

Discussion Paper Series – CRC TR 224

Discussion Paper No. 784
Project B 02

Optimal Delegation in a Multidimensional World

Andreas Kleiner¹

September 2026

¹Department of Economics, University of Bonn. Email: andreas.kleiner@uni-bonn.de.

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
through CRC TR 224 is gratefully acknowledged.

Optimal Delegation in a Multidimensional World

Andreas Kleiner*

September 2026

We study a model of delegation in which a principal takes a multidimensional action and an agent has private information about a multidimensional state. We provide necessary and sufficient conditions for an arbitrary mechanism to maximize the principal's expected payoff. A key step shows that a function is the indirect utility of an incentive-compatible mechanism if and only if it is convex and lies below the agent's full-discretion payoff. We apply our optimality conditions to convex delegation sets and characterize when a single budget optimally links decisions. Linking can preserve discretion even when separate delegation would eliminate it. Optimal linking can also require a nonconvex delegation set, even with uniformly distributed types. If the principal cares little about aligned decisions, every optimal mechanism randomizes: risk in an aligned decision provides incentives in the biased ones at little cost to her.

*Department of Economics, University of Bonn. Email: andreas.kleiner@uni-bonn.de. I thank Ricardo Alonso, Kyle Bagwell, Alex Frankel, Nima Haghpanah, Ravi Jagadeesan, Navin Kartik, Elliot Lipton, Alejandro Manelli, Lina Uhe, Kun Zhang, and various seminar audiences for useful comments and discussions. I acknowledge financial support from the German Research Foundation (DFG) through Germany's Excellence Strategy - EXC 2047/1-390685813, EXC 2126/1-390838866, and the CRC TR-224 (Project B02).

1. Introduction

In many economic and political environments, a principal faces a better-informed but biased agent. The principal can choose a permissible set of actions and ‘delegate’ the decision to the agent: a firm appoints a manager to select investment levels in different projects; the US Congress delegates power to federal agencies; a legislature forms a committee to draft bills; a regulator lets a monopolist choose prices. The principal may yield some discretion to the agent to utilize his informational advantage but may impose restrictions on the agent’s actions to counter his bias. Following Holmström (1977), an extensive literature models such delegation problems by assuming that both the action and the state of the world lie in a one-dimensional space. A main result of this literature characterizes when it is optimal for the principal to constrain the agent’s choice to lie in an interval, and this conclusion has been used to explain why managers face spending caps, regulators impose price ceilings, and trade agreements specify maximum tariff levels.

The assumption that the action and state spaces are one-dimensional is made for tractability. In many applications, the underlying states and actions are more complex and more realistically modeled as multidimensional: managers invest in several projects, Congress delegates many decisions to the EPA, and committees draft multiple bills. What mechanisms are optimal in such multidimensional settings? How robust are conclusions obtained for one-dimensional models? And can we still expect that relatively simple mechanisms are often optimal?

To study these questions, we consider a principal who takes a multidimensional action and faces an agent with private information about a multidimensional state of the world (the agent’s *type*). Payoffs depend on the action and the state of the world, and transfers are infeasible. The principal can design arbitrary mechanisms, including stochastic ones, to maximize her expected payoff. Our main result characterizes, for an arbitrary mechanism, when this mechanism is optimal. While our primary aim is to understand multidimensional models, our approach yields novel insights already for one-dimensional models because our conditions apply to any mechanism. For multidimensional convex delegation sets we obtain a simple optimality condition that naturally extends familiar conditions for interval delegation in one-dimensional models. We also obtain conditions under which linking all decisions through a single budget is optimal. As extensions, we allow more general payoffs for the principal and an agent with an outside option, covering veto bargaining and monopoly regulation.

To illustrate, consider the following example, to which we return throughout the paper. The state has two independent components, each uniformly distributed on $[0, 1]$, and payoffs

are quadratic and additively separable across the two dimensions. The agent's ideal action is the state, and the principal's ideal action is lower by a common bias $\beta > 0$ in each dimension. If the two decisions had to be delegated separately, the one-dimensional theory would apply to each of them: in the optimal mechanism the agent may choose any action up to a cap. However, a powerful insight in mechanism design is that linking decisions can be useful even if they are independent (Jackson and Sonnenschein, 2007).

Indeed, the optimal action cap with a single decision is so restrictive that in high states of the world, both the principal and the agent would prefer higher actions. But if the principal were to allow the agent to take higher actions, the agent would choose them too often and the principal would be better off with the restriction. With several dimensions, however, the principal can allow the agent to take a more extreme action in one dimension in exchange for a more moderate action in the other. How should she optimally link the two decisions? Figure 1 shows that even in this uniform example, the optimal delegation sets can take different forms as the bias varies.

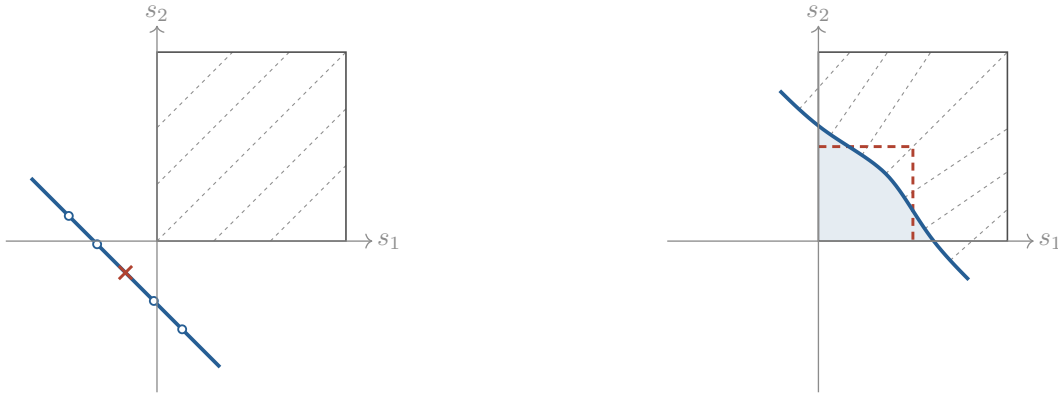


Figure 1: Optimal delegation sets in the two-dimensional uniform example: left for $\beta = 2/3$, right for $\beta = 1/4$. Optimal separate delegation is shown in red.

For every bias, it is optimal to delegate a set of actions to the agent, even though the principal could also use stochastic mechanisms, and our results determine the shape of that set. Moreover, linking remains valuable however large the bias: if $\beta \geq 1/2$, there would be no value to delegation in separate delegation problems, and the optimal delegation set would collapse to a single action. By bundling the two dimensions, the principal can still give the agent discretion, and the optimal delegation set is a line, as in the left panel for $\beta = 2/3$. It is then optimal to fix the *sum* of the two actions while leaving their *difference* to the agent; Corollary 6 shows that this is not special to the uniform distribution. For $\beta < 1/2$, by contrast, the optimal set is nonconvex and has a curved boundary, as the right panel of Figure 1 illustrates for $\beta = 1/4$. The rate at which the agent may trade a higher action in one dimension against a lower action in the other varies along the boundary, so optimal

linking is not simply a cap on the sum of the two decisions.

An important step is understanding which mechanisms are feasible for the principal in the first place. We characterize the feasible mechanisms by analyzing the *indirect utility* of a mechanism, which assigns to any type the payoff this type would get in the mechanism by choosing his report optimally. This payoff cannot exceed the *full discretion payoff*, i.e., the payoff this type would get if he could choose the action without any restrictions. Moreover, our assumption that the agent’s utility is an affine function of the state implies that the indirect utility must be a convex function. Lemma 1 shows that any function satisfying these two properties is the indirect utility of an incentive-compatible mechanism.

This characterization is easy to use and already helpful for one-dimensional delegation models. Our formulation differs from the previous literature, which often considered only deterministic mechanisms. The set of incentive-compatible deterministic mechanisms is not convex, since a convex combination of two such mechanisms is in general not incentive compatible.¹ Moreover, a common approach is to first treat the model as one with transfers and then impose that these transfers are zero. Compared to this approach, formulating the problem via indirect utilities is more direct and provides valuable geometric insights into which mechanisms can be optimal. For the multidimensional problem, working with indirect utilities allows us to express incentive compatibility through convexity, building on Rochet (1987).

To find the optimal mechanism, we formulate the principal’s problem in terms of indirect utilities. To illustrate and simplify the approach, we assume in Section 4 that the principal’s payoff has the same curvature as the agent’s payoff (we extend the analysis to more general preferences for the principal in Section 5). In this case, the problem becomes a linear program, and we use linear programming duality to derive necessary and sufficient conditions for a given mechanism to be optimal. Typically, optimal mechanisms pool certain types, and our main result shows that a mechanism is optimal if and only if, conditional on any pooling region, a stochastic dominance condition is satisfied. Intuitively, this condition requires that, restricted to the pooling region (where the indirect utility function is affine), no feasible indirect utility does better than the mechanism under consideration. If the pooling regions are at most one-dimensional, the stochastic dominance condition has a simple formulation in terms of majorization. Using this observation, we provide necessary and sufficient conditions for a convex delegation set with a smooth boundary to be optimal when payoffs are quadratic. These conditions are easy to check and extend those of Amador and Bagwell (2013) to the

¹ Some earlier papers also consider stochastic mechanisms (or allow for money burning or restricted transfers) and obtain a convex set of mechanisms; see, for example, Kováč and Mylovánov (2009), Amador and Bagwell (2013), Ambrus and Egorov (2017), Amador and Bagwell (2020), Kartik, Kleiner, and Van Weelden (2021), and Kleiner, Moldovanu, and Strack (2021).

multidimensional setting. We show that optimal mechanisms are deterministic if virtual values are positive in the interior of the type space and negative only on one of its faces. With more general preferences, however, every optimal mechanism randomizes if the principal cares little about decisions on which she agrees with the agent, since risk in these decisions provides incentives in the others.

The rest of the paper proceeds as follows. After reviewing the related literature, we present our model in Section 2, characterize incentive-compatible mechanisms in Section 3, and optimal mechanisms in Section 4. In Section 5 we consider more general preferences for the principal, and in Section 6 we incorporate outside options and apply them to veto bargaining and to the regulation of a monopolist.

Related Literature For the most part, the literature on delegation has focused on problems in which both the action and state spaces are one-dimensional; for important contributions, see Holmström (1977), Melumad and Shibano (1991), Alonso and Matouschek (2008), Kováč and Mylovanov (2009), Amador and Bagwell (2013), Kolotilin and Zapechelnuyk (2025), and Amador, Bagwell, and Carpizo (2025).

A few delegation papers do consider richer action and/or type spaces. Armstrong (1995) considers an agent with two-dimensional private information and discusses several applications. Since the principal’s action is assumed to be one-dimensional (and only interval delegation sets are considered), there is only limited scope to screen two-dimensional types in his analysis. Koessler and Martimort (2012) characterize the optimal mechanism in a setting where two decisions depend on a single-dimensional underlying state. Galperti (2019) studies personal budgeting for a consumer with self-control problems. He models the consumer’s problem as a delegation problem with multidimensional information and actions but restricts the principal’s choice to a particular class of “budgeting mechanisms”. The closest paper to ours is Frankel (2016), which studies the delegation of several independent decisions and therefore has multidimensional action and state spaces. For quadratic preferences with a constant bias, he shows that if the states are independently and identically distributed according to normal distributions then it is optimal to delegate a ‘half space’. Without the normality assumption, he shows that suitable half-space delegation sets yield a payoff per decision converging to the first-best as the number of independent decisions grows, provided the state distributions have uniformly bounded variances. Frankel (2014) also considers multidimensional delegation problems and characterizes the max-min optimal mechanism, which maximizes the principal’s payoff against the worst-case preference type of the agent. Recently, Alonso, Gan, and Hu (2026) show that convex delegation sets are robustly optimal if the designer faces uncertainty about the agent’s preferences and evaluates mechanisms using a max-min criterion.

The elicitation of multidimensional private information has also been studied in general mechanism design (e.g., Jackson and Sonnenschein, 2007) and cheap-talk environments (Chakraborty and Harbaugh, 2007; Lipnowski and Ravid, 2020). Jackson and Sonnenschein (2007) show that linking independent copies of a decision problem can yield expected payoffs per decision approaching those of an ex ante efficient allocation as the number of decisions grows. Our results can be used to study how the principal should optimally link decisions, which can be important if the number of decisions is limited.

On a methodological level, our work is related to the literature on multidimensional mechanism design, and in particular on multiproduct monopolists (see, e.g., Rochet, 1987; Rochet and Choné, 1998; Manelli and Vincent, 2007; Daskalakis, Deckelbaum, and Tzamos, 2017; Haghpour and Hartline, 2021; Lahr and Niemeyer, 2024). Rochet and Choné (1998) characterize optimality using mean-preserving spreads within bunching regions, and Daskalakis, Deckelbaum, and Tzamos (2017) obtain stochastic-dominance conditions for the optimality of finite-menu mechanisms. We develop related conditions for delegation without transfers.

2. Model

A principal chooses an action $a \in \mathbb{R}^n$. An agent is privately informed about the state of the world $s \in S$, where $S \subseteq \mathbb{R}^n$ is compact and convex and has non-empty interior. The agent's payoff $u_A(a, s)$ and the principal's payoff $u_P(a, s)$ depend on both the action and the state of the world. Specifically,

$$u_A(a, s) := a \cdot s + b(a), \tag{1}$$

where $b : \mathbb{R}^n \rightarrow \mathbb{R}$ is strictly concave and differentiable, and we assume $\max_{a \in \mathbb{R}^n} a \cdot s + b(a)$ has a solution for all $s \in \mathbb{R}^n$.

Example 1. *A common specification assumes that the agent's preferences are Euclidean, so that the payoff from action a in state s is $-\|a - s\|^2 = 2a \cdot s - \|a\|^2 - \|s\|^2$. After dividing by two and adding a term that does not depend on the action, and hence does not change preferences over actions, this is an instance of our payoffs.*

We assume that the principal's payoff $u_P(a, s)$ is continuous in a and s , concave and differentiable in a for each s , and the gradient $\nabla_a u_P(a, s)$ is differentiable in s for all a . In Section 4, we will use the additional structure that we obtain by imposing

Assumption 1. *The principal's payoff is*

$$u_P(a, s) := a \cdot g(s) + b(a), \tag{2}$$

where $g : S \rightarrow \mathbb{R}^n$ is continuously differentiable.

We assume that the state s is distributed according to a probability distribution F with continuously differentiable density f and support S . The principal aims to maximize her expected payoff and can design arbitrary mechanisms.

The revelation principle applies and we define a *mechanism* to be a function $m : S \rightarrow \Delta(\mathbb{R}^n)$ such that all expected payoffs are finite and integrable.² To simplify notation, we extend the domain of $b(\cdot)$ and $u_i(\cdot, s)$ linearly to include probability distributions over $\mathbb{R}^n$, so that $b(m(s)) = \mathbb{E}_{m(s)}[b(a)]$ and analogously for $u_i(\cdot, s)$. A mechanism is *incentive compatible* if for all s and s' in S ,

$$u_A(m(s), s) \geq u_A(m(s'), s).$$

Discussion To formulate a tractable model of multidimensional delegation, we opted for a formulation that yields a simple characterization of incentive compatible mechanisms instead of going for the most general formulation.

For example, parts of the literature on delegation impose exogenous constraints on the set of feasible actions. While such constraints can be relevant for some applications, we opted to not impose them to obtain a simpler characterization of feasible mechanisms. The general approach of characterizing feasible mechanisms via indirect utilities is still useful in the presence of such constraints, but the indirect utilities must then satisfy additional constraints.

Also, we allow for stochastic mechanisms even though from a practical perspective we might sometimes mainly be interested in deterministic mechanisms. Knowing that specific deterministic mechanisms are optimal in the larger class of stochastic mechanisms is interesting from a theoretical perspective and can be useful for applications of the framework to dynamic extensions (e.g., Ali, Kartik, and Kleiner, 2023). It also yields a clean characterization of incentive compatible mechanisms because it makes the set of feasible mechanisms convex. The lotteries used by stochastic mechanisms in our framework play a role similar to money burning that is allowed in parts of the literature (e.g., Amador and Bagwell, 2013; Ambrus and Egorov, 2017) and we would obtain the same characterization of feasible mechanisms if we instead allowed for money burning.

Finally, we assume that the agent's preference can be captured by a payoff that is affine in the state for each action. This assumption is common in delegation and mechanism design more generally. There is a sense in which the problem becomes a lot more difficult if preferences are not affine in the state: Kartik and Kleiner (2024) show that, when the agent chooses among lotteries over a multidimensional action space, this affine form is in a sense essential for local incentive constraints being sufficient for global incentive compatibility.

² We denote by $\Delta(\mathbb{R}^n)$ the set of Borel probability measures on $\mathbb{R}^n$.

More abstract action spaces can in principle be handled in a similar way, by relabeling each action according to the payoff it induces for the agent. This again yields a payoff of the form we assume, but additional complications arise: the relabeled actions need not fill out the whole space, and distinct actions may become indistinguishable for the agent while still mattering to the principal.

3. Characterizing incentive-compatible mechanisms

We follow Rochet (1987) and characterize the set of incentive-compatible mechanisms in terms of their indirect utilities. To any incentive-compatible mechanism m corresponds an *indirect utility* $U : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by

$$U(s) := \max_{s' \in S} \mathbb{E}[m(s')] \cdot s + b(m(s')).$$

Which indirect utilities correspond to some incentive-compatible mechanism? First, any indirect utility is convex as the maximum of a family of functions that are affine in the state s . In some mechanism design settings with transfers, this is all it takes to be the indirect utility of some incentive-compatible mechanism (e.g., Rochet and Choné, 1998). However, in the absence of transfers additional restrictions apply. In particular, if the principal cannot transfer money to the agent, the agent's utility cannot be higher than if he was free to choose his action. Defining the *full discretion payoff* $h : \mathbb{R}^n \rightarrow \mathbb{R}$ (i.e., the maximum payoff a type could ever get) by

$$h(s) := \max_{a \in \mathbb{R}^n} a \cdot s + b(a),$$

$U \leq h$ is therefore clearly necessary.³ The following result shows that these two conditions characterize the set of feasible indirect utilities.

Lemma 1. *An indirect utility U corresponds to an incentive-compatible mechanism if and only if U is convex and lies below the full discretion payoff: $U \leq h$.*

Intuitively, if U is convex then it would correspond to an incentive-compatible mechanism if transfers were available and the agent had quasi-linear preferences. If the required transfers are all negative (that is, payments from the agent to the principal) then we can use the agent's risk aversion (coming from the strict concavity of b) to simulate these transfers via stochastic actions. If $U \leq h$, one can indeed show that the required transfers are negative.

This last step relies on our choice that the domain of U and h is large enough. For example, Rochet and Choné (1998) define U only as a function on S ; however, requiring only that $U(s) \leq h(s)$ for all $s \in S$ is not sufficient for our conclusion. Figure 2 illustrates a

³ We denote the pointwise order by $\leq$, so $U \leq h$ means $U(s) \leq h(s)$ for all $s \in \mathbb{R}^n$.

