no higher than at $(a^{FB}, \overline{t}^C(a^{FB}))$ since social surplus is maximized whenever $a = a^{FB}$. Thus, all such points leave less surplus to the principal than does $(a^{FB}, \overline{t}^C(a^{FB}))$. The only feasible points that are potentially preferable feature actions and thresholds that are above a^{FB} and $\overline{t}^C(a^{FB}))$, respectively. Thus, if the second-best action is greater than $\underline{a}$ then it must be no smaller than the first-best action, $a^{SB} \geq a^{FB}$, and the optimal threshold must be larger than what is required

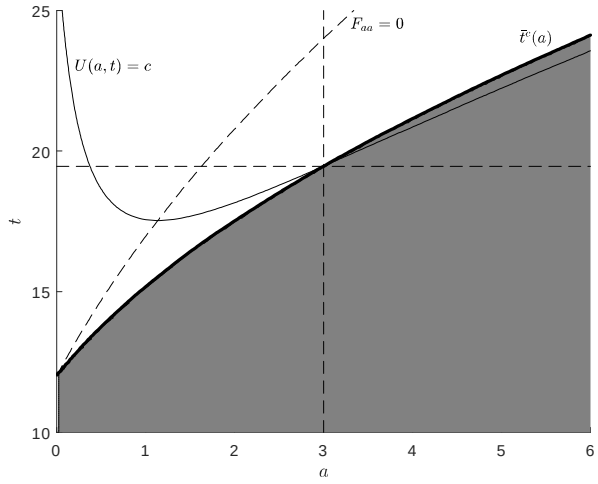

Figure 2: The feasible set for a regular distribution function.

to optimally implement the fist-best action, or $\bar{t}^C(a^{SB}) \geq \bar{t}^C(a^{FB})$. However, since implementation costs are discontinuous at $a = \underline{a}$, it cannot be ruled out that $\underline{a}$ is preferable to the principal. In conclusion, 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.*

EXAMPLE 3: Figure 2 illustrates Proposition 8 for the distribution function $F(x|a) = 1 - e^{-\frac{x}{4\sqrt{1+a}}}$, $x \geq 0$, $a \in [0, 6]$. 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, the contract $(a, t) = (3, 19.46)$ is the best feasible contract that induces a^{FB}. This contract yields profit of 4.71 whereas inducing $\underline{a} = 0$ yields profit of 4. It can be verified that inducing action $\overline{a} = 6$ 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. ▲

The fact that the second-best action is *greater* than the first-best action is in very stark contrast to the conclusion that obtains from a standard model in which the

threshold is fixed. If a^{FB} is implementable at the exogenous threshold, or $t \leq \bar{t}^C(a^{FB})$, then the second-best action must be *below* the first-best. The argument is by now familiar. Once again, the agent's expected utility is increasing in a. Hence, increasing a above a^{FB} decreases social surplus and increases the agent's surplus, thus leaving less surplus for the principal. Thus, in these cases, the standard model predicts that $a^{SB} \leq a^{FB}$. This is also the conclusion that was obtained when the threshold was made endogenous but the binding constraint was a budget constraint. Thus, even when the threshold is endogenous, it matter whether the budget constraint or the "implementability constraint" is binding. The reason is that the feasible set are shaped so differently; compare Figures 1 and 2.

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, 3, and 4 (to follow) provide examples of regular distribution functions. The appendix 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

The remainder of the paper 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, interior, and that it can feasibly be implemented in the second-best problem. It is also assumed that a solution to the second-best problem exists. It does not matter if feasibility is restricted by a budget constraint, implementability constraint, or a combination of the two. Thus, the main results in this subsection are driven by the interaction between a and t as benefits and implementation costs are traded off.

The assumption that (a^{FB}, t^{FB}) is feasible in the second-best problem merits discussion. First, this requires that the budget is large enough, or $\overline{b} \geq B(a^{FB}, t^{FB})$. Second, (a^{FB}, t^{FB}) must be incentive compatible, or $(a^{FB}, t^{FB}) \in \mathcal{I}$. While Section 4 describes the shape of $\mathcal{I}$ for any $F(x|a)$, the complication is that (a^{FB}, t^{FB}) depends not only on $F(x|a)$ but also on $\pi(a,t)$. The next result confirms that (a^{FB}, t^{FB}) is feasible in the second-best problem 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 fits a version of the salesman problem in which the customer's willingness-to-pay is observable and perfect price discrimination is possible, but where the agent can be paid only on the basis of whether a sale was made or not. In this case, $v(x) = x - c$, where c is once again production costs.

