Disentangling Moral Hazard and Adverse Selection

Hector Chade[*] and Jeroen Swinkels[†]

December 2016

Abstract

This paper analyzes a canonical principal-agent problem with moral hazard and adverse selection. The agent is risk averse and has private information about his disutility of taking an unobservable action. The principal is risk neutral and designs a menu of contracts consisting of a compensation scheme and a recommended action for each type of agent to maximize expected profit. We first derive a set of sufficient conditions for menus to be feasible (i.e., satisfy participation and incentive compatibility). Then we provide a method of solution, decoupling, consisting of first solving a cost minimization problem for a pure moral hazard problem for each type, action, and utility level of the agent, and then using the resulting cost function to solve a suitable adverse selection problem. This relaxed problem is both tractable and delivers a candidate solution for the original problem. We show several classes of primitives under which this candidate solution is in fact feasible and hence optimal. We also describe several properties that optimal menus exhibit when decoupling is valid.

Keywords. Moral Hazard, Adverse Selection, First-Order Approach, Incentive Compatibility, Principal-Agent Problem.

[*] Arizona State University, `hector.chade@asu.edu`

[†] Northwestern University, `j-swinkels@northwestern.edu`

We are especially grateful to Michael Powell, with whom a number of important conversations occurred early in the development of the paper. We would also like to thank Laura Doval, Andreas Kleiner, Alejandro Manelli, and seminar participants at Arizona State University, CEA-Universidad de Chile, and Boston College for helpful discussions and comments.

Contents

1 Introduction

In many settings of applied interest, there is both a screening and a moral hazard problem at play. A firm may want workers to self identify as more or less able, and to tailor the resulting incentive contract and implemented action accordingly. An insurance company may want to tailor the trade-off between risk sharing and incentives to avoid accidents to the privately known type of the customer. An investor may want an entrepreneur to both reveal the quality of the project and choose an appropriate level of effort. In each of these settings, an optimal contract needs to both elicit the agent's private information and provide him with incentives to take a suitable action contingent on his revealed information.

Although the design of contracts under either pure adverse selection or pure moral hazard is well understood, little is known about the case where the two of them are present simultaneously, especially when the agent is risk averse.[1] Partly, this is because the problem is innately a complicated one. In each of the pure cases, the set of deviations for the agent is one dimensional. But here, an agent can first misrepresent his type and then choose an action level different than the recommended one for the type announced. Hence, unlike the adverse selection case, where a sweeping incentive compatibility characterization exists (Mirrlees (1975), Myerson (1981)), or the moral hazard case, where the first-order approach (Rogerson (1985), Jewitt (1988)) simplifies the incentive constraints, there is no known analogous simplification in the combined case that handles the myriad of deviations available to the agent. The central purpose of this paper is to explore one potential such simplification.

This paper analyzes the optimal design of contracts with adverse selection and moral hazard in a canonical principal-agent model. A risk-averse agent has a type that represents his disutility of taking an unobservable action. A signal is generated that depends stochastically on the action of the agent. Both the type of the agent and his action lie in a continuum. A mechanism consists of two functions, where the first function specifies a recommended action for each announced type, and the second function specifies compensation to the agent that depends on both his announced type and the realization of the signal. Put differently, the agent, by announcing his type, is effectively choosing over a menu of compensation schemes. The problem facing the principal is to design the mechanism to maximize expected profit subject to guaranteeing that the agent is willing to participate, reveal his type, and take the recommended action.

We first provide a thorough analysis of incentive compatibility. We derive the necessary local conditions, and then turn to sufficient conditions for feasibility – that is, for the global incentive compatibility constraints to be satisfied. We derive two such sets of sufficient conditions.

[1] For adverse selection, see, e.g. Guesnerie and Laffont (1984) and the textbook treatment in Chapter 7 of Fudenberg and Tirole (1991). For moral hazard, see the seminal papers by Holmstrom (1979), Grossman and Hart (1983), and the textbook treatments in Laffont and Martimort (2002), and Bolton and Dewatripont (2005). The textbooks present examples with both moral hazard and adverse selection, with two types or with risk neutrality.

Our first condition centers around the action schedule. Consider two agents, a less able one, and a more able one. Holding fixed the action, the less able agent has a higher marginal cost of effort. But, as we shall see, the recommended effort must fall when the agent announces that he is less able, providing a force in the direction of lowering his marginal cost of effort. We show that a sufficient condition for feasibility is that the second force overwhelms the first, with the recommended action falling fast enough so that the marginal cost of effort at the recommended effort level falls when the agent is less able. We refer to this as the *diminishing marginal cost* condition (DMC).

Our second sufficient condition, the *single crossing condition* (SCC), focuses on the compensation scheme that different types of the agent face. If the compensation of the less able agent is lower powered in the sense that his compensation scheme, as a function of the signal, single-crosses that of the more able agent from above, then feasibility follows.

For each set of sufficient conditions, we also require that, holding fixed the type one has announced, the expected utility from income of the agent is concave in effort. That is, we invoke something similar to the first-order approach in the pure moral hazard setting.

The conditions just described are useful in checking the feasibility of a candidate mechanism. But they give no guidance as to how to derive such a candidate mechanism, nor do they give much insight into what an optimal mechanism might look like. What we desire is a way to solve the problem that is both computationally tractable and gives insight into the form of an optimal mechanism. To achieve both of these goals, we turn to the question of when the problem can be broken into two separate simpler problems, one of which is essentially a moral hazard problem, and the other of which is a screening problem.

In particular, we study the following decoupling program. First, fix any given combination of an effort level, required utility level, and type for the agent, and consider the relaxed moral hazard problem of minimizing the cost of implementing the given action for the agent subject to only the reservation utility and the first-order condition on effort. The solution to this problem is well understood, as it is the setting considered by Holmstrom (1979), Mirrlees (1975), and the enormous literature that follows. Solving this problem parametrically yields a cost and a compensation scheme for every triple of effort level, utility level, and type.

Second, using the resulting cost function, consider the pure adverse selection problem in which, subject to the cost function derived, one solves for the action schedule that maximizes the principal's expected profit. This problem, while lacking the quasi-linear structure commonly exploited in screening problems, remains tractable, and has a solution that shares many of the standard properties that we understand from the screening literature, e.g., Myerson (1981), Maskin and Riley (1984), and Guesnerie and Laffont (1984). In particular, knowing the action schedule ties down the requisite surplus of the agent as a function of his type, and the derivative of that function is related in the usual way to how the cost of effort of the agent varies with his type.

Assume one has gone through the decoupling program, where at the first stage, one solves the parameterized moral hazard problem to generate a cost function, and at the second stage, one solves for an optimal action schedule in the adverse selection problem with that cost function. Now, for each type, and for the associated action and surplus generated at the second stage, associate the compensation scheme given by the solution to the moral hazard problem for that action, surplus, and type. Taking these objects together defines a mechanism that is a candidate solution for the original problem.

If we knew that this candidate solution was in fact optimal, we would have gone a long way towards each of our two goals. First, the procedure described is one that can be implemented computationally. Rather than consider a double continuum of incentive constraints, we have a nested problem with each step well within the bounds of what can be handled numerically for any given specification of the primitives. Second, when the decoupled solution solves the original problem, we can say a great deal about the properties of an optimal menu. For example, the compensation scheme for each type has the form derived from a standard cost minimization problem with moral hazard. It hence, for example, satisfies well-known properties pinned down from the likelihood ratio. Also, the action schedule solves a screening problem and thus inherits many of its properties, such as that each type except for the least able receives an information rent pinned down by the incentives to misreport. Similarly, we can shed light on conditions for separation of types and exclusion thereof, exploiting the insights obtained in the benchmark case without moral hazard. In short, if decoupling works, then the insights we know from the moral hazard and the adverse selection problems considered individually extend to the combined case.

The question then comes down to whether the decoupled solution is in fact a solution to the original problem. Note first that the decoupled solution is a relaxed program: In among the various incentive constraints of the agent are that he should not want to deviate locally from his recommended action having announced his type honestly, and that he should not want to misrepresent his type conditional on following the recommended action for whatever type he states. So, if the decoupled solution is feasible, it is optimal.

We first point out that in numerically computed examples, one can use our sufficient conditions: the decoupled solution in any given setting serves as a candidate solution, and it is numerically straightforward to check whether either DMC or SCC is satisfied for that solution.

We next turn to the task of deriving primitives that guarantee *ex-ante* that the decoupling program will work. In particular, we look for primitives under which one or the other of DMC or SCC can be shown to hold. These include the case of quadratic utility, the two-outcome case, and a class of utility functions coupled with signal distributions that are exponential families. In the quadratic utility case, which delivers a closed form solution under moral hazard, the likelihood ratio enters the candidate mechanism through the Fisher Information. We provide conditions on the Fisher Information and on the disutility of effort that make the candidate mechanism feasible if the agent's outside option is high enough. The two-outcome case also delivers a closed-form solution under moral hazard, and we derive a class of disutility of effort functions that makes decoupling valid. Finally, we show that if the distribution of the signal is an exponential family, then under some plausible conditions on the agent's absolute coefficient of risk aversion and prudence the candidate mechanism obtained via decoupling is optimal. These cases subsume classes of primitives that are commonly used in economic applications of contract theory.

The assumptions underlying some of our sufficiency results are strong. One reason for this is simply that there are several contending forces to control as one tries to understand how the menu varies with the type of the agent. But, it is important to note that the answer to the question of whether decoupling works is *not* always yes. We present an example where it does not work, and provide intuition for its failure. So, while our results establish that in many interesting settings decoupling works, with all of its implications, one cannot simply start from the presumption that a model with both moral hazard and adverse selection is effectively one where the solution has to satisfy the conditions of both, but no more. Results of the form we derive are needed.

The literature on optimal contracts under adverse selection and moral hazard with a risk averse agent at the level of generality that we pursue is small. A closely related predecessor is an unpublished paper by Fagart (2002), who studies the same combination of moral hazard and adverse selection that we analyze, and discusses decoupling and some its implications. Fagart, however, does not tackle the crucial issue of under what conditions a solution to the first-order necessary conditions will also satisfy the global incentive compatibilty constraints, or of what primitives will ensure that the decoupled progam will yield a solution that satisfies these conditions. A second related paper is Gottlieb and Moreira (2013), who analyze a principal-agent problem with moral hazard and adverse selection but where the agent's private information is about the effect of effort on the distribution of a binary signal. In their set up, types are two-dimensional and effort takes on only two possible values, while in our case, types and actions are continuous, and private information is on the disutility of effort. They provide several insights about optimal menus, including distortion, pooling, and exclusion. Given the differences in the environments analyzed, ours and their paper are best viewed as complementary, with both shedding light on contracting problems with combined informational frictions. Finally, we mention the seminal paper by Laffont and Tirole (1986), which derives the optimality of linear contracts in a model with moral hazard and adverse selection with a risk neutral agent, one which also decouples in a straightforward way.[2]

We proceed as follows. After describing the model in Section 2, we analyze three benchmarks in Section 3: the first-best case without adverse selection or moral hazard, the pure adverse selection case, and the pure moral hazard case. For each benchmark we investigate the main features of the optimal contract, especially in the pure adverse selection case, since it does not

[2]There is also an emerging literature on dynamic contracts that in some instances combine adverse selection and moral hazard. Recent examples are Strulovici (2011), Williams (2015), and Halac, Kartik, and Liu (2016). For tractability, they impose more restrictive assumptions on primitives than we do in our general static model.

fit the standard quasilinear setup. For this case we provide a new result showing existence of a solution for a suitably relaxed version of the problem, where our method of proof lays the foundation for numerical analysis of the problem. Section 4 describes the necessary conditions for feasibility, i.e., the first- and second-order conditions that a menu must satisfy. Then we present two theorems stating sufficient conditions for global incentive compatibility. Section 5 describes in detail the decoupling program and the main features of its implied candidate mechanism, which is an optimal one if it happens to be feasible, and analyzes several classes of primitives that guarantee the validity of decoupling. Section 6 concludes the paper. Omitted proofs are in Appendix A. Appendix B establishes the existence of a solution to the relaxed pure adverse selection problem. Appendix C contains some technical results on existence in the pure moral hazard case, and also some differentiability results used in the analysis.

2 The Model

We analyze the following principal-agent problem with moral hazard and adverse selection. The agent has a type $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$, with θ distributed according to a cumulative distribution (cdf) H with positive and continuously differentiable density h. The agent exerts effort $a \geq 0$, possibly with upper bound $\overline{a} < \infty$, where effort has disutility given by $c(a,\theta)$ for every (a,θ). The function c is three times continously differentiable, with $c\left(0,\theta\right) = 0$, $c_a > 0$ for $a > 0$, $c_\theta > 0$, $c_{a\theta} > 0$, $c_{aa} > 0$, $c_{aa\theta} \geq 0$, and $c_{\theta\theta a} \geq 0$, for all (a,θ).

The agent is risk averse with strictly increasing, strictly concave, and thrice continuously differentiable utility function u over income.[3] If his type is θ, exerts effort a, and he obtains wage w, then his total utility is $u(w) - c(a,\theta)$. He has an outside option that yields utility $\bar{u}$. We briefly consider type-dependent reservation utility in Section 5.8.

Neither a nor θ are contractible, since θ is the agent's private information and he exerts effort unobservably. The principal only observes a signal x, where if the agent exerts a, then x is distributed according to a cdf $F\left(\cdot|a\right)$. For the most part we focus on the case where $x \in [\underline{x}, \bar{x}]$ and the cdf has a positive density $f\left(\cdot|a\right)$ that is twice (and at one point thrice) continuously differentiable in a. But we will also consider later the case where x is a discrete random variable and f is its probability distribution. We assume that f satisfies the monotone likelihood ratio property ($MLRP$), so that $l(\cdot|a) \equiv f_a\left(\cdot|a\right)/f\left(\cdot|a\right)$ is increasing in x. To avoid a nonexistence issue à la Mirrlees, we assume that l is bounded.

The principal is risk neutral and her expected utility if the agent exerts effort a and she pays a wage w is $B(a) - w$, where B, the expected benefit the principal derives from the agent's effort, is twice continuously differentiable, increasing, and concave in a. In some settings, the signal is

[3] We use increasing and decreasing in the weak sense of nondecreasing and nonincreasing, adding 'strictly' when needed, and similarly with positive and negative, and concave and convex.

output, and thus it is natural to assume that $B(a)$ is the expectation of x given a. But, in others, x is a signal distinct from the eventual profits that the principal will realize from the agent's effort, and so $B(a)$ need not be tied to the expectation of x.

The contracting problem unfolds as follows. The principal offers a menu of contracts that consists of a pair of functions (π, α), where $\pi : [\underline{x}, \overline{x}] \times [\underline{\theta}, \overline{\theta}] \to \mathbb{R}$ specifies the compensation the agent receives if he announces type θ, and signal x is observed, and $\alpha : [\underline{\theta}, \overline{\theta}] \to [0, \overline{a}]$ recommends an effort level to each type θ. Given the menu, the agent decides whether to accept or reject. If he accepts, then he reports a type θ' to the principal that may or may not be equal to the true type θ and chooses an effort level a' that may or may not equal $\alpha(\theta')$. The realization of x is then observed and the agent is paid $\pi(x, \theta')$. If the agent rejects the menu, then he takes his outside option which gives him utility $\bar{u}$.

Let $v(x, \theta') \equiv u(\pi(x, \theta'))$ be the agent's utility from income when he reports θ' and the observed signal is x, and let $\varphi \equiv u^{-1}$ be the inverse of u, which is strictly convex since u is strictly concave. As is standard in the moral hazard literature, it will be convenient to work with the utility of the compensation scheme instead of the wages. The principal is restricted to measurable functions v such that for each (θ, a), $\int_{\underline{x}}^{\overline{x}} |v(x, \theta)| f(x|a)\, dx$ is finite. We focus on deterministic menus, i.e., for each type θ, the menu specifies an action $a = \alpha(\theta)$ and a function $v(\cdot, \theta)$.[4]

By the extended revelation principle (Myerson (1982)), it is without loss of generality for the principal to restrict attention to menus of contracts that are incentive compatible (the agent reports his true type), and induce obedience (the agent chooses the recommended effort level). For the bulk of the paper, we simplify the exposition by assuming that the principal wishes all types of the agent to participate. We return to this issue briefly later.

The principal's problem is thus the following one:

$$\max_{(\alpha,v)} \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{x}}^{\overline{x}} \left[B(\alpha(\theta)) - \varphi(v(x,\theta))\right] f(x|\alpha(\theta))\, h(\theta)\, dx\, d\theta \tag{P}$$

$$s.t. \quad \int_{\underline{x}}^{\overline{x}} v(x,\theta) f(x|\alpha(\theta))\, dx - c(\alpha(\theta), \theta) \geq \bar{u} \qquad \forall\ \theta \tag{1}$$

$$(\alpha(\theta), \theta) \in \operatorname{argmax}_{(a',\theta')} \int_{\underline{x}}^{\overline{x}} v\left(x, \theta'\right) f\left(x|a'\right) dx - c\left(a', \theta\right) \quad \forall\ \theta. \tag{2}$$

That is, the principal chooses α and v to maximize her expected profit subject to the agent's participation and incentive compatibility constraints. For each type, the agent must be willing to accept the menu, report truthfully, and follow the recommended action. Any menu (α, v) that satisfies (1)–(2) is a *feasible* menu.

Problem (P) is, in general, quite intractable. In particular, one must deal with the double-

[4] We impose this restriction both for reasons of tractability and because in many settings it seems the economically relevant case.

continuum of deviations available to the agent, since he can both lie about his type, and then choose an action different than the one recommended.

3 Benchmark Cases

We begin with three simpler cases, which will serve as benchmarks for comparison. The first one is the complete information setting without adverse selection or moral hazard. The second one is the pure adverse selection case. Since the agent is risk averse, there are a few differences with the standard screening problem that we point out. Finally, we go over the pure moral hazard case.

3.1 The First Best

Consider first the case where both θ and a are observable and thus we do not have the incentive constraints (2). Then it is immediate that v is independent of x (since the agent is risk averse), and that (1) binds for all types (the principal extracts all the surplus). Hence, setting $v(x, \theta) \equiv v(\theta)$ for all θ, we have that $v(\theta) = \bar{u} + c(\alpha(\theta), \theta)$. In turn, α is uniquely determined at each θ by

$$B_a(\alpha(\theta)) = \varphi'(\bar{u} + c(\alpha(\theta), \theta))c_a(\alpha(\theta), \theta). \tag{3}$$

Since $c_{\theta a} > 0$ and $\varphi'' > 0$, it follows that the optimal α is strictly decreasing in θ. Regarding the optimal v, we have $v'(\theta) = dc(\alpha(\theta), \theta)/d\theta = c_a(\alpha(\theta), \theta)\alpha'(\theta) + c_\theta(\alpha(\theta), \theta)$, the sign of which is ambiguous, since as θ rises, the agent must be compensated for a higher disutility of effort for any given effort, but is required to exert lower effort.

3.2 Pure Adverse Selection

Consider now the case in which the action is contractible (no moral hazard) but θ is private information, i.e., the pure adverse selection case. The principal's problem is then

$$\begin{aligned} \max_{\alpha, v} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \left(B(\alpha(\theta)) - \varphi(v(\theta))\right) h(\theta) d\theta \\ \text{s.t.} \quad & v(\theta) - c(\alpha(\theta), \theta) \geq \bar{u} \ \ \forall \theta && (4) \\ & v(\theta) - c(\alpha(\theta), \theta) \geq v(\hat{\theta}) - c(\alpha(\hat{\theta}), \theta) \ \ \forall \theta, \hat{\theta}. && (5) \end{aligned}$$

Given the assumption $c_{a\theta} > 0$, the incentive compatibility constraints (5) imply that (α, v) must be decreasing in θ. Moreover, a standard argument shows that incentive compatibility is equivalent to α decreasing and

$$S(\theta) = S(\overline{\theta}) + \int_{\theta}^{\overline{\theta}} c_\theta(\alpha(s), s) ds, \tag{6}$$

where $S(\theta) = v(\theta) - c(\alpha(\theta), \theta)$. Thus, constraints (4) hold if and only if $S(\overline{\theta}) \geq \bar{u}$, and optimality demands that the participation constraint binds for the least able type, so $S(\overline{\theta}) = \bar{u}$.

To rewrite this problem in a way that will be convenient to us later when we turn to the decoupling program in the setting with both moral hazard and adverse selection, define $\hat{C}(a, u_0, \theta)$ to be the cost of implementing action a when the agent's type is θ and he has to be given a utility level u_0. When there is no moral hazard, v is independent of x as in the first-best case, and so $\hat{C}\,(a, u_0, \theta)$ is simply $\varphi(u_0 + c(a, \theta))$. When we turn to the decoupling program, $\hat{C}$ will reflect the cost, in the context of a Holmstrom-Mirlees style relaxed moral hazard problem, of inducing effort level a from type θ when surplus u_0 must be provided, a much more complicated object.

When the agent is risk neutral (e.g., Guesnerie and Laffont (1984)), $\hat{C}\,(a, u_0, \theta) = c\,(a, \theta) + u_0$, and so $\hat{C}$ is linear in the surplus u_0, where in equilibrium, u_0 is equal to $S\,(\theta)$. One can hence rewrite the objective function to eliminate $S\,(\theta)$, resulting in an integrand that depends at each θ only on the effort level at θ, and then maximize the integrand pointwise with respect to the effort level. But here, because the agent is risk averse, $\hat{C}$ is not linear in u_0. As such, the trade-off involved in asking extra effort from a particular type θ depends on $S\,(\theta)$, which by (6), depends on the effort level schedule for all types to the right of θ. When we turn to the decoupling program, there is, of course, no reason to expect $\hat{C}$ to be linear in u_0.