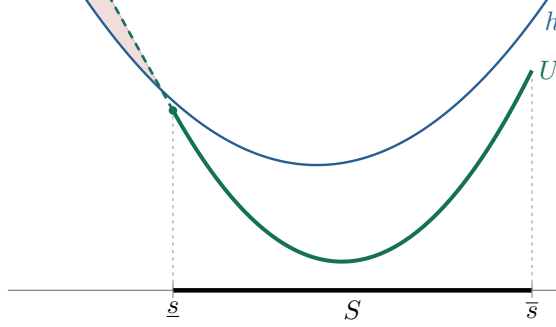


Figure 2: The function U satisfies $U(s) \leq h(s)$ for all $s \in S$ but does not correspond to a feasible mechanism because every convex extension of U lies above the dashed ray.

convex function U defined on S which lies below h on all of S , but which does not correspond to a mechanism because it cannot be extended to a convex function that lies below h on all of $\mathbb{R}^n$ (which must be satisfied for any indirect utility by the simple implication in Lemma 1).

Proof. We have argued for the necessity of these conditions before the statement of the lemma. For the converse direction, let us first recall basic observations from convex analysis. The convex conjugate of a function U is denoted by U^* and defined by $U^*(a) := \sup_{s \in \mathbb{R}^n} a \cdot s - U(s)$. We will use the following facts, which follow immediately from this definition: (i) $h = (-b)^*$, (ii) $U \leq h$ implies $h^* \leq U^*$, and (iii) $a \in \partial U(s)$ implies $U^*(a) = a \cdot s - U(s)$.⁴

Suppose U is convex and satisfies $U \leq h$. Fix $s \in S$ and $a \in \partial U(s)$. Facts (iii), (ii) and (i), together with $(-b)^{**} = -b$, give

$$U(s) = a \cdot s - U^*(a) \leq a \cdot s - h^*(a) = a \cdot s + b(a),$$

which is the payoff type s obtains from action a for sure. Because b is strictly concave, we can therefore assign to type s a lottery $m(s)$ with expected value a that yields payoff $u_A(m(s), s) = U(s)$. It remains to show that m is incentive compatible. For all s and s' ,

$$\begin{aligned} u_A(m(s), s) &= U(s) \geq U(s') + \mathbb{E}[m(s')] \cdot (s - s') \\ &= \mathbb{E}[m(s')] \cdot s' + b(m(s')) + \mathbb{E}[m(s')] \cdot (s - s') = u_A(m(s'), s), \end{aligned}$$

where the first inequality follows since $\mathbb{E}[m(s')] \in \partial U(s')$.

Q.E.D.

Figure 3 illustrates Lemma 1 for one-dimensional types and quadratic payoffs. It shows two indirect utilities that correspond to incentive-compatible mechanisms. In Figure 3a, all types between s_1 and s_2 obtain their full discretion payoff and U is affine below s_1 and above s_2 . This indirect utility can be obtained by letting types choose their preferred action from the interval of deterministic actions $[s_1, s_2]$. The indirect utility in Figure 3b contains an

⁴ Here, $\partial U(s)$ denotes the subdifferential of U at s . To see (iii), note that the definition of U^* implies $U^*(a) \geq a \cdot s - U(s)$. Conversely, $a \in \partial U(s)$ implies that for all s' , $a \cdot s - U(s) \geq a \cdot s' - U(s')$. Taking the supremum of the right-hand side with respect to s' yields $a \cdot s - U(s) \geq U^*(a)$.

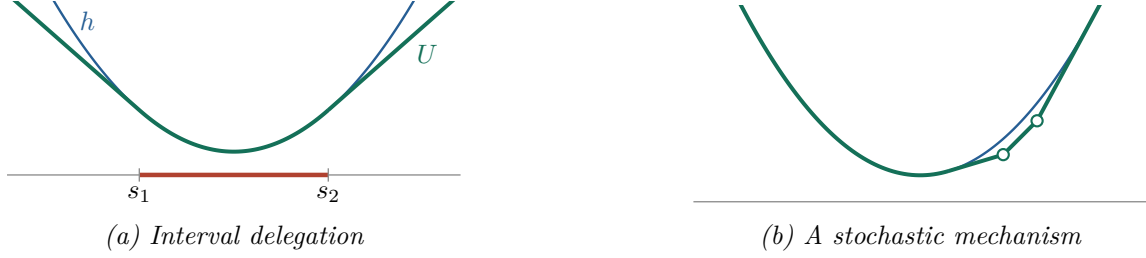


Figure 3: Examples of indirect utilities. The blue curves show the function h for one-dimensional types and quadratic payoffs. The green curves show indirect utilities corresponding to incentive-compatible mechanisms.

affine piece that lies strictly below the graph of h . This part of the indirect utility corresponds to types that obtain a stochastic action, which yields no type its full discretion payoff.

Lemma 1 also illuminates a connection between optimal delegation problems and certain Bayesian persuasion problems established in Kolotilin and Zapechelnyuk (2025) and Kleiner, Moldovanu, and Strack (2021). In a mean-measurable persuasion problem (see Gentzkow and Kamenica, 2016; Kolotilin, 2018), the designer is choosing a Blackwell experiment that reveals information about a one-dimensional state to a receiver. The designer's payoff is determined by the expected value of the receiver's posterior belief; the designer's problem is therefore to choose a distribution of posterior means to maximize her expected payoff.

A distribution of posterior means, G , can be induced by some experiment if and only if it is a mean-preserving contraction of the prior F (Kolotilin, 2018); if the support of F is contained in $[\underline{s}, \bar{s}]$ this in turn is equivalent to

$$U_G(s) := \int_{\underline{s}}^s G(x) dx \leq \int_{\underline{s}}^s F(x) dx =: U_F(s)$$

for all s and with equality for $s = \bar{s}$. Clearly, U_G and U_F are increasing convex functions (satisfying $U_G(\underline{s}) = U_F(\underline{s})$, $U'_G(\underline{s}) = U'_F(\underline{s})$, $U_G(\bar{s}) = U_F(\bar{s})$, and $U'_G(\bar{s}) = U'_F(\bar{s})$). Since any choice of U_G satisfying these properties determines a feasible choice of G , one can model the designer in the persuasion problem as choosing a convex function that lies below a given convex function (and coincides with this function at $\underline{s}$ and $\bar{s}$), see Gentzkow and Kamenica (2016). Lemma 1 illustrates therefore the close connection between the persuasion problem and the one-dimensional delegation problem. Kolotilin and Zapechelnyuk (2025) and Kleiner, Moldovanu, and Strack (2021) show how the additional requirement on U_G can be satisfied without loss of optimality (by choosing the interval $[\underline{s}, \bar{s}]$ a lot larger than the support of F), in which case the two problems become isomorphic. Turning to multidimensional problems, Lemma 1 shows that multidimensional delegation problems can also be formulated as the designer choosing a convex function below a given function. However, multidimensional versions of the persuasion problem do not correspond to choosing convex functions in general (Kleiner, Moldovanu, Strack, and Whitmeyer, 2024) and, therefore, the two problems are

not equivalent in higher dimensions.

4. Characterizing optimal mechanisms

We characterize optimal mechanisms in this section under the maintained assumption that the principal's preferences satisfy Assumption 1. To do so, we first formulate the principal's problem in terms of indirect utilities. We then state the main characterization of optimal mechanisms in Section 4.2 and illustrate the result for particular mechanisms. Finally, we outline the proof of the main result in Section 4.6.

4.1. Formulating the principal's problem

Consider an indirect utility U that corresponds to some incentive-compatible mechanism. In general, there are many incentive-compatible mechanisms that induce the same indirect utility; however, under our assumptions on payoffs, all such mechanisms induce the same payoff for the principal.

Example 2. *If the agent's payoff is quadratic (i.e., $u_A(a, s) = a \cdot s - \frac{1}{2}\|a\|^2$), an indirect utility determines, almost everywhere, the mean action and the total variance of each assigned lottery. Under Assumption 1, these two moments also determine the principal's payoff, so she is indifferent among all mechanisms that induce the same indirect utility. If instead she weighs deviations differently across dimensions, the direction of randomization matters; Section 5 studies that case.*

To show this more generally, let m be an incentive-compatible mechanism with corresponding indirect utility U . Using $\nabla U(s) = \mathbb{E}[m(s)]$ (by the envelope theorem) and $U(s) = \nabla U(s) \cdot s + b(m(s))$, the principal's payoff from mechanism m in state s is completely determined by U :

$$\mathbb{E}[m(s)] \cdot g(s) + b(m(s)) = \nabla U(s) \cdot [g(s) - s] + U(s). \quad (3)$$

This observation implies in addition that the principal's payoff is a linear function of U . Therefore, a solution to the principal's problem can be found at an extreme point of the feasible set. Since in higher dimensions the set of extremal convex functions is a lot richer than in one-dimensional settings (see Johansen, 1974; Bronshtein, 1978; Lahr and Niemeyer, 2024), we proceed to characterize optimal mechanisms.

One disadvantage of the reformulated objective in (3) is that the indirect utility U and its gradient ∇U appear. As is standard in multidimensional mechanism design (see, for example, Rochet and Choné, 1998), we can use the divergence theorem to reformulate the objective function so that only U appears. For intuition, suppose first $S = [s, \bar{s}]$. We can

use integration by parts to obtain

$$\begin{aligned} & \int U(s)f(s) + U'(s)[g(s) - s]f(s) \, ds \\ &= \int U(s) \left[f(s) - [[g(s) - s]f(s)]' \right] \, ds + U(\bar{s})(g(\bar{s}) - \bar{s})f(\bar{s}) - U(\underline{s})(g(\underline{s}) - \underline{s})f(\underline{s}). \end{aligned}$$

In higher dimensions, we can use the divergence theorem (or integration by parts dimension-by-dimension) to obtain the following analogue:

$$\begin{aligned} & \int \left[U(s) + \nabla U(s) \cdot [g(s) - s] \right] \, dF(s) \\ &= \int U(s) \left[f(s) - \operatorname{div}[(g(s) - s)f(s)] \right] \, ds + \int_{\operatorname{bd} S} U(s)[g(s) - s]f(s) \cdot \hat{\mathbf{n}}_S(s) \, d\mathcal{H}(s), \end{aligned}$$

where div denotes the divergence of a function, for any set A its boundary is denoted by $\operatorname{bd} A$, $\mathcal{H}$ denotes the $n - 1$ -dimensional Hausdorff measure on the boundary of S , and $\hat{\mathbf{n}}_S(s)$ denotes the unit outward normal vector to the convex set S at $s \in \operatorname{bd} S$. Note that the second term vanishes if the density f is zero on the boundary of the type space, in which case the following expressions simplify.

Collecting the two weights that multiply $U(s)$ in this identity, define

$$\nu(s) := \begin{cases} f(s) - \operatorname{div}[(g(s) - s)f(s)] & \text{if } s \in \operatorname{int} S \\ [g(s) - s]f(s) \cdot \hat{\mathbf{n}}_S(s) & \text{if } s \in \operatorname{bd} S, \end{cases} \quad (4)$$

and let λ be the Lebesgue measure on S plus the Hausdorff measure on the boundary of S , so that the principal's payoff from an indirect utility U is $\int U(s)\nu(s) \, d\lambda(s)$. Denoting by μ the (signed) measure with density ν relative to λ , that is, $\mu(E) = \int_E \nu(s) \, d\lambda(s)$, we can write the principal's problem as

$$\begin{aligned} & \max_{U \text{ convex}} \int U(s) \, d\mu(s) \\ & \text{s.t. } U \leq h. \end{aligned} \quad (\text{P})$$

Geometric intuition Since $\int U \, d\mu = \int U(s)\nu(s) \, d\lambda(s)$, the objective in (P) weights the utility of each type s by $\nu(s)$, and we can interpret ν as a *virtual value for utility*. Heuristically, $\nu(s)$ measures how much the principal's payoff increases if the indirect utility of type s is increased, but where types on the boundary get extra weight. Our formulation of the principal's problem suggests a geometric approach to think about optimal mechanisms. According to (P), the principal aims to give those types s with $\nu(s) > 0$ a utility as high as possible, where the full discretion payoff h provides an upper bound. Conversely, for types with $\nu(s) < 0$ the principal aims to assign utilities as low as possible, but convexity of U provides a bound on how low the utility of a given type can be given other types' utilities.

Let us illustrate with a one-dimensional example, where we assume that ν is positive

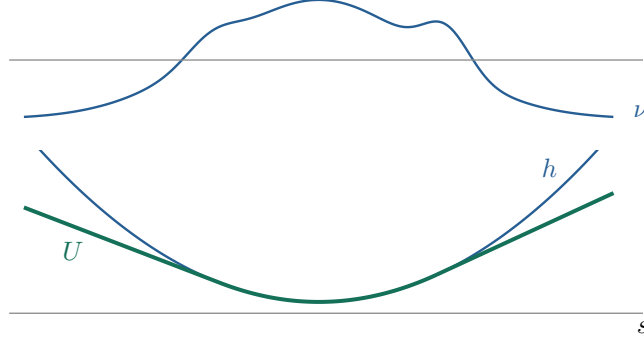


Figure 4: A one-dimensional example illustrating why interval delegation is optimal if ν is positive on some interval and negative otherwise.

on some interval of intermediate types and negative outside this interval (see Figure 4). As we discuss below, assumptions commonly used in the literature are sufficient for this. The principal will want to give low and high types—for whom $\nu(s) < 0$ —utilities as low as possible until the convexity constraint binds, which makes U affine below some cutoff and above some cutoff. For the types in between, $\nu(s) > 0$, and the principal sets their utility as high as possible, hence equal to their full discretion payoff. An optimal mechanism can therefore be chosen to be interval delegation, as illustrated in Figure 3a. We establish this formally in Corollary 2.

Example 3. We return to the example from the introduction. Let $S = [0, 1]^2$, let F be uniform on S , and let preferences be quadratic with a common bias in the two dimensions:

$$b(a) = -\frac{1}{2}\|a\|^2, \quad g(s) = s - (\beta, \beta), \quad \beta > 0.$$

The agent's ideal action is s , whereas the principal's is $s - (\beta, \beta)$. The full-discretion payoff is $h(s) = \|s\|^2/2$, and the virtual value for utility in (4) simplifies to

$$\nu(s) = \begin{cases} 1 & \text{if } s \in \text{int } S, \\ \beta & \text{if } s \text{ lies on the west or south edge,} \\ -\beta & \text{if } s \text{ lies on the north or east edge.} \end{cases} \quad (5)$$

The principal therefore wants to lower utility on the north and east edges and raise it elsewhere. This suggests pooling types near those edges with types in the interior.

4.2. Optimal mechanisms

We next characterize when a given mechanism is optimal. To do so, we introduce additional notation: Given an indirect utility U , we let $\mathcal{Q}$ denote the partition of $\mathbb{R}^n$ into maximal sets on which U is affine. Two types belong to the same partition element exactly when they can be assigned the same lottery over actions, so the partition elements are the pooling

regions of the mechanism. This description is unambiguous at every type at which U is differentiable. At a kink, U is affine on several maximal sets containing that type, and we specify in Section A.1 which of them the type is assigned to.⁵

We denote by $\{\mu|_Q\}_{Q \in \mathcal{Q}}$ a conditional measure of μ given Q . Given $\mathcal{Q}$, we say that some property holds for $|\mu|$ -a.e. $Q \in \mathcal{Q}$ if it holds for every Q outside a subcollection of $\mathcal{Q}$ whose union has $|\mu|$ -measure zero.

For any $Q \in \mathcal{Q}$, there is at most one $s \in Q$ such that $U(s) = h(s)$: U is affine on a convex set containing Q , $U \leq h$, and h is strictly convex.⁶ If such a point exists, we let δ_Q be the point mass of mass $\mu|_Q(Q)$ at it; if $U(s) < h(s)$ for all $s \in Q$, then δ_Q is the zero measure. For any two signed measures α and β , we write $\alpha \leq_{cx} \beta$ and say that α is smaller than β in the *convex order* if $\int c(s) d\alpha \leq \int c(s) d\beta$ for every convex function $c : \mathbb{R}^n \rightarrow \mathbb{R}$. We can now state the main characterization result:

Theorem 1. *Suppose the principal's preferences satisfy Assumption 1 and let U be a feasible indirect utility. Then U is optimal if and only if for $|\mu|$ -a.e. $Q \in \mathcal{Q}$, $\mu|_Q(Q) \geq 0$ and $\mu|_Q \leq_{cx} \delta_Q$.*

Why are the conditions in Theorem 1 sufficient for U to be optimal? Consider a partition element $Q \in \mathcal{Q}$ and suppose δ_Q is a positive point mass at s^* . Then any feasible indirect utility V will be convex, lie below U at s^* with $U(s^*) = h(s^*)$, and U will be an affine function on Q . Together with the fact $\mu|_Q \leq_{cx} \delta_Q$ this implies

$$\int V d\mu|_Q \leq \int V d\delta_Q \leq \int U d\delta_Q = \int U d\mu|_Q.$$

Indeed, the first inequality holds because V is convex and $\mu|_Q \leq_{cx} \delta_Q$, the second because $V(s^*) \leq U(s^*)$ and δ_Q is a positive point mass at s^* , and the equality because U is affine on Q and $\mu|_Q \leq_{cx} \delta_Q$. It follows that, restricted to Q , the principal's objective is at least as large under U as under any feasible V . If this holds for almost all Q , then U is optimal. The conclusion that these conditions are also necessary shows that the problem can in some sense be decomposed: whenever the principal can improve U conditional on Q , she can extend this improved version to a feasible indirect utility that yields unconditionally a higher payoff.

Pooling and the sign of virtual values. The characterization also helps us find candidate optimal mechanisms. Pooling must combine types with positive and negative virtual values: the principal sacrifices utility for the former in order to reduce utility for the latter,

⁵ The set of kinks is Lebesgue-null since U is convex, but it can still carry positive mass under μ because of the extra measure on the boundary of S . If the density f vanishes on $\text{bd } S$, the issue does not arise: μ is then absolutely continuous, U is differentiable at $|\mu|$ -almost every type, and $\mathcal{Q}$ is the partition into the level sets of ∇U .

⁶ Since $h = (-b)^*$ and conjugacy exchanges differentiability and strict convexity, h is differentiable and strictly convex; see Hiriart-Urruty and Lemaréchal (2004, E.4.1).

and this trade-off must be justified within each pooling region, not just across the mechanism as a whole. In particular, when ν is positive in the interior and negative only on parts of the boundary, every nontrivial pool containing interior types must reach a negatively weighted part of the boundary. This suggests starting from those boundary types and constructing pooling regions that extend into the interior. The sign pattern thus narrows the search for an optimum before its allocation has been determined, and we use this insight below.

4.3. Optimal delegation sets

A particularly simple mechanism has the principal delegate the decision to the agent, possibly restricting the agent's action to some subset of actions. Note that any deterministic mechanism can be implemented as an indirect mechanism in this way. For a nonempty closed set $A \subseteq \mathbb{R}^n$, *delegating to A* means that the agent chooses his preferred action among the ideal actions of types in A ; the types indexing the permitted actions need not belong to S . We say that delegating to A is optimal if this mechanism is optimal. In the following we specialize Theorem 1 to these mechanisms.