The second case is also 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 licensing examples in the introduction. In those cases, however, the principal typically controls the bonus at best in an indirect way. For instance, the value of a driver's license depends on taxes and fees associated with car ownership as well as the quality of the road network compared to the availability of public transit. In the medical license example, there are jurisdictions around the world where health care is publicly funded and physicians' compensation is more heavily regulated.

Now, to understand how the first-best and second-best solutions compare in situations like this, imagine that the principal in the second-best problem contemplates

inducing (a^{FB}, t^{FB}). From this starting point, what are the consequences of changing the contract? First, any departure from (a^{FB}, t^{FB}) strictly lowers social surplus. Likewise, under the MLRP, any departure to another 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 (a, t) 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.*[10] *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.*

Corollary 2 describes the possible ways in which the second-best is distorted away from the first-best when the principal is intrinsically interested in the criterion for success. It is important to realize that both the action and the criterion for success are distorted. The moral hazard problem implies a welfare loss along both dimensions.

More precise predictions depend on the properties of the benefit function, and possibly which kind of feasibility constraint binds, if any. The easiest setting to understand is when the benefit function is differentiable and additively separable in a and t. Thus, there is no interaction between the two variables, or $\pi_{at} = 0$. The iso-surplus curve is downward-sloping to the south-west and north-east of (a^{FB}, t^{FB}), and upwards-sloping to the north-west and to the south-east of this point. If the MLRP holds and the second-best is in the interior of the feasible set (i.e. the budget constraint and the implementability constraint are not binding) then the second-best must be found at a point of tangency between the iso-surplus curve and the agent's indifference curve. The latter slopes upwards. Corollary 2 implies that the second-best cannot be to the south-east of the first-best. This leaves only the possibility that the second-best is to the north-west of the first-best, or $t^{SB} > t^{FB}$ and $a^{SB} < a^{FB}$.

When a and t interact in $\pi(a, t)$, or if one of the feasibility constraints bind, then matters are more complicated. The next example shows that *both* the action and the threshold may be distorted downwards compared to the first-best.

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

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 the appendix, 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 $\overline{t}^C(a) = \frac{1}{\beta}a^\beta$. Note that this is greater than a^β, thus confirming that the first-best solution is feasible in the second-best problem. The 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 $t^{SB} < \overline{t}^C(a^{SB})$. 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. This is despite the

fact that the threshold may be lower in the second-best. In such cases, however, the second-best action is much lower too. ▲

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, $t^{SB} \geq \widehat{t}(a^{SB})$. This follows from the fact that given 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. 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. An implication is that the price in the salesman example is set above the monopoly price for the demand curve described by $1 - F(\cdot|a^{SB})$.

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. The appendix 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

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 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 (or cost) that depends only on the probability that the agent fails or succeeds. This formulation 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. The results in Sections 3 and 4 are then relevant. 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, assume that there is a fixed cost of processing the payment of the bonus. Then, $\kappa(F(t|a))$ 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. A similar situation occurs if the principal derives status from overseeing a tough test with a high failure rate. In these cases, $t^{FB} = \overline{x}$. Conversely, κ is decreasing if a failure means that the principal will have to incur fixed costs of restarting a research endeavour or incur

search costs to replace the agent. Then, $t^{FB} = \underline{x}$.

Alternatively, consider some professional body that controls the admission of candidates or apprentices into a "club" (e.g. a guild or other professional organization, or the tenure committee of a university department). This body may have in mind an ideal size, or pass-rate. In this case, t^{FB} is interior.

The first-best problem is to maximize

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

which evidently means that 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, any first-best action must maximize $\nu(a) - a$, or $a^{FB} \in \arg\max_a (\nu(a) - a)$.