To proceed, we will formulate the principal's problem as the following optimal control problem:

$$\begin{aligned} \max_{\alpha, S} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \Big(B(\alpha(\theta)) - \hat{C}(\alpha\,(\theta)\,, S(\theta), \theta)) \Big)\, h(\theta) d\theta \\ \text{s.t.} \quad & \alpha \ \text{ decreasing} && (7) \\ & S'(\theta) = -c_\theta(\alpha(\theta), \theta) \ \text{ for almost all } \ \theta && (8) \\ & S(\overline{\theta}) = \bar{u}, \quad S(\underline{\theta}) \ \text{ free} && (9) \end{aligned}$$

Let $\eta(\theta)$ be the co-state variable associated with (8), and ignore constraint (7) (we will check it ex-post). Then the Hamiltonian of this relaxed problem is:

$$\mathcal{H} = \Big(B(\alpha(\theta)) - \hat{C}(\alpha\,(\theta)\,, S(\theta), \theta)) \Big)\, h(\theta) - \eta(\theta) c_\theta(\alpha(\theta), \theta)$$

The optimality conditions are $\partial \mathcal{H} / \partial a = 0$, $\eta'(\theta) = -\partial \mathcal{H} / \partial S$, and the transversality condition $\eta(\underline{\theta}) = 0$. Simple algebra yields that if $\alpha\,(\theta)$ is interior, then it satisfies

$$(B_a(\alpha(\theta)) - \hat{C}_a(\alpha\,(\theta)\,, S(\theta), \theta))h(\theta) = c_{a\theta}(\alpha(\theta), \theta) \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0}(\alpha\,(t)\,, S(t), t))h(t) dt, \tag{10}$$

which reflects the standard efficiency versus information-rents trade-off. We will assume $\hat{C}_{aa} > 0$. In the case of pure adverse selection, this is trivially satisfied, since $\hat{C}_{aa} = \varphi'' c_a^2 + \varphi' c_{aa}$, where φ

is strictly increasing and strictly convex. Given $\hat{C}_{aa} > 0$, and given our other assumptions, the left-hand-side (*lhs*) of (10) is decreasing in a, and the right-hand-side (*rhs*) is increasing in a, and so if (10) yields a solution $\alpha(\theta)$, then this solution is unique.

A standard property of screening problems is that effort is equal to the first best level for the most capable agent $\underline{\theta}$. Here, however, $\hat{C}_a$ will in general depend on u_0. In particular, if $\hat{C}_{au_0} > 0$, as in the simplest case where $\hat{C}(a, u_0, \theta) = \varphi(u_0 + c(a, \theta))$, then even the effort of $\underline{\theta}$ will be distorted downward from the first best. The effort level of less capable agents will be distorted downwards both because $S(\theta) > \overline{u}$ and because the *rhs* of (10) is positive.

The validity of this solution is predicated on two properties. First, we need to know that a solution to equations (8)–(10) exists, and that such a solution does indeed solve the relaxed problem. We address this issue in Appendix B where we show that it is sufficient that $\hat{C}$ is jointly convex in (a, u_0) for each θ and satisfies an appropriate boundary condition at $a = 0$ and $a = \overline{a}$. Our method of proof is constructive and hence points the way to numerical analysis of this problem when (8)–(10) are too complicated to admit a closed form solution.[5]

Second, for the case of pure adverse selection, we need the omitted monotonicity constraint (7) to be satisfied (in the decoupling program, we will see that monotonicity is not enough, and indeed a major focus of our analysis will be to find a tractable replacement for (7)). We now search for sufficient conditions under which this is the case. In fact, we will provide conditions under which α is *strictly* decreasing, and thus the optimal menu completely sorts types.

Totally differentiating (10) with respect to θ yields, after some algebra,

$$\alpha' = \frac{c_{a\theta\theta} \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h + c_{a\theta} \hat{C}_{u_0} h - c_{a\theta} \frac{h'}{h} \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h - \hat{C}_{au_0} c_\theta h + \hat{C}_{a\theta} h}{(B_{aa} - \hat{C}_{aa}) h - c_{aa\theta} \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h}.$$

Thus, given our assumptions, the denominator is negative, and so $\alpha' < 0$ if and only if

$$c_{a\theta\theta} \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h + c_{a\theta} \hat{C}_{u_0} h - c_{a\theta} \frac{h'}{h} \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h - \hat{C}_{au_0} c_\theta h + \hat{C}_{a\theta} h > 0$$

which rearranges to

$$\frac{c_{a\theta\theta}}{c_{a\theta}} - \frac{h'}{h} + \left(\frac{c_{a\theta} \hat{C}_{u_0} + \hat{C}_{a\theta} - \hat{C}_{au_0} c_\theta}{c_{a\theta} \hat{C}_{u_0}} \right) \frac{\hat{C}_{u_0} h}{\int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h} > 0. \tag{11}$$

Since in the pure adverse selection case $\hat{C}(a, u_0, \theta) = \varphi(u_0 + c(a, \theta))$, we have that $\hat{C}_a = \varphi' c_a$ and $\hat{C}_{u_0} = \varphi'$, and hence $\hat{C}_{a\theta} = \varphi'' c_\theta c_a + \varphi' c_{a\theta}$, and $\hat{C}_{au_0} = \varphi'' c_a$. From this, it is easy to verify that

[5] Because our analysis in Appendix B covers a general $\hat{C}$, it subsumes the case of pure adverse selection where the agent's utility is not additively separable in income and effort.

the term in parenthesis is equal to 2, and thus $\alpha' < 0$ for any given θ if and only if

$$\frac{c_{a\theta\theta}}{c_{a\theta}} - \frac{h'}{h} + \frac{2\varphi' h}{\int_{\underline{\theta}}^{\theta} \varphi' h} > 0. \tag{12}$$

We are now ready to prove the following result:

Proposition 1 *(i) There is sorting at the top, i.e., $\alpha'(\theta) < 0$ in a right neighborhood of $\underline{\theta}$;*
(ii) If h is log-concave and $c_{a\theta}$ is log-convex in θ, then $\alpha'(\theta) < 0$ for all θ.

The proof of part (i) is straightforward: (12) clearly holds at $\theta = \underline{\theta}$, for the first two terms are bounded while the last term grows without bound as θ goes to $\underline{\theta}$. Hence, by continuity $\alpha'(\theta) < 0$ in a right neighborhood of $\underline{\theta}$. This shows that at the optimum there is complete sorting at the top. The proof of part (ii) is slightly more involved and is in the Appendix.

We now turn briefly to optimal exclusion. It is immediate from incentive compatibility that if any exclusion occurs, then it occurs for some interval of the form $(\theta^*, \overline{\theta}]$, where evaluated at θ^*, (10) holds for $\alpha(\theta^*) = 0$. It follows that if $c_a(0,\theta) = c_{a\theta}(0,\theta) = 0$ for all θ, then no type is excluded from the optimal contract, since under these conditions, the *lhs* of (10) is strictly positive at $a = 0$, while the *rhs* is 0. In turn, one can show that type θ is excluded if $c_{a\theta}(0,\theta) > 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, and $h(\theta)/H(\theta)$ is sufficiently small.

3.3 Pure Moral Hazard

The last benchmark case is the one where θ is observable but the action is not, i.e., there is moral hazard but no adverse selection. Begin by defining, for any given action a, type θ, and reservation level u_0, $C(a, u_0, \theta)$ as the cost of implementing action a for type θ given that he needs to receive utility level u_0. That is,

$$C(a, u_0, \theta) = \min_{\hat{v}} \int_{\underline{x}}^{\overline{x}} \varphi\left(\hat{v}\left(x\right)\right) f\left(x|a\right) dx \tag{13}$$

$$s.t. \quad \int_{\underline{x}}^{\overline{x}} \hat{v}\left(x\right) f\left(x|a\right) dx - c\left(a, \theta\right) \geq u_0 \tag{14}$$

$$a \in \operatorname{argmax}_{a'} \int_{\underline{x}}^{\overline{x}} \hat{v}\left(x\right) f\left(x|a'\right) dx - c\left(a', \theta\right). \tag{15}$$

We will assume in what follows that the *First-Order Property* (*FOP*) is valid, by which we mean that $\int \hat{v} f$ is concave in a for any $\hat{v}$ that solves the relaxed problem in which (15) is replaced by the condition $\int_{\underline{x}}^{\overline{x}} \hat{v}\left(x\right) f_a\left(x|a\right) dx - c_a\left(a, \theta\right) = 0$. Primitives for this are well understood.[6] Following

[6]See Rogerson (1985), Jewitt (1988), and a large literature that follows thereon. In some of what follows, we will rely on the primitives given by Chade and Swinkels (2016).

Holmstrom (1979) and Mirrlees (1975), if we let λ and μ be the Lagrange multipliers for the participation and the agent's first-order condition in the relaxed problem, we obtain that the cost-minimizing compensation scheme $\hat{v}$ solves

$$\varphi'(\hat{v}\,(x)) = \lambda + \mu l(x|a) \ \forall \ x,$$

where λ and μ are functions of a, u_0, and θ, and where, as usual, since l is increasing in x, so is $\hat{v}$.[7] We denote the compensation scheme that solves this program by $\hat{v}(\cdot, a, u_0, \theta)$.

There are well-known problems regarding the existence of a solution to the relaxed problem. The assumptions made on l rule out one of those issues, namely, the one pointed out by Mirrlees (1975). The other issue, pointed out by Moroni and Swinkels (2014), is driven by the properties of u. In Appendix C we tackle this issue and provide sufficient conditions to guarantee that the problem has a well-defined solution. This appendix also provides a formal justification for the interchanging of differentiation and integration that we perform at several points of the analysis. We henceforth ignore these issues and refer the reader to the appendix for further details.

The solution to the principal's problem is then, for each type θ, to choose a to maximize $B\,(a) - C(a, \bar{u}, \theta)$. Since the type of the agent is observable, i.e., there is no adverse selection, two properties are true. First, there are no information rents, and so each type of agent receives $\bar{u}$. Second, the problem separates in the sense that there is no interaction between the moral hazard problem the principal faces at one type versus another. The thrust of much of what follows is to explore when a version of this sort of decoupling is valid even with adverse selection.

In what follows, we will assume that $C_{aa} > 0$. It is intuitive that this should be true, but primitives are not trivial to find. One such set is provided by Jewitt, Kadan, and Swinkels (2008). Given $C_{aa} > 0$, we have that a is implicitly defined as a function of $\bar{u}$ and θ, by $B_a - C_a = 0$. In particular, if $C_{a\theta} > 0$, then the optimal action the principal implements is decreasing in θ, and if $C_{au_0} > 0$, then the optimal action decreases in $\bar{u}$. The second result is significant, because as we will see, given decoupling, the principal will indeed treat each type θ as facing a pure moral hazard problem, but one in which the agent, except of the highest type, has information rents, and hence has to be given utility higher than $\bar{u}$. The implication is that when moral hazard is combined with screening, the principal will distort effort *down* from the pure moral hazard case.

Given this, let us dive for a moment into when C_{au_0} and $C_{a\theta}$ are positive. The Envelope Theorem yields $C_{u_0} = \lambda$, and hence, since the Lagrange multipliers are continuously differentiable (see Jewitt, Kadan, and Swinkels (2008)), $C_{u_0 a} = \lambda_a$. Similarly, $C_\theta = \lambda c_\theta + \mu c_{a\theta}$, and hence $C_{\theta a} = \lambda_a c_\theta + \lambda c_{\theta a} + \mu_a c_{a\theta} + \mu c_{aa\theta}$, where we note that given our assumptions, it is thus sufficient

[7] For some utility functions, as for example, $u = \sqrt{w}$, there is the implicit constraint $w \geq 0$, and for some specificiations of the outside option and cost function, this constraint will bind at the optimal contract. While our analysis could be extended to this case, this adds complexity which is distracting for the purposes of this paper. For $\bar{u}$ sufficiently large, this constraint will not bind, and we will henceforth assume that this is so.

to show that each of λ_a and μ_a are positive. The next lemma provides sufficient conditions for each of these to be true. Before stating it, we need a definition. Let ψ map $1/u'$ into money. That is, ψ solves $1/u'(\psi(\tau)) = \tau$. Define also $\rho(\tau) = u(\psi(\tau))$, so ρ carries $1/u'$ into utility.

Lemma 1 *Let l be submodular in x and a, i.e., $l_{xa} \leq 0$. Then, $\mu_a \geq 0$. If in addition f is log-concave in a and ρ is concave, then $\lambda_a \geq 0$ as well.*

4 Adverse Selection and Moral Hazard

We now return to the general problem with both moral hazard and adverse selection. The main difficulty is the lack of a sweeping incentive compatibility characterization as in the pure adverse selection case, where things boil down to the monotonicity and integral condition, or the pure moral hazard case, where one can usefully simplify the problem using the first-order approach.

The roadmap of our analysis of the general problem is as follows. In this section, we will first derive some necessary conditions that feasible menus must satisfy, and then derive a set of sufficient conditions for feasibility. In the next section, we introduce the decoupling method to simplify the analysis of the problem, which consists of solving a suitably relaxed version of the problem. A crucial issue is to check that the solution to the decoupled problem satisfies the sufficient conditions for feasibility. We show that this is the case in a number of classes of problem and derive implications for optimal menus.

4.1 Feasible Menus: Necessary Conditions

Consider a candidate menu (α, v). We begin by deriving a set of necessary conditions for (α, v) to be feasible. To do so, we will need to assume that a variety of derivatives are well-defined at points where α is differentiable.

Definition 1 *A menu (α, v) is regular if (i) everywhere that α is differentiable in θ, so is v; (ii) for all (x, θ), $\overline{v}(x, \theta) \equiv \lim_{\varepsilon \downarrow 0} v(x, \theta - \varepsilon)$ and $\underline{v}(x, \theta) \equiv \lim_{\varepsilon \downarrow 0} v(x, \theta + \varepsilon)$ are well defined.*

We will henceforth assume regularity. At many places in what follows, regularity is more than we need, but imposing it simplifies and clarifies the exposition. When we turn to the decoupling program in the next section, one thing we will do is provide primitives for the candidate (α, v) generated by the decoupling program to be regular.

In this section and for the sake of simplicity, we will also assume that anywhere that α is differentiable each of $\int v f_a$, $\int v_\theta f$, $\int v_\theta f_a$, $\int v_{\theta\theta} f$ and $\int v f_{aa}$ are well defined and finite.[8]

Let $S(\theta) = \int v(x, \theta) f(x|\alpha(\theta)) dx - c(\alpha(\theta), \theta)$ be the surplus of the agent with type θ.

[8]See Appendix C for conditions that ensure the validity of passing the derivative through the integral.

Lemma 2 *Any feasible menu* (α, v) *satisfies, for all* θ,

$$\int v(x,\theta) f_a(x|\alpha(\theta))dx = c_a(\alpha(\theta),\theta) \tag{16}$$

$$S(\theta) = S(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} c_{\theta}(\alpha(s),s)ds. \tag{17}$$

If α *is continuous at* θ *then*

$$\int v_{\theta}(x,\theta) f(x|\alpha(\theta))dx = 0. \tag{18}$$

Proof Equation (16) is the first-order condition for the agent's choice of effort, while (18) is the one for his choice of what type to report. They ensure that locally, the agent does not want to report truthfully but deviate from the recommended action, or misreport his type while following the recommended action. Condition (17) captures that for the agent not to want to locally misrepresent his type and change his action along the locus $(\theta, \alpha(\theta))$, it must be that $S'(\theta) = -c_{\theta}(\alpha(\theta),\theta)$. Formally, since c is continuously differentiable in (a,θ), the conditions of Theorem 2 of Milgrom and Segal (2002) hold, and so S is absolutely continuous and hence differentiable a.e. and thus satisfies the integral condition (17). $\square$

We will show in a moment that α is strictly decreasing, with $\alpha' < 0$ at any point of differentiability. It can easily then be shown that wherever α is differentiable, any two of (16)–(18) imply the third. This is intuitive. The agent chooses his report and action within the two dimensional space of actions cross reports and (using that $\alpha' \neq 0$) each of the three directions of local deviation (change action holding fixed report, change report holding fixed type, and change both action and report along the locus $(\alpha(\cdot), \cdot)$) can be expressed as a linear combination of the other two.

Another necessary condition for feasibility is that α is decreasing in θ. The proof is standard: for any two types $\theta' > \theta$ incentive compatibility implies

$$\int v(x,\theta') f(x|\alpha(\theta'))dx - c(\alpha(\theta'),\theta') \geq \int v(x,\theta) f(x|\alpha(\theta))dx - c(\alpha(\theta),\theta'), \text{ and} \tag{19}$$

$$\int v(x,\theta) f(x|\alpha(\theta))dx - c(\alpha(\theta),\theta) \geq \int v(x,\theta') f(x|\alpha(\theta'))dx - c(\alpha(\theta'),\theta). \tag{20}$$

Adding these inequalities yields $c(\alpha(\theta'),\theta') + c(\alpha(\theta),\theta) \leq c(\alpha(\theta'),\theta) + c(\alpha(\theta),\theta')$, which, since c is supermodular in (a,θ), implies that $\alpha(\theta') \leq \alpha(\theta)$.

Now, let us turn to the second-order conditions for local optimality. Differentiating the agent's objective function twice and evaluating the derivatives at the feasible menu, we obtain the following matrix:

$$G = \begin{bmatrix} \int v_{\theta\theta}(x,\theta) f(x|\alpha(\theta))\, dx & \int v_{\theta}(x,\theta) f_a(x|\alpha(\theta))\, dx \\ \int v_{\theta}(x,\theta) f_a(x|\alpha(\theta))\, dx & \int v(x,\theta) f_{aa}(x|\alpha(\theta))\, dx - c_{aa}(\alpha(\theta),\theta) \end{bmatrix}.$$

The second-order conditions require that the diagonal elements of G are negative, and the determinant of G is positive. A first implicaton of the second-order conditions is that α' is in fact strictly negative anywhere that α is differentiable, and so in particular, α is strictly decreasing.

Lemma 3 *Let (α, v) be feasible. Then, anywhere that α is differentiable, $\alpha' < 0$.*

Proof Given that feasibility implies that α is decreasing, we need only to rule out that $\alpha' = 0$. To see this, begin by noting that (16) and (18) are identities in θ. Differentiating (18) yields

$$\int v_{\theta\theta}(x,\theta) f(x|\alpha(\theta))\,dx = -\alpha'(\theta)\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx. \tag{21}$$

Hence, if $\alpha' = 0$, then $\int v_{\theta\theta}(x,\theta) f(x|\alpha(\theta))\,dx = 0$. Similarly, differentiating (16) yields

$$\alpha'\left(\int v(x,\theta) f_{aa}(x|\alpha(\theta))\,dx - c_{aa}(\alpha(\theta),\theta)\right) = c_{a\theta}(\alpha(\theta),\theta) - \int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx, \tag{22}$$

and so if $\alpha' = 0$, then, $\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx = c_{a\theta}(\alpha(\theta),\theta) > 0$. But then,

$$\det G = -\left(c_{a\theta}(\alpha(\theta),\theta)\right)^2 < 0,$$

violating the second-order conditions. □

We can now provide a simple characterization of the second order necessary conditions.

Lemma 4 *Let (α, v) be a menu satisfying the first-order conditions, and let α be differentiable at θ. Then, the second-order necessary conditions are satisfied at $(\theta, \alpha(\theta))$ if and only if*

$$\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx \leq 0. \tag{23}$$

Proof Since $\alpha' < 0$, it follows from (21) that $\int v_{\theta\theta}(x,\theta) f(x|\alpha(\theta))\,dx \leq 0$ if and only if (23) holds. Similarly, from (22), $\int v f_{aa} - c_{aa} \leq 0$ when evaluated at (v, α) if and only if $c_{a\theta} - \int v_\theta f_a \geq 0$, and since c is supermodular in (a, θ), it suffices for this that (23) holds.

Finally, note that using (21) and (22) we can rewrite G as

$$G = \begin{bmatrix} -\alpha'(\theta)\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx & \int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx \\ \int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx & \frac{1}{\alpha'(\theta)}\left(c_{a\theta}(\alpha(\theta),\theta) - \int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx\right) \end{bmatrix},$$

and so $\det G = -c_{a\theta}(\alpha(\theta),\theta)\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx$. Since $c_{a\theta} > 0$, the determinant is nonnegative if and only if $\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx \leq 0$, thereby completing the proof. □

The second-order conditions thus pivot entirely around whether $\int v_\theta(x,\theta) f_a(x|\alpha(\theta))\,dx \leq 0$. One way to interpret this condition is that if a given type θ who truthfully reports his type were to raise his announcement by a little bit, then the contract he would face would change in such a way as to weaken his marginal incentive to work.