On the real line, convex order can be checked by comparing total masses, first moments, and the integrals of the convex functions $\max\{s - t, 0\}$ for all t (Shaked and Shanthikumar, 2007, Theorem 3.A.1). This simplifies Theorem 1 whenever U is affine on at most one-dimensional sets; as we will see, this is useful even if the type space is multidimensional. We first consider interval delegation with one-dimensional types (for earlier characterizations, see Alonso and Matouschek, 2008; Amador and Bagwell, 2013).

Corollary 1. *Suppose $n = 1$ and $s_1, s_2 \in S$ with $s_1 < s_2$. Delegating to $[s_1, s_2]$ is optimal if and only if*

- (i) $\nu(s) \geq 0$ for all $s \in [s_1, s_2]$,
- (ii) $\int_{[s, \bar{s}]} (x - s) d\mu(x) \leq 0$ for all $s \in [s_2, \bar{s}]$ with equality for $s = s_2$, and
- (iii) $\int_{[\underline{s}, s]} (s - x) d\mu(x) \leq 0$ for all $s \in [\underline{s}, s_1]$ with equality for $s = s_1$.

The partition induced by this mechanism consists of the two tails $(-\infty, s_1]$ and $[s_2, \infty)$ and the singletons between them. Condition (i) gives the comparison on the singletons, and conditions (ii) and (iii) express the convex-order comparisons on the two pooling regions.

Figure 5 illustrates condition (ii) of Corollary 1. For $s > s_2$, a small upward tilt of the indirect utility starting at s raises the utility of every type $x \geq s$ in proportion to $x - s$. Its effect on the principal's expected payoff is therefore proportional to $\int_{[s, \bar{s}]} (x - s) d\mu(x)$, and condition (ii) rules out a profitable tilt. Equality at s_2 expresses in addition optimality with respect to tilting the utility downwards at s_2 , which is feasible since U is strictly convex in a neighborhood to the left of s_2 .

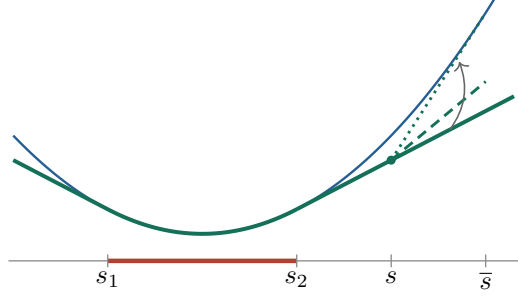


Figure 5: Optimality of interval delegation

Interestingly, the conditions identified in Corollary 1 are in our setting equivalent to the ones obtained in Amador and Bagwell (2013, Proposition 2a). At first glance, this might be surprising since we characterize optimality of interval delegation in the class of stochastic mechanisms and Amador and Bagwell characterize optimality in the class of deterministic mechanisms (and stochastic mechanisms can do strictly better in general). Figure 5 illustrates why the conditions are the same: Suppose the principal strictly benefits from deviating to the dashed indirect utility, which represents a stochastic mechanism. Since her payoff is linear in U , the arguments in the previous paragraph imply that she also benefits from deviating to the dotted indirect utility. Since the dotted linear utility corresponds to a deterministic mechanism, we conclude that conditions (ii) and (iii) in Corollary 1 are necessary for interval delegation to be optimal in the class of deterministic mechanisms (and necessity of condition (i) can be shown easily).

For applications, it can be useful to have simple sufficient conditions that ensure interval delegation is optimal for some interval. The following corollary provides such a sufficient condition.

Corollary 2. *Suppose $n = 1$ and there is an interval $I \subseteq S$ such that the restriction of μ to I is nonnegative and its restriction to $S \setminus I$ is nonpositive. Then delegating to a closed interval, possibly a singleton, is optimal.*

The proof starts from an optimal indirect utility and uses Theorem 1 to show that at most two of its pools carry negative virtual weight. Alternatively, one can prove it using the geometric ideas discussed in Section 4.1.

In the present payoff setting, the hypothesis of Corollary 2 includes the global sign condition in Amador, Bagwell, and Frankel (2018, Proposition 2(a)), which requires ν to be nonnegative throughout the interior of S . A simple implication of our result is the following, which can be useful for applications (see Kováč and Mylovanov, 2009, for a related condition for settings with quadratic preferences).

Corollary 3. *Suppose $n = 1$ and $g(s) = s + \beta$ for some $\beta \in \mathbb{R}$. If f is log-concave, then*

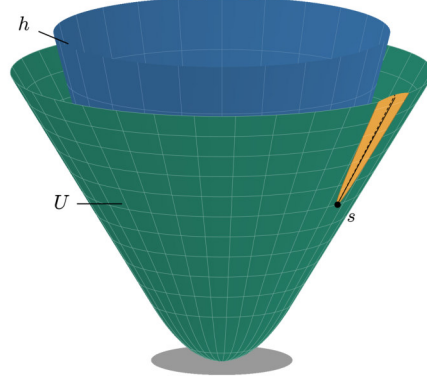


Figure 6: Indirect utility of a convex delegation set and a stochastic perturbation. Adding a lottery raises utility to the orange planar patch; the dashed line shows the tilt along the ray.

delegating to a closed interval, possibly a singleton, is optimal.

For a one-dimensional type space, the approach used in Corollary 1 simplifies the conditions in Theorem 1 for any mechanism, not just interval delegation. The same idea is useful with multidimensional types. Suppose $b(a) = -\|a\|^2/2$, so each type's ideal action equals his type, and let A be nonempty, closed and convex. Delegating to A assigns each type his Euclidean projection onto A . For each $a \in \text{bd } A$, let $Q_a := a + N_A(a)$ denote the set of types whose projection is a , where $N_A(a)$ is the normal cone to A at a . If the boundary of A is differentiable, Q_a is an outward normal ray. For any $s \in Q_a$, the translated ray $s + N_A(a)$ consists of the types on this pooling ray that lie beyond s , moving away from A .

Corollary 4. Suppose $b(a) = -\|a\|^2/2$ and $A \subseteq S$ is closed, convex, has nonempty interior and a differentiable boundary. Delegating to A is optimal if and only if

- (i) $\nu(s) \geq 0$ for all $s \in \text{int } A$, and
- (ii) for $|\mu|$ -almost every normal ray Q_a , with $a \in \text{bd } A$, we have $\mu|_{Q_a}(Q_a) \geq 0$ and

$$\int_{s+N_A(a)} \|x - s\| d\mu|_{Q_a}(x) \leq 0 \quad \text{for all } s \in Q_a,$$

with equality for $s = a$.

The conditions in Corollary 4 closely resemble those in Corollary 1. Indeed, Condition (i) in either case requires that ν is nonnegative on the set of types that obtain their full discretion payoffs, and Condition (ii) (and Conditions (ii) and (iii), respectively) imposes that for each point on the boundary the analogous stochastic dominance condition holds.

The economic interpretation of condition (ii) is analogous to that of the interval delegation conditions. Along a pooling ray, an upward tilt starting at s raises utility in proportion to the distance from s , and the integral in (ii) gives the corresponding conditional change in the principal's payoff. By adding a suitable lottery to the menu, the principal can produce such

a tilt with an arbitrarily small utility increase confined to a small neighborhood of the black line segment in Figure 6. No deterministic mechanism can reproduce this change in indirect utility: an affine patch generated by a deterministic mechanism must extend to a plane tangent to h , whereas the orange patch extends to a plane strictly below h everywhere. This illustrates the role of stochastic mechanisms in the multidimensional perturbation argument.

Budget delegation. We call a set of the form $\{a : w \cdot a \leq c\}$ a *budget set*: it lets the agent choose any action whose weighted sum does not exceed c . With quadratic payoffs and constant bias β , an optimal budget must have weights positively proportional to β . Other directions either restrict actions the wrong way or leave the agent unrestricted in a direction of disagreement, where an additional cap would improve the principal's payoff. We therefore use weights β : the budget restricts the component of the action along β and leaves the orthogonal components unrestricted. With equal biases, this is simply a cap on the sum of the actions. The nontrivial pooling regions in a budget set are rays parallel to w , and, as for the normal rays in Corollary 4, the comparisons required by Theorem 1 are one-dimensional.

Write $Z = \beta \cdot s$ and let Y be the component of the state orthogonal to β . Types pooled together have the same Y and differ only in Z . Optimality can therefore be characterized using the distribution of Z conditional on Y .

Corollary 5. *Suppose $b(a) = -\|a\|^2/2$ and $g(s) = s - \beta$ for some $\beta \neq 0$, and fix $c \in \mathbb{R}$. Let $Z = \beta \cdot s$ and let Y be the component of s orthogonal to β .*

Delegating to the budget set $\{a : \beta \cdot a \leq c\}$ is optimal if and only if, conditional on almost every Y , the cap c is optimal in the one-dimensional quadratic problem with state Z and principal's ideal action $Z - \|\beta\|^2$.

A single budget must solve all these conditional problems with the same cap. Even with a large bias, a budget that always binds need not be optimal, because the principal may prefer different pooled choices on different lines.

For two independent decisions, this yields a simple sufficient condition for budget delegation:

Corollary 6. *Suppose $n = 2$, $b(a) = -\|a\|^2/2$, $g(s) = s - (\beta, \beta)$ for a scalar $\beta > 0$, and $S = [\underline{s}, \bar{s}]^2$, and let the two coordinates of the state be independent draws from a common log-concave density that is symmetric about the midpoint $m = (\underline{s} + \bar{s})/2$. If $\beta \geq \frac{\bar{s} - \underline{s}}{2}$, then budget delegation is optimal.*

If $\beta \geq \frac{\bar{s} - \underline{s}}{2}$, it can be verified that in the one-dimensional problem the principal would not give any discretion to the agent and optimally decide without consulting him. In the two-dimensional problem, budget delegation is then optimal.

4.4. The uniform example

We return to our uniform example to illustrate. Recall that $b(a) = -\|a\|^2/2$ and s is uniformly distributed on $[0, 1]^2$. Figure 1 shows the optimal delegation sets in the two bias regimes, which we now derive.

Large bias ($\beta \geq 1/2$). At these bias levels separate delegation would eliminate all discretion, and applying Corollary 6 implies that budget delegation is optimal. To illustrate the conditions of Theorem 1, we use it to argue directly that budget delegation with the budget set $a_1 + a_2 \leq 1 - 2\beta$ is optimal. For this delegation set, the agent chooses his perpendicular projection onto its boundary, so types are pooled along segments parallel to $(1, 1)$. Recall that ν equals one in the interior of the square, β on the west and south edges, and $-\beta$ on the north and east edges. Since these segments are parallel, the conditional interior density is constant along each of them. The positive and negative endpoint weights cancel in total mass, but they shift the signed barycenter southwest by (β, β) relative to the segment's midpoint. All these midpoints lie on $s_1 + s_2 = 1$, so the signed barycenters lie exactly on the budget line and each assigned action is its pool's signed barycenter. The remaining convex-order comparisons required by Theorem 1 hold because ν is negative only on the north and east edges and the budget line lies south-west of the type space.⁷

Small bias ($0 < \beta < 1/2$). Here the budget line defined above passes through the type space and the convex-order conditions fail. We construct instead a curved boundary, giving types in a southwest region their full-discretion payoff. By symmetry it suffices to describe the branch above the diagonal.

To determine the boundary of the delegation set, we parametrize it by $t \mapsto a(t)$, where $a(t)$ is the action that type $(t, 1)$ chooses. Consider first a smooth decreasing portion with $a(t)$ in $[0, 1]^2$ and $a'_1(t) > 0$. Since the agent picks the closest point of the delegation set, the types pooled at $a(t)$ form the segment from $a(t)$ to $(t, 1)$, and this segment is normal to the boundary:

$$a'(t) \cdot [(t, 1) - a(t)] = 0. \quad (6)$$

By Theorem 1, $a(t)$ must be the signed barycenter of the conditional virtual measure on that segment.

Consider the thin band T_ε between the pools labelled t and $t + \varepsilon$, shown in Figure 7. The pool labelled t' leaves the north edge at $(t', 1)$ and is normal to the boundary at $a(t')$. Simple geometric arguments then yield that the horizontal width of T_ε at the height of $a(t)$

⁷ Indeed, the segment from the assigned action to the pool's north-east endpoint contains the pool within S . On this segment, a convex function lies below the affine function that agrees with it at the two ends. Since the only negative mass of $\mu|_Q$ is at the north-east endpoint, the convex function has the smaller integral against $\mu|_Q$.

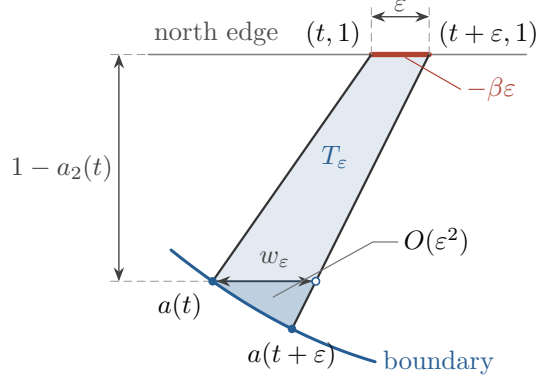


Figure 7: The thin band between neighboring pools.

is

$$w_\varepsilon = \varepsilon + [1 - a_2(t)] \left[\frac{a'_2(t + \varepsilon)}{a'_1(t + \varepsilon)} - \frac{a'_2(t)}{a'_1(t)} \right], \quad (7)$$

while at the north edge the width is ε . Up to a region of area $O(\varepsilon^2)$, T_ε is therefore a trapezoid of vertical extent $1 - a_2(t)$ with parallel sides w_ε and ε . Its interior carries virtual mass $\frac{1}{2}[1 - a_2(t)](w_\varepsilon + \varepsilon) + O(\varepsilon^2)$, and the vertical coordinate of its barycenter is

$$a_2(t) + [1 - a_2(t)] \frac{w_\varepsilon + 2\varepsilon}{3(w_\varepsilon + \varepsilon)} + O(\varepsilon).$$

The north-edge segment carries virtual mass $-\beta\varepsilon$ at height one. Balancing the vertical first moments at the height of $a(t)$ and dividing by $1 - a_2(t)$ gives

$$\frac{1 - a_2(t)}{6} (w_\varepsilon + 2\varepsilon) = \beta\varepsilon + O(\varepsilon^2). \quad (8)$$

Dividing by ε and letting ε vanish, the boundary solves

$$[1 - a_2(t)] \left(3 + [1 - a_2(t)] \frac{d}{dt} \frac{a'_2(t)}{a'_1(t)} \right) = 6\beta. \quad (9)$$

Thus the barycenter condition yields a second-order differential equation that determines how the slope of the boundary changes. Section B.4 fixes the two constants of its solution by imposing that the boundary condition holds at the pool that collapses to the northwest corner, and that it meet its reflection on the diagonal. It also verifies that the convex-order conditions are satisfied and the delegation set is optimal.

Two features are worth noting. First, wherever $1 - a_2(t) < 2\beta$, the equation requires the slope a'_2/a'_1 to increase with t , and hence with a_1 , which implies that the delegation set is not convex. The appendix establishes the stronger conclusion that no closed convex delegation set is optimal for $0 < \beta < 1/2$. Second, the boundary also has a portion outside the square, where a pool reaches the west edge before reaching its assigned action. Therefore, some types near the northwest corner choose an action with a negative first coordinate.

4.5. When is randomization unnecessary?

This subsection uses Theorem 1 to give restrictions on where an optimal mechanism can randomize and where the action it assigns can jump. Both are governed by where virtual values are negative. Randomization and jumps are confined to the convex hull of the types whose utility the principal would like to hold down, so when those types lie on a single face of the type space, an optimal mechanism is deterministic and the corresponding delegation set is connected.

Common regularity conditions require that the bias and density do not vary too quickly, so that virtual values ν are positive in the interior of S . Write $K = \text{co}(\text{supp } \mu^-)$ for the convex hull of the support of the negative part of μ .

Proposition 1. *Suppose $\nu > 0$ in $\text{int } S$. Then every optimal indirect utility U is continuously differentiable on $\text{int } S \setminus K$, and each type in this region receives the deterministic action $\nabla U(s)$, which varies continuously with the type. Thus, up to a set of types of probability zero, randomization can occur only in K .*

To see why, keep U fixed on the types that carry negative virtual value and raise it everywhere else as far as convexity and the bound h permit. This does not lower the principal's payoff, so by optimality it leaves her payoff unchanged; since ν is strictly positive in the interior, an optimal U therefore already coincides with the raised function. That function is built from the fixed values together with points on the graph of h . A type outside K cannot be reached from the fixed values alone, so a plane supporting the raised function there must also touch h and hence correspond to a deterministic action. For a convex function, differentiability on an open set also makes the gradient continuous, which gives the continuous action rule.

If K is contained in a proper face of S , we get the stronger result:

Corollary 7. *Suppose $\nu > 0$ in $\text{int } S$ and $\text{supp } \mu^-$ is contained in a proper face of S . Then every optimum is deterministic almost everywhere and has a continuous action rule on $\text{int } S$, and there is a connected optimal delegation set.*

The corollary follows because a proper face does not meet the interior of S , so the proposition applies throughout a connected set of types, over which the chosen action varies continuously.

4.6. Proof sketch

To simplify the exposition, we assume throughout this subsection that the density vanishes on the boundary of the type space, so $f(s) = 0$ for all $s \in \text{bd } S$. Then ν vanishes on $\text{bd } S$,

so μ is absolutely continuous with respect to Lebesgue measure. Since a convex function is differentiable Lebesgue-almost everywhere, U is differentiable at $|\mu|$ -almost every type, and $\mathcal{Q}$ is simply the partition of $\mathbb{R}^n$ into the level sets of ∇U : no choice has to be made at kinks.

To prove Theorem 1, we use duality in linear programming. To formulate the dual program, it is more convenient to work with indirect utilities that are defined on a compact domain. But recall that it is not enough in Lemma 1 to only require that $U(s) \leq h(s)$ for all $s \in S$ (see Figure 2). The following technical result ensures that we can restrict the indirect utilities to have a compact domain as long as this domain is chosen large enough.

Lemma 2. *There is a compact convex $X \subseteq \mathbb{R}^n$ such that the principal's problem can be written as $\max\{\int U d\mu | U : X \rightarrow \mathbb{R}, U \text{ convex}, U(x) \leq h(x) \forall x \in X\}$.*

Formally, we choose X large enough that the restriction to S of every solution of the problem on X extends to a convex function on $\mathbb{R}^n$ that still lies below h ; by Lemma 1 this extension is a feasible indirect utility with the same payoff. The choice of X has two parts. First, X contains a neighborhood of S and of $g(S)$, on which a solution is bounded above by h ; and since a solution does at least as well as always choosing the action 0, it cannot be too low on S either. Together these two bounds limit how steep a solution can be on S . Second, for each slope a that remains possible, X contains the type whose ideal action is a . This is the type at which a plane of slope a comes closest to h , so checking the plane there places it below h everywhere; taking the supremum of the supporting planes of a solution at types in S then gives the required extension.