The second-best problem is to maximize

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

Assuming the MLRP is satisfied, $U(a,t)$ is strictly decreasing in t on $T^C(a)$ for all $a \in (\underline{a}, \overline{a})$. To continue, assume that $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t on $T^C(a)$ for all $a \in (\underline{a}, \overline{a})$. This holds if κ is either increasing or if it is not too sensitive to changes in the probability of failure. Then, if the second-best action is interior, the accompanying threshold, t^{SB}, must be the highest feasible threshold, whether this is determined by a budget constraint or an implementability constraint. The arguments in Sections 3 and 4 are now relevant, as they describe in which direction along the boundary of the feasible set to move in order to increase $\nu(a) - a - U(a,t)$. The next question is then whether a move in said direction also changes $F(t|a)$ in such a way that $\kappa(F(t|a))$ increases, whereas a move in the opposite direction decreases $\kappa(F(t|a))$. If so, the two effects reinforce each other.

To illustrate, consider the upper boundary or prong of the feasible set described in Section 3. Moving leftward along this boundary both increases the threshold and decreases the action. Thus, such a move increases $F(t|a)$. This is beneficial if κ is increasing, such as when there is a cost of processing the bonus. A similar argument holds when κ is u-shaped and the budget is sufficiently large.

Corollary 4 *Assume that the MLRP and CDFC hold and that the principal faces*

a budget constraint, $\overline{b} < \infty$. Assume the benefit function takes the form in (1) and that $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t for all $t \in (\underline{x}, \overline{x})$ and all $a \in (\underline{a}, \overline{a}]$. Assume that a^{FB} is unique and interior and that either

1. *κ is weakly increasing, or*
2. *κ is u-shaped and minimized whenever $F(t|a) = q \in (0,1)$, while $\overline{b}$ is so large relative to q that $F(t|a) \geq q$ along the upper boundary of the feasible set.*[11]

In either case, any second-best action is no greater than a^{FB}.

Turning to implementability constraints in the sense of Section 4, environments in which $F(x|\underline{a})$ is degenerate have particularly interesting and clear-cut properties. Since the agent receives a constant amount of rent on the boundary of the feasible set, the principal can focus on maximizing social surplus, subject to feasibility.

Corollary 5 *Assume that $F(x|a)$ is regular and that $F(x|\underline{a}) = 1$ for all $x \in [\underline{x}, \overline{x}]$. Assume that the benefit function takes the form in (1) and that $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t in the interior of $\mathcal{I}$. Then, absent budget constraints, any interior second-best action maximizes $S(a, \overline{t}^C(a))$.*

EXAMPLE 6: Assume once again that $F(x|a)$ is an exponential distribution with mean $h(a)$. Assume first that $h(a) = a^\beta$, $\beta \in (0,1)$ and $\underline{a} = 0$. Then, as can be seen from Example 4, $F(\overline{t}^C(a)|a)$ is constant.[12] In other words, $\kappa(F(t|a))$ does not change along the boundary of $\mathcal{I}$. Thus, a^{SB} and a^{FB} coincide. The welfare cost of moral hazard in this case comes only from the fact that t^{SB} is distorted away from t^{FB}.

More generally, for any concave $h(\cdot)$ for which $h(0) = 0$ and $F(x|a)$ is regular, it holds that

$$F(\overline{t}^C(a)|a) = 1 - e^{-\frac{h(a)}{ah'(a)}} \geq 1 - e^{-1} > \frac{1}{2}.$$

Assume next that $\frac{h(a)}{ah'(a)}$ is increasing. Then, $F(\overline{t}^C(a)|a)$ increases and moves further away from $\frac{1}{2}$ as a increases. For example, $F(x|a)$ is regular and $\frac{h(a)}{ah'(a)}$ is strictly increasing if $h(a) = \ln(1+a)$. In such cases, if κ is either (i) increasing or (ii) u-shaped and minimized at $F(t|a) = \frac{1}{2}$ – which is plausible in Li and Yang's (2020) problem – then a^{SB} exceeds a^{FB} when the latter is interior. ▲

[11] In Example 1, $F(t|a) \geq F(x_0(a)|a) = e^{-1}$ on the upper prong of the feasible set, regardless of how large or small $\overline{b}$ is. More generally, once $\overline{b}$ is large enough that $\overline{a}$ can be implemented, a further increase in $\overline{b}$ causes the upper prong to shift upward, and $F(t|a)$ thus approaches 1 as $\overline{b} \to \infty$.