Note that by differentiating (16), one arrives at

$$\int v_\theta (x,\theta)\, f_a (x|\alpha(\theta))\, dx = \frac{d}{d\theta}\left(c_a (\alpha(\theta),\theta)\right) - \alpha'(\theta) \int v(x,\theta)\, f_{aa}(x|\alpha(\theta))\, dx. \tag{24}$$

By Lemma 4, the *lhs* of (24) is negative. But then, we can rearrange the *rhs* of (24) to obtain

$$\alpha'(\theta) \leq \frac{c_{a\theta}(\alpha(\theta),\theta)}{\int v(x,\theta) f_{aa}(x|\alpha(\theta))dx - c_{aa}(\alpha(\theta),\theta)} < 0. \tag{25}$$

Hence unlike the pure adverse selection case, not only must α be strictly decreasing, but α' is bounded away from 0.

Inequality (25) helps to reinforce the point that, absent the decoupling procedure we analyze below, the problem is likely to be very difficult. Note in particular that the object $\int v(x,\theta)f_{aa}(x|\alpha(\theta))dx$ depends on the fine details of how v and f_{aa} vary with x.

4.2 Feasible Menus: Sufficient Conditions

Thus far we have focused on necessary conditions for (α, v) to be feasible. Assume that a candidate menu (α, v) satisfies that α is strictly decreasing, that the first-order conditions (16) – (18) hold, that $S(\overline{\theta}) \geq \bar{u}$, and the *FOP*.[9] In this section, we provide two sufficient conditions from which one can conclude that (α, v) is globally incentive compatible, and hence feasible.

Let the graph of α be denoted by $L \equiv \{(\theta, \alpha(\theta)) : \theta \in [\underline{\theta}, \overline{\theta}]\}$. We begin by showing that 'on locus' (i.e., on the graph of α) type θ is strictly harmed by misstating his type.

Lemma 5 *For all $\hat{\theta} \neq \theta$, $\int v(x,\hat{\theta})f(x|\alpha(\hat{\theta}))dx - c(\alpha(\hat{\theta}),\theta) < S(\theta)$.*

Proof Assume wlog that $\hat{\theta} > \theta$. Then using (17) we obtain

$$S(\theta) - S(\hat{\theta}) = \int_\theta^{\hat{\theta}} c_\theta(\alpha(s), s)ds > \int_\theta^{\hat{\theta}} c_\theta(\alpha(\hat{\theta}), s)ds = c(\alpha(\hat{\theta}),\hat{\theta}) - c(\alpha(\hat{\theta}),\theta),$$

where the inequality follows since α is strictly decreasing and c is strictly supermodular in (a, θ). Hence,

$$S(\theta) > S(\hat{\theta}) + c(\alpha(\hat{\theta}),\hat{\theta}) - c(\alpha(\hat{\theta}),\theta) = \int v(x,\hat{\theta})f(x|\alpha(\hat{\theta}))dx - c(\alpha(\hat{\theta}),\theta),$$

and so, on locus, the agent is strictly best off to report his true type. □

Our first sufficient condition for feasibility focuses on $c_a(\alpha(\theta),\theta)$, the marginal cost of effort of a compliant agent as a function of his type. Say that the menu (α, v) satisfies the *diminishing*

[9] Recall that a menu (α, v) satisfies the *FOP* if $\int v(x,\theta)f_a(x|\cdot)dx$ decreases in a for all θ.

marginal cost condition (DMC) if $c_a(\alpha(\cdot),\cdot)$ is decreasing. As θ increases, there are two forces. First, holding fixed effort, since c is supermodular, c_a increases. But, since α is strictly decreasing in θ and c is strictly convex in a, there is also a force in the direction of c_a decreasing. Thus, *DMC* holds if and only if α decreases fast enough that the second term overwhelms the first. *DMC* captures one specific sense of the idea that less capable agents face weaker marginal incentives for effort. Our first sufficiency theorem states that if *DMC* holds, then in fact feasibility holds.

Theorem 1 *If a menu* (α, v) *satisfies DMC, then it is feasible.*

We will describe here how the proof works when α is continuous. The proof in Appendix A allows for the possibility that α has (downward) discontinuities.

Given Lemma 5, it suffices to show that for any off-locus deviation $(\hat{\theta},\hat{a}) \notin L$, there is an on-locus point that the agent at least weakly prefers. We focus on deviations with $\hat{a} > \alpha(\hat{\theta})$ (the other case is similar). Assume first that $\hat{a} > \alpha(\underline{\theta})$. Then, by *FOP*, type $\underline{\theta}$ prefers $\alpha(\underline{\theta})$ to $\hat{a}$. But then, whatever the true type $\theta_T \geq \underline{\theta}$ of the deviating agent, the agent, with a higher marginal cost of effort, will *a fortiori* prefer $\alpha(\underline{\theta})$ to $\hat{a}$. We can hence restrict attention to the case where $\hat{a} \in (\alpha(\hat{\theta}), \alpha(\underline{\theta})]$, as illustrated in Figure 1.

For $a \in \left[\alpha\left(\bar{\theta}\right), \alpha\left(\underline{\theta}\right)\right]$, let $\gamma(a)$ be the inverse of α. We will show that, holding fixed the action $\hat{a}$, the agent is better off to announce type $\gamma(\hat{a})$ than type $\hat{\theta}$, establishing the needed domination of $(\hat{a},\hat{\theta})$ by an on-locus point. To do so, start from the point $(\hat{\theta},\alpha(\hat{\theta}))$ in Figure 1, and compare moving vertically versus moving along the locus L as one raises a from $\alpha(\hat{\theta})$ to $\hat{a}$. At any intermediate a, we have that the utility of income of the agent as we move vertically is rising with a at rate

$$\int v(x,\hat{\theta}) f_a(x|a)\,dx \leq \int v(x,\hat{\theta}) f_a(x|\alpha(\hat{\theta}))dx = c_a(\alpha(\hat{\theta}),\hat{\theta}),$$

where the inequality follows by FOP. In comparison, as a increases along the locus, we have that the utility from income rises at rate

$$\begin{aligned}
\frac{d}{da}\int v(x,\gamma(a)) f(x|a)\,dx &= \gamma'(a)\int v_\theta(x,\gamma(a)) f(x|a)\,dx + \int v(x,\gamma(a)) f_a(x|a)\,dx \\
&= \int v(x,\gamma(a)) f_a(x|a)\,dx \\
&= c_a(a,\gamma(a)),
\end{aligned}$$

where the first equality follows by (18), and the second by (16). But, by hypothesis, since $a > \alpha(\hat{\theta})$, and hence $\gamma(a) < \hat{\theta}$, we have $c_a(a,\gamma(a)) \geq c_a(\alpha(\hat{\theta}),\hat{\theta})$. Thus, the utility of income of the agent rises faster with a as one moves along the locus than vertically from the locus, and we have

$$\int v(x,\gamma(\hat{a})) f(x|\hat{a})\,dx \geq \int v(x,\hat{\theta}) f(x|\hat{a})\,dx.$$

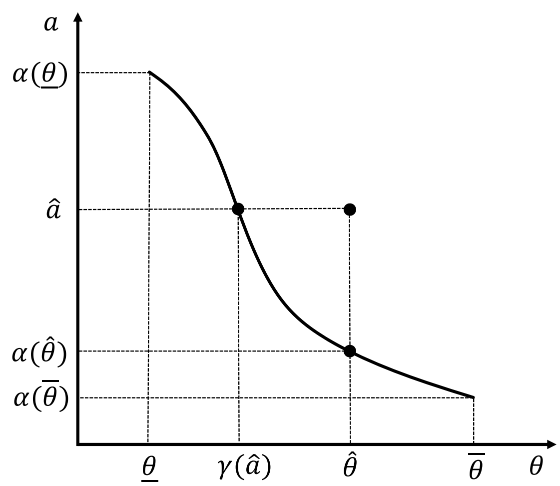

Figure 1: **DMC with continuous** α. Under DMC, a deviation by θ_T to $(\hat{\theta}, \hat{a})$ is dominated by the on-locus deviation $(\gamma(\hat{a}), \hat{a})$, which in turn is worse than telling the truth and being obedient.

Figure.pdf

Since the agent incurs cost of effort $c(\hat{a}, \theta_T)$ in either case, the agent is better off to announce type $\gamma(\hat{a})$ than type $\hat{\theta}$.

How much slack is in condition DMC? For the case with $f_{aa} = 0$, there is none whatsoever, as the following corollary shows.

Corollary 1 *Let $f_{aa}(x|a) = 0$ for all (x, a). Then (α, v) is feasible if and only if DMC holds.*

Proof The if direction follows from Theorem 1. For necessity, note that since (α, v) is feasible, α is decreasing and hence differentiable almost everywhere. Where α is discontinuous, and hence drops, c_a clearly drops as well. Thus, it is enough to show that where α is differentiable $dc_a(\alpha(\theta), \theta)/d\theta \leq 0$. But since (16) holds as an identity in θ, we can differentiate both sides to arrive at

$$\frac{d}{d\theta} c_a(\alpha(\theta), \theta) = \frac{d}{d\theta} \int v(x, \theta) f_a(x|\alpha(\theta)) dx = \int v_\theta(x, \theta) f_a(x|\alpha(\theta)) dx \leq 0,$$

where the last inequality follows from $f_{aa} = 0$ and the second-order condition using Lemma 4. □

An alternative sufficient condition for feasibility is based on a single crossing property of v.

We say that v satisfies the *single crossing condition* (*SCC*) if for all $\theta > \theta'$, $v(\cdot,\theta')$ single-crosses $v(\cdot,\theta)$ from below. This can again be interpreted as a condition that as θ increases, incentives get weaker. Neither *SCC* nor *DMC* implies the other. First, it may well be that *DMC* holds, but that the contracts at different θ cross more than once. But, conditional on knowing that contracts at different θ cross at only once, *SCC* is weaker than *DMC*. To see this, consider for simplicity the differentiable case, and assume that *SCC* fails. Since contracts cross at most once by hypothesis, v_θ has sign pattern $-/+$ (strictly), and so, since by (18) , $\int v_\theta(x,\theta) f(x|\alpha(\theta)) = 0$, and since $\frac{f_a}{f}$ is increasing, it follows from Theorem II in Beesack (1957) (henceforth Beesack's inequality), that

$$\int v_\theta(x,\theta) f(x|\alpha(\theta)) \frac{f_a(x|\alpha(\theta))}{f(x|\alpha(\theta))} dx = \int v_\theta(x,\theta) f_a(x|\alpha(\theta)) dx > 0.$$

But, noting that

$$\int v(x,\theta) f_a(x|\alpha(\theta)) dx = c_a(\alpha(\theta),\theta),$$

holds as an identity, we have (since $\alpha' < 0$, and by *FOP*)

$$(c_a(\alpha(\theta),\theta))_\theta = \int v_\theta(x,\theta) f_a(x|\alpha(\theta)) dx + \alpha' \int v(x,\theta) f_{aa}(x|\alpha(\theta)) dx > 0.$$

Hence, conditional on a single crossing, if *SCC* fails, then so does *DMC*.

Our next theorem establishes that *SCC* also implies feasibility. We will find this especially useful when we turn to the decoupling program and exponential families, a class of distributions for which contracts can only cross in simple ways as θ varies.

Theorem 2 *If v satisfies SCC, then it is feasible.*

As before, we describe the proof here for the case that α is continuous, and refer the reader to Appendix A for a proof of the case where α can jump downward. As for the previous theorem, consider a deviation $(\hat{\theta},\hat{a})$, where $\hat{a} \in (\alpha(\hat{\theta}), \alpha(\underline{\theta})]$. Now, for any $\theta \in (\gamma(\hat{a}), \hat{\theta}]$ the payoff to the agent, holding fixed his action at $\hat{a}$, changes at rate $\int v_\theta(x,\theta) f(x|\hat{a}) dx$ with his announcement. If this is everywhere negative then the agent is at least as well off to pair the action $\hat{a}$ with announcement $\gamma(\hat{a})$, and we have again established an on-locus point that dominates $(\hat{\theta},\hat{a})$. But, note that $\int v_\theta(x,\theta) f(x|\alpha(\theta)) dx = 0$ by (18), and that $v_\theta(\cdot,\theta)$ has sign pattern $+/-$ by *SCC*. It follows from Beesack's inequality that

$$\int v_\theta(x,\theta) f(x|\hat{a}) dx = \int v_\theta(x,\theta) \frac{f(x|\hat{a})}{f(x|\alpha(\theta))} f(x|\alpha(\theta)) dx \leq 0,$$

using that $\hat{a} > \alpha(\theta)$, and so $\frac{f(\cdot|\hat{a})}{f(\cdot|\alpha(\theta))}$ is an increasing function of x.

5 Decoupling

5.1 The Decoupling Formulation

The sufficient conditions for feasibility of the previous section provide an important tool for checking whether a candidate mechanism satisfies all the incentive compatibility constraints embedded in the principal's contract design problem. But problem (13) is still an exceedingly complex one to solve. In this section we describe a decoupling procedure which, when it works, not only makes the analysis tractable but also yields a number of insights about the optimal mechanism.

To this end, consider again the principal's problem (13). Denote by G the set of feasible mechanisms, i.e., those that satisfy the global incentive constraints (1)–(2). Recall that $C(a, u_0, \theta)$ is the cost, in the pure moral hazard problem, of implementing action a by type θ with outside option u_0, and that $\hat{v}(\cdot, a, u_0, \theta)$ is the associated compensation scheme. Then, the principal's optimized profits are

$$\max_{(\alpha,v)\in G} \int \left(B(\alpha(\theta)) - \int \varphi(v(x,\theta)) f(x|\alpha(\theta)) dx \right) h(\theta) d\theta \leq \max_{\alpha} \int \left(B(\alpha(\theta)) - C(\alpha(\theta), S(\theta), \theta) \right) h(\theta) d\theta, \tag{26}$$

where $S(\theta) = \bar{u} + \int_{\theta}^{\bar{\theta}} c_{\theta}(\alpha(s), s)\, ds$, and where our notation supresses the fact that S depends on the function α. To see this inequality, note we have already shown that any feasible mechanism must satisfy (16), so that the agent does not want to locally change his action having honestly reported his type, and (17), so that $S(\theta) = \bar{u} + \int_{\theta}^{\bar{\theta}} c_{\theta}(\alpha(s), s)\, ds$ (where we recall that any optimal solution will set $S(\bar{\theta}) = \bar{u}$). But, $C(\alpha(\theta), S(\theta), \theta)$ is precisely the cost of the optimal compensation scheme satisfying these two conditions for type θ.

The *rhs* expression effectively nests a moral hazard problem within an adverse selection one. As such, it suggests the following three-step *decoupling* scheme.

Step 1(Moral Hazard): Cost Minimization for Each (a, u_0, θ). For each type θ, action a, and utility u_0, solve the moral-hazard problem as described in Section 3.3, thus constructing the function C. Note that by Kadan and Swinkels (2013), finding the optimal $\hat{v}$ for any given (a, u_0, θ) is the solution to a simple numerical procedure, and hence constructing C is numerically straightforward when a closed form solution is not available, as is checking whether C is convex in a, as our analysis requires.

Of course, for decoupling to work, it has to be that the moral hazard subproblem has a solution for each θ, which we assume here but it is formally justified in Appendix C.

Step 2 (Adverse Selection): Profit Maximization with Respect to α. Having constructed the cost function C, solve the relaxed pure adverse selection problem

$$\max_{\alpha} \int \left(B(\alpha(\theta)) - C(\alpha(\theta), S(\theta), \theta) \right) h(\theta) d\theta,$$

subject to the conditions (8)–(9) on the surplus function S, but where we initially ignore any condition on α. Hence we are back to the problem analyzed in Section 3.2. This problem is again amenable to numerical solution.[10]

Step 3 (Verification): Construct and Verify The Feasibility of the Candidate Solution. For each θ, let $v(\cdot,\theta) = \hat{v}(\cdot,\alpha(\theta), S(\theta,\alpha),\theta)$. By (26), if the mechanism (α, v) so constructed satisfies feasibility, then it solves the general problem. In particular, this will be true if the candidate solution satisfies either of the sufficient conditions *DMC* or *SCC*.

It is easy to see that decoupling need not work in every setting. In particular, DMC, which is necessary when $f_{aa} = 0$, is arbitrarily demanding when c_{aa} is small. Consider for example, the cost function $c(a,\theta) = \theta a^{1+\omega}$, for which we require that $\alpha'(\theta) \leq -\alpha(\theta)/\theta\omega$. For ω small, this bound explodes.[11] But this is not the only reason why decoupling may fail. In Section 5.6 below, we present a fully worked example that does not depend on c being near-linear in a, but in which decouping indeed fails.

5.2 Decoupling and the Properties of Optimal Menus

In the next section we provide classes of problems where decoupling works. But before doing so, we will go over some properties that optimal menus under decoupling satisfy.

We first show that the menu obtained via decoupling is regular, as defined in Section 4.2, as long as α is strictly decreasing. This property will follow as a corollary to the next two lemmata.

Lemma 6 *The functions $\hat{v}(x,a,u_0,\theta)$, and $C(a,u_0,\theta)$ are twice continuously differentiable in their arguments.*

Lemma 7 *Assume $C_{aa} > 0$. Then, α is continuously differentiable.*

Corollary 1 *Let (α, v) be a solution to the decoupling problem. If $C_{aa} > 0$, and α is strictly decreasing, then (α, v) is regular.*

Proof This follows from Lemmas 6 and 7, since for all (x,θ), $v(x,\theta) = \hat{v}(x,\alpha(\theta),S(\theta,\alpha),\theta)$. □

Since the previous result requires that α be strictly decreasing in θ, we start by providing sufficient conditions for this to be true.

Lemma 8 *Suppose that h is decreasing and l is submodular in (a,x). Then the function α that solves the decoupling problem is strictly decreasing.*

[10] In a numerical implementation of this procedure, one could solve for C on an "as-needed" basis as one solves the adverse selection problem.

[11] The problem here is not that C is becoming near linear in the effort implemented. In the square-root utility case discussed below, for example, C will typically be convex in effort even if $f_{aa} = c_{aa} = 0$.

Proof From the argument given in the pure adverse selection case, we know that the function α delivered by the second step of the decoupling argument is strictly decreasing if and only if

$$\frac{h'}{h} - \frac{c_{a\theta\theta}}{c_{a\theta}} - \frac{(c_{a\theta}C_{u_0} + C_{a\theta} - C_{au_0}c_\theta)\, h}{c_{a\theta}\int_{\underline{\theta}}^{\theta} C_{u_0} h ds} < 0.$$

Since h is decreasing and given our assumptions on c, it follows that a sufficient condition for $\alpha' < 0$ is $D \equiv c_{a\theta}C_{u_0} + C_{a\theta} - C_{au_0}c_\theta \geq 0$. We will provide conditions for this to hold.

Using repeatedly the Envelope Theorem on the cost minimization problem in step 1 above yields $C_{u_0} = \lambda$, $C_{u_0 a} = \lambda_a$, and $C_{\theta a} = \lambda_a c_\theta + \lambda c_{\theta a} + \mu_a c_{a\theta} + \mu c_{aa\theta}$. Thus,

$$D = c_{a\theta}C_{u_0} + C_{a\theta} - C_{au_0}c_\theta = 2\lambda c_{a\theta} + \mu c_{aa\theta} + \mu_a c_{a\theta}.$$

But $\lambda \geq 0$ and $\mu \geq 0$, and since l is submodular, Lemma 1 implies that $\mu_a \geq 0$. Hence, $D \geq 0$. $\square$

Lemma 8 holds if $l_{xa} = 0$, as in the oft-used class $F(x|a) = G^a(x)$, $a \in [0, 1]$. It also holds in the exponential case, $f(x|a) = e^{-x/a}/a$, $x > 0$, $a > 0$, which has $l_{xa} = -2/a^3 < 0$.

Next, we show that under some conditions, the shadow value of the participation constraint (14) is monotonically decreasing in θ.

Lemma 9 *Assume that f is log-concave in a, l is submodular in (x, a), ρ is concave, and $\alpha' \leq -c_{a\theta}/c_{aa}$. Then $C_{u_0}(S(\cdot), \cdot, \alpha(\cdot))$ is decreasing in θ.*

Besides regularity, the decoupling menu, when optimal, exhibits a number of interesting properties. Most centrally is the simple point that, for each θ, the resulting compensation scheme v is a Holmstrom-Mirlees contract. Thus, v has the same standard properties such as monotonicity and curvature pinned down by the behavior of the likelihood ratio, solving the standard risk sharing versus incentives trade-off. At the same time, all types except for the highest one obtain an information rent, which is pinned down exactly as in the pure adverse selection case, i.e., due to the incentive compatibility problem of inducing the agent to reveal his type.

Regarding α, notice that it exhibits the decreasing property of the pure adverse selection case, but in a strict sense to account for the moral hazard problem, the presence of which enlarges the agent's set of possible deviations. Suppose that C is convex in a (sufficient conditions for this appear in Jewitt, Kadan, and Swinkels (2008)). Since every type except for the highest obtains utility higher than $\bar{u}$, it follows that under the conditions in Lemma 1 (which imply that $C_{u_0 a} \geq 0$) and from the first-order condition in Step 2 that the action of each type implemented by the principal is lower than that in the pure moral hazard case (where all types obtain $\bar{u}$). That is, the presence of adverse selection lowers the effort asked from the agent.