Now let X be as in the above lemma and denote by $\mathcal{U}$ the set of convex continuous functions that map X to $\mathbb{R}$ and by $\mathcal{M}_+$ the set of positive measures on X . We can formulate the principal's problem as follows:

$$\begin{aligned} & \max_{U \in \mathcal{U}} \int U(s) d\mu(s) \\ & \text{s. t. } U \leq h \end{aligned} \tag{P}$$

The dual problem We will show that the following problem is the dual problem:

$$\begin{aligned} & \inf_{\gamma \in \mathcal{M}_+} \int h(s) d\gamma(s) \\ & \text{s. t. } \gamma \geq_{cx} \mu. \end{aligned} \tag{D}$$

Note that h is a convex function; therefore, if μ was a positive measure, this would be a trivial problem with solution $\gamma = \mu$. However, since μ is a signed measure and γ has to be a positive measure, μ is not feasible in general.

It is easy to see that weak duality holds, i.e., that the value of the primal problem (P) is

always below the value of the dual problem (D). Indeed, for feasible U and γ ,

$$\int U(s) d\mu(s) \underbrace{\leq}_{(i)} \int U(s) d\gamma(s) \underbrace{\leq}_{(ii)} \int h(s) d\gamma(s) \quad (10)$$

since (i) U is convex and $\mu \leq_{cx} \gamma$ and (ii) γ is a positive measure and $U \leq h$. The following result shows that strong duality holds, that is the optimal values of both problems are equal and the dual problem has a solution.

Lemma 3 (Strong duality). *A feasible indirect utility U is optimal if and only if there exists a positive measure $\gamma \geq_{cx} \mu$ such that*

$$U(s) = h(s) \text{ for } \gamma\text{-almost every } s \quad (11)$$

$$\int U(s) d\mu(s) = \int U(s) d\gamma(s). \quad (12)$$

This result is an analogue of a result in the revenue-maximization problem of a multi-product monopolist (see Theorem 2 in Daskalakis, Deckelbaum, and Tzamos, 2017). Our formulation of the delegation problem allows us to easily deduce strong duality. Note that there is a convex function U such that $h(x) - U(x) > 0$ for all $x \in X$. Therefore, Slater's constraint qualification is satisfied and standard results in linear programming imply that the dual problem has a solution and that the optimal solutions of the primal and dual problems achieve the same value. Since both inequalities in (10) have to hold as equalities, Lemma 3 follows.

Proof idea for Theorem 1. Sufficiency does not use duality. Applied to any feasible indirect utility V , the conditions of the theorem give $\int V d\mu|_Q \leq \int U d\mu|_Q$ for almost every $Q \in \mathcal{Q}$, exactly as in the discussion following the theorem; integrating over partition elements yields $\int V d\mu \leq \int U d\mu$.

For the converse direction, suppose U is optimal. By Lemma 3, there is a positive measure γ satisfying the complementary slackness conditions (11) and (12) and $\gamma \geq_{cx} \mu$; this last condition is equivalent to $\gamma + \mu^- \geq_{cx} \mu^+$, where μ^+ and μ^- denote the positive and negative parts of μ , respectively. Strassen's theorem then implies that $\gamma + \mu^-$ is a mean-preserving spread of μ^+ : one can obtain the measure $\gamma + \mu^-$ by taking, for every s , the mass μ^+ puts on s and spreading it according to a probability measure D_s with expected value s . Since U is convex, Jensen's inequality implies $U(s) \leq \int U(x) dD_s(x)$, and (12) forces equality. But equality means that U is affine on the convex hull of the support of D_s .

We check that the spread does not move mass between partition elements: Since the spread produces the measure $\gamma + \mu^-$, it suffices that U be differentiable at μ^- -almost every point and at γ -almost every point. The first holds because μ^- is absolutely continuous and a convex function is differentiable almost everywhere. The second holds because (11) gives

$U(x) = h(x)$ for γ -almost every x , and a convex function that touches the differentiable h from below is differentiable at the point of contact. At such a point x in the support of D_s , the supporting plane that makes Jensen's inequality an equality at s touches U both at s and at x , so $\nabla U(x) = \nabla U(s)$ and x lies in the same partition element as s .

Conditioning γ and μ on the partition elements therefore yields, for almost every $Q \in \mathcal{Q}$, a positive measure $\gamma|_Q$ with $\gamma|_Q \geq_{cx} \mu|_Q$ and $\gamma|_Q(Q) = \mu|_Q(Q)$; in particular $\mu|_Q(Q) \geq 0$. By (11), $\gamma|_Q$ is concentrated on the types in Q at which U meets h , and there is at most one such type because U is affine on Q and h is strictly convex. So $\gamma|_Q$ is a point mass or the zero measure, that is, $\gamma|_Q = \delta_Q$, which is the assertion of the theorem. The Appendix carries out these steps without assuming that f vanishes on the boundary of S .

5. More general payoffs for the principal

In this section, we consider a more general specification of the principal's preferences and assume the principal's payoff from action a in state s is $u_P(a, s)$. We assume that u_P is continuous, that $u_P(\cdot, s)$ is concave and differentiable for each s , and that the partial derivatives of $\nabla_a u_P$ with respect to s exist and are jointly continuous.

Our earlier specification that the principal's preferences satisfy Assumption 1 ensures that the principal's expected payoff is completely determined by the agent's indirect utility U and, moreover, is a linear function of U . For the more general preferences considered here, neither conclusion holds in general. To illustrate the first issue, suppose the agent has quadratic preferences and that m is an incentive-compatible mechanism that imposes nondegenerate lotteries for a positive measure of types. Then every mechanism that imposes, for every type, a lottery with the same expected value and variance as m induces the same indirect utility as m , but the principal will not be indifferent between these mechanisms in general. Hence, to find an optimal mechanism, in addition to finding an optimal indirect utility U , one would have to optimize over which lotteries are used to induce U . However, for deterministic mechanisms the indirect utility still determines the principal's expected payoff even with the more general preferences (since the gradient of U almost everywhere determines the deterministic action chosen by the corresponding mechanism). We use this insight to provide sufficient conditions under which a given deterministic mechanism is optimal. We then show that when the principal's preferences are quadratic with different weights across dimensions, the best lotteries inducing a given indirect utility can be identified explicitly, so that the problem is again one over indirect utilities.

Deterministic mechanisms. To state the result, let

$$\kappa := \sup\{\kappa' \in \mathbb{R} : u_P(a, s) - \kappa' b(a) \text{ is concave in } a \text{ for all } s\}$$

be a measure of concavity for the principal's payoff relative to the agent's payoff; the supremum is attained, since concavity is preserved under pointwise limits, and we assume it is finite.

For a feasible indirect utility U , let

$$\Phi(U) := \int \left[u_P(\nabla U(s), s) + \kappa [U(s) - u_A(\nabla U(s), s)] \right] dF(s). \quad (13)$$

Every mechanism inducing U gives the principal at most $\Phi(U)$: at almost every type its lottery has mean $\nabla U(s)$, so applying Jensen's inequality to the concave function $u_P(\cdot, s) - \kappa b(\cdot)$ and substituting $\mathbb{E}[b] = U(s) - \nabla U(s) \cdot s$ for the remaining term leaves exactly the bracket in (13). Since $u_P(\cdot, s) - \kappa b(\cdot)$ is concave and $U \mapsto \nabla U$ is linear, Φ is concave. For a deterministic mechanism, $U(s) = u_A(\nabla U(s), s)$, so $\Phi(U)$ is the principal's expected payoff. Such a mechanism is therefore optimal whenever its indirect utility maximizes Φ among feasible indirect utilities.

Given a feasible indirect utility U , let $\tilde{\mu}$ denote the finite signed measure on S satisfying

$$\int V(s) d\tilde{\mu}(s) = \int \left[\kappa V(s) + \left[\nabla_a u_P(\nabla U(s), s) - \kappa s - \kappa \nabla b(\nabla U(s)) \right] \cdot \nabla V(s) \right] dF(s) \quad (14)$$

for every continuously differentiable V . Such a measure exists, and (14) continues to hold when V is a difference of feasible indirect utilities; in particular U need not be differentiable, since its gradient enters only F -almost everywhere.

If the principal's payoff satisfies Assumption 1, then $\kappa = 1$ and $\tilde{\mu}$ equals μ , independently of U . With the more general payoff, $\tilde{\mu}$ measures how, starting at indirect utility U , marginally increasing the agent's payoff changes the principal's payoff.

To state conditions under which a deterministic mechanism with indirect utility U is optimal, let $\mathcal{Q}$ be the partition induced by U as in Section 4.2; note that it depends on U alone. For any $Q \in \mathcal{Q}$, if there is $s \in Q$ with $U(s) = h(s)$ we denote by δ_Q a point mass of mass $\tilde{\mu}|_Q(Q)$ at s . If $U(s) < h(s)$ for all $s \in Q$, let δ_Q denote the zero-measure.

Theorem 2. *Let U be a feasible indirect utility induced by a deterministic mechanism m . Then m is optimal if for $|\tilde{\mu}|$ -a.e. $Q \in \mathcal{Q}$, $\tilde{\mu}|_Q(Q) \geq 0$ and $\tilde{\mu}|_Q \leq_{cx} \delta_Q$.*

To prove Theorem 2, we compare m with an arbitrary incentive-compatible mechanism $\tilde{m}$ with indirect utility $\tilde{U}$. Since $u_P(a, s) - \kappa b(a)$ is concave in a , replacing it by its tangent at the action chosen under m yields

$$\int \left[u_P(\tilde{m}(s), s) - u_P(m(s), s) \right] dF(s) \leq \int (\tilde{U} - U) d\tilde{\mu}.$$

Indeed, $\Phi(\tilde{U})$ bounds the principal's payoff from $\tilde{m}$, $\Phi(U)$ equals her payoff from m , and $\Phi(\tilde{U}) - \Phi(U) \leq \int (\tilde{U} - U) d\tilde{\mu}$ by concavity of Φ . It therefore suffices that U maximize $\int \tilde{U} d\tilde{\mu}$ among feasible indirect utilities, and the conditions of the theorem deliver this by the same

comparison on each partition element that establishes sufficiency in Theorem 1. Only that comparison is used: the necessity part of Theorem 1 rests on properties of μ that $\tilde{\mu}$ need not share.⁸

Since $\tilde{\mu}$ coincides with μ from Section 4 if the principal's preferences take the form in Assumption 1, it follows that these conditions are necessary in that case. Also, for one-dimensional types and general payoffs, one can verify that the conditions of Theorem 2 specialize for interval delegation to the sufficient conditions obtained in Amador and Bagwell (2013).

Differently weighted quadratic preferences. In general, a stochastic mechanism need not attain the bound $\Phi(U)$. With differently weighted quadratic preferences, however, every feasible U is induced by a mechanism that attains it. Suppose

$$b(a) = -\frac{1}{2}\|a\|^2, \quad u_P(a, s) = a \cdot g(s) - \frac{1}{2}a \cdot Wa,$$

for a fixed symmetric positive definite matrix W . The principal's ideal action is $W^{-1}g(s)$ and $W = I$ recovers Assumption 1. Here $u_P - \kappa'b$ is concave in a exactly when $W - \kappa'I$ is positive semidefinite, so $\kappa > 0$ is the smallest eigenvalue of W .

For a feasible U , any mechanism inducing U assigns a lottery with mean $\nabla U(s)$ and total variance $\sigma^2(s) := 2[u_A(\nabla U(s), s) - U(s)]$, twice the agent's utility loss from randomization at s ; the indirect utility pins down these two moments but not the direction in which the variance is spread. Changing that direction leaves the mean and the expected squared norm of the lottery unchanged, hence also every type's evaluation of every report, so the principal is free to place the required variance where she penalizes it least.

Proposition 2. *For every feasible U , the highest expected payoff among mechanisms inducing U equals $\Phi(U)$. It is attained by assigning, at almost every type, the two actions $\nabla U(s) \pm \sigma(s)e$ with equal probabilities, where e is a unit eigenvector of W for the eigenvalue κ .*

Thus, for every feasible U , $\Phi(U)$ is the highest payoff the principal can obtain from a mechanism inducing U . Expanding (13) gives

$$\Phi(U) = \int \left[\kappa U(s) + [g(s) - \kappa s] \cdot \nabla U(s) - \frac{1}{2} \nabla U(s) \cdot (W - \kappa I) \nabla U(s) \right] dF(s). \quad (15)$$

The principal's problem is to maximize this concave functional over convex $U : \mathbb{R}^n \rightarrow \mathbb{R}$ with

⁸ Concavity enters this argument only through the tangent inequality, and only at the action the mechanism assigns. Since $u_P(a, s) - \kappa b(a)$ and $u_P(a, s) - \kappa u_A(a, s)$ differ by a term that is linear in a , Theorem 2 therefore also holds whenever $a \mapsto u_P(a, s) - \kappa u_A(a, s)$ admits a concave upper bound that is tight at the action assigned to F -almost every type, with the coefficient of ∇V in (14) read as the gradient of that bound at the assigned action. We use this form in Section 6, where a fixed cost makes payoffs discontinuous, so that the difference need not be concave for any κ .

$U \leq h$. Its quadratic term is negative semidefinite; it vanishes when W is a positive multiple of the identity, which returns the linear problem of Section 4, and when W is diagonal an optimal mechanism randomizes only in a least-weighted dimension.

Since $\tilde{\mu}$ is the derivative of Φ at U , a feasible U is optimal if and only if $\int (V - U) d\tilde{\mu} \leq 0$ for every feasible V . In particular, the conditions of Theorem 2 are sufficient for a stochastic mechanism as well, provided it uses the lotteries of Proposition 2.

Biased and aligned decisions. Randomization lowers the agent's utility and works like money burning, but it is costly for the principal as well: under Assumption 1, her cost of a given risk equals the agent's. Something that is specific to multidimensional problems is that the principal can place the risk in a decision she cares little about, while the agent still bears its full cost. The following makes this precise.

Write the state as $s = (x, z) \in S \subseteq \mathbb{R}^k \times \mathbb{R}^{n-k}$ and the action as $a = (a^x, a^z)$, with $1 \leq k < n$. The first k decisions are biased and the remaining $n - k$ decisions are aligned: the agent's ideal action is s , while the principal's payoff is

$$u_P(a, s) = -\frac{1}{2}\|a^x - (x - \beta)\|^2 - \frac{\kappa}{2}\|a^z - z\|^2, \quad 0 < \kappa < 1,$$

where $\beta \in \mathbb{R}^k \setminus \{0\}$ collects the possibly unequal biases. The principal thus agrees with the agent on the aligned decisions but cares less about them. Up to a term that does not depend on a , this is the weighted quadratic payoff above with weight 1 on the biased and weight κ on the aligned decisions.

Proposition 3. *Fix a nonzero bias vector β . There exists $\kappa_0 > 0$ such that, whenever $0 < \kappa < \kappa_0$, every optimal mechanism uses nondegenerate lotteries on a positive-probability set of types. An optimal mechanism can concentrate all randomization in a single aligned decision.*

No independence assumption is required for the proposition. If the biased and aligned states are independent, an optimal mechanism can additionally be chosen so that $\mathbb{E}[a^z \mid x, z] = z$. The expected aligned action is thus undistorted, and randomization in the aligned decisions provides incentives in the biased decisions.

Randomization here does what money burning does when transfers are available, except that the instrument is a decision that has to be taken in any case. Placing the required variance $\sigma^2(s)$ in an aligned decision costs the agent $\sigma^2(s)/2$ but costs the principal only $\kappa\sigma^2(s)/2$, so incentives become arbitrarily cheap relative to their value as κ falls. Deterministic distortions of the aligned decisions would be equally cheap. However, in a deterministic mechanism the action is ∇U , so a distortion of the aligned decisions that varies with x forces the biased decisions to vary with z , whereas the risk of a lottery can vary with x freely.

6. Outside option for the agent

In some economic applications, outside options for the agent arise naturally. As examples, we discuss below applications to veto bargaining and monopoly regulation. Often, these outside options can be formulated as a simple condition on the indirect utility of the agent, and the indirect utility approach extends easily.

To model this outside option, consider our general model above and suppose $h \geq 0$ and every type of the agent has an outside option that yields utility 0. A mechanism is *feasible* for this model if it is incentive compatible and yields each type $s \in S$ a nonnegative indirect utility: $u_A(m(s), s) \geq 0$ for all $s \in S$. Clearly, any feasible mechanism induces an indirect utility that satisfies $U(s) \geq 0$ for all $s \in S$. Conversely, the arguments used to prove Lemma 1 can be used to prove the following (we omit the straightforward argument).

Lemma 4. *An indirect utility U corresponds to a feasible mechanism for the model with outside option if and only if U is convex, $U(s) \geq 0$ for all $s \in S$, and $U \leq h$.*

Let $\mathcal{U}_0 := \{U \in \mathcal{U} : \forall s \in S, U(s) \geq 0\}$ denote the subset of convex functions that are nonnegative for all types in S . Under Assumption 1, the principal's problem is then

$$\begin{aligned} & \max_{U \in \mathcal{U}_0} \int U \, d\mu \\ & \text{s. t. } U \leq h. \end{aligned}$$

An analogue of Theorem 1 holds in this setting, with two modifications. First, the convex order $\leq_{cx}$ is replaced by the order $\leq_{\mathcal{U}_0}$, where $\alpha \leq_{\mathcal{U}_0} \beta$ if and only if $\int V \, d\alpha \leq \int V \, d\beta$ for all $V \in \mathcal{U}_0$. Second, the mass of δ_Q is no longer determined. Since the outside option prevents the principal from lowering the utility of types with $U(s) = 0$, the dual solution may put more mass on a contact point than $\mu|_Q(Q)$. In this section we therefore let δ_Q denote any nonnegative measure concentrated on $Q \cap \{s : U(s) = h(s)\}$, with its mass left free; as before there is at most one such contact point, so δ_Q is a point mass or the zero measure.

Theorem 3. *Consider the model with outside option and let U be a feasible indirect utility. Then U is optimal if and only if for $|\mu|$ -a.e. $Q \in \mathcal{Q}$ there is such a measure δ_Q with*

$$\mu|_Q \leq_{\mathcal{U}_0} \delta_Q \quad \text{and} \quad \int U \, d\mu|_Q = \int U \, d\delta_Q.$$

In Theorem 1 the second condition was automatic: the convex order compares affine functions in both directions and thereby fixes the mass of δ_Q at $\mu|_Q(Q)$. The order $\leq_{\mathcal{U}_0}$ does not, so the equality must be imposed, and it is what now pins the mass down: if Q contains a contact point s with $h(s) > 0$, that mass is $\int U \, d\mu|_Q / h(s)$.

Sufficiency follows from the same payoff comparison as in Theorem 1, and that comparison

uses no property of μ beyond the two displayed conditions; we apply it with $\tilde{\mu}$ in place of μ in the monopoly application below. For necessity one has in addition to account for the types whose outside option binds; the argument is given in the appendix and again relies on strong duality. The compact-domain reduction of Lemma 2 applies here as well, since extending an indirect utility preserves its values on S and hence the outside-option constraint, and on this domain the dual problem is to minimize $\int h d\gamma$ over positive measures γ satisfying $\mu \leq_{\mathcal{U}_0} \gamma$. The proof of Lemma 3 does not apply, however, because Slater’s constraint qualification is not necessarily satisfied (there might be no $U \in \mathcal{U}_0$ with $U < h$). Strong duality can still be established by showing that the value function has bounded slope.