[12] This is also the case for the Kumaraswamy distribution in Example 2 when $\beta > 1$ and $\underline{a} = 0$.

6 Discussion

6.1 Binary versus continuous actions

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.

Corollary 6 *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.

6.2 More information gathering

In the salesman example, the agent is a middleman between the firm (principal) and the customer. The latter is a sentient being that can perhaps be persuaded to reveal more information. Thus, it is worthwhile considering more refined ways in which the principal can gather information about the agent's performance.

For instance, the principal can simply ask the customer to report his willingness-to-pay (promising that it will not impact the price that he pays) and let the agent's compensation be based on this report. Strictly speaking, in the model the customer would have no incentive to misreport, but in reality it is not hard to imagine that the customer would be concerned that a truthful report would lead to price discrimination in future interactions. Moreover, since the salesman is the middleman, he would have an incentive to misrepresent the customer's report.

Similarly, the principal could instruct the salesman to present the customer with a menu of options. Self-selection then reveals more detailed information about the customer's willingness-to-pay. The menu would consist of combinations of prices and accompanying probabilities of acquiring the good. This may once again give the agent an incentive to misrepresent his interaction with the customer. Putting aside that possibility, the fundamental fact remains that the principal is once again distorting the offer that is made to the customer away from the offer that would have been made in the absence of moral hazard concerns. When faced with a privately informed customer, a fixed monopoly price is normally optimal. The menu of options is desirable only because it may help reveal more information about the agent's performance and thus alleviate the moral hazard problem. This paper focuses on the simplest possible kind of distortions, which involves manipulating a threshold.

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 model is sufficiently tractable that the first-order approach is not needed. Indeed, 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. Thus, the solution is sensitive to the properties of the performance technology in a way that is absent when the first-order approach is valid.

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

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

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

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.

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

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 Proposition 3. In text. ■

Proof of Lemma 1. Fix some $t \in (\underline{x}, \overline{x})$. The contract (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 $a \in A^C(t)$. This establishes the feasible set on $(\underline{a}, \overline{a}) \times (\underline{x}, \overline{x})$. 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. The result is trivial if $A^C(t) = (\underline{a}, \overline{a})$. Thus, assume $A^C(t) \neq (\underline{a}, \overline{a})$. Given an interior threshold, t, an action $a^* \in (\underline{a}, \overline{a})$ is in A^C if and only if

$$F(t|a^*) + (a - a^*)\,F_a(t|a^*) \leq F(t|a)$$

or

$$F(t|a^*) - F(t|a) + (a - a^*)\,F_a(t|a^*) \leq 0$$

for all $a \in A$. Conversely, a^* is not in $A^C(t)$ if there exists any $a \in [\underline{a}, \overline{a}]$ for which the inequality is violated.

Since it is assumed that $A^C(t) \neq (\underline{a}, \overline{a})$, there must be some interior action that is not in $A^C(t)$. Select a^* to be one such point, $a^* \notin A^C(t)$. Then, there exists some other action, a, for which the above inequality is violated, or

$$F(t|a^*) - F(t|a) + (a - a^*)\,F_a(t|a^*) > 0. \tag{2}$$

Differentiate the left hand side with respect to t to get

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

Recall that $F_a(t|z)<0$. By NUC$_x$, $\frac{f_a(x|z)}{F_a(x|z)}$ is increasing in z. Thus, if $a^*>a$ then the term in the parenthesis is negative. Hence, the integral is positive. Similarly, if $a^*<a$, then

$$\frac{\partial\left(F(t|a^*)-F(t|a)+(a-a^*)\,F_a(t|a^*)\right)}{\partial t}=-\int_{a^*}^{a}F_a(t|z)\left(\frac{f_a(t|z)}{F_a(t|z)}-\frac{f_a(t|a^*)}{F_a(t|a^*)}\right)dz.$$

In this case, the term in the parenthesis is positive, and the right hand side is therefore positive as well. In other words, regardless of whether a is smaller or larger than a^*, an increase in t causes the term on the left in (2) to weakly increase. Thus, (2) is still satisfied. This means that a^* is still not implementable when t increases; it remains better to deviate to a. Thus, the set of implementable actions cannot grow as t increases. This proves the result. ■

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. Combined with NUC$_x$, $F_{aa}(\cdot|a)$ is therefore either always negative or it is first-positive-then-negative in t.