The analysis of exclusion for the pure adverse selection case can be adapted to the combined case as well. In particular, if the principal excludes some types, then he does so for an interval of the highest ones, i.e., the least efficient ones. Moreover, if C is convex in a, then from the first-order condition of Step 2, type θ is excluded if and only if

$$B_a(0)h(\theta) < c_{a\theta}(0,\theta)\int_{\underline{\theta}}^{\theta} C_{u_0}(S(t),t,\alpha(t))h(t)dt.$$

If $1/u'(w)$ is convex in w, then it follows from standard arguments that

$$C_{u_0} = \lambda = \int \frac{1}{u'(\pi(x))} f(x|a)\,dx \geq \frac{1}{u'\left(\int \pi(x) f(x|a)\,dx\right)} \geq \frac{1}{u'(\varphi(\bar{u}))},$$

and hence if $c_{a\theta}(0,\theta) > 0$, and $h(\theta)/H(\theta)$ is sufficiently small, then type θ will be excluded.

5.3 Decoupling under Square Root Utility

It is in general difficult to obtain a closed form expression for C from the cost minimization problem. One case that is tractable and can serve as a testbed for our decoupling procedure is when the agent has square-root utility $u(w) = \sqrt{2w}$, which has the convenient property that $1/u' = u$ and $\varphi(v) = v^2/2$. In this case we obtain the following cost function:

$$C(a,u_0,\theta) = \frac{1}{2}\left((u_0 + c(a,\theta))^2 + \frac{c_a^2(a,\theta)}{I(a)}\right), \tag{27}$$

where $I(a) = \int l^2 f$ is the Fisher Information that the output carries about a. The v associated with this cost function is given by[12]

$$\hat{v}(x,a,u_0,\theta) = u_0 + c(a,\theta) + \frac{c_a(a,\theta)}{I(a)} l(x|a).$$

We are now ready for the following result:

Proposition 2 *Assume $u(w) = \sqrt{2w}$, $\bar{a} < \infty$, I twice continuously differentiable and bounded away from zero with $1/I$ increasing and convex, $c_a c_{aa}$ increasing in a for each θ, and $c_{a\theta}/(c_{aa}h)$ is increasing in θ for all a. Then there is a threshold $\bar{u}^*$ such that for all $\bar{u} > \bar{u}^*$ the decoupling problem delivers an optimal solution to the principal's problem.*

The proof of this proposition centers on DMC. The assumptions made also guarantee that C is convex in a, since $1/I$ increasing and convex and $\partial(c_a c_{aa})/\partial a \geq 0$ for all (a,θ) imply that c_a^2/I is convex in a, and thus (27) is a sum of convex functions and is therefore convex.

The following is a simple example that satisfies all the conditions of the proposition:

[12]Substituting $v = \lambda + \mu l$ into the constraints yields $\lambda = u_0 + c$ and $\mu = c_a/I$. Inserting these expressions into v and then into the objective function gives C.

Example 1 *Let* $u(w) = \sqrt{2w}$*,* $\overline{a} < \infty$*, h uniform, and* $c(a, \theta) = a\theta + 0.5a^2$*,* $f(x|a) = e^{-x/(a+\kappa)}/(a+\kappa)$*,* $\kappa > 0$*. Then the conditions of Proposition 2 are all satisfied. Moreover,* $I(a) = 2/(a+\kappa)^2$ *and hence* $I_a = -4/(a+\kappa)^3 < 0$ *and* $I_a/I^2 = -(a+\kappa)$*, which is decreasing in a. As a result, for* $\bar{u}$ *sufficiently large, decoupling yields an optimal solution to the principal's problem.*

We have shown that for large enough $\bar{u}$ it must be the case that $dc_a(\alpha(\theta), \theta)/d\theta \leq 0$. If in addition l_x/I increases in a, (which holds, for example, if $I_a \leq 0$ and l is supermodular) then as a by-product we obtain that $\hat{v}_{x\theta} \leq 0$, so that in a very strong sense, higher types receive 'lower-powered' contracts. To see this, differentiate (42) with respect to x and then θ to obtain

$$\hat{v}_{x\theta} = (c_{aa}\alpha' + c_{a\theta})\frac{l_x}{I} + c_a\left(\frac{l_x}{I}\right)_a \alpha',$$

where both terms are negative by hypothesis. It can be checked that in the example above, $(l_x/I)_a = 0$ and hence $\hat{v}_{x\theta} \leq 0$.

We close with a remark about existence of an optimal contract in the pure moral hazard case. This is not guaranteed in this case a the lowest value of the domain of u is finite (indeed, zero). By Lemma 14 in Appendix C, existence is guaranteed for $\bar{u}$ sufficiently large, say above a threshold $\bar{u}'$. Hence, by taking $\bar{u} \geq \max\{\bar{u}^*, \bar{u}'\}$ both existence and the validity of decoupling hold.

5.4 Decoupling in the Two-Outcome Case

Another case where a closed-form expression for C can be obtained for any u is when there are only two outcomes, i.e., $x \in \{x_\ell, x_h\}$, $x_h > x_\ell \geq 0$. We denote by $v_h(\theta)$ and $v_\ell(\theta)$ the utility levels that the agent receives after x_h and x_ℓ, respectively, when his type is θ. Let $p(a) \in [0, 1]$ be the probability that the output is x_h when the agent takes action a. We will assume that $p(a) = a$, folding any 'curvature' assumptions into $c(a, \theta)$. In this section we use the standard interpretation for B as the expected value of output, i.e. $B(a) = ax_h + (1-a)x_\ell$, so $B_a = x_h - x_\ell$ and $B_{aa} = 0$.

Consider the case in which the principal wants to implement action $\alpha(\theta)$ for type θ. The cost-minimizing contract solves the participation constraint (with $S(\theta) = \bar{u} + \int_\theta^{\overline{\theta}} c_\theta(\alpha(s), s)ds$ as the reservation utility) and first-order incentive constraint of the agent. That is,

$$\begin{aligned} v_h(\theta) &= S(\theta) + c(\alpha(\theta), \theta) + (1-\alpha(\theta))c_a(\alpha(\theta), \theta) && (28) \\ v_\ell(\theta) &= S(\theta) + c(\alpha(\theta), \theta) - \alpha(\theta)c_a(\alpha(\theta), \theta). && (29) \end{aligned}$$

Hence, the cost function is

$$\begin{aligned} C(\alpha(\theta), S(\theta), \theta) &= \alpha(\theta)\varphi(v_h(\theta)) + (1-\alpha(\theta))\varphi(v_\ell(\theta)) \\ &= \alpha(\theta)\varphi\left(S(\theta) + c(\alpha(\theta),\theta) + (1-\alpha(\theta))c_a(\alpha(\theta),\theta)\right) \\ &\quad +(1-\alpha(\theta))\varphi\left(S(\theta) + c(\alpha(\theta),\theta) - \alpha(\theta)c_a(\alpha(\theta),\theta)\right), \end{aligned} \tag{30}$$

From this, we can derive simple conditions for C to be convex in a.

Lemma 10 *If $c_{aaa} \geq 0$, and $\bar{a} \leq 2/3$, then C is strictly convex in a for all θ and u_0.*

With this in hand, we can provide a set of sufficient conditions for decoupling to be valid.

Proposition 3 *Decoupling delivers an optimal solution to the principal's problem if $\bar{a} \leq 2/3$, h is decreasing, $c_{aaa} \geq 0$, and $c_{a\theta}/c_{aa}$ is increasing in a and θ.*

Once again the proof centers on the sufficient condition *DMC*.

Example 2 *Let h be uniform on $[0,1]$, let $\bar{a} = 2/3$, and let $c(a,\theta) = a^2 e^{\theta}$. Then the conditions of Proposition 3 hold. If instead, $c(a,\theta) = a^2 + a\theta$, the conditions again hold.*

Example 3 *The class of cost functions in Proposition 3 is far from necessary. Consider the following cost function:*

$$c(a,\theta) = (1-a)\log(1-a) + a\log a + a\theta, \tag{31}$$

which is positive for all $a \in [0.5, 1)$ as long as $\underline{\theta}$ is large enough. Moreover, it satisfies $c_a > 0$, $c_{aa} = 1/(a(1-a))$, $c_{aaa} > 0$ for $a \in [0.5,1)$, and $c_{a\theta} = 1$, $c_{aa\theta} = 0$, and $c_{a\theta\theta} = 0$. This does not satisfy $c_{a\theta}/c_{aa} = a(1-a)$ increasing in a. Decoupling, however, delivers an optimal solution to the principal's problem when h is decreasing in θ.[13]

5.5 Decoupling with Exponential Families

Another tractable class of problems where we can check decoupling involves logarthmic utility and exponential families. Recall that f is an exponential family if it can be written as $f(x|a) = m(a)g(x)e^{K(a)j(x)}$. We will denote by $k > 0$ the derivative of K. It is immediate that $l(x|a) = k(a)j(x) + m'(a)/m(a)$, so that *MLRP* holds if and only if j is increasing in x. Also, $l_{xa}(x|a) = k'(a)j'(x)$ is nonnegative if k is increasing in a. In the next result we will assume that B (the

[13]To show it, all we need is to check that (45) in Appendix A holds; and since the *lhs* of (45) is positive for this c (and with h decreasing), it suffices that the *rhs* of (45), $c_{a\theta}C_{aa} - c_{aa}D$, is negative. Using the properties of c one obtains $c_{a\theta}C_{aa} - c_{aa}D = \left(-a\varphi_h' - \varphi_\ell'\right)/(a(1-a)) < 0$, and decoupling yields an optimal solution.

gross benefit to the principal of the agent's effort) is linear in a, while the expected value of the signal given action can have an arbitrary shape.

Recall that ψ solves $1/u'(\psi(\tau)) = \tau$, and that $\rho(\tau) = u(\psi(\tau))$. The next result imposes conditions on both of these mappings. We will also use in the next result a sufficient condition that ensures that $\int v f_{aa} \leq 0$ and hence the FOP, namely, either $CDFC$, or that $\int l_x F_{aa} \geq 0$ (see Chade and Swinkels (2016), Section 6.3, for an explanation and for weaker conditions).

Proposition 4 *Let f be an exponential family with $l_{ax} \geq 0$. Let $c_{aaa} \leq 0$, $c_{a\theta\theta} \geq 0$, and $c_{aa\theta} = 0$, let $B_{aa} = 0$, and let $C_{aa} \geq 0$. Assume that $F_{aaa} \leq 0$, and one of the following holds:*

(i) $\int l_x F_{aa} \geq 0$, and $u(w) = \log w$,

(ii) $F_{aa} \geq 0$, and $u(w) = \sqrt{2w}$, or,

(iii) $F_{aa} \geq 0$, $\psi'' \geq 0$, and $-\frac{\rho'''}{\rho''} + 2\frac{\rho''}{\rho'} < 0$.

Then, any solution to the decoupling problem with $\alpha' < 0$ is an optimal solution to the principal's problem.

The proof of the proposition centers around establishing that SCC must be satisfied under the stated conditions. The following gives examples both where the conditions of Proposition 4 are satisfied and where they fail.

Example 4 *Let $F(x|a) = x^a$, $x, a \in [0,1]$, $B(a) = a$, and $c = \theta a^2/2$, and $u = (w^{1-\sigma} - 1)/(1-\sigma)$, $\sigma \in (0,1]$. Then the conditions of Proposition 4 hold only for logarithmic or square-root utility. To see this, notice that CDFC and the conditions on c and B hold. Also, If $\sigma = 1$ we obtain the log-utility case, while if $\sigma = 0.5$ we obtain an affine transformation of the square root case, both covered by the proposition. For any other value we have that $\psi(\tau) = \tau^{\frac{1}{\sigma}}$, for which $\rho(\tau) = u(\psi(\tau)) = (\tau^{\frac{1}{\sigma}-1} - 1)/(1-\sigma)$, and hence $-(\rho'''/\rho'') + 2(\rho''/\rho') = (1-\sigma)/(\sigma\tau) > 0$ for all $\sigma \in (0,1)$. (And for $\sigma > 1$, ψ becomes strictly concave, and again the proposition does not apply.)*

Proposition 4 provides a class of primitives under which, if the solution to the pure adverse selection step of the decoupling program yields $\alpha' < 0$, then decoupling delivers an optimal menu. The only hard condition to gauge is (iii), since it involves a complex inequality involving the derivatives of ρ plus a convexity condition on ψ. It is instructive to relate these conditions to properties on u. To this end, let $R = -u''/u'$ and $P = -u'''/u''$ be the coefficients of absolute risk aversion and prudence, respectively. We first note that the inequality involving ρ in (iii) is related to the curvature of $1/\rho'$. In fact, it is easy to check that

$$\frac{d^2(1/\rho')}{d\tau^2} = \frac{\rho''}{\rho'^2}\left(-\frac{\rho'''}{\rho''} + 2\frac{\rho''}{\rho'}\right),$$

which is positive (negative) when ρ is concave (convex) if and only if $-(\rho'''/\rho'') + 2(\rho''/\rho') < 0$. Thus, (iii) is equivalent under ρ concave or convex to the convexity or concavity of $1/\rho'$. Second, notice that since $\rho(\tau) = u(\psi(\tau))$, we have $\rho'(\tau) = u'(\psi(\tau))\psi'(\tau) = (u')^2 (\psi(\tau))/R(\psi(\tau))$ and hence $1/\rho' = R/(u')^2$. As a result, differentiating twice with respect to τ yields, after simplification,

$$\frac{d^2(1/\rho')}{d\tau^2} = 4(1.5R - P) - \frac{P'}{R}.$$

The first term on the right is negative by the assumption that $\psi'' \geq 0$, which holds if and only if $P \geq 2R$ (to see this, differentiate $\psi' = R/u'$). Therefore, we obtain that if ρ is strictly convex (which implies ψ convex) then condition (iii) holds if $P' \geq 0$, i.e., if absolute prudence increases in income. And if ρ is strictly concave and ψ is convex, then (iii) holds if P' is sufficiently negative, i.e., if absolute prudence decreases rapidly enough in income.

Example 5 *Consider the same assumptions as in Example 4 except that* $u(w) = \delta w - e^{-\beta w}$, *where* $\beta > 0$ *and* $\delta > 0$. *Then* $R = \beta^2 e^{-rw}/(\delta + \beta e^{-\beta w})$ *and* $P = \beta$, *and thus* $P' = 0$. *Finally,* ρ *(and hence* ψ*) is strictly convex if* $\delta < 2\beta e^{-\beta w}$ *for all* w, *which can be ensured by choice of* δ *and* β *and if other primitives are such that* π *can be bounded above. Thus,* (iii) *obtains.*

5.6 An Example Where Decoupling Does Not Work

A convenient feature of the two-outcome case analyzed above is that, due to the linearity of the probability distribution in the agent's action, the sufficient condition for feasibility $dc_a(\alpha(\theta), \theta)/d\theta \leq 0$ is also necessary (see Corollary 1). We will exploit this feature to provide an example where decoupling does not deliver an optimal solution to the principal's problem.

Suppose $h = 1$ on $[0, 1]$, $c(a, \theta) = a^2\theta/2$ and $\phi(v) = v^2/2$ (square-root utility). Then, one can show that $\alpha' \leq -c_{a\theta}/c_{aa} = -\alpha(\theta)/\theta$ if and only if $\xi(\theta) \geq 0$ everywhere, where[14]

$$\xi(\theta) \equiv \theta^2\alpha^2(\theta) + \theta\left(S(\theta) + \frac{\alpha^2(\theta)\theta}{2}\right) - \int_0^\theta \left(S(t) + \frac{\alpha^2(t)t}{2}\right) dt \geq 0. \tag{32}$$

This clearly holds at $\theta = 0$, since $\xi(0) = 0$. Consider any point at which $\xi(\theta) \leq 0$. Then, since $\alpha'(\theta) \leq -\alpha(\theta)/\theta$, we have that

$$\xi'(\theta) = \theta\alpha(\theta)(3\alpha'(\theta)\theta + 2\alpha(\theta)) \underset{s}{=} \alpha'(\theta) + \frac{2}{3}\frac{\alpha(\theta)}{\theta} \leq -\frac{1}{3}\frac{\alpha(\theta)}{\theta} < 0.$$

But then in particular, $\xi'(0) < 0$, and so for θ near 0, ξ is negative and decoupling fails to deliver a feasible solution to the principal's problem.

[14]Using (28) and (29), plus (47) and (46) in Appendix A, one can show that (45) (also in Appendix A) reduces to the expression (32).

5.7 A Note on the Linear Probability Case

Even when the fully-decoupled problem does not yield an α with the appropriate slope, Corollary 1 points the way to a considerable simplification of the problem when $f_{aa} = 0$. In particular, consider any solution (α, v) to problem (P). For each θ, let $\hat{v}(\cdot, \theta)$ be the Holmstrom-Mirrlees contract corresponding to the triple $(\alpha(\theta), S(\theta), \theta)$. Given that $v(\cdot, \theta)$ has to satisfy IC (and so in specific, the first-order conditions) and has to give surplus $S(\theta)$ for type θ, $\hat{v}(\cdot, \theta)$ must cost the principal no more than $v(\cdot, \theta)$, and will cost strictly less unless $v(\cdot, \theta)$ and $\hat{v}(\cdot, \theta)$ coincide (since the optimal solution to the Holmstrom-Mirrlees problem is unique). Note, however, that the mechanism $(\alpha, \hat{v})$ continues to satisfy DMC, and so is feasible. We conclude that when f_{aa} is linear, then the optimal mechanism must have the property that for each type θ, $v(\cdot, \theta)$ is of the standard Holmstrom-Mirrlees form. Thus, even if the decoupling program does not yield a feasible mechanism (because α fails DMC) finding the optimal mechanism is reasonably tractable, as the problem can be written simply as

$$\max_{\alpha} \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{x}}^{\overline{x}} \left[B(\alpha(\theta)) - C(\alpha(\theta), S(\theta), \theta)\right] f(x|\alpha(\theta))\, h(\theta)\, dx\, d\theta$$

$$s.t. \qquad \alpha'(\theta) \leq -\frac{c_{a\theta}(\alpha(\theta), \theta)}{c_{aa}(\alpha(\theta), \theta)} \ \forall\ \theta, \text{ and}$$

$$S(\theta) = \bar{u} + \int c_{\theta}(\alpha(t), t)\, dt\, \forall\ \ \theta.$$

This is a problem to which ironing techniques can be applied, and then fill in the requisite v using the relevant Holmstrom-Mirrlees contract at each θ.

Thus, in the case of linear probabilities, even if full decoupling does not work, one still has the implication that an optimal mechanism consists of a menu over Holmstrom-Mirrlees contracts. An interesting problem for future research is to explore whether there are other settings where a similar simplification can be performed. This seems in general quite difficult, as the simple constraint DMC is too strong, while the constraint given by inequality (25) is both not strong enough and involves the seemingly intractable object $\int v f_{aa}$.

5.8 A Type-Dependent Outside Option

For simplicity, we took the outside option $\bar{u}$ of the agent to be independent of type. One could easily imagine that in many economically interesting applications, agents who are more capable in the setting at hand are also more capable in other settings, so that $\bar{u}$ declines in θ. It is important to note that in such a setting, it remains the case that $S'(\theta) = c_{\theta}(\alpha(\theta), s)$ holds. The only significant difference is that the constant in (17) is no longer necessarily $S\left(\bar{\theta}\right)$, but may instead be $S(\theta^{*})$ for some type $\theta^{*} \neq \bar{\theta}$. While this somewhat complicates the derivation of, for example, the candidate contract under decoupling, it does not change the basic necessary or sufficient conditions for feasibility, and it remains true that if effort is falling fast enough in effort so that contracts become lower powered either in the sense of DMC or of SCC, then the decoupled solution will be feasible and hence optimal. In particular, one can easily modify the statement and proof of Proposition 2 to say that there is $\bar{u}^*$ such that if $\bar{u}\left(\bar{\theta}\right) \geq \bar{u}^*$, then the solution to the decoupled program is guaranteed to be feasible. Propositions 3 and 4 hold unchanged.

As a simple example of a setting with type-dependent reservation utility, consider an insurance customer who can exert effort to reduce the probability of an accident. Then absent insurance, a customer with a lower cost of effort will have a higher utility than one with a higher cost. Indeed, by the Envelope Theorem, the outside option of the agent satisfies $\overline{u}'(\theta) = -c_\theta(a_n(\theta), \theta)$, where $a_n(\theta)$ is the optimal action for the agent of type θ without insurance. Since a partially insured agent will choose a lower effort level, it follows that $S'(\theta) = -c_\theta(\alpha(\theta), \theta) > \overline{u}'(\theta)$ in any optimal mechanism. Hence, the binding participation constraint will be $S(\underline{\theta}) = \overline{u}(\underline{\theta})$. In general, the relationship between $S'(\theta)$ and $\overline{u}'(\theta)$ depends on whether the agent is exerting more effort inside or outside of the relationship with the principal.

6 Concluding Remarks

We study a canonical problem with both moral hazard and screening. We derive necessary conditions for a menu (α, v) to be feasible. In contrast to the standard screening problem it is not enough that the recommended action falls as the agent becomes less capable. Rather, our necessary conditions require that the recommended action falls fast enough that, in a specific weak sense, less capable agents face weaker incentives. We also provide two different sufficient conditions for a menu to be feasible. The first requires that as the agent becomes less capable, the recommended action falls fast enough so that his marginal cost of effort falls. The second asks that as the agent becomes less capable, his contract becomes flatter in the sense of single crossing. Each of our two sufficient conditions can thus be interpreted as a stronger version of the requirement that less-capable agents face weaker incentives.