Lemma 5. *Strong duality holds for the problem with outside options, and the dual problem has a solution.*

Veto bargaining To illustrate this model variant, consider a model of incomplete information veto bargaining in which a proposer can make a policy proposal which the veto player can either accept or reject (see, e.g., Kartik, Kleiner, and Van Weelden, 2021). After a rejection, a status quo policy remains in effect. The principal and agent have quadratic preferences over the policy that is implemented, where the principal’s ideal policy is known to be a_P and the agent’s ideal policy s is his private information. Since the principal does not know which policies are acceptable to the veto player—that is, which policies the veto player prefers to the status quo policy—the principal generally benefits from eliciting the agent’s information by committing to a mechanism.

This model can be interpreted as a delegation problem in which the agent has an outside option of obtaining the status quo and where the principal’s preferences are independent of the state. Let $u_A(a, s) = a \cdot s - \frac{1}{2}\|a\|^2$ denote the agent’s payoff if policy a is implemented in state s and $u_P(a) = a \cdot a_P - \frac{1}{2}\|a\|^2$ the principal’s payoff. Let $a_{SQ} = 0$ denote the status quo action. Then any mechanism satisfying the outside option has an indirect utility satisfying $U(s) \geq 0$ for all $s \in S$ and any $U \in \mathcal{U}_0$ with $U \leq h$ corresponds to some feasible mechanism.

The indirect utility approach can be used to solve the one-dimensional version of this problem analyzed in Kartik, Kleiner, and Van Weelden (2021). This approach simplifies their arguments because it directly yields necessary and sufficient conditions. More importantly, the indirect utility approach is a useful tool to analyze incomplete-information veto bargaining over multidimensional policies.

Monopoly regulation with outside option. The method also yields new sufficient conditions in the canonical one-dimensional problem of regulating a privately informed monopolist without transfers (Amador and Bagwell, 2022). A firm observes its productivity

$s \in S = [s, \bar{s}]$ and chooses output $a \geq 0$. Its profit and the regulator's payoff are

$$u_A(a, s) = sa + b(a) - c\mathbf{1}_{\{a>0\}}, \quad u_P(a, s) = u_A(a, s) + v(a),$$

where $c \geq 0$ is a fixed production cost, v is weighted consumer surplus, and $b(0) = v(0) = 0$. The firm can shut down for a payoff of zero. Feasibility is characterized as in Lemma 1, except that the fixed cost changes which lotteries implement a given supporting line: a lottery between shutdown and a positive output can be needed even when expected output is small. The regularity assumptions and formal results are in Section B.3.

A *cap allocation* offers the menu $\{0\} \cup [\underline{a}, \infty)$: a producing firm must supply at least $\underline{a}$, which corresponds to a price cap. For the nondegenerate caps considered in the appendix, types below $s_1 = [c - b(\underline{a})]/\underline{a}$ shut down, types between s_1 and $s_2 = -b'(\underline{a})$ produce $\underline{a}$, and higher types choose their unconstrained monopoly output. Expected profit is therefore zero on the shutdown region, affine on the pooling region, and equal to the full-discretion profit h on the unconstrained region, as Figure 8 illustrates.

A tighter price cap raises output among constrained firms but can also induce shutdown. When both cutoffs lie in the interior of S , the first-order condition for the output floor is

$$\int_{s_1}^{s_2} [s - s_2 + v'(\underline{a})] f(s) ds = \frac{v(\underline{a})}{\underline{a}} (s_2 - s_1) f(s_1). \quad (16)$$

The left-hand side is the marginal welfare effect of expanding output among the constrained firms. The right-hand side is the welfare lost from additional shutdowns: the marginal firm earns zero profit, so its exit costs the regulator the weighted consumer surplus $v(\underline{a})$.

Combining the payoff bound of Section 5 with the pooling comparisons above, Proposition 4 gives sufficient conditions for a particular cap to dominate every incentive-compatible mechanism, including lotteries over output and shutdown. The conditions compare alternatives on the unconstrained, pooling, and shutdown regions; in the interior-cutoff case, their equality on the pooling region is precisely (16). They allow both output and the set of producing types to change, and the comparison is with a larger mechanism class than the deterministic quantity allocations studied by Amador and Bagwell (2022).

7. Conclusion

Our paper studies a multidimensional model of optimal delegation. A key step to analyze the model is a simple characterization of incentive-compatible mechanisms. Using this characterization, we derive conditions that are necessary and sufficient for an arbitrary mechanism to be optimal and apply them to convex delegation sets and to budgets that link the decisions. Optimal mechanisms are deterministic when virtual values are positive in the interior of the type space and negative only on one of its faces. With more general preferences for the prin-

principal, however, randomization can be strictly optimal, because the principal can place the risk in decisions she cares little about. We also discuss how to incorporate outside options.

There are various directions that may be fruitful for future research. On the theoretical side, it would be of interest to find conditions that are simple to check and ensure that a mechanism in some class of specific mechanisms is optimal (for example, conditions on the primitives ensuring that some convex delegation set is optimal). Regarding applications, it would be of interest to understand how the multidimensional nature of information and actions affects conclusions that have been obtained in one-dimensional models. Examples could include bargaining over policies with rich policy spaces, or regulation of a monopolist that has private information about both fixed and marginal costs.

References

- Ali, S. Nageeb, Navin Kartik, and Andreas Kleiner (2023). Sequential Veto Bargaining with Incomplete Information, *Econometrica* 91.4, pp. 1527–1562.
- Alonso, Ricardo, Tan Gan, and Ju Hu (2026). Robust Delegation. unpublished, CEPR Discussion Paper 21382.
- Alonso, Ricardo and Niko Matouschek (2008). Optimal Delegation, *Review of Economic Studies* 75.1, pp. 259–293.
- Amador, Manuel and Kyle Bagwell (2013). The Theory of Optimal Delegation with an Application to Tariff Caps, *Econometrica* 81.4, pp. 1541–1599.
- (2020). Money burning in the theory of delegation, *Games and Economic Behavior* 121, pp. 382–412.
- (2022). Regulating a Monopolist With Uncertain Costs Without Transfers, *Theoretical Economics* 17, pp. 1719–60.
- Amador, Manuel, Kyle Bagwell, and Luiz Carpio (2025). Generalized Delegation. unpublished, Stanford University.
- Amador, Manuel, Kyle Bagwell, and Alex Frankel (2018). A Note on Interval Delegation, *Economic Theory Bulletin* 6.2, pp. 239–249.
- Ambrus, Attila and Georgy Egorov (2017). Delegation and Nonmonetary Incentives, *Journal of Economic Theory* 171, pp. 101–135.
- Armstrong, Mark (1995). Delegating Decision-Making to an Agent with Unknown Preferences. unpublished.
- Bogachev, Vladimir Igorevich (2007). Measure theory. Vol. 2. Springer.
- Bronshtein, Efim Mikhailovich (1978). Extremal convex functions, *Siberian Mathematical Journal* 19.1, pp. 6–12.
- Chakraborty, Archishman and Rick Harbaugh (2007). Comparative cheap talk, *Journal of Economic Theory* 132.1, pp. 70–94.
- Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos (2017). Strong Duality for a Multiple-Good Monopolist, *Econometrica* 85.3, pp. 735–767.

- Dunford, Nelson and Jacob T. Schwartz (1988). *Linear Operators, Part 1: General Theory*. John Wiley and Sons.
- Frankel, Alex (2014). Aligned delegation, *American Economic Review* 104.1, pp. 66–83.
- (2016). Delegating multiple decisions, *American Economic Journal: Microeconomics* 8.4, pp. 16–53.
- Galperti, Simone (2019). A theory of personal budgeting, *Theoretical Economics* 14.1, pp. 173–210.
- Gentzkow, Matthew and Emir Kamenica (2016). A Rothschild-Stiglitz approach to Bayesian persuasion, *American Economic Review* 106.5, pp. 597–601.
- Gretsky, N. E., J. M. Ostroy, and William R. Zame (2002). Subdifferentiability and the Duality Gap, *Positivity* 6, pp. 261–274.
- Haghpanah, Nima and Jason Hartline (2021). When is pure bundling optimal?, *The Review of Economic Studies* 88.3, pp. 1127–1156.
- Hiriart-Urruty, Jean-Baptiste and Claude Lemaréchal (2004). *Fundamentals of convex analysis*. Springer Science & Business Media.
- Holmström, Bengt (1977). *On Incentives and Control in Organizations*. Ph.D. Thesis, Stanford University.
- Jackson, Matthew O. and Hugo F. Sonnenschein (2007). Overcoming incentive constraints by linking decisions, *Econometrica* 75.1, pp. 241–257.
- Johansen, Søren (1974). The extremal convex functions, *Mathematica Scandinavica* 34.1, pp. 61–68.
- Kartik, Navin and Andreas Kleiner (2024). *Convex Choice*. unpublished, arXiv:2406.19063.
- Kartik, Navin, Andreas Kleiner, and Richard Van Weelden (2021). Delegation in veto bargaining, *American Economic Review* 111.12, pp. 4046–87.
- Kleiner, Andreas, Benny Moldovanu, and Philipp Strack (2021). Extreme points and majorization: Economic applications, *Econometrica* 89.4, pp. 1557–1593.
- Kleiner, Andreas, Benny Moldovanu, Philipp Strack, and Mark Whitmeyer (2024). *The Extreme Points of Fusions*. unpublished, arXiv:2409.10779.
- Koessler, Frédéric and David Martimort (2012). Optimal delegation with multi-dimensional decisions, *Journal of Economic Theory* 147.5, pp. 1850–1881.
- Kolotilin, Anton (2018). Optimal Information Disclosure: A Linear Programming Approach, *Theoretical Economics* 13.2, pp. 607–635.
- Kolotilin, Anton and Andriy Zapechelnjuk (2025). Persuasion meets Delegation, *Econometrica* 93.1, pp. 195–228.
- Kováč, Eugen and Tymofiy Mylovanov (2009). Stochastic Mechanisms in Settings Without Monetary Transfers: The Regular Case, *Journal of Economic Theory* 144.4, pp. 1373–1395.
- Lahr, Patrick and Axel Niemeyer (2024). *Extreme Points in Multi-Dimensional Screening*. unpublished, arXiv:2412.00649.
- Lipnowski, Elliot and Doron Ravid (2020). Cheap talk with transparent motives, *Econometrica* 88.4, pp. 1631–1660.
- Manelli, Alejandro M. and Daniel R. Vincent (2007). Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly, *Journal of Economic Theory* 137.1, pp. 153–185.

- Melumad, Nahum and Toshiyuki Shibano (1991). Communication in Settings with No Transfers, *RAND Journal of Economics* 22.2, pp. 173–198.
- Phelps, Robert R. (2001). Lectures on Choquet’s theorem. Springer.
- Rochet, Jean-Charles (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context, *Journal of Mathematical Economics* 16.2, pp. 191–200.
- Rochet, Jean-Charles and Philippe Choné (1998). Ironing, sweeping, and multidimensional screening, *Econometrica* 66.4, pp. 783–826.
- Shaked, Moshe and J. George Shanthikumar (2007). Stochastic orders. Springer.
- Shapiro, Alexander (2010). On duality theory of conic linear problems. In: *Semi-Infinite Programming: Recent Advances*. Ed. by M. A. Goberna and M. A. Lopez. Springer, U.S.A.

A. Omitted proofs

A.1. Defining the partition $\mathcal{Q}$

Throughout, U is a feasible indirect utility, that is, by Lemma 1, a convex function $U : \mathbb{R}^n \rightarrow \mathbb{R}$ with $U \leq h$. In particular, $\partial U(s)$ is a nonempty, compact, convex set for every $s \in \mathbb{R}^n$.

By the construction in the proof of Lemma 1, each $p \in \partial U(s)$ is the expected action of a lottery that type s can be assigned, and the partition element containing s is determined by which of them is chosen. Where U is differentiable there is only one, namely $\nabla U(s)$, and nothing has to be decided. At a kink we assign s the subgradient that governs U just inside S : fix $\hat{s} \in \text{int } S$ and let

$$M_U(s) := \arg \min_{p \in \partial U(s)} p \cdot (s - \hat{s}), \quad T_U(s) := \arg \min_{p \in M_U(s)} \|p\|^2.$$

We then define

$$Q_p := \{s \in \mathbb{R}^n : T_U(s) = p\}, \quad \mathcal{Q} := \{Q_p : Q_p \neq \emptyset\},$$

so $\mathcal{Q}$ is a measurable partition of $\mathbb{R}^n$, and we write $Q(s)$ for the element containing s .

A.2. Proofs for Section 4

Proof of Lemma 2. It suffices to consider candidates with payoff at least $b(0)$, which is attainable by always choosing action 0. Choose a compact convex neighborhood K of $S \cup g(S)$ large enough to contain every point $2g(s) - t$ with $s, t \in S$. For any convex candidate $U \leq h$ on a domain containing K , convexity gives

$$b(0) \leq \int U \, d\mu \leq \int U(g(s)) \, dF(s) \leq \frac{U(t) + \max_K h}{2} \quad (t \in S).$$

For the middle inequality, recall that the principal’s payoff at s is $U(s) + \nabla U(s) \cdot [g(s) - s]$, the value at $g(s)$ of the tangent to U at s , which by convexity is at most $U(g(s))$. The last inequality uses that $g(s)$ is the midpoint of t and $2g(s) - t$, both of which lie in K . Thus

$U \geq 2b(0) - \max_K h$ on S . A convex function that is bounded below on S and bounded above on a neighborhood of S has uniformly bounded subgradients on S ; since K is such a neighborhood and $U \leq h$ there, the supporting slopes on S are uniformly bounded.

By continuity of ∇b , we can choose a compact convex $X \supseteq K$ containing $-\nabla b(a)$ for every slope a within this bound. Consider any such candidate U on X . Evaluating a supporting plane of slope a at $-\nabla b(a)$, where action a is ideal, shows that its intercept is at most $b(a)$. Hence the plane lies below the payoff plane of action a , and therefore below h everywhere. Taking the supremum of the supporting planes at points of S gives a convex extension that agrees with U on S and lies below h globally. By Lemma 1, this extension is feasible and gives the same payoff. Conversely, every globally feasible indirect utility restricts to a feasible candidate on X , so the two problems have the same value.

Finally, these extensions are uniformly bounded and equi-Lipschitz on X . Applying Arzelà–Ascoli to the extensions of a maximizing sequence gives a uniformly convergent subsequence. Its limit is convex, lies below h , and attains the limiting payoff because μ is finite. Thus the maximum is attained. Q.E.D.

Proof of Lemma 3. Let $\mathcal{C}(X)$ denote the vector space of continuous functions on X with the supremum norm and recall that its dual space is the space of (Radon) measures on X , which we denote by $\mathcal{M}(X)$. Let $\mathcal{V} := \{g \in \mathcal{C}(X) : \forall x \in X, g(x) \geq 0\}$; the polar cones of $\mathcal{U}$ and $\mathcal{V}$ are defined by

$$\begin{aligned}\mathcal{U}^* &:= \{\gamma \in \mathcal{M}(X) : \forall U \in \mathcal{U}, \int U d\gamma \geq 0\} \\ \mathcal{V}^* &:= \{\gamma \in \mathcal{M}(X) : \forall g \in \mathcal{V}, \int g d\gamma \geq 0\}.\end{aligned}$$

The principal’s problem can be written as $\max_{U \in \mathcal{U}} \int U d\mu$ subject to $h - U \in \mathcal{V}$. This is a conical linear program and its dual is $\inf_{\gamma \in \mathcal{V}^*} \int h d\gamma$ subject to $\gamma - \mu \in \mathcal{U}^*$ (e.g., Shapiro, 2010). Since $\mathcal{V}^* = \mathcal{M}_+(X)$ by the Riesz representation theorem (Dunford and Schwartz, 1988, p. 265) and $\gamma - \mu \in \mathcal{U}^*$ is equivalent to $\gamma \geq_{cx} \mu$, (D) is the dual problem.

Since there is $U \in \mathcal{U}$ such that $h - U$ is in the interior of $\mathcal{V}$, Slater’s condition is satisfied and standard results imply that strong duality holds (e.g., Shapiro, 2010, Proposition 2.8). Given (10), it follows that U is optimal if and only if there is a positive measure $\gamma \geq_{cx} \mu$ such that $\int U d\mu = \int U d\gamma$ and $\int U d\gamma = \int h d\gamma$, which implies the result. Q.E.D.

As explained in the proof sketch in the main text, in the proof of Theorem 1 we spread mass from a type s to types t that lie in the same affine piece of U as s and the argument requires that such t belong to the same partition element as s . Where U is differentiable this is immediate, since the supporting plane is then unique. At a kink the selection makes

a choice, and two types that share a supporting plane need not be assigned the same one. The next lemma rules this out on a set of types that carries all the mass of $|\mu|$.

Let G consist of the types $s \in \text{int } S$ at which U is differentiable, together with the types $s \in \text{bd } S$ at which S has a unique outward normal and the restriction of U to $\text{bd } S$ is differentiable.

Lemma 6. *The following statements hold.*

$$(i) \quad |\mu|(S \setminus G) = 0.$$

$$(ii) \quad \text{If } s, t \in G \text{ and } T_U(s) \in \partial U(t), \text{ then } Q(s) = Q(t).$$

Proof. (i) A convex function is differentiable Lebesgue-almost everywhere, so U is differentiable at almost every type in $\text{int } S$. Since S is convex, $\text{bd } S$ is locally the graph of a convex function, which is differentiable almost everywhere; hence the outward normal is unique at $\mathcal{H}$ -almost every boundary type. Similarly, U is Lipschitz and therefore differentiable along $\text{bd } S$ at $\mathcal{H}$ -almost every boundary type. Hence $\lambda(S \setminus G) = 0$, and $|\mu|(S \setminus G) = 0$ because $|\mu| \ll \lambda$.

(ii) By definition of the partition $\mathcal{Q}$, $Q(s) = Q(t)$ is equivalent to $T_U(s) = T_U(t)$, which is what we show. Note first that at a type $z \in G \cap \text{bd } S$ the outward normal is unique and differentiability of U along $\text{bd } S$ fixes the tangential component of every subgradient, so $\partial U(z)$ is a segment in the direction $\hat{\mathbf{n}}_S(z)$; since $\hat{s}$ lies in the interior of S , the selection $T_U(z)$ is the endpoint of that segment with the smallest component along $\hat{\mathbf{n}}_S(z)$.

If U is differentiable at t , then $\partial U(t) = \{T_U(t)\}$ and the hypothesis $T_U(s) \in \partial U(t)$ gives $T_U(s) = T_U(t)$. Otherwise $t \in G \cap \text{bd } S$, and the preceding observation gives

$$T_U(s) = T_U(t) + r \hat{\mathbf{n}}_S(t) \quad \text{for some } r \geq 0.$$

Suppose $r > 0$. Since $T_U(s)$ is a subgradient at both s and t , we have $U(s) - U(t) = T_U(s) \cdot (s - t)$, while $T_U(t) \in \partial U(t)$ gives $U(s) - U(t) \geq T_U(t) \cdot (s - t)$. Subtracting, $r \hat{\mathbf{n}}_S(t) \cdot (s - t) \geq 0$, whereas convexity of S gives $\hat{\mathbf{n}}_S(t) \cdot (s - t) \leq 0$. Hence $\hat{\mathbf{n}}_S(t) \cdot (s - t) = 0$, so s lies in the supporting hyperplane of S at t . That hyperplane therefore also supports S at s , so $s \in \text{bd } S$ and, since $s \in G$, $\hat{\mathbf{n}}_S(s) = \hat{\mathbf{n}}_S(t)$. It also follows that $T_U(t) \cdot (s - t) = T_U(s) \cdot (s - t)$, so the plane with slope $T_U(t)$ touches U at s as well, that is, $T_U(t) \in \partial U(s)$. But then $T_U(t) = T_U(s) - r \hat{\mathbf{n}}_S(s)$ is a subgradient at s with a strictly smaller component along $\hat{\mathbf{n}}_S(s)$ than $T_U(s)$, contradicting the selection at s . Hence $r = 0$. Q.E.D.