Given NUC$_x$, recall 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 $A^C(t)=\{\underline{a},\overline{a}\}$ because

the extreme actions are the only actions on the convex hull of $F(t|\cdot)$ in this case. 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, the set of actions on the convex hull of $F(t|\cdot)$ 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 $\underline{a}^C(t) \in (\underline{a}, \overline{a})$. ■

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 is outlined in the text, but it remains to prove that the indifference curve crosses $\overline{t}^C(a)$ at most once, and if so from above. It has already been shown that expected utility is globally decreasing in t. Similarly, expected utility is increasing in a on the feasible set, and indeed on the superset where $F_{aa}(t|a) > 0$. Thus, on this superset of the feasible set, the slope of the indifference

curve is positive and equal to

$$\frac{dt}{da}_{|U(a,t)=c} = \frac{F_{aa}(t|a)(1-F(t|a))}{-f(t|a)F_a(t|a) - f_a(t|a)\,(1-F(t|a))} > 0,$$

where the denominator is positive by the fact that the MLRP implies that the survival function $1-F(t|a)$ is log-supermodular in (a,t).

Next, any point on the boundary of the feasible set, $(a, \bar{t}^C(a))$, is characterized by the condition that

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

This boundary has also been proven to have positive slope. The slope equals

$$\frac{d\bar{t}^C(a)}{da} = -\frac{(\underline{a}-a)\,F_{aa}(t|a)}{f(t|a) - f(t|\underline{a}) + (\underline{a}-a)\,f_a(t|a)} > 0.$$

Now compare the slopes at any point of intersection. Since it turns out to be easier to compare the inverse functions, consider

$$\begin{aligned}
T(a,t) &= \frac{da}{dt}_{|U(a,t)=c} - \frac{d\underline{a}^C(t)}{dt} \\
&= F_{aa}(t|a)\left[\left(-\frac{f(t|a)}{(1-F(t|a)}F_a(t|a) - f_a(t|a)\right) - \left(-\frac{f(t|a)-f(t|\underline{a})}{(\underline{a}-a)} - f_a(t|a)\right)\right] \\
&= F_{aa}(t|a)\left[-\frac{f(t|a)}{(1-F(t|a)}F_a(t|a) + \frac{f(t|a)-f(t|\underline{a})}{(\underline{a}-a)}\right] \\
&= F_{aa}(t|a)\left[-\frac{f(t|a)}{1-F(t|a)}F_a(t|a) + \frac{(f(t|a)-f(t|\underline{a}))\,F_a(t|a)}{F(t|\underline{a})-F(t|a)}\right] \text{ by definition of } (\underline{a}^C(t),t) \\
&= -F_a(t|a)F_{aa}(t|a)\left[\frac{f(t|a)}{1-F(t|a)} - \frac{f(t|a)-f(t|\underline{a})}{F(t|\underline{a})-F(t|a)}\right] \\
&= -\frac{F_a(t|a)F_{aa}(t|a)\times(f(t|a)\,(F(t|\underline{a})-F(t|a)) - (f(t|a)-f(t|\underline{a}))\,(1-F(t|a)))}{(1-F(t|a))\,(F(t|\underline{a})-F(t|a))} \\
&= -\frac{F_a(t|a)F_{aa}(t|a)\times(f(t|a)\,(F(t|\underline{a})-1) + f(t|\underline{a})\,(1-F(t|a)))}{(1-F(t|a))\,(F(t|\underline{a})-F(t|a))} \\
&= -\frac{F_a(t|a)F_{aa}(t|a)\,(1-F(t|\underline{a}))}{F(t|\underline{a})-F(t|a)}\left[\frac{f(t|\underline{a})}{1-F(t|\underline{a})} - \frac{f(t|a)}{1-F(t|a)}\right].
\end{aligned}$$

Since $F_a < 0, F_{aa} > 0$ and $F(t|\underline{a}) - F(t|a) > 0$, the term in front of the brackets in the last line is positive. By 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 positive too.