We then turn to the question of when one can decouple the problem by first solving the moral hazard problem for each action-surplus-type triple and then solving a (non-linear) screening problem for the resultant cost function. We provide several cases where decoupling is guaranteed to work, and also cases where it will not. When decoupling does work, we have the economically important implication that the optimal mechanism has the form that the agent, by choice of his announcement of type, is choosing from a menu over Holmstrom-Mirrlees contracts. Thus, insights that we learn from the pure moral hazard problem carry over into the setting which also includes adverse selection, and similarly, insights that we learn from the pure adverse selection problem carry through when there is also moral hazard.

Appendix A Proofs

A.1 Proof of Proposition 1 *(ii)*

We must show that (12) holds for all θ under the premise. For convenience, rewrite (12) as

$$z(\theta) \equiv \frac{h'}{h} - \frac{2\phi' h}{\int_{\underline{\theta}}^{\theta} \phi' h} - \frac{c_{a\theta\theta}}{c_{a\theta}} < 0. \tag{33}$$

Since $\alpha'(\underline{\theta}) < 0$, there is a largest type θ_0 such that $\alpha'(\theta) < 0$ for all $\theta < \theta_0$, where clearly, $\theta_0 > \underline{\theta}$.

Towards a contradiction, assume that $\theta_0 < \overline{\theta}$. Then $z(\theta_0) = 0$ (equivalently, $\alpha'(\theta_0) = 0$), and $z'(\theta_0) \geq 0$ (since $z(\theta_0) = 0$ and $\alpha'(\theta) < 0$ for all $\theta < \theta_0$). We will show that the last two properties cannot hold simultaneously under the stated assumptions on h and $c_{a\theta}$. We will thus conclude that $\alpha'(\theta) < 0$ for all θ.

Assume that $z(\theta_0) = 0$ and consider $z'(\theta_0)$. The first term in (33) is decreasing in θ since h is log-concave. Notice next that

$$\left(\frac{c_{a\theta\theta}}{c_{a\theta}}\right)_\theta = \frac{\partial}{\partial a}\left(\frac{c_{a\theta\theta}}{c_{a\theta}}\right)\alpha' + \frac{\partial}{\partial \theta}\left(\frac{c_{a\theta\theta}}{c_{a\theta}}\right),$$

where we use $(\cdot)_\theta$ as shorthand for the total derivative with respect to θ. When we evaluate this expression at $\theta = \theta_0$, the first term vanishes since $\alpha'(\theta_0) = 0$, and only the second term remains, which is nonnegative since $c_{a\theta}$ is log-convex. Hence, a necessary condition for $z'(\theta_0) \geq 0$ is that the derivative of the second term of (33) is nonpositive at $\theta = \theta_0$, which holds if and only if

$$\phi'' c_a \alpha' h \int_{\underline{\theta}}^{\theta_0} \phi' h + \phi' h' \int_{\underline{\theta}}^{\theta_0} \phi' h - \phi'^2 h^2 \leq 0$$

when evaluated at $\theta = \theta_0$. Since the first term vanishes at that point, we obtain

$$\phi' h' \int_{\underline{\theta}}^{\theta} \phi' h - \phi'^2 h^2 \leq 0,$$

which holds if and only if at $\theta = \theta_0$

$$\frac{h'}{h} - \frac{\phi' h}{\int_{\underline{\theta}}^{\theta} \phi' h} \leq 0.$$

But this implies that at $\theta = \theta_0$

$$\frac{h'}{h} - \frac{2\phi' h}{\int_{\underline{\theta}}^{\theta} \phi' h} < 0,$$

and therefore

$$z(\theta_0) = \left(\frac{h'(\theta_0)}{h(\theta_0)} - \frac{2\phi' h(\theta_0)}{\int_{\underline{\theta}}^{\theta_0} \phi' h} - \frac{c_{a\theta\theta}(\alpha(\theta_0), \theta_0)}{c_{a\theta}(\alpha(\theta_0), \theta_0)} \right) < 0,$$

a contradiction. Hence, $z(\theta_0) = 0$ and $z'(\theta_0) \geq 0$ cannot hold simultaneously, and we are done. □

A.2 Proof of Lemma 1

From the first-order condition of the cost-minimization problem plus the binding participation and incentive constraints we obtain the following system of equations in λ and μ:

$$\int \rho(\lambda + \mu l(x|a)) f(x|a) dx = c(a, \theta) + u_0 \tag{34}$$

$$\int \rho(\lambda + \mu l(x|a)) f_a(x|a) dx = c_a(a, \theta). \tag{35}$$

Differentiating (34)–(35) with respect to a and using $\int v f_a = c_a$ yields:

$$\begin{aligned} \lambda_a \int \rho' f + \mu_a \int \rho' l f &= -\mu \int \rho' l_a f \\ \lambda_a \int \rho' f_a + \mu_a \int \rho' l f_a &= -\mu \int \rho' l_a f_a + c_{aa} - \int \rho f_{aa}. \end{aligned}$$

Therefore,

$$\mu_a = \frac{\mu(\int \rho' f_a \int \rho' l_a f - \int \rho' f \int \rho' l_a f_a) + (\int \rho' f)(c_{aa} - \int \rho f_{aa})}{\int \rho' f \int \rho' l f_a - (\int \rho' f_a)^2}, \tag{36}$$

and λ_a can be written as

$$\lambda_a = \frac{-\mu \int \rho' f l_a - \mu_a \int \rho' f l}{\int \rho' f}. \tag{37}$$

We must show that both μ_a and λ_a are positive.

Consider first μ_a. To show that it is positive under our assumptions we will appeal to the the following Chebyshev's order inequality: *If f and g are two functions that are integrable and monotone in the same sense on $[a, b]$, and p is a positive and integrable function on $[a, b]$, then*

$$\int_a^b p(x) f(x) g(x) dx \int_a^b p(x) \geq \int_a^b p(x) f(x) dx \int_a^b p(x) g(x) dx,$$

with equality if and only if one of the functions f, g reduces to a constant. If f and g are monotone in the opposite sense, then the inequality reverses.[15]

The second term in the numerator of (36) is positive by *FOP*, so $\mu_a \geq 0$ if the first term in the numerator and the denominator are both nonnegative.

[15]Mitrinovic (1970), Section 2.6, Theorem 10, p. 40.

Consider the denominator: setting $\gamma \equiv \rho' f$ and multiplying and dividing by f wherever necessary allows us to rewrite it as

$$\int l^2\gamma \int \gamma - \int l\gamma \int l\gamma,$$

and so Chebyshev's inequality with $p = \gamma$, and $f = g = l$ shows that the denominator of (36) is positive.

Consider now the term in parenthesis in the first term of the numerator. Once again, using γ and multiplying and dividing by f wherever necessary, we obtain the following rewriting:

$$-\left[\int \gamma l_a l \int \gamma - \int \gamma l \int \gamma l_a\right].$$

We know that l increases in x. Since $l_{xa} \le 0$, l_a decreases in x. Then by the reverse case of Chebyshev's inequality with $p = \gamma$, $f = l_a$, and $g = l$, we obtain that the expression in square brackets is negative and hence the first term in the numerator is positive. This shows that $\mu_a \ge 0$.

Turning to λ_a, notice that $\int \rho' f l = \int \rho' f_a$, which is negative as $\int f_a = 0$ and ρ' is decreasing in x. Since $\mu_a \ge 0$, it follows that $\lambda_a \ge 0$ if $\int \rho' f l_a \le 0$. But this holds since f log-concave in a, which is equivalent to $l_a \le 0$. Thus, $\lambda_a \ge 0$ and the proof of the lemma is complete. □

A.3 Proof of Theorem 1

We proceed in several steps. Denote by γ the generalized inverse of α (α need not be continuous everywhere; it can jump down a countable number of times).

Step 1. Since $S(\overline{\theta}) \ge \overline{u}$, (17) implies that $S(\theta) \ge \overline{u}$ for all θ, and thus participation holds.

Step 2. It follows from Lemma 5 that it suffices to rule out deviations of θ to $(\hat{\theta}, \hat{a}) \notin L$, i.e., 'off-locus' deviations. We focus on deviations with $\hat{a} \ge \alpha(\hat{\theta})$ (the other case is similar).

Let us denote by θ_T the agent's true type. If $\theta_T \ge \hat{\theta}$, then

$$\int v(x,\hat{\theta})f(x|\hat{a})dx - \int v(x,\hat{\theta})f(x|\alpha(\hat{\theta}))dx \le c(\hat{a},\hat{\theta}) - c(\alpha(\hat{\theta}),\hat{\theta}) \le c(\hat{a},\theta_T) - c(\alpha(\hat{\theta}),\theta_T),$$

where the first inequality follows from FOP and $\hat{a} \ge \alpha(\hat{\theta})$, and the second since c is supermodular. Rearranging yields

$$\int v(x,\hat{\theta})f(x|\hat{a})dx - c(\hat{a},\theta_T) \le \int v(x,\hat{\theta})f(x|\alpha(\hat{\theta}))dx - c(\alpha(\hat{\theta}),\theta_T),$$

and thus the agent is better off with an on-locus deviation, which in turn is strictly dominated by truthtelling by Lemma 5.

STEP 3. If $\hat{a} > \alpha(\theta_T)$ and $\hat{\theta} \geq \theta_T$, then deviation $(\hat{\theta}, \hat{a})$ is dominated for type θ_T by $(\hat{\theta}, \alpha(\theta_T))$. To see this, consider any action $a \in [\alpha(\theta_T), \hat{a}]$. Then

$$\int v(x,\hat{\theta}) f_a(x|a)dx \leq \int v(x,\hat{\theta}) f_a(x|\alpha(\hat{\theta}))dx = c_a(\alpha(\hat{\theta}),\hat{\theta}) \leq c_a(\alpha\left(\theta_T\right),\theta_T) \leq c_a(a,\theta_T),$$

where the first inequality follows from the *FOP*, the equality follows by the first-order condition (16), the second inequality follows by *DMC,* and the third by convexity of c in a. Hence, $c_a(a,\theta_T) \geq \int v(x,\hat{\theta}) f_a(x|\hat{a})dx$ for any $a \in [\alpha(\theta_T),\hat{a}]$, which implies that θ_T's expected utility is decreasing in a in that range, and so θ_T is at least well off with deviation $(\hat{\theta},\alpha(\theta_T))$ as with $(\hat{\theta},\hat{a})$.

From Steps 2 and 3 we can restrict attention to deviations $(\hat{\theta},\hat{a})$ with $\hat{\theta} \geq \theta_T$ and $\hat{a} \leq \alpha(\theta_T)$.

STEP 4. Let $(\hat{\theta},\hat{a})$ be such that $\hat{a} > \alpha(\hat{\theta})$ and $(\gamma(\hat{a}),\hat{a}) \in L$, i.e., $\hat{a} = \alpha(\gamma(\hat{a}))$. We will show that $\int v(x,\gamma(\hat{a})) f(x|\hat{a})dx \geq \int v(x,\hat{\theta}) f(x|\hat{a})dx$ and hence

$$\int v(x,\gamma(\hat{a})) f(x|\hat{a})dx - c(\hat{a},\theta_T) \geq \int v(x,\hat{\theta}) f(x|\hat{a})dx - c(\hat{a},\theta_T),$$

showing that $(\hat{\theta},\hat{a})$ is dominated for θ_T by an on-locus deviation.

Define the function Q by

$$Q(a) \equiv S(\gamma(a)) + c(a,\gamma(a)) - \int v(x,\hat{\theta}) f(x|a)dx. \tag{38}$$

It is immediate from the definition that $Q(\alpha(\hat{\theta})) = 0$, since $S(\gamma(\alpha(\hat{\theta}))) = S(\hat{\theta})$ and the second and third terms of (38) become $-S(\hat{\theta})$. Also, Q is differentiable a.e. since S and γ are, and has derivative given by

$$\begin{aligned} Q'(a) &= S_\theta(\gamma(a))\gamma'(a) + c_a(a,\gamma(a)) + c_\theta(a,\gamma(a))\gamma'(a) - \int v(x,\hat{\theta}) f_a(x|a)dx \\ &= c_a(a,\gamma(a)) - \int v(x,\hat{\theta}) f_a(x|a)dx, \end{aligned}$$

where the equality follows from $S_\theta = -c_\theta$. Now, for any $a \in \left(\alpha(\hat{\theta}),\hat{a}\right)$,

$$Q'(a) = c_a(a,\gamma(a)) - \int v(x,\hat{\theta}) f_a(x|a)dx \geq c_a(\alpha(\hat{\theta}),\hat{\theta}) - \int v(x,\hat{\theta}) f_a(x|\alpha(\hat{\theta}))dx = 0,$$

where to understand the inequality, notice that since $a > \alpha(\hat{\theta})$, then $\gamma(a) \leq \hat{\theta}$. But then $c_a(a,\gamma(a)) \geq c_a(\alpha(\hat{\theta}),\hat{\theta})$ by *DMC*, while $\int v(x,\hat{\theta}) f_a(x|a)dx \leq \int v(x,\hat{\theta}) f_a(x|\alpha(\hat{\theta}))dx$ by *FOP*. The last equality is simply the first-order condition (16) for type $\hat{\theta}$. Hence, in particular, $Q(\hat{a}) \geq 0$. But, by definition and since $(\gamma(\hat{a}),\hat{a}) \in L$, $\int v(x,\gamma(\hat{a})) f(x|\hat{a})dx = S(\gamma(\hat{a})) + c(\hat{a},\hat{\gamma}(\hat{a}))$, and hence $Q\left(\hat{a}\right) \geq 0$ is equal to $\int v(x,\gamma(\hat{a})) f(x|\hat{a})dx \geq \int v(x,\hat{\theta}) f(x|\hat{a})dx$, completing the proof of this step.

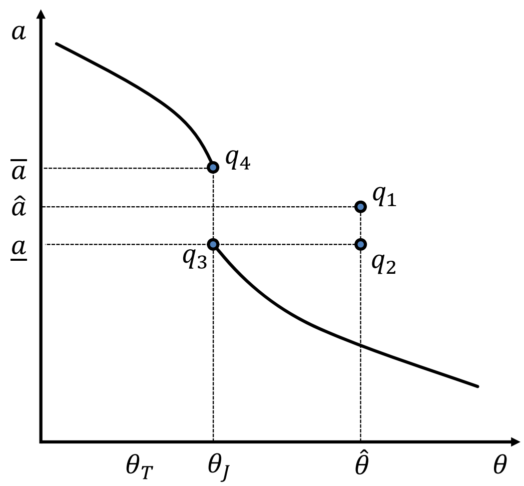

Figure 2: **DMC**. Under DMC, a deviation by θ_T to q_1 is dominated by one to q_2, which in turn is worse as a deviation to q_3 that yields the same expected utility as a deviation to q_4. In sum, a deviation to q_1 is dominated by q_4, and this in turn yields a lower payoff than telling the truth and being obedient.

Step 5. Let $(\hat{\theta}, \hat{a})$ be such that $\hat{a} > \alpha(\hat{\theta})$ and $(\theta_J, \hat{a}) \notin L$, i.e., $\hat{a} \neq \alpha(\theta_J)$. Then α jumps at $\theta_J = \gamma(\hat{a})$ with the jump endpoints being denoted by $\underline{a}$ and $\overline{a}$, respectively, i.e. $\hat{a} \in [\underline{a}, \overline{a}]$. Note that since $\hat{a} \leq \alpha(\theta_T)$, it follows that $\theta_T \leq \theta_J$. We will show that $(\theta_T, \alpha(\theta_T))$ dominates deviation $(\hat{\theta}, \hat{a})$ for θ_T. To this end, define the following expressions:

$$\begin{aligned}
\tilde{S}_1 &= \int v(x, \hat{\theta}) f(x|\hat{a}) dx - c(\hat{a}, \theta_J) \\
\tilde{S}_2 &= \int v(x, \hat{\theta}) f(x|\underline{a}) dx - c(\underline{a}, \theta_J) \\
\tilde{S}_3 &= \int \underline{v}(x, \theta_J) f(x|\underline{a}) dx - c(\underline{a}, \theta_J) \\
\tilde{S}_4 &= \int \overline{v}(x, \theta_J) f(x|\overline{a}) dx - c(\overline{a}, \theta_J),
\end{aligned}$$

which give the surplus of θ_J at points q_i, $i = 1, ..., 4$, in Figure 2.

We first claim that Step 3 shows that $\tilde{S}_1 \leq \tilde{S}_2$. If $\alpha(\theta_J) = \underline{a}$, then this is immediate from Step

3. And if $\alpha(\theta_J) \neq \underline{a}$, we can apply Step 3 at $\theta_J + \varepsilon$ for any $\varepsilon > 0$ to obtain that $(\hat{\theta}, \alpha(\theta_J + \varepsilon))$ dominates $(\hat{\theta}, \hat{a})$ for type $\theta_J + \varepsilon$. Taking the limit as ε goes to zero completes the proof of the claim.

Notice also that, by Step 4, $\tilde{S}_2 \leq \tilde{S}_3$ (again taking a limit if $\alpha(\theta_J) \neq \underline{a}$).

Finally, since S is continuous, $\tilde{S}_3 = \tilde{S}_4$. Therefore, we have $\tilde{S}_4 = \tilde{S}_3 \geq \tilde{S}_2 \geq \tilde{S}_1$, which yields

$$\int \overline{v}(x, \theta_J) f(x|\overline{a}) dx - \int v(x, \hat{\theta}) f(x|\hat{a}) dx \geq c(\overline{a}, \theta_J) - c(\hat{a}, \theta_J) \geq c(\overline{a}, \theta_T) - c(\hat{a}, \theta_T),$$

where the first inequality is equivalent to $\tilde{S}_4 \geq \tilde{S}_1$, and the second follows from supermodularity of c and $\theta_T \leq \theta_J$. Rearranging yields

$$\int \overline{v}(x, \theta_J) f(x|\overline{a}) dx - c(\overline{a}, \theta_T) \geq \int v(x, \hat{\theta}) f(x|\hat{a}) dx - c(\hat{a}, \theta_T). \tag{39}$$

If $\theta_T = \theta_J$, then (39) reveals that $(\theta_T, \alpha(\theta_T))$ dominates $(\hat{\theta}, \hat{a})$. And if $\theta_T \neq \theta_J$, then S continuous and Lemma 5 imply that $S(\theta_T) > \int \overline{v}(x, \theta_J) f(x|\overline{a}) dx - c(\overline{a}, \theta_T)$, and hence once again $(\theta_T, \alpha(\theta_T))$ dominates $(\hat{\theta}, \hat{a})$. This completes the proof of Step 5 and the theorem. □

A.4 Proof of Theorem 2

We proceed in several steps.

STEP 1. Since $S(\overline{\theta}) \geq \overline{u}$, (17) implies that $S(\theta) \geq \overline{u}$ for all θ, and thus participation holds.

STEP 2. Recall from Lemma 5 that conditional on being on the graph of α, the agent strictly prefers to announce his true type. It thus suffices to show that any deviation to $(\hat{\theta}, \hat{a})$ for θ_T, with $\hat{a} \neq \alpha(\hat{\theta})$ is dominated by a deviation such that the action and the report are on the graph of α. We will show that this holds for any deviation that is above the graph of α, i.e., with $\hat{a} > \alpha(\hat{\theta})$. A symmetric argument handles deviations below the graph.

STEP 3. We will show that fixing $\hat{a}$, and for any $\tilde{\theta}$ that the agent is contemplating announcing with $\hat{a} > \alpha(\tilde{\theta})$, the agent is better off by modifying his deviation so as to slightly lower θ from $\tilde{\theta}$. To do so, consider first the case where v is differentiable in θ at $\tilde{\theta}$. Notice that $\int v_\theta(x, \tilde{\theta}) f(x|\alpha(\tilde{\theta})) dx = 0$ by the first-order condition (18), and v_θ has sign pattern $+/-$ by hypothesis. Hence

$$\int v_\theta(x, \tilde{\theta}) f(x|\hat{a}) dx = \int v_\theta(x, \tilde{\theta}) f(x|\alpha(\tilde{\theta})) \frac{f(x|\hat{a})}{f(x|\alpha(\tilde{\theta}))} dx \leq 0,$$

where we have used *MLRP*, $\hat{a} > \alpha(\tilde{\theta})$, and Beesack's inequality. Thus, the agent's expected utility is decreasing in θ at $(\tilde{\theta}, \hat{a})$.

Consider now a jump point at $\tilde{\theta}$ with endpoints $\overline{a} = \lim_{\varepsilon \downarrow 0} \alpha(\tilde{\theta} - \varepsilon)$ and $\underline{a} = \lim_{\varepsilon \downarrow 0} \alpha(\tilde{\theta} + \varepsilon)$, and where $\hat{a} \geq \overline{a}$. It is enough to show that

$$\int \left(\overline{v}(x,\tilde{\theta}) - \underline{v}(x,\tilde{\theta})\right) f(x|\overline{a})dx \geq 0 \tag{40}$$

for then, since $\bar{v}(\cdot,\tilde{\theta}) - \underline{v}(\cdot,\tilde{\theta})$ has sign pattern $+/-$ and since as before $f(\cdot|\hat{a})/f(\cdot|\overline{a})$ is increasing in x, we have

$$\int \left(\overline{v}(x,\tilde{\theta}) - \underline{v}(x,\tilde{\theta})\right) f(x|\overline{a})\frac{f(x|\hat{a})}{f(x|\overline{a})}dx \geq 0.$$

Thus, the agent is again better off by modifying his deviation by lowering the report of this type. To show (40), notice that

$$\int \overline{v}(x,\tilde{\theta})f(x|\overline{a})dx - c(\overline{a},\tilde{\theta}) = \int \underline{v}(x,\tilde{\theta})f(x|\underline{a})dx - c(\underline{a},\tilde{\theta}) \geq \int \underline{v}(x,\tilde{\theta})f(x|\overline{a})dx - c(\overline{a},\tilde{\theta}),$$

where the equality follows since $\tilde{\theta}$ must be indifferent between the actions at the jump. To see the inequality, note that by regularity, $\int \underline{v}(x,\tilde{\theta})f_a(x|\underline{a})dx - c_a(\underline{a},\tilde{\theta}) = 0$, and so by the *FOP,* $\underline{a}$ is a global maximizer of $\int \underline{v}(x,\tilde{\theta})f(x|\cdot)dx - c(\cdot,\tilde{\theta})$.