Proof of Theorem 1. Throughout, U is a convex function on $\mathbb{R}^n$ satisfying $U \leq h$ (Lemma 1); the compact domain X of Lemma 2 is used only to construct the dual variable γ .

Let $\{\mu|_Q\}_{Q \in \mathcal{Q}}$ be regular, proper conditional measures of μ given the partition $\mathcal{Q}$ (Bogachev, 2007, Example 10.4.11). They are concentrated on $Q \cap S$ for $|\mu|$ -almost every Q and

satisfy

$$\mu = \int \mu|_{Q(s)} d|\mu|(s).$$

Since $d\mu/d|\mu|$ takes the values ± 1 , the positive and negative parts of $\mu|_Q$ are the conditional measures of μ^+ and μ^- , with the same outer weighting.

Sufficiency. Suppose the conditional inequalities hold, and let V be any feasible indirect utility. For $|\mu|$ -almost every $Q \in \mathcal{Q}$,

$$\int V d\mu|_Q \leq \int V d\delta_Q \leq \int U d\delta_Q = \int U d\mu|_Q.$$

The first inequality is convex order, and the second uses $\delta_Q \geq 0$ and its concentration on $\{s : U(s) = h(s)\}$. The equality follows because U agrees with an affine function on Q and convex order preserves mass and first moments. It also holds when $\delta_Q = 0$. Integrating over partition elements with respect to $|\mu|$ proves optimality. All the integrals are finite: subgradients of U on the compact set S are bounded, so the types whose ideal actions equal these subgradients, including the relevant contact points, lie in a compact set.

Necessity. By Lemma 3, optimality gives a positive measure γ satisfying (11), (12), and $\gamma \geq_{cx} \mu$. Strassen's theorem (see, for example, Phelps, 2001, pp. 93–94) supplies a probability kernel D_s such that

$$\gamma + \mu^- = \int D_s d\mu^+(s), \quad \int t dD_s(t) = s \quad \text{for } \mu^+\text{-almost every } s.$$

Combining the first identity with (12) gives

$$\int \left[\int U(t) dD_s(t) - U(s) \right] d\mu^+(s) = 0.$$

Each bracket is nonnegative by Jensen's inequality, so it vanishes for μ^+ -almost every s . Let p be the subgradient selected at such a point s . Since D_s has barycenter s ,

$$\int [U(t) - U(s) - p \cdot (t - s)] dD_s(t) = 0.$$

The integrand is nonnegative because $p \in \partial U(s)$, so it vanishes D_s -almost everywhere: the plane selected at s touches U at D_s -almost every t , that is, $p \in \partial U(t)$.

We show next that D_s is concentrated on $Q(s)$ for μ^+ -almost every s . Since $\mu^+ \leq |\mu|$, Lemma 6(i) gives $s \in G$ for μ^+ -almost every s , and likewise $\mu^-(S \setminus G) = 0$. By (11), γ is concentrated on $\{t : U(t) = h(t)\}$. Since $\int D_s d\mu^+(s) = \gamma + \mu^-$, it follows that for μ^+ -almost every s , D_s -almost every t either lies in G or satisfies $U(t) = h(t)$. Fix such an s and t . If $t \in G$, then by the preceding paragraph the plane selected at s touches U at t , so Lemma 6(ii) gives $Q(t) = Q(s)$. If $U(t) = h(t)$, then $U \leq h$ and differentiability of h imply that U has a unique supporting plane at t ; by the preceding paragraph it is the plane selected at s , so again $Q(t) = Q(s)$.

For each partition element, define

$$\alpha_Q := \int D_s d\mu|_Q^+(s) - \mu|_Q^-.$$

Integrating over the partition and using the disintegrations of μ^+ and μ^- recorded above,

$$\int \alpha_{Q(s)} d|\mu|(s) = \int D_t d\mu^+(t) - \mu^- = \gamma.$$

Since D_s is concentrated on $Q(s)$, each α_Q is concentrated on Q , so the α_Q are conditional measures of γ given $\mathcal{Q}$, with the same outer weighting by $|\mu|$. Because γ is positive, α_Q is positive for $|\mu|$ -almost every Q , and by (11) it is concentrated on $Q \cap \{U = h\}$.

Because D_s has barycenter s , Jensen's inequality gives $\mu|_Q \leq_{cx} \alpha_Q$. Moreover, $\alpha_Q(Q) = \mu|_Q(Q)$. There is at most one contact point in Q , so $\alpha_Q = \delta_Q$. If there is no contact point, $\alpha_Q = 0$ and $\mu|_Q(Q) = 0$. This proves the necessary conditions. Q.E.D.

Proof of Corollary 1. The partition consists of $(-\infty, s_1]$, $[s_2, \infty)$, and the singletons in (s_1, s_2) . On a singleton cell, $\mu|_Q$ is a point mass, so the convex-order comparison holds automatically and only nonnegative mass is required. This gives (i).

On the upper tail, convex order compares μ restricted to $[s_2, \bar{s}]$ with a point mass of the same total mass at s_2 . Equality in (ii) matches their first moments, and the inequalities for $s > s_2$ give the remaining convex-order comparisons, since the functions $(x - s)^+$ with $s \leq s_2$ are affine on $[s_2, \bar{s}]$. It remains to check that the tail mass is nonnegative. If $s_2 < \bar{s}$, the expression in (ii) is zero at s_2 and nonpositive to its right, so its right derivative at s_2 satisfies

$$-\mu([s_2, \bar{s}]) \leq 0,$$

where we use that μ has no atom at the interior point s_2 . If $s_2 = \bar{s}$, the tail consists only of the endpoint mass, which is nonnegative by (i). The lower-tail comparison follows analogously from (iii). Thus the conditions are precisely those of Theorem 1. Q.E.D.

Proof of Corollary 2. Enlarging I by an adjacent interval on which μ vanishes preserves the hypothesis, so we may assume that I admits no such enlargement. Let U be an optimal indirect utility. By Theorem 1, every pool has nonnegative mass, so a pool that contains negative weight also contains positive weight and therefore meets I . Since pools are disjoint intervals, all negative weight below I lies in a single pool Q_1 , and all negative weight above I in a single pool Q_2 .

If some pool carries all virtual weight, assigning its lottery's mean to every type yields an optimal singleton delegation set. Otherwise, suppose there is negative weight below I . The weight of Q_1 is nonpositive below I and nonnegative in I . Its total mass is positive: with zero mass, $\mu|_{Q_1} \leq_{cx} \delta_{Q_1}$ would force a zero first moment, whereas the positive weight

lies to the right of the negative weight. So δ_{Q_1} is a point mass at the signed barycenter s_1 of $\mu|_{Q_1}$, where U is tangent to h . This barycenter lies to the right of the mean of the positive weight, so it is not below I . It is not above I either: by maximality, either types arbitrarily close above I carry negative weight and belong to Q_2 , or I extends to $\bar{s}$ and Q_1 would carry all virtual weight. If there is no negative weight below I , maximality implies that I extends to $\underline{s}$, and we set $s_1 = \underline{s}$. Define s_2 symmetrically.

The indirect utility of delegating to $[s_1, s_2]$ is the tangent to h at s_1 below s_1 , h on $[s_1, s_2]$, and the tangent to h at s_2 above s_2 . Below s_1 it therefore agrees with U on Q_1 , which, when it exists, contains all weight below s_1 , and similarly above s_2 . On $[s_1, s_2] \subseteq I$ it equals $h \geq U$, and μ is nonnegative there. Hence it yields at least $\int U d\mu$, and delegating to $[s_1, s_2]$ is optimal. Q.E.D.

Proof of Corollary 3. By (4), $\nu/f = 1 - \beta(\log f)'$ in the interior of the support. If $\beta > 0$, this expression is nondecreasing, while the endpoint masses are $-\beta f(\underline{s}) \leq 0$ and $\beta f(\bar{s}) \geq 0$. Thus μ is nonnegative on an upper interval and nonpositive below it, as required by Corollary 2. For $\beta < 0$ the signs and direction reverse. If $\beta = 0$, then $\mu = F$ and full discretion is optimal. Q.E.D.

Proof of Corollary 4. Since $b(a) = -\|a\|^2/2$, each type chooses his Euclidean projection onto A . This choice is unique, so the indirect utility is differentiable and its partition consists of the singletons in $\text{int } A$ and the outward normal rays $Q_a = a + N_A(a)$ for $a \in \text{bd } A$. Each ray meets h only at a .

On a singleton cell, $\mu|_Q$ is a point mass, so the convex-order comparison holds and Theorem 1 only requires nonnegative mass, which is equivalent to (i).

Fix a normal ray Q_a and write its points as $x = a + r \hat{n}_A(a)$ with $r \geq 0$, where $\hat{n}_A(a)$ is the outward unit normal at a . The theorem compares $\mu|_{Q_a}$ with the point mass δ_{Q_a} of mass $\mu|_{Q_a}(Q_a)$ at a . These measures are one-dimensional, and by the characterization recalled before Corollary 1 it amounts to equal masses, equal first moments, and the comparisons for the functions $(r - t)^+$.

For $s = a + t \hat{n}_A(a)$ with $t \geq 0$, we have $x \in s + N_A(a)$ exactly when $r \geq t$, and then $\|x - s\| = r - t$. The integral in (ii) is therefore $\int (r - t)^+ d\mu|_{Q_a}$. At $t = 0$ it is the first moment of $\mu|_{Q_a}$ about a , so the equality in (ii) matches first moments. For $t > 0$, the inequalities in (ii) are the comparisons with δ_{Q_a} , which gives $(r - t)^+$ integral zero. For $t < 0$, $(r - t)^+$ is affine on the ray, so matching masses and first moments suffice. Hence (ii) is equivalent to $\mu|_{Q_a}(Q_a) \geq 0$ and $\mu|_{Q_a} \leq_{cx} \delta_{Q_a}$. Together with (i), these are precisely the conditions of Theorem 1. Q.E.D.

Proof of Corollary 5. The budget leaves the agent unrestricted orthogonally to β and caps his choice in the remaining direction. Types therefore separate below the cap and pool along rays parallel to β above it. Conditional on Y , the agent's ideal action s has $\beta \cdot s = Z$ and the principal's ideal action $s - \beta$ has $\beta \cdot (s - \beta) = Z - \|\beta\|^2$. Since the bias has no orthogonal component, (3) gives exactly the virtual-value comparisons of this one-dimensional problem. Thus the singleton and pooling comparisons in Theorem 1 coincide with those for the conditional cap, proving the claim. Q.E.D.

Proof of Corollary 6. Let f_0 be the common density. Conditional on $s_1 - s_2 = d$, the density of $s_1 + s_2$ at w is proportional to $f_0((w+d)/2)f_0((w-d)/2)$, so it is log-concave and symmetric about $2m$. Conditioning on $s_1 - s_2$ amounts to conditioning on Y , so $Z = \beta(s_1 + s_2)$ is conditionally log-concave with $\mathbb{E}[Z \mid Y] = 2\beta m$.

By Corollary 5, it suffices to show that, for almost every Y , the cap $c = 2\beta(m - \beta)$ is optimal in the one-dimensional problem with state Z and principal's ideal action $Z - \|\beta\|^2$, where $\|\beta\|^2 = 2\beta^2$. By (3), the virtual measure of this problem integrates a function V to $\mathbb{E}[V(Z) \mid Y] - \|\beta\|^2 \mathbb{E}[V'(Z) \mid Y]$. It therefore has mass one and first moment $\mathbb{E}[Z \mid Y] - \|\beta\|^2 = c$, and it integrates $(Z - t)^+$ to

$$\mathbb{E}[(Z - t)^+ \mid Y] - \|\beta\|^2 \mathbb{P}(Z > t \mid Y).$$

The bias bound gives $m - \beta \leq \underline{s}$, so $c \leq 2\beta\underline{s} \leq Z$ almost surely. For $t < c$, the function $(Z - t)^+$ is affine on the support, so its integral is $c - t$, as for a unit point mass at c . For $t \geq c$ with $\mathbb{P}(Z > t \mid Y) > 0$, the integral equals $\mathbb{P}(Z > t \mid Y)(\mathbb{E}[Z - t \mid Z > t, Y] - \|\beta\|^2)$. The mean residual life $\mathbb{E}[Z - t \mid Z > t, Y]$ of a log-concave distribution is nonincreasing in t , and at $t = c$, where $c \leq Z$, it equals $\mathbb{E}[Z \mid Y] - c = \|\beta\|^2$. Hence the integral is nonpositive for $t \geq c$, where a unit point mass at c gives zero. The virtual measure is thus smaller in convex order than a unit point mass at c , and Theorem 1 makes pooling at c optimal conditional on almost every Y . So the budget $a_1 + a_2 \leq 2(m - \beta)$ is optimal and binds almost surely.

It remains to show that, with a single decision, complete pooling is optimal exactly when $\beta \geq (\bar{s} - \underline{s})/2$. By the first-moment condition, the pooled action must be the principal's ex ante ideal action $m - \beta$. If $m - \beta \leq \underline{s}$, the argument above, with Z replaced by s and $\|\beta\|^2$ by β , proves optimality. Otherwise $\underline{s} < m - \beta < m$, so $\mathbb{E}[s \mid s > m - \beta] > m$, and the comparison for $(s - (m - \beta))^+$ fails. Q.E.D.

Proof of Proposition 1. Fix an optimal U , and let $\bar{U}$ be the convex envelope of the function ψ equal to U on $\text{supp } \mu^- \subseteq \text{bd } S$ and to h elsewhere. Then $U \leq \bar{U} \leq h$, $\bar{U} = U$ on $\text{supp } \mu^-$, and $\bar{U}$ is feasible by Lemma 1. Since $\bar{U} - U \geq 0$ vanishes on the negative support, $\int (\bar{U} - U) d\mu \geq 0$. Optimality, strict positivity of ν almost everywhere in $\text{int } S$, and continuity

therefore imply $\bar{U} = U$ throughout $\text{int } S$.

Since $\text{supp } \mu^-$ is compact and $U \leq h$, ψ is lower semicontinuous. Because b is finite on $\mathbb{R}^n$, $h = (-b)^*$ and hence ψ are superlinear. Its convex envelope therefore admits an attained Carathéodory representation. Fix $x \in \text{int } S \setminus K$ and a minimizing convex combination with positive weights,

$$x = \sum_i \theta_i z_i, \quad \bar{U}(x) = \sum_i \theta_i \psi(z_i).$$

Some z_j lies outside $\text{supp } \mu^-$, since $x \notin K$. For any affine support p of $\bar{U}$ at x , $p \leq \bar{U} \leq \psi$ and equality in this combination force $p(z_i) = \psi(z_i)$ for each i . In particular, $p(z_j) = h(z_j)$ and $p \leq h$, so p is the unique tangent to h at z_j . Hence $\bar{U}$, and thus U , is differentiable at x , with gradient $a = \nabla h(z_j)$, the ideal action of type z_j . Any mechanism inducing U assigns x a lottery with mean a and expected value of b equal to $U(x) - a \cdot x = p(x) - a \cdot x = b(a)$, so strict concavity of b forces the lottery to be degenerate. Q.E.D.

Proof of Corollary 7. The convex hull of the negative support stays on that face and therefore does not meet the interior of S , so Proposition 1 applies throughout the interior. As the type varies over this connected set, the chosen action varies continuously, so the set of actions used by interior types is connected, and its closure is compact and connected as well. Offering exactly these actions reproduces the optimal utility for interior types, and by continuity also for boundary types. Relabeling them by their ideal types, as in our definition of delegation, preserves compactness and connectedness. Q.E.D.

B. Supplementary appendix

B.1. Proofs for Section 5

Proof of Theorem 2. Let $\tilde{m}$ be an incentive-compatible mechanism with indirect utility $\tilde{U}$. At almost every s , the expected action under $\tilde{m}(s)$ is $\nabla\tilde{U}(s)$ and $b(\tilde{m}(s)) = \tilde{U}(s) - s \cdot \nabla\tilde{U}(s)$; since m is deterministic it chooses the action $\nabla U(s)$ and satisfies the same identity. Applying the tangent inequality for the concave function $a \mapsto u_P(a, s) - \kappa b(a)$ at $\nabla U(s)$, averaging over $\tilde{m}(s)$, and substituting these identities gives

$$\begin{aligned} u_P(\tilde{m}(s), s) - u_P(m(s), s) &\leq \kappa [\tilde{U}(s) - U(s)] \\ &\quad + [\nabla_a u_P(\nabla U(s), s) - \kappa s - \kappa \nabla b(\nabla U(s))] \cdot [\nabla\tilde{U}(s) - \nabla U(s)]. \end{aligned}$$

Integrating and applying (14) to $\tilde{U} - U$ yields

$$\int [u_P(\tilde{m}(s), s) - u_P(m(s), s)] dF(s) \leq \int (\tilde{U} - U) d\tilde{\mu}.$$

It remains to show that the right-hand side is not positive. For a.e. $Q \in \mathcal{Q}$,

$$\int \tilde{U} d\tilde{\mu}|_Q \leq \int \tilde{U} d\delta_Q \leq \int U d\delta_Q = \int U d\tilde{\mu}|_Q,$$

where the first inequality holds because $\tilde{U}$ is convex and $\tilde{\mu}|_Q \leq_{cx} \delta_Q$, the second because δ_Q is nonnegative and concentrated where $U = h$ while $\tilde{U} \leq h$, and the equality because U is affine on Q and convex order preserves the integrals of affine functions; the case $\delta_Q = 0$ is covered as well. Integrating over the partition elements gives $\int (\tilde{U} - U) d\tilde{\mu} \leq 0$, so no incentive-compatible mechanism yields the principal a higher payoff than m . Q.E.D.

Proof of Proposition 2. Let m be a mechanism inducing U . At almost every type, $\mathbb{E}[m(s)] = \nabla U(s)$ by the envelope theorem and $U(s) = \nabla U(s) \cdot s - \frac{1}{2} \mathbb{E}\|m(s)\|^2$, so the covariance matrix $\Sigma(s)$ of $m(s)$ satisfies

$$\text{tr } \Sigma(s) = \mathbb{E}\|m(s)\|^2 - \|\nabla U(s)\|^2 = \sigma^2(s).$$

The principal's expected payoff at s is $u_P(\nabla U(s), s) - \frac{1}{2} \text{tr}(W\Sigma(s))$, and since $W - \kappa I$ and $\Sigma(s)$ are positive semidefinite,

$$\text{tr}(W\Sigma(s)) = \text{tr}((W - \kappa I)\Sigma(s)) + \kappa \text{tr } \Sigma(s) \geq \kappa \sigma^2(s).$$

That payoff is therefore at most $u_P(\nabla U(s), s) - \kappa \sigma^2(s)/2$, the integrand of (13). The lottery in the statement has mean $\nabla U(s)$ and places all of its variance in the direction e , so $\text{tr } \Sigma(s) = \sigma^2(s)$ and $\text{tr}(W\Sigma(s)) = \sigma^2(s) e \cdot W e = \kappa \sigma^2(s)$, and it attains equality. Replacing each assigned lottery in this way leaves its mean and expected squared norm unchanged,

hence every type's payoff from every report; the modified mechanism is therefore incentive compatible and still induces U . Integrating gives $\Phi(U)$.