Thus,

$$\frac{da}{dt}_{|U(a,t)=c} - \frac{d\underline{a}^C(t)}{dt} \geq 0$$

or

$$\frac{d\overline{t}^C(a)}{da} - \frac{dt}{da}_{|U(a,t)=c} \geq 0,$$

at any point of intersection. Thus, the indifference curve is flatter than $\overline{t}^C(a)$ in (a,t) space at any point of intersection.

The rest of the proof follows the argument in the text. ■

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 be on the convex hull of $F(t^{FB}|\cdot)$. 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

$$\int_c^{\overline{x}} v'(x)\left(F(x|a) - \left(F(x|a^{FB}) + \left(a - a^{FB}\right)F_a(x|a^{FB})\right)\right)dx \geq 0 \text{ for all } a \in [\underline{a}, \overline{a}]. \tag{3}$$

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

$$F(c|a') - \left(F(c|a^{FB}) + \left(a' - a^{FB}\right)F_a(c|a^{FB})\right) \geq 0 \text{ for all } a \in [\underline{a}, \overline{a}]. \tag{4}$$

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 (see Figure 2) 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 (3). Thus, (4) must hold and it now follows that $(a^{FB}, t^{FB}) \in \mathcal{I}$. ■

Proof of Corollary 2. The corollary follows from the proof 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. In text. ■

Proof of Corollary 4. The corollary is trivial if a^{FB} is not implementable. Hence, assume a^{FB} is implementable. The corollary is also trivial if the second best action is $\underline{a}$, so assume that there is a second-best action that is larger than $\underline{a}$. Then, since $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t, the optimal contract is on the upper boundary of the feasible set. By the argument leading to Proposition 3, any rightward move along this boundary starting from $a = a^{FB}$ must decrease $\nu(a) - a - U(a,t)$. Such a change also decreases $F(t|a)$ and therefore $\kappa(F(t|a))$ if the latter is weakly increasing. Thus, both effects work in the same direction and it follows that $a \geq a^{FB}$ cannot be second-best in this case. If $\kappa(F(t|a))$ is u-shaped but $F(t|a) \geq q$ along the upper boundary of the feasible set, then it once again holds that a rightward move along this boundary lowers $\kappa(F(t|a))$. The conclusion is again that $a \geq a^{FB}$ cannot be second-best. ■

Proof of Corollary 5. Since $\kappa(F(t|a)) - U(a,t)$ is strictly increasing in t, any solution to the second-best problem in which the action is interior must take the form $(a, \bar{t}^C(a))$. By the argument leading to Proposition 7, $U(a, \bar{t}^C(a))$ is constant. Hence, $\pi(a, \bar{t}^C(a)) - W(a, \bar{t}^C(a))$ is proportional to $S(a, \bar{t}^C(a))$ and the corollary follows. ■

Proof of Corollary 6. 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 $\overline{a}$ is large enough that $h(\overline{a}) \leq \overline{a}$. Since

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

this 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 $a \in [0,1]$.

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

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

The first term inside the brackets is negative, while the second is positive. Thus, $F_{aa}(\cdot|a)$ changes sign as x increases. Hence, the CDFC is not satisfied, but NUC_x and CAT are. Since the second term is decreasing in a, NDC_a automatically holds if the first term does not increase to fast in a. It can be verified that this is true if

$h(a) = a^\beta$, $\beta \in (0,1)$ and $a \in [0,1]$.

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 $c \geq \overline{c}$. 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" in the set of actions that can be implemented. 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.

Note that in the $h(a) = \sqrt{a}$ example, Proposition 4 implies that $\bar{t}^C(a)$ is u-shaped in a. Thus, it has a downwards-sloping portion (as in Figure 1) and an upwards-sloping portion (as in Figure 2). As in the case with budget constraints in Figure 1, it is thus possible that $a^{SB} < a^{FB}$ and $t^{SB} > t^{FB}$. Section 4 focused on the NDC_a assumption in part to highlight that budget constraints and implementability constraints can lead to opposite conclusions.

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 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. ▲