STEP 4. Suppose that $\hat{a} > \alpha(\underline{\theta})$. We will show that the agent is better off with a deviation to $(\underline{\theta},\alpha(\underline{\theta}))$. To see this, notice that the previous step shows that

$$\int v(x,\underline{\theta})f(x|\hat{a})dx \geq \int v(x,\hat{\theta})f(x|\hat{a})dx,$$

so the agent prefers deviation $(\underline{\theta},\hat{a})$ to $(\hat{\theta},\hat{a})$. By *FOP* and c supermodular we obtain

$$\begin{aligned}\int v(x,\underline{\theta})f(x|\hat{a})dx - \int v(x,\underline{\theta})f(x|\alpha(\underline{\theta}))dx &\leq c(\hat{a},\underline{\theta}) - c(\alpha(\underline{\theta}),\underline{\theta}) \\ &\leq c(\hat{a},\theta_T) - c(\alpha(\underline{\theta}),\theta_T),\end{aligned}$$

which rearranges to

$$\int v(x,\underline{\theta})f(x|\alpha(\underline{\theta}))dx - c(\alpha(\underline{\theta}),\theta_T) \geq \int v(x,\underline{\theta})f(x|\hat{a})dx - c(\hat{a},\theta_T),$$

and thus $(\underline{\theta},\alpha(\underline{\theta}))$ is a better deviation for the agent. Deviations with $\hat{a} < \alpha(\overline{\theta})$ are similarly ruled out and we will hence restrict attention to deviations with $\hat{a} \in [\alpha(\overline{\theta}),\alpha(\underline{\theta})]$.

STEP 5. Assume $\hat{a} > \alpha(\hat{\theta})$ and $\theta_T \geq \hat{\theta}$. Then by *FOP*, the agent is better off at $(\hat{\theta},\alpha(\hat{\theta}))$. To see this, notice that

$$\int v(x,\hat{\theta})f(x|\hat{a})dx - \int v(x,\hat{\theta})f(x|\alpha(\hat{\theta}))dx \leq c(\hat{a},\hat{\theta}) - c(\alpha(\hat{\theta}),\hat{\theta}) \leq c(\hat{a},\theta_T) - c(\alpha(\hat{\theta}),\theta_T),$$

where the first inequality follows from *FOP* and the second by supermodularity of c. Rearranging yields the result.

STEP 6. Assume $\theta_T < \hat{\theta}$, and that there is a $\tilde{\theta}$ such that $\alpha(\tilde{\theta}) = \hat{a}$. Then by Step 3, the agent is better off with deviation $(\tilde{\theta}, \alpha(\tilde{\theta}))$.

STEP 7. Finally, let us consider the case where there is a jump at θ_J containing $\hat{a}$, i.e., $\underline{a} \leq \hat{a} \leq \overline{a}$. Assume first that $\theta_J < \theta_T$. Then, by Step 3, $\int v(x, \theta_T) f(x|\hat{a}) dx \geq \int v(x, \hat{\theta}) f(x|\hat{a}) dx$, and so type θ_T prefers the deviation $(\theta_T, \hat{a})$ to $(\hat{\theta}, \hat{a})$. But, by *FOP,* $(\theta_T, \alpha(\theta_T))$ is better still. So, assume $\theta_J \geq \theta_T$.

We will use the following notation:

$$\begin{aligned}
\hat{S}_1 &= S(\theta_T) \\
\hat{S}_{2,\varepsilon} &= \int v(x, \theta_J - \varepsilon) f(x|\alpha(\theta_J - \varepsilon)) dx - c(\alpha(\theta_J - \varepsilon), \theta_T) \\
\hat{S}_{3,\varepsilon} &= \int v(x, \theta_J + \varepsilon) f(x|\hat{a}) dx - c(\hat{a}, \theta_T) \\
\hat{S}_4 &= \int v(x, \hat{\theta}) f(x|\hat{a}) dx - c(\hat{a}, \theta_T).
\end{aligned}$$

These are the expected utilities for type θ_T associated with the announcement-action pairs corresponding to the points q_i, $i = 1, 2, 3, 4$, in Figure 3.

Notice that, by definition of S, we have

$$\hat{S}_{2,\varepsilon} = S(\theta_J - \varepsilon) + c(\alpha(\theta_J - \varepsilon), \theta_J - \varepsilon) - c(\alpha(\theta_J - \varepsilon), \theta_T),$$

and thus

$$\hat{S}_2 \equiv \lim_{\varepsilon \downarrow 0} \hat{S}_{2,\varepsilon} = S(\theta_J) + c(\overline{a}, \theta_J) - c(\overline{a}, \theta_T).$$

Now, by *FOP*,

$$\int v(x, \theta_J + \varepsilon) f(x|\hat{a}) dx \leq S(\theta_J + \varepsilon) + c(\hat{a}, \theta_J + \varepsilon),$$

otherwise type $\theta_J + \varepsilon$ would have a profitable deviation to $\hat{a}$. Therefore,

$$\hat{S}_{3,\varepsilon} \leq S(\theta_J + \varepsilon) + c(\hat{a}, \theta_J + \varepsilon) - c(\hat{a}, \theta_T),$$

and so

$$\hat{S}_3 \equiv \lim_{\varepsilon \downarrow 0} \hat{S}_{3,\varepsilon} \leq S(\theta_J) + c(\hat{a}, \theta_J) - c(\hat{a}, \theta_T). \tag{41}$$

By Lemma 5, $\hat{S}_1 \geq \hat{S}_2$, while by Step 3, $\hat{S}_3 \geq \hat{S}_4$. But

$$\hat{S}_2 - \hat{S}_3 \geq (c(\overline{a}, \theta_J) - c(\overline{a}, \theta_T)) - (c(\hat{a}, \theta_J) - c(\hat{a}, \theta_T)) \geq 0,$$

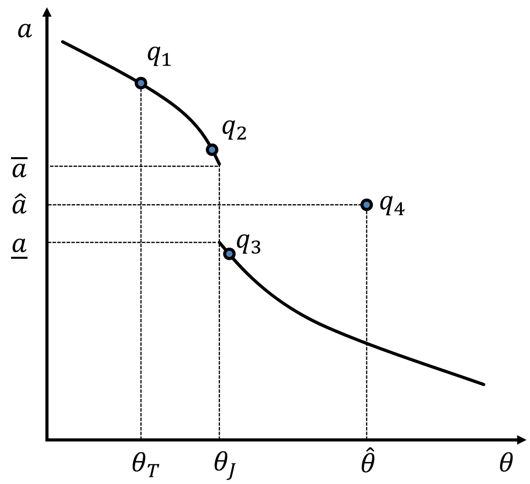

Figure 3: **SCC**. Under SCC, a deviation by θ_T to q_4 is dominated by one to q_3 (slightly to the right of the jump). But that deviation yields a lower payoff than q_3, which in turn is worse than a deviation to q_2 (slightly to the left of the jump) and, since it is on locus, it is dominated by telling the truth and being obedient at point q_1.

where we have used $\theta_J \geq \theta_T$ and that c is strictly supermodular in (a, θ). Hence, $\hat{S}_1 \geq \hat{S}_4$, and deviation $(\hat{a}, \hat{\theta})$ is dominated. Hence, the menu (α, v) is incentive compatible. $\square$

A.5 Proof of Lemma 6

Lemma 11 in Appendix C provides conditions under which a solution to the system (34)–(35) is a pair $\lambda(a, \theta, u_0)$ and $\mu(a, \theta, u_0)$ that is twice continuously differentiable. And since $\hat{v}\left(x, a, u_0, \theta\right)) = \rho\left(\lambda + \mu l(x|a)\right)$, it follows that $\hat{v}$ is twice continuously differentiable as well, and thus so is $C\left(a, u_0, \theta\right) = \int \varphi\left(\hat{v}\left(x, a, u_0, \theta\right)\right) f\left(x|a\right) dx$. $\square$

A.6 Proof of Lemma 7

Recall that for each θ, α is defined by

$$\kappa(\alpha(\theta), \theta) - \zeta(\theta) = 0,$$

where

$$\kappa(a,\theta) = \frac{(B_a - C_a)h}{c_{a\theta}}, \text{ and } \zeta(\theta) = \int_{\underline{\theta}}^{\theta} C_{u_0} h.$$

Consider any point (a, θ) where $\kappa(a,\theta) - \zeta(\theta) = 0$. Then, $B_a - C_a > 0$, and $B_a - C_a$ is strictly decreasing in a, given the assumption that $C_{aa} > 0$, and that $c_{aa\theta} \geq 0$. Hence, $\kappa_a < 0$, and, since κ and ζ are continuous in θ, it follows that α is continuous in θ.

Now, the fact that α is continuous implies that $S(\theta) = \int_{\theta}^{\bar{\theta}} c_\theta(\alpha(t), t)\, dt$ is continuously differentiable. Hence, ζ is continuously differentiable, since the integrand $C_{u_0}(\alpha(t), S(t), t)\, h(t)$ is continuous by Lemma 6. But, κ is continuously differentiable as well, and so as we have already established that $\kappa_a < 0$, α is continuously differentiable by the Implicit Function Theorem. $\square$

A.7 Proof of Lemma 9

Notice that

$$\frac{dC_{u_0}(S(\theta), \theta, \alpha(\theta))}{d\theta} = -C_{u_0u_0}c_\theta + C_{u_0\theta} + C_{u_0a}\alpha' \leq -C_{u_0u_0}c_\theta + C_{u_0\theta} - C_{u_0a}\frac{c_{a\theta}}{c_{aa}},$$

where the inequality uses part (i) and the additional premise. From the Envelope Theorem $C_{u_0u_0} = \lambda_{u_0}$ and $C_{u_a\theta} = \lambda_\theta$, so

$$-C_{u_0u_0}c_\theta + C_{u_0\theta} - C_{u_0a}\frac{c_{a\theta}}{c_{aa}} = -\lambda_{u_0}c_\theta + \lambda_\theta - \lambda_a\frac{c_{a\theta}}{c_{aa}}.$$

The goal is to show that this expression is nonpositive. Differentiating the system (34)–(35) with respect to u_0 and θ, respectively, yields

$$\begin{aligned} \lambda_{u_0} &= \frac{\int \rho'^2 f}{\int l^2\gamma \int \gamma - \int l\gamma \int l\gamma} \\ \lambda_\theta &= \frac{c_\theta \int \rho'^2 f - c_{a\theta}\int \rho' f_a}{\int l^2\gamma \int \gamma - \int l\gamma \int l\gamma} \end{aligned}$$

Using these expressions and (37) we obtain

$$-\lambda_{u_0}c_\theta + \lambda_\theta - \lambda_a\frac{c_{a\theta}}{c_{aa}} = \frac{-c_{a\theta}\int \rho' f_a}{\int l^2\gamma \int \gamma - \int l\gamma \int l\gamma} - \frac{c_{a\theta}}{c_{aa}}\left(\frac{-\mu \int \rho' f l_a - \mu_a \int \rho' f l}{\int \rho' f}\right).$$

Since $l_a \leq 0$, it follows that $\int \rho' f l_a \leq 0$. Hence, to show that $-\lambda_{u_0}c_\theta + \lambda_\theta - \lambda_a(c_{a\theta}/c_{aa}) \leq 0$ it suffices to show that

$$\frac{-c_{a\theta}\int \rho' f_a}{\int l^2\gamma \int \gamma - \int l\gamma \int l\gamma} + \frac{c_{a\theta}}{c_{aa}}\left(\frac{\mu_a \int \rho' f l}{\int \rho' f}\right) \leq 0.$$

Moreover, since the first term in the numerator of (36) is nonnegative, it follows that that term times $\int \rho' f l \leq 0$ is nonpositive. As a result, to show that $-\lambda_{u_0}c_\theta + \lambda_\theta - \lambda_a(c_{a\theta}/c_{aa}) \leq 0$ it suffices to show that

$$\frac{-c_{a\theta}\int\rho' f_a}{\int l^2\gamma\int\gamma-\int l\gamma\int l\gamma}+\frac{c_{a\theta}}{c_{aa}}\left(\frac{(\int\rho' f)(c_{aa}-\int\rho f_{aa})\int\rho' fl}{\int\rho' f\left(\int l^2\gamma\int\gamma-\int l\gamma\int l\gamma\right)}\right)=_s-\int\rho f_{aa}\int\rho' fl\leq 0.$$

But this holds since $\int\rho f_{aa}\leq 0$ by *FOP* and $\int\rho' fl\leq 0$ since ρ' is decreasing. Therefore, $C_{u_0}(S(\cdot),\cdot,\alpha(\cdot))$ decreases in θ. □

A.8 Proof of Proposition 2

It suffices to verify that the (α, v) that obtains through decoupling is such that α satisfies *DMC*. Using the surplus function S, the decoupling contract is given by

$$v(x,\theta)=S(\theta)+c(\alpha(\theta),\theta)+\frac{c_a(\alpha(\theta),\theta)}{I(\alpha(\theta))}l(x|\alpha(\theta)), \tag{42}$$

where α is implicitly defined by (see (10))

$$(B_a(\alpha(\theta))-C_a(\alpha(\theta),S(\theta),\theta))-\frac{c_{a\theta}(\alpha(\theta),\theta))}{h(\theta)}\int_{\underline{\theta}}^{\theta}C_{u_0}(\alpha(s),S(s),s)h(s)ds=0.$$

Differentiating this expression with respect to θ yields (arguments omitted for simplicity)

$$\underbrace{\left(B_{aa}\alpha'-\frac{dC_a}{d\theta}\right)}_{\mathcal{A}}-\underbrace{\left(\left(\frac{d}{d\theta}\left(\frac{c_{a\theta}}{h}\right)\right)\int_{\underline{\theta}}^{\theta}C_{u_0}h+c_{a\theta}C_{u_0}\right)}_{\mathcal{B}}=0. \tag{43}$$

We will show that for any given θ, and under the assumption that $dc_a(\alpha(\theta),\theta)/d\theta>0$ or, equivalently, that $\alpha'(\theta)>-c_{a\theta}(\alpha(\theta),\theta)/c_{aa}(\alpha(\theta),\theta)$, term $\mathcal{A}$ is bounded above by a bound that is independent of $\bar{u}$ and θ, while term $\mathcal{B}$ is bounded below by an expression that is independent of θ but diverges to infinity as $\bar{u}$ goes to infinity. As a result, there is a threshold $\bar{u}^*$ such that, for all values of $\bar{u}>\bar{u}^*$, (43) fails at any θ where $\alpha'>-c_{a\theta}/c_{aa}$. Hence, for any such $\bar{u}$ we have that $\alpha'\leq -c_{a\theta}/c_{aa}$ for all θ, proving the result.

Consider term $\mathcal{B}$ first. Using (27), we obtain that $C_{u_0}=u_0+c(a,\theta)=S(\theta)+c(\alpha(\theta),\theta)\geq\bar{u}$. Hence, the second term of $\mathcal{B}$ is bounded below by $\bar{u}\min_{a,\theta}c_{a\theta}$, which is independent of θ and diverges to infinity as $\bar{u}$ diverges. Since $\int_{\underline{\theta}}^{\theta}C_{u_0}h\geq 0$, it is thus enough so show that $d(c_{a\theta}/h)/d\theta$ is nonnegative, which follows from $c_{a\theta}/(c_{aa}h)$ is increasing in θ for all a and the premise that

$\alpha' > -c_{a\theta}/c_{aa}$. To see this, notice that

$$\begin{aligned}
\frac{d}{d\theta}\left(\frac{c_{a\theta}}{h}\right) &\underset{s}{=} \frac{c_{aa\theta}\alpha' + c_{a\theta\theta}}{c_{a\theta}} - \frac{h'}{h} \\
&\geq \left(\frac{c_{a\theta\theta}}{c_{a\theta}} - \frac{c_{aa\theta}}{c_{aa}}\right) - \frac{h'}{h} \\
&\underset{s}{=} \frac{\partial}{\partial\theta}\left(\frac{c_{a\theta}}{c_{aa}h}\right) \\
&\geq 0,
\end{aligned}$$

where the first equality in sign follows by differentiation, the first inequality by the premise and $c_{aa\theta} \geq 0$, the second equality in sign by differentiation, and the last inequality by assumption.

Consider now term $\mathcal{A}$. Rewrite (27) as $C = 0.5(u_0 + c)^2 + g$, where $g = 0.5c_a^2/I$. Then $C_a = (u_0 + c)c_a + g_a$ and hence

$$C_a(\alpha(\theta), S(\theta), \theta) = (S(\theta) + c(\alpha(\theta), \theta))\, c_a(\alpha(\theta), \theta) + g_a(\alpha(\theta), \theta).$$

Differentiating with respect to θ and recalling that $S' + c_\theta = 0$ yields

$$\begin{aligned}
\frac{d}{d\theta}C_a(\alpha(\theta), S(\theta), \theta) \quad &= \quad (S + c)(c_{aa}\alpha' + c_{a\theta}) + c_a^2\alpha' + g_{aa}\alpha' + g_{a\theta} \\
&\geq \quad \min_{\theta}\left(-(c_a^2 + g_{aa})\frac{c_{a\theta}}{c_{aa}} + g_{a\theta}\right)
\end{aligned}$$

where the inequality uses the premise $c_{aa}\alpha' + c_{a\theta} > 0$, and that under the assumptions made $g_{aa} \geq 0$. Since a and θ have compact support, it follows that this expression acheives a minimum over θ that is independent of $\bar{u}$. Similarly,

$$B_{aa}\alpha' \leq -B_{aa}\frac{c_{aa}}{c_{a\theta}} \leq \max_{a,\theta}\left(-B_{aa}\frac{c_{aa}}{c_{a\theta}}\right) < \infty.$$

Consequently, term $\mathcal{A}$ has a finite upper bound independent of $\bar{u}$ and we are done. $\square$

A.9 Proof of Lemma 10

Differentiating (30) twice with respect to a yields

$$C_{aa} = c_{aa}\left[(1-a)\varphi_h' + a\varphi_\ell'\right] + (\varphi_h' - \varphi_\ell')\left[(1-2a)c_{aa} + a(1-a)c_{aaa}\right] + a(1-a)c_{aa}^2\left[(1-a)\varphi_h'' + a\varphi_\ell''\right],$$

where we have set $\varphi(v_i) \equiv \varphi_i$, $i = \ell, h$.

Using $c_{aaa} \geq 0$, $\bar{a} \leq 2/3$, and discarding the (positive) terms involving $a(1-a)$ we obtain

$$C_{aa} \geq c_{aa}\left((2-3a)\varphi_h' + (3a-1)\varphi_\ell'\right) > c_{aa}(2-3a+3a-1)\varphi_\ell' > 0,$$

and hence $C_{aa} > 0$, showing that C is strictly convex in a. □

A.10 Proof of Proposition 3

By Theorem 1, it suffices to find conditions under which $\alpha' \leq -c_{a\theta}/c_{aa}$ for all θ. Recall that under decoupling α solves

$$(B_a - C_a)h = c_{\theta a}\int_{\underline{\theta}}^{\theta} C_{u_0} h ds.$$

Differentiating with respect to θ and recalling that $B_{aa} = 0$ yields

$$\alpha' = \frac{c_{a\theta\theta}\int_{\underline{\theta}}^{\theta} C_{u_0}h - c_{a\theta}\frac{h'}{h}\int_{\underline{\theta}}^{\theta} C_{u_0}h + Dh}{-C_{aa}h - c_{aa\theta}\int_{\underline{\theta}}^{\theta} C_{u_0}h}, \tag{44}$$

where $D \equiv c_{\theta a}C_{u_0} + C_{a\theta} - c_\theta C_{u_0 a}$. Since $\bar{a} \leq 2/3$ and $c_{aaa} \geq 0$, it follows that $C_{aa} > 0$ and thus the denominator of (44) is negative. As a result, $\alpha' \leq -c_{a\theta}/c_{aa}$ if and only if for all θ,

$$\left(c_{aa}c_{\theta\theta a} - c_{a\theta}c_{aa\theta} - c_{aa}c_{a\theta}\frac{h'}{h}\right)\int_{\underline{\theta}}^{\theta} C_{u_0} h ds \geq h\left(c_{a\theta}C_{aa} - c_{aa}D\right). \tag{45}$$

We will show that the conditions assumed in the proposition imply that the *rhs* of (45) is negative while the *lhs* is positive, and hence we will be done.