Expanding Φ gives (15), which is quadratic in U , so for every feasible V

$$\Phi(V) - \Phi(U) = \int (V - U) d\tilde{\mu} - \frac{1}{2} \int \nabla(V - U) \cdot (W - \kappa I) \nabla(V - U) dF.$$

The last term is nonpositive, so $\int (V - U) d\tilde{\mu} \leq 0$ for every feasible V suffices for optimality. It is also necessary: $U + \varepsilon(V - U)$ is feasible for $\varepsilon \in [0, 1]$, and the derivative of Φ along this segment at $\varepsilon = 0$ equals $\int (V - U) d\tilde{\mu}$. Q.E.D.

Proof of Proposition 3. Write $a = (a^x, a^z)$ and measure the principal's payoff by her expected loss relative to full information:

$$L_\kappa(m) := \frac{1}{2} \mathbb{E} \left[\|a^x - (x - \beta)\|^2 + \kappa \|a^z - z\|^2 \right], \quad a \sim m(x, z).$$

The expectation includes the state and the mechanism's randomization.

An inexpensive stochastic implementation. Let X be the projection of S onto the biased coordinates, choose $x_0 \in \arg \min_{x \in X} \beta \cdot x$, and let e be a coordinate unit vector in the aligned block. Define m_0 to assign each of the actions

$$\left(x - \beta, z \pm \sqrt{2\beta \cdot (x - x_0)} e \right)$$

with probability $1/2$. Its variance is nonnegative by the choice of x_0 . On S , its truthful utility is

$$U_0(x, z) = \frac{1}{2} (\|x\|^2 + \|z\|^2) - \frac{1}{2} \|\beta\|^2 - \beta \cdot (x - x_0).$$

For any $s, s' \in S$, direct substitution gives

$$u_A(m_0(s), s) - u_A(m_0(s'), s) = \frac{1}{2} \|s - s'\|^2.$$

Thus m_0 is incentive compatible. Its global indirect utility is the supremum of the offered lotteries' payoff planes; the displayed formula for U_0 is asserted only on S . Setting $C := \mathbb{E}[\beta \cdot (x - x_0)] > 0$, we obtain

$$\inf_m L_\kappa(m) \leq L_\kappa(m_0) = \kappa C. \tag{17}$$

No independence assumption is used here.

Deterministic mechanisms cannot match this order of loss. We show the stronger statement

$$\lim_{\kappa \downarrow 0} \frac{\inf_{m \text{ deterministic and IC}} L_\kappa(m)}{\kappa} = +\infty. \tag{18}$$

Otherwise there would be $\kappa_j \downarrow 0$, a finite M , and deterministic incentive-compatible action

rules a_j satisfying $L_{\kappa_j}(a_j) \leq M\kappa_j$. Consequently,

$$\mathbb{E}\|a_j^x - (x - \beta)\|^2 \longrightarrow 0, \quad \mathbb{E}\|a_j^z - z\|^2 \leq 2M,$$

so the second moments of a_j are uniformly bounded. Let U_j be their global indirect utilities. On S , $U_j(s) = s \cdot a_j(s) - \|a_j(s)\|^2/2$, so $\mathbb{E}[U_j(s)]$ is bounded below uniformly in j . For any $r \in \mathbb{R}^n$, convexity and $U_j \leq h$ give

$$U_j(s) \leq \frac{1}{2}U_j(r) + \frac{1}{2}h(2s - r).$$

Integrating over s bounds $U_j(r)$ below uniformly on each compact set. Together with $U_j \leq h$, these bounds imply local equicontinuity. Passing to a subsequence, U_j converges locally uniformly to a finite convex function $U \leq h$, and $\nabla U_j \rightarrow \nabla U$ at almost every point where U is differentiable. Since F has a density, we obtain, F -almost everywhere,

$$\nabla_x U(x, z) = x - \beta, \quad U(s) = s \cdot \nabla U(s) - \frac{1}{2}\|\nabla U(s)\|^2.$$

Choose an open product box inside S on which the density is positive; such a box exists by continuity of the density. The first identity implies there that $U(x, z) = \|x\|^2/2 - \beta \cdot x + w(z)$. Substituting into the second identity yields

$$-\beta \cdot x + w(z) = -\frac{1}{2}\|\beta\|^2 + z \cdot \nabla w(z) - \frac{1}{2}\|\nabla w(z)\|^2$$

almost everywhere on that box, which is impossible for $\beta \neq 0$. This proves (18). Together with (17), it gives a strict advantage over all deterministic mechanisms for sufficiently small κ , without independence.

Optimal implementation under independence. Suppose now that x and z are independent. For any feasible U , define, on the whole space,

$$v(x) := \mathbb{E}_z \left[U(x, z) - \frac{1}{2}\|z\|^2 \right], \quad \bar{U}(x, z) := v(x) + \frac{1}{2}\|z\|^2.$$

The function v is convex and satisfies $v(x) \leq \|x\|^2/2$, so $\bar{U}$ is feasible. In (15), the coefficient of $\nabla_z U$ is zero and the quadratic term is $-(1 - \kappa)\|\nabla_x U\|^2/2$. Independence therefore gives

$$\Phi(\bar{U}) - \Phi(U) = \frac{1 - \kappa}{2} \mathbb{E} \left[\|\nabla_x U(x, z) - \mathbb{E}_z[\nabla_x U(x, z)]\|^2 \right] \geq 0.$$

An optimum can consequently be chosen with indirect utility $v_\kappa(x) + \|z\|^2/2$. Applying Proposition 2 gives

$$q_\kappa(x) = \nabla v_\kappa(x), \quad \sigma_\kappa(x)^2 = 2 \left[x \cdot q_\kappa(x) - \frac{1}{2}\|q_\kappa(x)\|^2 - v_\kappa(x) \right].$$

The resulting mechanism assigns $(q_\kappa(x), z \pm \sigma_\kappa(x)e)$ with equal probabilities, so every aligned action has conditional mean equal to its common ideal. Without independence, the conclu-

sion that all randomization can be concentrated in a single aligned decision follows directly from Proposition 2.

For completeness, an optimum exists for each fixed $\kappa > 0$: bounded loss bounds second moments, so the compactness argument above applies also to stochastic mechanisms. After choosing the least-cost lotteries of Proposition 2, their means and total variances converge almost everywhere along the resulting subsequence. Fatou's lemma for the nonnegative quadratic loss then gives a feasible minimizer. Q.E.D.

B.2. Proofs for Section 6

Proof of Lemma 5. We parametrize the primal problem: for $g \in C(X)$, let $V(g)$ denote the value of the primal problem with upper bound g and let

$$\mathcal{F}(g) := \{U \in \mathcal{U}_0 : U \leq g\}$$

denote its feasible set. The value $V(h)$ is finite because $0 \leq U \leq h$ on S for every $U \in \mathcal{F}(h)$. It is enough to show that $\frac{V(g)-V(h)}{\|g-h\|}$ is bounded above (see, e.g., Condition 3, p. 266, in Gretskey, Ostroy, and Zame, 2002), since this yields a subgradient of $-V$ at h and hence the result (Shapiro, 2010, Proposition 2.5).

Given $\tilde{U} \in \mathcal{F}(g)$, define $U := \max\{\tilde{U} - \|g - h\|, 0\}$. This function is convex and nonnegative, and it satisfies $U \leq h$ on X because $\tilde{U} \leq g \leq h + \|g - h\|$ and $h \geq 0$; hence $U \in \mathcal{F}(h)$. On S we have $\tilde{U} \geq 0$ and therefore $0 \leq \tilde{U} - U \leq \|g - h\|$. Since μ is concentrated on S , it follows that

$$\int \tilde{U} d\mu - \int U d\mu \leq \|\mu\|_{TV} \|g - h\|.$$

Taking the supremum over $\tilde{U} \in \mathcal{F}(g)$ bounds $\frac{V(g)-V(h)}{\|g-h\|}$ by $\|\mu\|_{TV}$, which gives strong duality and dual attainment. Q.E.D.

Proof of Theorem 3. *Sufficiency.* Let V be any feasible indirect utility. For $|\mu|$ -a.e. $Q \in \mathcal{Q}$,

$$\int V d\mu|_Q \leq \int V d\delta_Q \leq \int U d\delta_Q = \int U d\mu|_Q,$$

where the first inequality holds because $V \in \mathcal{U}_0$, the second because δ_Q is nonnegative, $V \leq h$, and $U = h$ on the support of δ_Q , and the equality is the second condition of the theorem. Integrating over the partition elements shows that U is optimal.

Necessity. Suppose U is optimal. By Lemma 5 there is a positive measure γ with $\mu \leq_{\mathcal{U}_0} \gamma$ such that $U = h$ γ -almost everywhere and $\int U d\mu = \int U d\gamma$.

The order $\leq_{\mathcal{U}_0}$ can be converted into the convex order at the cost of an additional measure. Consider maximizing $\int V d(\mu - \gamma)$ over convex V that are nonnegative on S . Its value is zero because $\mu \leq_{\mathcal{U}_0} \gamma$, and the constant function 1 satisfies the constraint strictly, so there

is a positive measure β on S with $\mu + \beta \leq_{cx} \gamma$ (Shapiro, 2010, Proposition 2.8). Since $\int U d\mu = \int U d\gamma$ and $U \geq 0$ on S , this forces $\int U d\beta = 0$, so β is concentrated on the types with $U(s) = 0$.

By Strassen's theorem (see, for example, Phelps, 2001, pp. 93–94), $\gamma + \mu^-$ is a mean-preserving spread of $\mu^+ + \beta$, and as in the proof of Theorem 1 the equality of the integrals of U makes Jensen's inequality tight at almost every source.

We may take β to be a part of μ^- . Indeed, every source in β has $U(s) = 0$, every target satisfies $U \geq 0$ (targets lie in S , or in the support of γ , where $U = h \geq 0$), and Jensen's inequality is tight, so all mass originating in β lands where $U = 0$. Delete this part of the spread together with its targets, split the latter into their γ and μ^- parts, and reinstate the deleted μ^- part by leaving that mass in place as a source. Writing γ and β for the measures that result, the spread now runs from $\mu^+ + \beta$ to $\gamma + \mu^-$, with $\beta \leq \mu^-$ concentrated on $\{s : U(s) = 0\}$. Since the deleted part of γ also sits where $U = 0$, the equality $\int U d\mu = \int U d\gamma$ is preserved.

All sources are now dominated by $|\mu|$, so the argument in the proof of Theorem 1 applies unchanged: the spread does not move mass across partition elements, even where U has boundary kinks. Conditioning as in the proof of Theorem 1 and letting δ_Q denote the conditional measure of γ , which is nonnegative and concentrated on $Q \cap \{s : U(s) = h(s)\}$, we obtain

$$\mu|_Q + \beta|_Q \leq_{cx} \delta_Q$$

for almost every Q . For $V \in \mathcal{U}_0$ the term $\int V d\beta|_Q$ is nonnegative and can be dropped, which gives $\mu|_Q \leq_{\mathcal{U}_0} \delta_Q$; for $V = U$ it vanishes, which gives the asserted equality. Q.E.D.

B.3. Monopoly regulation

This subsection gives the assumptions and formal results for the monopoly regulation application in Section 6. A monopolist is privately informed about his one-dimensional productivity parameter $s \in S = [\underline{s}, \bar{s}]$, which determines the constant marginal cost of production. A monopolist with type s who produces output $a > 0$ obtains profits $sa + b(a) - c$, where $c \geq 0$ is a fixed cost of production. Alternatively, the monopolist can shut down and produce nothing, which yields profits 0.

The regulator's payoff is a weighted sum of producer and consumer surplus,

$$u_P(a, s) := sa + b(a) - c\mathbf{1}_{a>0} + v(a), \quad a \geq 0,$$

where $v(a)$ is the weighted consumer surplus resulting from quantity a . We assume b is strictly concave and continuously differentiable on $[0, \infty)$, satisfies $b(0) = 0$, and has $\lim_{a \rightarrow \infty} b(a)/a = -\infty$. We assume v is continuous on $[0, \infty)$ and continuously differentiable

on $(0, \infty)$, satisfies $v(0) = 0$, and that $b + v$ is concave. The regulator's payoff is therefore concave in the output produced, but jumps at shutdown whenever $c > 0$.

The regulator chooses an incentive-compatible mechanism to maximize her expected payoff. As in the general model, mechanisms may assign lotteries, and feasibility requires that every type obtain a nonnegative expected profit. Writing $u_A(a, s) = sa + b(a) - c\mathbf{1}_{a>0}$ and including the shutdown option when extending the indirect utility to fictitious types,

$$U(s) := \max \left\{ 0, \sup_{s' \in S} u_A(m(s'), s) \right\}, \quad h(s) := \sup_{a \geq 0} \{ sa + b(a) - c\mathbf{1}_{a>0} \}.$$

This extension agrees with the monopolist's expected profit on S ; it is convex and satisfies $0 \leq U \leq h$ on $\mathbb{R}$. Let $s_0 := \inf\{s : h(s) > 0\}$ denote the lowest type for whom production is profitable if no quantity constraints are imposed.

Lemma 7. *A function on S is induced by a feasible mechanism if and only if it has a convex extension U to $\mathbb{R}$ satisfying $0 \leq U \leq h$.*

The proof follows the supporting-line argument in Lemma 1. The fixed cost changes which lotteries implement a given supporting line: a lottery between shutdown and a positive output can be needed even when the expected output is small.

We illustrate the approach by giving sufficient conditions for a *cap allocation* (see Amador and Bagwell, 2022) to be optimal. Given $\underline{a} \geq \lim_{s \downarrow s_0} h'(s)$, the cap allocation $q_{\underline{a}}$ lets the monopolist choose from $\{0\} \cup [\underline{a}, \infty)$: types that strictly prefer shutdown produce nothing, and the remaining types choose their preferred output above the cap. This output floor corresponds to a price cap. Its indirect utility is illustrated in Figure 8.

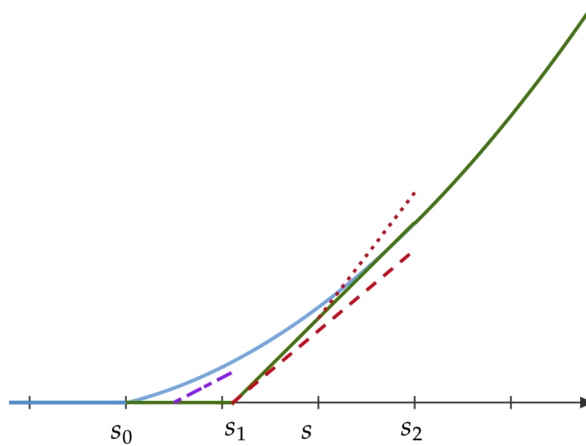


Figure 8: Perturbations of a cap allocation. The blue curve is h and the green curve the indirect utility of the cap allocation; the red and purple curves illustrate perturbations on the pooling and shutdown regions, respectively.

Fix a cap allocation with $\underline{a} > \lim_{s \downarrow s_0} h'(s)$ and indirect utility U , and let

$$s_1 := \frac{c - b(\underline{a})}{\underline{a}}, \quad s_2 := -b'(\underline{a}).$$

Then $s_0 < s_1 < s_2$: the monopolist shuts down below s_1 , produces $\underline{a}$ between s_1 and s_2 , and chooses his unconstrained output above s_2 . These thresholds need not lie in S ; in particular, s_2 remains the contact point with h even when all producing types in S choose the cap.

Because the regulator's payoff jumps at shutdown, the linearization of Section 5 must be replaced by a bound that is tight at shutdown as well. For this application, set

$$\hat{\kappa} := 1 + \frac{v(\underline{a}) - \underline{a}v'(\underline{a})}{b(\underline{a}) - \underline{a}b'(\underline{a}) - c}$$

and assume that $a \mapsto (1 - \hat{\kappa})b(a) + v(a)$ is concave on $(0, \infty)$. The denominator equals $\underline{a}(s_2 - s_1)$ and is therefore positive. This coefficient is specific to the cap and need not agree with the coefficient κ of Section 5; it may be negative. Its defining property is that

$$(1 - \hat{\kappa})[b(\underline{a}) - c] + v(\underline{a}) = \underline{a}[(1 - \hat{\kappa})b'(\underline{a}) + v'(\underline{a})],$$

so that the tangent at $\underline{a}$ to the adjusted payoff $a \mapsto (1 - \hat{\kappa})[b(a) - c] + v(a)$ passes through the shutdown payoff at $a = 0$. Following this tangent below $\underline{a}$ and the adjusted payoff above it gives a concave upper bound that is tight at every output the cap uses.

The measure of Section 5 takes the following form here, with $\hat{\kappa}$ in place of κ and the concave bound in place of the adjusted payoff: $\tilde{\mu}$ is the signed measure supported on S with

$$\begin{aligned} \int V d\tilde{\mu} &= \int_S \hat{\kappa} V(s) dF(s) \\ &+ \int_S \left[(1 - \hat{\kappa}) \left(s + b'(\max\{\underline{a}, U'(s)\}) \right) + v'(\max\{\underline{a}, U'(s)\}) \right] V'(s) dF(s) \end{aligned} \quad (19)$$

for every continuously differentiable V on a neighborhood of S .⁹ The coefficient of V' is the slope of the concave bound at the output the cap assigns: it equals $(1 - \hat{\kappa})(s + b'(\underline{a})) + v'(\underline{a})$ below s_2 and $v'(h'(s))$ above s_2 , and the two agree at s_2 .

Proposition 4. *Under the curvature condition above, the cap allocation is optimal if*

- (i) $\tilde{\mu}$ is nonnegative on $[s_2, \infty)$,
- (ii) $\int_{(s_1, s)} (s - x) d\tilde{\mu}(x) \leq 0$ for all $s \in [s_1, s_2]$ with equality for $s = s_2$, and
- (iii) $\int_{[s, s_1]} (x - s) d\tilde{\mu}(x) \leq 0$ for all $s \in [s_0, s_1]$.

The interpretation is as before. Condition (i) ensures that the regulator benefits from

⁹ The measure is well defined: before multiplying by f , the coefficient of V' is the sum of a linear function of s and the derivative of a concave function evaluated at a nondecreasing output, and therefore has bounded variation on S . Integrating by parts gives a finite signed measure, including its endpoint masses, and smooth approximation extends the identity to differences of feasible indirect utilities.

giving types above s_2 their full-discretion payoff. Condition (ii) is the comparison on the pooling interval, illustrated by the red curve in Figure 8, and the equality at s_2 ensures that the affine piece of U has the same value as the corresponding mass at its contact point. Condition (iii) rules out gains from raising the utility of the shutdown types, as illustrated by the purple curve. The comparisons use the entire measure, so any mass at an endpoint of S is included; a mass at s_1 is assigned to the shutdown region and a mass at s_2 to the region above it.