Consider first the *rhs*. The derivatives of C that we need to compute are C_{u_0}, $C_{u_0 a}$, and $C_{a\theta}$. Omitting arguments, and setting $\varphi(v_i) \equiv \varphi_i$, $i = \ell, h$, we obtain

$$\begin{aligned}
C_{u_0} &= \alpha\varphi_h' + (1-\alpha)\varphi_\ell', \qquad (46)\\
C_{u_0 a} &= \left(\varphi_h' - \varphi_\ell'\right) + \alpha(1-\alpha)c_{aa}\left(\varphi_h'' - \varphi_\ell''\right),\\
C_{a\theta} &= \varphi_h'(c_\theta + (1-\alpha)c_{a\theta}) - \varphi_\ell'(c_\theta - \alpha c_{a\theta}) + \alpha(1-\alpha)c_{aa}(\varphi_h''(c_\theta + (1-\alpha)c_{a\theta})\\
&\quad -\varphi_\ell''(c_\theta - \alpha c_{a\theta})) + \alpha(1-\alpha)c_{aa\theta}(\varphi_h' - \varphi_\ell').
\end{aligned}$$

Using these expressions, we can show that $D > 0$. To see this, notice that since $c_{a\theta}C_{u_0} > 0$, it suffices to show that $C_{a\theta} - c_\theta C_{u_0 a} > 0$. But,

$$C_{a\theta} - c_\theta C_{u_0 a} = c_{a\theta}\left((1-\alpha)\varphi_h' + \alpha\varphi_\ell'\right) + \alpha(1-\alpha)\left(c_{aa}c_{a\theta}\left((1-\alpha)\varphi_h'' + \alpha\varphi_\ell''\right) + c_{aa\theta}\left(\varphi_h' - \varphi_\ell'\right)\right),$$

which is positive since φ is strictly convex and $v_h \geq v_\ell$.

From the definition of D and the derivatives of C we obtain

$$c_{a\theta}C_{aa} - c_{aa}D = c_{a\theta}c_{aa}((\varphi_h' - \varphi_\ell')(1-3a) - \varphi_\ell') + a(1-a)(\varphi_h' - \varphi_\ell')(c_{a\theta}c_{aaa} - c_{aa}c_{aa\theta}). \quad (47)$$

We claim that if $c_{a\theta}c_{aaa} - c_{aa}c_{aa\theta} \le 0$, then $c_{a\theta}C_{aa} - c_{aa}D \le 0$ and thus the *rhs* of (45) is negative. To see this, note that since $c_{a\theta}c_{aaa} - c_{aa}c_{aa\theta} \le 0$, the last term of (47) is negative and thus

$$c_{a\theta}C_{aa} - c_{aa}D \le c_{a\theta}c_{aa}\left((1-3a)\varphi_h' - (2-3a)\varphi_\ell'\right) < c_{a\theta}c_{aa}\varphi_h'(1-3a-2+3a) = -c_{a\theta}c_{aa}\varphi_h' < 0,$$

where the second inequality uses the assumption that $\bar{a} \le 2/3$ and $\varphi_\ell' \le \varphi_h'$. But

$$c_{a\theta}c_{aaa} - c_{aa}c_{aa\theta} =_s \frac{c_{aaa}}{c_{aa}} - \frac{c_{aa\theta}}{c_{a\theta}} =_s -\frac{\partial}{\partial a}\frac{c_{a\theta}}{c_{aa}} \le 0,$$

and hence the premise $c_{a\theta}c_{aaa} - c_{aa}c_{aa\theta} \le 0$ holds by hypothesis.

Consider now the *lhs* of (45). Notice that

$$c_{aa}c_{a\theta\theta} - c_{a\theta}c_{aa\theta} =_s \frac{c_{a\theta\theta}}{c_{a\theta}} - \frac{c_{aa\theta}}{c_{aa}} = \frac{\partial}{\partial\theta}\frac{c_{a\theta}}{c_{aa}} \ge 0,$$

which coupled with h decreasing in θ implies that the left side is positive. Hence, (45) holds and decoupling solves the principal's problem. □

A.11 Proof of Proposition 4

Since f is an exponential family, v_θ (unless it is identically zero) can only cross zero once (Chade and Swinkels (2016), henceforth CS, Section 6.3). We will show that in the decoupling menu, v_θ has sign pattern $+/-$. Since α is strictly decreasing, it follows by Theorem 2 that the decoupling menu is feasible and hence, as the solution to a relaxed program, optimal.

To show that v_θ has sign pattern $+/-$, note first that under the stated conditions, any solution to $(B_a - C_a)\, h = c_{a\theta} \int C_{u_0} h$ has $(C_a)_\theta \le 0$ (to see this, use $B_{aa} = 0$, h decreasing, and $c_{aa\theta} = 0$). We will show that if v_θ has sign pattern $-/+$ (strictly) then in fact $(C_a)_\theta > 0$, a contradiction.

Note first that for an exponential family, for any a and a', $l_x\,(x|a')\,/l_x\,(x|a)$ is independent of x (CS, Section 6.2). Thus, fixing some $\hat{a}$ and letting $\hat{l} = l\,(\cdot|\hat{a})$, we can express all relevant contracts as being of the form

$$\frac{1}{u'\,(\pi\,(\theta, x))} = m\,(\theta) + s\,(\theta)\,\hat{l}\,(x) > 0,$$

where $s\,(\theta) > 0$, and where the condition that v_θ has sign pattern $-/+$ is equivalent to $s_\theta > 0$.[16]

[16]To see this, note that $v = \rho(m + s\hat{l})$ implies $v_\theta = \rho'(m_\theta + s_\theta\hat{l})$ and $v_{\theta x} = \rho''(m_\theta + s_\theta\hat{l})s\hat{l}_x + \rho' s_\theta \hat{l}_x$. Hence, at any point where $v_\theta = 0$, we have $v_{\theta x} = \rho' s_\theta \hat{l}_x$, which is positive if and only if $s_\theta > 0$.

From the Envelope Theorem, $C_a = \int \pi f_a + \mu \left(c_{aa} - \int v f_{aa}\right)$. We will show that under the stated conditions, and under the assumption that v_θ has sign pattern $-/+$, each of μ, $\int \pi f_a$, and $c_{aa} - \int v f_{aa}$ increases with θ, with $\int \pi f_a$ increasing strictly. Since $c_{aa} - \int v f_{aa} \geq 0$ (as $c_{aa} \geq 0$ and since, by CS, under each set of conditions, the FOP holds, and hence $\int v f_{aa} \leq 0$), it will then follow that $(C_a)_\theta > 0$, giving the desired contradiction to v_θ being $-/+$.

Consider first μ. Since $s(\theta)\,\hat{l}_x(x) = \mu(\theta)\,l_x(x|\alpha(\theta))$ for all θ, we have

$$s_\theta(\theta)\,\hat{l}_x(x) = \mu_\theta(\theta)\,l_x(x|\alpha(\theta)) + \mu(\theta)\,\alpha'(\theta)\,l_{xa}(x|\alpha(\theta)).$$

By the premise, $s_\theta > 0$, and thus $s_\theta(\theta)\,\hat{l}_x(x) > 0$. But the second term on the right side is negative by the assumption that $l_{ax} \geq 0$ and $\alpha' < 0$. Hence, $\mu_\theta > 0$.

Let us show next that $\int \pi f_a$ is strictly increasing in θ. Note that

$$\left(\int \pi f_a\right)_\theta = \int \pi_\theta f_a + \alpha' \int \pi f_{aa}.$$

Let us first show that the second term is positive. Since $\alpha' < 0$, it would be enough to show that $-\int \pi f_{aa} = \int \pi_x F_{aa} \geq 0$. In case (i), we have $u = \log(w)$, and so $\psi(\tau)$ is the identity function, and hence $\pi_x = \psi'\left(m + s\hat{l}\right) s l_x = s l_x$. Thus, $\int \pi_x F_{aa} = s \int l_x F_{aa} \geq 0$. In the other two cases, the result is immediate since $\pi_x > 0$, and $F_{aa} \geq 0$.

It is thus enough to show that $\int \pi_\theta f_a > 0$. To see this, note first that by construction, $\int v f_a = c_a$, and $S' = \left(\int v f - c\right)_\theta = -c_\theta$. But, then,

$$-c_\theta = \left(\int v f - c\right)_\theta = \left(\int v f_a - c_a\right)\alpha' + \int v_\theta f - c_\theta = \int v_\theta f - c_\theta,$$

and so it follows that $\int v_\theta f = 0$. Thus, by the premise that v_θ is $-/+$ (strictly), and since $1/u'$ is strictly increasing, we have by Beesack's inequality that

$$\int \pi_\theta f = \int \frac{1}{u'} v_\theta f > 0.$$

But, since $\pi_x = \psi'\left(m + s\hat{l}\right) s\hat{l}_x$, we have that

$$\pi_{x\theta} = \psi''\left(m + s\hat{l}\right) s\hat{l}_x \left(m_\theta + s_\theta \hat{l}\right) + \psi'\left(m + s\hat{l}\right) s_\theta \hat{l}_x.$$

Now, for $u = \log(w)$, $\psi(\tau) = \tau$, while for $u = \sqrt{2w}$, $\psi(\tau) = \tau^2$, and so, in both cases, $\psi'' \geq 0$. It thus follows that $\pi_{x\theta} > 0$ whenever $\pi_\theta =_s m_\theta + s_\theta \hat{l} \geq 0$. Let $\hat{m} > 0$ be such that $\int (\pi_\theta - \hat{m}) f = 0$, where since $\int \pi_\theta f > 0$, we have $\hat{m} > 0$. Thus, since $\pi_{x\theta} > 0$ whenever $\pi_\theta \geq 0$, it follows that $\pi_\theta - \hat{m}$ single-crosses 0 from below (and does so strictly, in the sense that $\pi_\theta - \hat{m}$ is zero at only one point). Hence, since f_a/f is strictly increasing in x, Beesack's inequality implies that

$$\int \pi_\theta f_a = \int (\pi_\theta - \hat{m})\, f_a > 0,$$

and it follows that $\int \pi f_a$ strictly increases in θ.

Finally, consider $c_{aa} - \int v f_{aa}$. Since $(c_{aa})_\theta = c_{aaa}\alpha' + c_{aa\theta} \geq 0$, using that $c_{aaa} \leq 0$, it suffices that

$$\left(\int v f_{aa}\right)_\theta = \int v_\theta f_{aa} + \alpha' \int v f_{aaa} \leq 0.$$

But,

$$\alpha' \int v f_{aaa} = -\alpha' \int v_x F_{aaa} \leq 0.$$

since $\alpha' < 0$, $v_x > 0$, and $F_{aaa} \leq 0$. It thus suffices to show that $\int v_\theta f_{aa} \leq 0$, or equivalently, that $\int v_{\theta x} F_{aa} \geq 0$.

Consider first $u = \log w$. Then, ρ is the identity function, which implies that $v = \log\left(m + s\hat{l}\right)$, and so $v_\theta = (m_\theta + s_\theta \hat{l})/(m + s\hat{l})$. Differentiating with respect to x yields

$$v_{\theta x} = \frac{s_\theta \hat{l}_x \left(m + s\hat{l}\right) - \left(m_\theta + s_\theta \hat{l}\right) s\hat{l}_x}{\left(m + s\hat{l}\right)^2} = \frac{ms_\theta - sm_\theta}{\left(m + s\hat{l}\right)^2}\hat{l}_x.$$

But, given that $\int v_\theta f = 0$, and given that v_θ is assumed to have sign pattern $-/+$, Beesack's inequality implies that

$$0 \leq \int v_\theta f_a = \int v_{\theta x} (-F_a) = (ms_\theta - sm_\theta) \int \frac{1}{\left(m + s\hat{l}\right)^2} \hat{l}_x (-F_a)\,,$$

and so, since $\hat{l}_x (-F_a) / \left(m + s\hat{l}\right)^2$ is everywhere positive, we have $ms_\theta - sm_\theta \geq 0$. But then,

$$\int v_{\theta x} F_{aa} = (ms_\theta - sm_\theta) \int \frac{1}{\left(m + s\hat{l}\right)^2} \hat{l}_x F_{aa} \geq 0,$$

where the inequality follows first because $\int \hat{l}_x F_{aa} \geq 0$ by assumption; second, because f is an exponential family, and so (CS, Section 6.3) $\hat{l}_x (-F_{aa})$ has sign pattern $+/-$,; and third, from $m + s\hat{l}\,(x) > 0$ for all x, which implies that $1/\left(m + s\hat{l}\right)^2$ is positive and decreasing. As a result, Beesack's inequality applies, and the inequality obtains.

Consider now case (ii) where $u\,(w) = \sqrt{2w}$. Then, ρ is the identity, so that $v = m + s\hat{l}$, and thus $v_{\theta x} = s_\theta \hat{l}_x > 0$, where we use that $s_\theta > 0$ by assumption. Thus, since $F_{aa} \geq 0$, $\int v_{\theta x} F_{aa} \geq 0$.

Finally, consider case (iii). As before, $\int v_\theta f_a \geq 0$, and so, integrating by parts, $\int v_{\theta x} (-F_a) \geq$

0. We would be done if we could show that $v_{\theta x}$ has sign pattern $+/-$. To see that this would suffice, notice that since we are working with an exponential family, we have that $-F_a$ is lsm, or equivalently, that F_{aa}/F_a is increasing in x (see CS, Proposition 2 and Corollary 3). But then, since $F_{aa} \geq 0$ by assumption, $-F_{aa}/F_a$ is positive and decreasing, and so $\int v_{\theta x} F_{aa} \geq 0$, as desired.

It remains to show that $v_{\theta x}$ has sign pattern $+/-$. But, $v_\theta = \rho'\left(m + s\hat{l}\right)\left(m_\theta + s_\theta \hat{l}\right)$, and so $v_{\theta x} = z\hat{l}_x$, where

$$z = \rho''\left(m + s\hat{l}\right) s \left(m_\theta + s_\theta \hat{l}\right) + \rho'\left(m + s\hat{l}\right) s_\theta. \tag{48}$$

But, $v_{\theta xx} = z_x \hat{l}_x + z\hat{l}_{xx}$, and so where $v_{\theta x} = 0 = z$, $v_{\theta xx} = z_x \hat{l}_x =_s z_x$. Now,

$$z_x = \left(\rho'''\left(m + s\hat{l}\right) s^2 \left(m_\theta + s_\theta \hat{l}\right) + 2\rho''\left(m + s\hat{l}\right) s s_\theta\right) \hat{l}_x.$$

Since $\rho'\left(m + s\hat{l}\right) s_\theta > 0$, where $z = 0$, we have by (48) that $\rho'' \neq 0$, and so

$$m_\theta + s_\theta \hat{l} = -\frac{\rho'\left(m + s\hat{l}\right)}{\rho''\left(m + s\hat{l}\right)} \frac{s_\theta}{s}.$$

Inserting this expression into z_x yields

$$\begin{aligned} z_x &= \left(\rho'''\left(m + s\hat{l}\right) s^2 \left(-\frac{\rho'\left(m + s\hat{l}\right)}{\rho''\left(m + s\hat{l}\right)} \frac{s_\theta}{s}\right) + 2\rho''\left(m + s\hat{l}\right) s s_\theta\right) \hat{l}_x \\ &=_s -\frac{\rho'''\left(m + s\hat{l}\right)}{\rho''\left(m + s\hat{l}\right)} + 2\frac{\rho''\left(m + s\hat{l}\right)}{\rho'\left(m + s\hat{l}\right)} < 0, \end{aligned}$$

where we use that $s_\theta > 0$ by assumption. Therefore, $z_x < 0$ where $v_{\theta x} = 0$, and so $v_{\theta xx} = z_x \hat{l}_x < 0$ as well. Hence, $v_{\theta x}$ has sign pattern $+/-$ as required, and we are done. $\square$

Appendix B Existence in the Relaxed Pure Adverse Selection Problem

Proposition 5 *Let $\hat{C}$ be continuous, jointly convex in (a, u_0) for each θ, and satisfy $\hat{C}_a(0, u_0, \theta) = 0$ and $\lim_{a \to \overline{a}} \hat{C}_a(a, u_0, \theta) = \infty$ for all (u_0, θ). Let $\overline{u}$ be in the interior of the range of u. Then* *there exists a solution to the relaxed pure adverse selection problem*

$$\begin{aligned}
\max_{\alpha,S} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \left(B(\alpha(\theta)) - \hat{C}(\alpha\left(\theta\right), S(\theta), \theta)\right) h(\theta) d\theta \\
s.t. \quad & S'(\theta) = -c_\theta(\alpha(\theta), \theta) \ \text{for almost all } \theta, \text{ and} && (49)\\
& S(\overline{\theta}) = \bar{u}. && (50)
\end{aligned}$$

Before proving the proposition, we comment on the assumptions on $\hat{C}$. For the canonical framework without moral hazard, $\hat{C}(a, u_0, \theta) = \varphi(u_0 + c(a, \theta))$, where $\varphi = u^{-1}$. In this case, $c_a(0, \theta) = 0$ and $\lim_{a \to \overline{a}} c_a(a, \theta) = \infty$ imply the boundary conditions on $\hat{C}_a$, and convexity follows from $u_0 + c$ convex in (a, u_0) and φ strictly increasing and convex.

The situation is more complicated in the decoupling program with $\hat{C} = C$, where C comes from the cost minimization step of the pure moral hazard problem. The boundary conditions can be ensured as above with assumptions on c_a. But although primitives for C convex in a are known, ensuring joint convexity in (a, u_0) is much harder. For the square root utility case analyzed in Section 5.3, all the assumptions are easily satisfied. Moreover, checking the convexity of a numerically generated C for any given set of primitives is straightforward.

Proof Recall that the Hamiltonian associated with the problem is $\mathcal{H} = (B - \hat{C})h - \eta c_\theta$, where η is the co-state variable. To see that $\mathcal{H}$ is concave in (a, u_0), notice that joint concavity requires (i) $\mathcal{H}_{aa} \leq 0$, which follows from B concave in a, $c_{\theta aa} \geq 0$, and $\hat{C}_{aa} \geq 0$; (ii) $\mathcal{H}_{u_0 u_0} \leq 0$, which follows since $\hat{C}_{u_0 u_0} \geq 0$; and (iii) $\mathcal{H}_{aa}\mathcal{H}_{u_0 u_0} - \mathcal{H}_{a u_0}^2 \geq 0$, which follows since $\hat{C}_{aa}\hat{C}_{u_0 u_0} - \hat{C}_{a u_0}^2 \geq 0$.

Given the boundary conditions on $\hat{C}_a$, the optimality conditions are $\partial\mathcal{H}/\partial a = 0$, $\eta'(\theta) = -\partial\mathcal{H}/\partial S$, and $\eta(\underline{\theta}) = 0$, from which we obtain

$$B_a - \hat{C}_a = \frac{c_{a\theta}}{h} \int_{\underline{\theta}}^{\theta} \hat{C}_{u_0} h, \tag{51}$$

plus (49)–(50). The concavity of $\mathcal{H}$ ensures that (49)–(51) are also sufficient. As a result, we will focus on them in our search for a solution (α, S) to the problem.

Define $a^*(s, z, \theta)$ as the solution to

$$B_a(a) - \hat{C}_a(a, s, \theta) = \frac{c_{a\theta}(a, \theta)}{h(\theta)} z \tag{52}$$

where a^* exists from the boundary conditons on $\hat{C}_a$, and is unique from the convexity of $\hat{C}$ and c_θ in a and the concavity of B. We will then be done if we find a solution to the following system of ordinary differential equations:

$$\begin{bmatrix} S'(\theta) \\ Z'(\theta) \end{bmatrix} = \begin{bmatrix} g^S(S(\theta), Z(\theta), \theta) \\ g^Z(S(\theta), Z(\theta), \theta) \end{bmatrix}.$$

with boundary conditions $S(\overline{\theta}) = \overline{u}$ and $Z(\underline{\theta}) = 0$, where

$$\begin{bmatrix} g^S(S(\theta), Z(\theta), \theta) \\ g^Z(S(\theta), Z(\theta), \theta) \end{bmatrix} = \begin{bmatrix} -c_\theta(a^*(S(\theta), Z(\theta), \theta), \theta) \\ C_{u_0}(a^*(S(\theta), Z(\theta), \theta), S(\theta), \theta)h(\theta) \end{bmatrix}.$$

Indeed if we take $\alpha(\theta) = a^*(S(\theta), Z(\theta), \theta)$ then $Z(\theta) = \int_{\underline{\theta}}^{\theta} C_{u_0}(\alpha(t), S(t), t)h(t)dt$. Hence, by definition of a^* and comparing (51) and (52), (α, S) satisfies the relevant conditions.

Define

$$\begin{aligned} u_{\max} &= (\overline{\theta} - \underline{\theta}) \max_{(a,\theta)\in[0,\overline{a}]\times[\underline{\theta},\overline{\theta}]} c_\theta(a, \theta), \text{ and} \\ z_{\max} &= (\overline{\theta} - \underline{\theta}) \max_{(a,\theta)\in[0,\overline{a}]\times[\overline{u},u_{\max}]\times[\underline{\theta},\overline{\theta}]} C_{u_0}(a, s, \theta). \end{aligned}$$

Choose $\delta > 0$ such that $\overline{u} - \delta$ remains in the domain of u, and let $R = [\overline{u}, u_{\max}] \times [0, z_{\max}]$ and $R_\delta = [\overline{u} - \delta, u_{\max} + \delta] \times [-\delta, z_{\max} + \delta]$. Then a^* is Lipschitz on $R_\delta \times [\underline{\theta}, \overline{\theta}]$, and hence so are g^S and g^Z.