The cap's first-order condition. To relate condition (ii) to (16), suppose $\underline{s} < s_1 < s_2 < \bar{s}$ and write $J(\underline{a}) := \int_S u_P(q_{\underline{a}}(s), s) dF(s)$. The shutdown threshold satisfies $ds_1/d\underline{a} = (s_2 - s_1)/\underline{a}$. At s_1 , operating rather than shutting down yields the regulator $v(\underline{a})$; at s_2 , constrained and unconstrained output coincide. Differentiating the cap's expected payoff therefore gives

$$J'(\underline{a}) = \int_{s_1}^{s_2} [s - s_2 + v'(\underline{a})] f(s) ds - \frac{v(\underline{a})}{\underline{a}} (s_2 - s_1) f(s_1).$$

For $s \leq s_2$, let $\ell(s) := (1 - \hat{\kappa})(s - s_2) + v'(\underline{a})$ denote the coefficient of V' in (19). The definition of $\hat{\kappa}$ gives $\ell(s_1) = v(\underline{a})/\underline{a}$. Integration by parts then yields

$$\begin{aligned} \int_{(s_1, s_2)} (s_2 - s) d\tilde{\mu}(s) &= \int_{s_1}^{s_2} [s_2 - s - v'(\underline{a})] f(s) ds + (s_2 - s_1) \ell(s_1) f(s_1) \\ &= -J'(\underline{a}). \end{aligned}$$

Thus the equality in condition (ii) is exactly the cap's first-order condition. The proposition also applies when a cutoff lies outside S , without imposing the interior-cutoff restriction used for this interpretation.

Proof of Lemma 7. For a feasible mechanism, the supremum of the payoff lines of its lotteries and of the shutdown option gives a convex extension satisfying $0 \leq U \leq h$.

Conversely, let U be such an extension and choose a supporting slope $a \in \partial U(s)$ for each $s \in S$. Since $h = 0$ on $(-\infty, s_0]$, we have $U = 0$ there, so every supporting slope is nonnegative. If $a = 0$, then $U(s) = 0$, and we assign shutdown.

Suppose $a > 0$. The largest expected value of $b(a) - c\mathbf{1}_{a>0}$ at a fixed mean is given by its concave hull, which for $c > 0$ joins the origin tangentially to $b - c$ and then follows $b - c$, and which for $c = 0$ is simply b . Since the supporting line lies below $U \leq h$, its intercept $U(s) - sa$ is at most this hull evaluated at a . We can therefore choose $A \geq a$ with

$$\frac{b(A) - c}{A} = \frac{U(s) - sa}{a}.$$

Indeed, the largest value of the left-hand side over $A \geq a$, multiplied by a , is the hull evaluated at a , and the ratio decreases to $-\infty$ thereafter.

Assign output A with probability a/A and shutdown otherwise. This lottery has expected output a and yields type s the profit $U(s)$. Its payoff to any other type t is $U(s) + a(t - s)$, a supporting line below $U(t)$, so the mechanism is incentive compatible, and these choices can be made measurably. Finally, $U(s) \geq 0$ bounds the target ratio below by $-s$, so the chosen outputs can be bounded uniformly on S and the resulting payoffs are integrable. Q.E.D.

Proof of Proposition 4. Consider the adjusted positive-output payoff $a \mapsto (1 - \hat{\kappa})[b(a) - c] + v(a)$. By the curvature condition and the choice of $\hat{\kappa}$, its tangent at $\underline{a}$ passes through the origin. Following this tangent on $[0, \underline{a}]$ and the adjusted payoff above $\underline{a}$ gives a concave upper bound which is tight at shutdown and at every positive output used by the cap. Its slope at shutdown is $(1 - \hat{\kappa})b'(\underline{a}) + v'(\underline{a})$, and at any output $a \geq \underline{a}$ it is $(1 - \hat{\kappa})b'(a) + v'(a)$; these are the slopes appearing in (19).

Since

$$u_P(a, s) - \hat{\kappa}u_A(a, s) = (1 - \hat{\kappa})sa + (1 - \hat{\kappa})[b(a) - c\mathbf{1}_{a>0}] + v(a),$$

this bound is the hypothesis of Theorem 2 in the form discussed in Section 5, with $\hat{\kappa}$ in the role of κ . Hence, for any feasible mechanism $\tilde{m}$ with indirect utility $\tilde{U}$,

$$\int_S [u_P(\tilde{m}(s), s) - u_P(q_{\underline{a}}(s), s)] dF(s) \leq \int (\tilde{U} - U) d\tilde{\mu}.$$

It remains to show that the last integral is nonpositive. We use the sufficiency comparison of Theorem 3 with $\tilde{\mu}$ in place of μ : for each of the three regions Q below we exhibit a nonnegative measure δ_Q with $U = h$ on its support such that

$$\int_Q \tilde{U} d\tilde{\mu} \leq \int \tilde{U} d\delta_Q \leq \int U d\delta_Q = \int_Q U d\tilde{\mu},$$

where the second inequality uses $\tilde{U} \leq h$ together with the nonnegativity of δ_Q .

On $[s_2, \infty)$ take $\delta_Q = \tilde{\mu}|_Q$, which is nonnegative by (i) and lives where $U = h$; the first inequality is then an equality.

On (s_1, s_2) take δ_Q to be the point mass at s_2 of mass $\tilde{\mu}((s_1, s_2))$. This mass is nonnegative, being the left derivative at s_2 of the expression in (ii), which is nonpositive and vanishes there, and $U(s_2) = h(s_2)$. Testing the functions $(s - x)^+$ yields precisely the inequalities in (ii), and the equality at s_2 matches the first moments, so $\tilde{\mu}|_Q \leq_{cx} \delta_Q$ and the first inequality holds. Since U is affine on this interval and convex order preserves the integrals of affine functions, the final equality holds as well.

On the shutdown region take $\delta_Q = 0$; the final equality holds because U vanishes there. Any feasible $\tilde{U}$ is nonnegative, convex, and vanishes on $(-\infty, s_0]$, so on this region it is a limit of positive linear combinations of the functions $(x - s)^+$ with $s \geq s_0$ (Shaked and Shanthikumar, 2007, p. 183), and condition (iii) therefore gives $\int_{(-\infty, s_1]} \tilde{U} d\tilde{\mu} \leq 0$.

Adding the three comparisons shows that the cap allocation is optimal.

Q.E.D.

B.4. The uniform-square example

We first record a fact about signed barycenters that makes it sufficient to check the first-moment conditions instead of the full convex-order conditions.

From barycenters to convex order. Every pool constructed below has the same shape: all of its negative virtual mass sits at a single point, and all of its positive mass lies on the segment between that point and the assigned action a . For a pool Q of this shape the signed-barycenter condition already delivers the full convex-order comparison required by Theorem 1,

$$\int (s - a) d\mu|_Q(s) = 0 \implies \mu|_Q(Q) \geq 0 \quad \text{and} \quad \mu|_Q \leq_{cx} \delta_Q, \quad (20)$$

with δ_Q the point mass at a of mass $\mu|_Q(Q)$, as in the main text. Moreover, $\mu|_Q(Q) > 0$ unless all of the positive mass sits at the negatively weighted point.

Small bias: the conditional measures. Fix $0 < \beta < 1/2$ and retain the parametrization $a(t)$ of the main text, writing $q(t) := a'_2(t)/a'_1(t)$ for the slope of the boundary; the constructed branch has $a'_1(t) > 0$ and $-1 \leq q(t) < 0$. Orthogonality (6) gives

$$a_1(t) = t + (1 - a_2(t))q(t), \quad a'_1(t) = \frac{1 + (1 - a_2(t))q'(t)}{1 + q(t)^2}. \quad (21)$$

The part of the pool labelled t inside S is the segment of the normal through $a(t)$ given by

$$Q_t \cap S = \left\{ (t + (1 - y)q(t), y) : \underline{y}(t) \leq y \leq 1 \right\}, \quad \underline{y}(t) := \max\{a_2(t), y_0(t)\},$$

where $y_0(t) := 1 + t/q(t)$ is the height at which it meets the vertical axis. By (21), $y_0(t) > a_2(t)$ exactly when $a_1(t) < 0$, and the types then stop at the west edge rather than at the assigned action. At height y neighboring pools are $\varepsilon + (1 - y)[q(t + \varepsilon) - q(t)]$ apart, so the Jacobian is $1 + (1 - y)q'(t)$, verified positive below; and $y'_0(t) = [1 + (1 - y_0(t))q'(t)]/q(t) < 0$, so the west edge contributes positive mass $-\beta y'_0(t)$ per unit of t . Conditional measures are written against the pool label t ; this differs from the normalization in Theorem 1 only by a positive rescaling, which leaves the convex-order comparisons unchanged. Writing the conditional measure in the height coordinate and integrating against dt therefore gives

$$d\mu|_{Q_t}(y) = [1 + (1 - y)q'(t)] \mathbf{1}_{[\underline{y}(t), 1]}(y) dy - \beta y'_0(t) \mathbf{1}_{\{y_0(t) > a_2(t)\}} \delta_{y_0(t)}(dy) - \beta \delta_1(dy), \quad (22)$$

the limit behind the thin band of the main text.

A type $s \in Q_t$ satisfies $s - a(t) = (s_2 - a_2(t))(-q(t), 1)$, so all of this mass lies on one

ray from $a(t)$ and the signed-barycenter condition is the scalar equation

$$\int (y - a_2(t)) d\mu|_{Q_t}(y) = 0. \quad (23)$$

For a pool inside the square, $y(t) = a_2(t)$, the atom at $y_0(t)$ is absent and (23) reads

$$\frac{(1 - a_2(t))^2}{2} + \frac{q'(t)(1 - a_2(t))^3}{6} = \beta(1 - a_2(t)),$$

which is precisely the differential equation (9) of the main text. Solving it for $q'(t)$ and using (21) gives the interior system

$$q'(t) = \frac{3(2\beta - 1 + a_2(t))}{(1 - a_2(t))^2}, \quad a'_2(t) = \frac{2q(t)(3\beta - 1 + a_2(t))}{(1 - a_2(t))(1 + q(t)^2)}. \quad (24)$$

For a pool meeting the west edge, $y(t) = y_0(t)$ and the atom there enters as well. Carrying out the integration in (23) and substituting $y'_0(t)$ leaves a linear equation in $q'(t)$, so that

$$q'(t) = \frac{\beta(1 - a_2(t)) - \int_{y_0(t)}^1 (y - a_2(t)) dy + \frac{\beta}{q(t)}(y_0(t) - a_2(t))}{\int_{y_0(t)}^1 (y - a_2(t))(1 - y) dy - \frac{\beta}{q(t)}(1 - y_0(t))(y_0(t) - a_2(t))}, \quad (25)$$

together with (21). Its denominator is strictly positive for $a_2(t) \leq y_0(t) < 1$ and $q(t) < 0$, and at $y_0(t) = a_2(t)$ it reduces to (24), so the two systems agree at the axis crossing.

The exterior branch. Choose a starting depth $\ell \in (0, \beta)$ and impose

$$a(0) = (-\ell, 1 - \ell), \quad q(0) = -1. \quad (26)$$

At $t = 0$ the pool collapses to the northwest corner, where the west- and north-edge weights $-\beta/q$ and $-\beta$ cancel exactly when $q(0) = -1$. Equation (25) has the form $0/0$ there, but substituting $q(t) = -1 + tz(t)$ and cancelling the common factor t turns it into a regular singular point, so standard arguments give a unique local solution, depending continuously on ℓ , with

$$q'(0) = \frac{\beta - \ell}{2\beta\ell} > 0, \quad a'_2(0) = -\frac{3\beta - \ell}{4\beta}. \quad (27)$$

The slope increases until the boundary reaches the vertical axis.¹⁰ Hence $-1 < q < 0$, $a'_1 > 1/2$ and $a'_2 < 0$; as $da_2/da_1 = q > -1$ the depth stays below 2ℓ , and $a_1 < 0$ forces $q < -t/(1 - a_2)$, so the slope cannot reach zero first. The branch therefore reaches $a_1 = 0$

¹⁰Let N be the numerator of (25), so that q' has the sign of N . At a first zero of q' we would have $-1 < q < 0$, and $N = 0$ would force $1 + y_0 - 2a_2 < 2\beta$; since $a'_1 = 1/(1 + q^2)$ and $y'_0 = 1/q$ there, differentiating gives $N' = [q^2(1 - a_2 - \beta) + (y_0 - a_2) + \beta/q]/[q(1 + q^2)] > 0$, because $q(1 + q^2) < 0$ while $q^2(1 - a_2) + (y_0 - a_2) \leq \frac{1+q^2}{2}(1 + y_0 - 2a_2) < \beta(1 + q^2) < \beta(q^2 - 1/q)$. So N cannot reach zero from above.

in finite time without a singularity, at a height $a_{2,c}$ and slope q_c with

$$\ell < 1 - a_{2,c} < 2\ell < 2\beta, \quad -1 < q_c < 0. \quad (28)$$

Matching on the diagonal. Continue with the interior system (24); dividing its two equations and integrating gives the first integral

$$(1 + q(t)^2)(1 - a_2(t))^2(3\beta - 1 + a_2(t)) = C, \quad C = (1 + q_c^2)(1 - a_{2,c})^2(3\beta - 1 + a_{2,c}), \quad (29)$$

which depends continuously on ℓ . By (28), $C \rightarrow 0$ as $\ell \downarrow 0$ and $C < 8\beta^3$ throughout, while as $\ell \uparrow \beta$ the exterior solution converges to $q \equiv -1$, $a_2(t) = 1 - \beta - t/2$, which reaches the axis at $t = 2\beta$, so that $C \rightarrow 8\beta^3$.

Consider starting values with $4\beta^3 < C < 8\beta^3$, so that $q = -[C/((1 - a_2)^2(3\beta - 1 + a_2)) - 1]^{1/2}$ on the interior branch. The slope increases until $a_2 = 1 - 2\beta$ and then decreases, returning to -1 at the unique height $a_{2,+} \in (1 - 3\beta, 1 - 2\beta)$ with $2(1 - a_{2,+})^2(3\beta - 1 + a_{2,+}) = C$. Since $da_1/da_2 = 1/q$ and $a_1 = 0$ at the axis crossing, the north-edge label at this return is

$$T(\ell) = 1 - a_{2,+} + \int_{a_{2,+}}^{a_{2,c}} \left(\frac{C}{(1 - y)^2(3\beta - 1 + y)} - 1 \right)^{-1/2} dy, \quad (30)$$

and we select ℓ so that $T(\ell) = 1$. On the connected component of $\{\ell \in (0, \beta) : C(\ell) > 4\beta^3\}$ whose closure contains β , the function T is continuous. As $\ell \uparrow \beta$ the crossing and return heights both tend to $1 - 2\beta$, so $T(\ell) \rightarrow 2\beta < 1$. At its other endpoint $C \downarrow 4\beta^3$, and since $(1 - y)^2(3\beta - 1 + y)$ has a strict quadratic maximum $4\beta^3$ at $y = 1 - 2\beta$, the integrand increases there to a limit with a nonintegrable singularity, so $T(\ell) \rightarrow \infty$. The intermediate value theorem gives the required starting value, for which $q(1) = -1$ and $a(1) = (a_{2,+}, a_{2,+})$; by (30), $a_{2,+} > 0$, so the contact lies strictly inside the square and the boundary joins its reflection smoothly on the diagonal.

The menu and the agent's incentives. Orthogonality is only a first-order condition along each normal. It remains to show that the pools partition the type space outside the full-discretion region, so that U is well defined, and that U is convex, so that each type's assigned action is optimal among all actions in the menu. For the selected solution $a'_1 > 0$, $a'_2 < 0$ and $-1 \leq q < 0$, and the Jacobian $1 + (1 - y)q'$ is positive on every pool: on the exterior branch because $q' > 0$, and on the interior branch because it is affine in y , equals one at $y = 1$ and equals $2(3\beta - 1 + a_2)/(1 - a_2) > 0$ at $y = a_2$. At each height the coordinate $t + (1 - y)q(t)$ is therefore strictly increasing in t , and $\underline{y}(t)$ decreases with t , so the upper pools do not cross and, together with their reflections, fill the part of S outside the full-discretion region, the closed southwest region bounded by the coordinate axes, the interior portion of $a(t)$ and its reflection; $a_{2,+} > 0$ keeps these segments inside the square. The menu is this

region together with the two exterior arcs.

Set $U = h$ on the full-discretion region and $U(s) = a(t) \cdot s - \frac{1}{2}\|a(t)\|^2$ for $s \in Q_t$, with the reflected definition below the diagonal. Orthogonality gives $\nabla U(s) = a(t)$ and

$$D^2U(s) = \frac{a'_1(t)}{1 + (1 - s_2)q'(t)} \begin{pmatrix} 1 & q(t) \\ q(t) & q(t)^2 \end{pmatrix} \succeq 0, \quad (31)$$

while $D^2U = I$ on the full-discretion region. Utilities and gradients agree across the boundary and the diagonal and these derivatives are bounded, so the piecewise smooth gradient is Lipschitz and U is convex on S . Every offered action, including the limiting ones at the ends of the exterior arcs, generates a supporting payoff plane, so the menu implements the stated assignments, and the supremum of those planes is convex, lies below h on $\mathbb{R}^2$ and equals h at every offered action.

Optimality and the necessity of nonconvexity. On each nontrivial pool the positive part of $\mu|_{Q_t}$ lies between the assigned action and the north- or east-edge endpoint carrying its negative mass, and (23) is its signed-barycenter condition, so (20) gives $\mu|_{Q_t}(Q_t) > 0$ and $\mu|_{Q_t} \leq_{cx} \delta_{Q_t}$. Hence every feasible indirect utility V satisfies

$$\int V d\mu|_{Q_t} \leq \int V d\delta_{Q_t} \leq \int U d\delta_{Q_t} = \int U d\mu|_{Q_t},$$

while on the remaining types the virtual measure is nonnegative and $U = h$. Integrating proves optimality among all feasible mechanisms, including stochastic ones.

The same comparison gives uniqueness: equality forces any other optimum V to satisfy $V(a(t)) = h(a(t))$ for almost every pool and to be affine on the segment from $a(t)$ to its negatively weighted endpoint, and since $V \leq h$ that affine restriction is the tangent plane of h at $a(t)$, hence U ; on the full-discretion region equality requires $V = h$. So $V = U$ on S by continuity, and the quadratic payoff identity forces every optimal mechanism to assign a lottery with zero variance to almost every type.

Finally, $1 - a_{2,c} < 2\beta$ by (28), so $q' > 0$ on an interior arc just after the boundary crosses the vertical axis, where $d^2a_2/da_1^2 = q'/a'_1 > 0$. Take two nearby contact points on this arc, both strictly inside S ; their midpoint lies above the boundary, where $U < h$. A closed convex delegation set implementing an optimum would contain both contact points, since the unique optimal utility equals h there, hence also their midpoint, and would give that type full discretion, contradicting $U < h$. No closed convex delegation set is therefore optimal for $0 < \beta < 1/2$.