Let $\zeta : \mathbb{R}^2 \to [0, 1]$ be a Lipschitz function such that $\zeta(s, z) = 1$ if $(s, z) \in R$ and $\zeta(s, z) = 0$ if $(s, z) \notin R_\delta$. Then $(\zeta g^S, \zeta g^Z)$ is Lipschitz on $\mathbb{R}^2 \times [\underline{\theta}, \overline{\theta}]$, and so by standard results in the theory of differential equations (see, e.g., Theorems 2.3 and 2.6 in Khalil (1992)) there exist continuous functions $\hat{S}$ and $\hat{Z}$ mapping $\mathbb{R} \times [\underline{\theta}, \overline{\theta}]$ into $\mathbb{R}$ such that $\hat{S}(u_0, \underline{\theta}) = u_0$, $\hat{Z}(u_0, \underline{\theta}) = 0$, and

$$\begin{bmatrix} \hat{S}_\theta(u_0, \theta) \\ \hat{Z}_\theta(u_0, \theta) \end{bmatrix} = \begin{bmatrix} \left(\zeta g^S\right)(\hat{S}(u_0, \theta), \hat{Z}(u_0, \theta), \theta) \\ \left(\zeta g^Z\right)(\hat{S}(u_0, \theta), \hat{Z}(u_0, \theta), \theta) \end{bmatrix}.$$

Note that $\hat{S}(u_{\max}, \overline{\theta}) \geq \overline{u}$ by the definition of g^S and $u_{\max}$. Similarly, since $\hat{S}_\theta \leq 0$, $\hat{S}(\overline{u}, \overline{\theta}) \leq \overline{u}$. Hence, by continuity, there exists u^* such that $\hat{S}(u^*, \overline{\theta}) = \overline{u}$. But since $\hat{S}_\theta \leq 0$ and $\hat{Z}_\theta \geq 0$, and using the definition of $z_{\max}$, $(\hat{S}(u^*, \theta), \hat{Z}(u^*, \theta)) \in R$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, and so since $\zeta = 1$ on R, the pair $(S(\cdot), Z(\cdot)) = (\hat{S}(u^*, \cdot), \hat{Z}(u^*, \cdot))$ satisfies the required conditions, and we are done. □

Appendix C Existence and Differentiability in the Moral Hazard Problem

Let W be the domain of the utility function, an interval with infimum $\underline{w}$ and supremum $\bar{w}$. Let $\underline{v} = \lim_{w\to\underline{w}} u(w)$, and let $\bar{v} = \lim_{w\to\bar{w}} u(w)$. Let $\mathcal{E}$ be the set of (a, u_0, θ) such that the relaxed moral hazard problem in Section 3.3 admits a solution $\hat{v}$ where $\hat{v}(\underline{x}) > \underline{v}$ and $\hat{v}(\bar{x}) < \bar{v}$. If we let $\underline{\tau} = \lim_{w \to \underline{w}} \frac{1}{u'(w)}$, and $\bar{\tau} = \lim_{w \to \bar{w}} \frac{1}{u'(w)}$, then it is easy to show that $\hat{v}(\underline{x}) > \underline{v}$ if and only if $\lambda + \mu l(\underline{x}|a) > \underline{\tau}$ for the associated Lagrange multipliers, and similarly, that $\hat{v}(\bar{x}) < \bar{v}$ if and only if $\lambda + \mu l(\bar{x}|a) < \bar{\tau}$.

Lemma 11 *The set $\mathcal{E}$ is open. The multipliers λ and μ are twice continuously differentiable functions of (a, u_0, θ) on $\mathcal{E}$.*

Proof Let

$$G(\lambda, \mu, a, u_0, \theta) = \begin{pmatrix} g_1(\lambda, \mu, a, u_0, \theta) \\ g_2(\lambda, \mu, a, u_0, \theta) \end{pmatrix},$$

where

$$\begin{aligned} g_1(\lambda, \mu, a, u_0, \theta) &= \int \rho(\lambda + \mu l(x|a)) f(x|a)\, dx - c(a, \theta) - u_0, \\ g_2(\lambda, \mu, a, u_0, \theta) &= \frac{\partial}{\partial a} \int \rho(\lambda + \mu l(x|a)) f(x|a)\, dx - c(a, \theta) - u_0. \end{aligned}$$

Let $(a^0, u_0^0, \theta^0) \in \mathcal{E}$, let λ^0 and μ^0 be the associated Lagrange multipliers, and let $\kappa^0 = (\lambda^0, \mu^0, a^0, u_0^0, \theta^0)$. Then, $G(\kappa^0) = 0$, and by definition of $\mathcal{E}$, $\lambda^0 + \mu^0 l(\underline{x}|a^0) > \underline{\tau}$, and $\lambda^0 + \mu^0 l(\bar{x}|a^0) < \bar{\tau}$. We need to show that λ and μ are implicitly defined as C^1 functions of (a, u_0, θ) on a neighborhood of (a^0, u_0^0, θ^0). Since $\lambda + \mu l(\underline{x}|a)$ and $\lambda + \mu l(\bar{x}|a)$ are continuous in (λ, μ, a), it would follow from this that $\mathcal{E}$ is open. We proceed in several steps.

Step 1. We first show that $g_{1\lambda}$ exists at κ^0, and is equal to $\int \rho'(\lambda^0 + \mu^0 l(x|a^0)) f(x|a^0)\, dx$. To show this, we must first show that it is valid to differentiate under the integral. This requires that $\rho(\lambda + \mu l(x|a)) f(x|a)$ be integrable. Since f is continuous on the compact interval $[\underline{x}, \bar{x}]$, it is bounded, and so it is enough to show that $|\rho(\lambda + \mu l(x|a))|$ is bounded. But,

$$\rho(\lambda + \mu l(x|a)) \leq \rho(\lambda^0 + \mu^0 l(\bar{x}|a^0)) < \infty,$$

where we use that $\lambda^0 + \mu^0 l(\bar{x}|a^0) < \bar{\tau}$ by hypothesis, and similarly, $\rho(\lambda + \mu l(x|a)) \geq \rho(\lambda^0 + \mu^0 l(\underline{x}|a^0)) > \infty$, and we are done. Another requirement for passing the derivative through the integral is that $\rho'(\lambda^0 + \mu^0 l(x|a^0)) f(x|a^0)$ is bounded above by an integrable function on some neighborhood of (λ^0, μ^0, a^0). To see this, choose $\underline{\delta}$ and $\bar{\delta}$ such that $\underline{\tau} < \underline{\delta} < \lambda^0 + \mu^0 l(\underline{x}|a^0)$ and $\lambda^0 + \mu^0 l(\bar{x}|a^0) < \bar{\delta} < \bar{\tau}$. Then, since $\lambda + \mu l(\underline{x}|a)$ and $\lambda + \mu l(\bar{x}|a)$ are continuous in (λ, μ, a), there is a neighborhood N of (λ^0, μ^0, a^0) such that

$$\underline{\delta} < \rho(\lambda + \mu l(\underline{x}|a)) < \rho(\lambda + \mu l(\bar{x}|a)) < \bar{\delta}$$

on N. But then, for all x, and everywhere on N,

$$\rho'(\lambda + \mu l(x|a)) \leq \max_{\sigma \in [\underline{\delta}, \bar{\delta}]} \rho'(\sigma) < \infty,$$

where the second inequality follows since ρ is continuously differentiable (with $\rho'(\sigma) = \frac{(u')^3}{-u''}(\psi(\sigma))$) and $[\underline{\delta}, \bar{\delta}]$ is compact. Thus, by Corollary 5.9 in Bartle (1966) (and Billingsley (1995), problem 16.5), we can pass the derivative through the integral and this provides an expression for $g_{1\lambda}$.

Step 2. $g_{1\lambda} = \int \rho'(\lambda + \mu l(x|a)) f(x|a)\, dx$ is itself continuous in (λ, μ, a) at (λ^0, μ^0, a^0). This follows since $\lambda + \mu l(x|a)$ is, under our conditions, uniformly continuous in (λ, μ, a), and ρ' is uniformly continuous in its argument on $[\underline{\delta}, \bar{\delta}]$.

Step 3. By similarly tedious arguments, $g_{1\mu}$, $g_{1a}, g_{2\lambda}, g_{2\mu}$, and g_{2a} are defined as the integral of the relevant derivative, and are continuous. Finally, $g_{i\theta}$ and g_{iu_0} are trivially continuous. Hence, G is continuously differentiable on a neighborhood of κ^0. Indeed, by similar arguments, G is twice continuously differentiable, noting in specific that

$$\rho''(\sigma) = \frac{(u')^3}{-u''}\left[3\frac{u''}{u'} - \frac{u'''}{u''}\right](\psi(\sigma)),$$

and so since u is C^3, ρ'' is continuous on the compact interval $[\underline{\delta}, \bar{\delta}]$, and hence it is bounded.

Step 4. By the argument in Jewitt et al. (2008), $\nabla G(\kappa^0) \neq 0$. Hence, by the Implicit Function Theorem for C^k functions (Fiacco (1983), Theorem 2.4.1), λ and μ are twice continuously differentiable functions of (a, u_0, θ) in a neighborhood of (a^0, u_0^0, θ^0). $\square$

The reader may wonder at the level of detail displayed in this proof. To see that there is something to prove, consider $u = \log w$. Then (see Moroni and Swinkels (2014) for details), it is easy to exhibit first, combinations of c_a, c, and u_0 for which no optimal contract exists, and second, combinations of c_a, c, and u_0 for which the optimal contract has $v(\underline{x}) = -\infty$, and at which the relevant integrals cease to be continuous (let alone differentiable) in the relevant parameters.

Lemma 11 implies that the cost function C is twice differentiable on $\mathcal{E}$, and also that α is continuously differentiable.

Another differentiability argument we have used in the text is about the integrals $\int v_\theta f$ and $\int v f_a$. It can be justified as follows:

Lemma 12 *Let $(\alpha(\theta^0), S(\theta^0), \theta^0) \in \mathcal{E}$. Then, for all a, $\int v(x,\theta) f(x|a)\, dx$ is differentiable in θ at θ^0, with*

$$\frac{\partial}{\partial\theta}\int v(x,\theta^0) f(x|a)\, dx = \int v_\theta(x,\theta^0) f(x|a)\, dx,$$

and similarly, $\int v\left(x,\theta^0\right) f(x|a)\,dx$ *is differentiable in* a *at* a*, with*

$$\frac{\partial}{\partial a}\int v\left(x,\theta^0\right) f(x|a)\,dx = \int v\left(x,\theta^0\right) f_a(x|a)\,dx.$$

Proof We will show the result for the case of differentiation by θ since the other case is similar. We must show first that $v\left(x,\theta^0\right) f(x|a)$ is integrable. This follows as before since

$$\left|v\left(x,\theta^0\right)\right| \leq \max\left(\left|v\left(\underline{x},\theta^0\right)\right|, \left|v\left(\bar{x},\theta^0\right)\right|\right) < \infty.$$

Next we show that, under decoupling, v_θ exists and it is uniformly bounded. To see this, note first that

$$v(x,\theta) = \rho\left(\lambda(\theta) + \mu(\theta)\, l\left(x|\alpha(\theta)\right)\right)$$

and so

$$v_\theta(x,\theta) = \rho'\left(\lambda(\theta) + \mu(\theta)\, l\left(x|\alpha(\theta)\right)\right)\left(\lambda'(\theta) + \mu'(\theta)\, l\left(x|\alpha(\theta)\right) + \mu(\theta)\, l_a\left(x|\alpha(\theta)\right)\alpha'(\theta)\right).$$

As before, let $\underline{\tau} < \underline{\delta} < \lambda^0 + \mu^0 l\left(\underline{x}|a^0\right)$, and let $\lambda^0 + \mu^0 l\left(\bar{x}|a^0\right) < \bar{\delta} < \bar{\tau}$. Since α is continuous, for all θ sufficiently close to θ^0, $\lambda(\theta) + \mu(\theta)\, l\left(x|\alpha(\theta)\right) \in \left[\underline{\delta}, \bar{\delta}\right]$, and so, as before, $\rho'\left(\lambda(\theta) + \mu(\theta)\, l\left(x|\alpha(\theta)\right)\right)$ is uniformly bounded on a neighborhood of θ^0. Also, since α and S are C^1, $\lambda(\theta)$ and $\mu(\theta)$ are continuously differentiable on some neighborhood of θ^0. But then, since l and l_a are uniformly bounded, we can also uniformly bound $\left(\lambda'(\theta) + \mu'(\theta)\, l\left(x|\alpha(\theta)\right) + \mu(\theta)\, l_a\left(x|\alpha(\theta)\right)\alpha'(\theta)\right)$ on the relevant neighborhood. It follows that v_θ is uniformly bounded on the neighborhood, and the lemma follows from Bartle (1966), Corollary 5.9. $\square$

Of course, for decoupling to work, it has to be that the resultant moral-hazard subproblem has a solution for each θ. That is, we need to know that $\left(\alpha(\theta), S(\theta), \theta\right) \in \mathcal{E}$ for all θ. By Moroni and Swinkels (2014), one set of conditions is given by the following lemma.

Lemma 13 *Assume that* $\bar{w} = \bar{v} = \infty$, $\underline{w} = \underline{v} = -\infty$, $\underline{\tau} = 0$, *and* $\bar{\tau} = \infty$. *Then, for all* (a, u_0, θ), $(a, u_0, \theta) \in \mathcal{E}$.

Proof Direct from Moroni and Swinkels (2014). $\square$

This Lemma, however, does not cover important cases such as $u = \ln(w)$ or $u = \sqrt{w}$, because in each case, $\underline{w} = 0 > -\infty$. Our next lemma covers $u = \sqrt{w}$, but does not cover $u = \ln w$.

Lemma 14 *Let* $\bar{w} = \bar{v} = \infty$, $\underline{w} = 0$, *and* $\bar{\tau} = \infty$. *Assume further that* $\rho'(\tau)\,\tau$ *is increasing and diverges in* τ. *Then, there is a threshold* $\hat{u}$ *such that for all* $\bar{u} \geq \hat{u}$, $\left(\alpha(\theta), S(\theta), \theta\right) \in \mathcal{E}$ *for all* θ.

Proof For any given a, and $\mu > 0$, let $i(\mu, a) = \int \rho(\mu(l(x|a) - l(\underline{x}|a))) f_a(x|a)\, dx$. Note that

$$\begin{aligned} i(\mu, a) &= \int \rho'(\mu(l(x|a) - l(\underline{x}|a))) \mu l_x(x|a)(-F_a(x|a))dx \\ &= \int \frac{1}{l(x|a) - l(\underline{x}|a)} [\rho'(\mu(l(x|a) - l(\underline{x}|a))) \mu(l(x|a) - l(\underline{x}|a))] l_x(x|a)(-F_a(x|a))dx, \end{aligned}$$

and so, since $\rho'(\tau)\tau$ is increasing in τ, it follows that the bracketed term, and hence $i(\cdot, a)$, is increasing in μ. Let $m = \min_a l(\bar{x}|a) - l(\underline{x}|a) > 0$, and let

$$\sigma = -\min_{\left\{(x,a)| \frac{m}{2} \leq l(x|a) - l(\underline{x}|a) \leq \frac{3m}{4}\right\}} l_x(x|a) F_a(x|a) > 0.$$

Then,

$$\begin{aligned} i(\mu, a) &\geq \int_{\left\{x| \frac{m}{2} \leq l(x|a) - l(\underline{x}|a) \leq \frac{3m}{4}\right\}} \frac{\rho'(\mu(l(x|a) - l(\underline{x}|a))) \mu(l(x|a) - l(\underline{x}|a))}{l(x|a) - l(\underline{x}|a)} l_x(x|a)(-F_a(x|a))dx \\ &\geq \frac{4\sigma}{3m} \int_{\left\{x| \frac{m}{2} \leq l(x|a) - l(\underline{x}|a) \leq \frac{3m}{4}\right\}} \rho'(\mu(l(x|a) - l(\underline{x}|a))) \mu(l(x|a) - l(\underline{x}|a))\, dx \\ &\geq \frac{4\sigma}{3m} \rho'\left(\mu \frac{m}{2}\right) \mu \frac{m}{2} \int_{\left\{x| \frac{m}{2} \leq l(x|a) - l(\underline{x}|a) \leq \frac{3m}{4}\right\}} dx \\ &\geq \frac{4\sigma}{3m} \frac{m}{4 \max_{\{x,a\}} l_x(x|a)} \rho'\left(\mu \frac{m}{2}\right) \mu \frac{m}{2} \\ &= \frac{\sigma}{3 \max_{\{x,a\}} l_x(x|a)} \rho'\left(\mu \frac{m}{2}\right) \mu \frac{m}{2}, \end{aligned}$$

where the first inequality follows from the fact that the integrand is positive, the second from $l(x|a) - l(\underline{x}|a) \leq 3m/4$, the third from the monotonicity of $\rho'(\tau)\tau$, and the fourth by integration. Notice that the lower bound on $i(\mu, a)$ thus obtained diverges in μ. Hence, there exists $\hat{\mu}$ such that $i(\mu, a) > c_a(a, \bar{\theta})$ for all a, and $\mu > \hat{\mu}$. Let

$$\hat{u} = \max_a \int \rho(\hat{\mu}(l(x|a) - l(\underline{x}|a))) f(x|a)\, dx \leq \rho\left(\hat{\mu} \max_a (l(\bar{x}|a) - l(\underline{x}|a))\right) < \infty.$$

It follows from Proposition 1 of Moroni and Swinkels (2014), along with $i(\cdot, a)$ increasing, that $(\alpha(\theta), S(\theta), \theta) \in \mathcal{E}$ for all θ for any $\bar{u} > \hat{u}$. In particular, at any θ, $S(\theta) + c(\alpha(\theta), \theta) > \bar{u} > \hat{u}$. $\square$

Finally, let us consider the case $u = \log w$ (for which $\rho'(\tau)\tau$ is identically 1, so the previous result does not apply). Then, as in the proof of the previous lemma,

$$\begin{aligned} i(\mu, a) &\geq \frac{4\sigma}{3m} \int_{\left\{x \mid \frac{m}{2} \leq l(x|a) - l(\underline{x}|a) \leq \frac{3m}{4}\right\}} \rho'\left(\mu\left(l\left(x|a\right) - l\left(\underline{x}|a\right)\right)\right) \mu\left(l\left(x|a\right) - l\left(\underline{x}|a\right)\right) dx \\ &= \frac{4\sigma}{3m} \int_{\left\{x \mid \frac{m}{2} \leq l(x|a) - l(\underline{x}|a) \leq \frac{3m}{4}\right\}} dx \\ &\geq \frac{4\sigma}{3m} \frac{m}{4 \max_{\{x,a\}} l_x\left(x|a\right)} \\ &\equiv s, \end{aligned}$$

and so, if we assume that $c_a\left(\bar{a}, \bar{\theta}\right) \leq s$, then Proposition 1 of Moroni and Swinkels (2014) applies.

References

Bartle, R. G. (1966): *The Elements of Integration*. John Wiley & Sons.

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

Billingsley, P. (1995): *Probability and Measure*. John Wiley & Sons.

Bolton, P., and M. Dewatripont (2005): *Contract Theory*. MIT Press.

Chade, H., and J. Swinkels (2016): "The No-Upward-Crossing Condition and the Moral Hazard Problem," Arizona State University Working Paper.

Fagart, M.-C. (2002): "Wealth Effects, Moral Hazard and Adverse Selection in a Principal-Agent Model," INSEE.

Fiacco, A. (1983): *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Press, Orlando, FL.

Fudenberg, D., and J. Tirole (1991): *Game Theory*. Cambridge University Press.

Gottlieb, D., and H. Moreira (2013): "Simultaneous Adverse Selection and Moral Hazard," Mimeo.

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

Guesnerie, R., and J.-J. Laffont (1984): "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *The Journal of Public Economics*, 25, 329–369.

Halac, M., N. Kartik, and Q. Liu (2016): "Optimal Contracts for Experimentation," *Review of Economic Studies*, 83, 1040–1091.

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

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

Jewitt, I., O. Kadan, and J. Swinkels (2008): "Moral Hazard with Bounded Payments," *Journal of Economic Theory*, 143, 59–82.

Kadan, O., and J. Swinkels (2013): "On the Moral Hazard Problem without the First Order Approach," *Journal of Economic Theory*, 148(6), 2313–2343.

Khalil, H. (1992): *Nonlinear Systems*. Macmillan Publishing Company.

Laffont, J.-J., and D. Martimort (2002): *title=The theory of incentives: the principal-agent model, author=Laffont, Jean-Jacques and Martimort, David, year=2009, The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Laffont, J.-J., and J. Tirole (1986): "Using Cost Observations to Regulate Firms," *Journal of Political Economy*, pp. 614–641.

Maskin, E., and J. Riley (1984): "Monopoly with Incomplete Information," *The RAND Journal of Economics*, 15(2), 171–196.

Milgrom, P., and I. Segal (2002): "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70, 583–601.

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

Mitrinovic, D. (1970): *Analytic Inequalities*. Springer.

Moroni, S., and J. Swinkels (2014): "Existence and Non-Existence in the Moral Hazard Problem," *Journal of Economic Theory*, 150, 668–682.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

——— (1982): "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics*, 10(1), 67–81.

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

Strulovici, B. (2011): "Contracts, Information Persistence, and Renegotiation," Northwestern Unverisity.

Williams, N. (2015): "A Solvable Continuous Time Principal Agent Model," *Journal of Economic Theory*, 159, 989–1015.