Comparisons of Experiments in Moral Hazard Problems

Zizhe Xia*

July 5, 2025

Abstract

I use a novel geometric approach to compare information in moral hazard problems. I study three nested geometric orders on information, namely the column space, the conic span, and the zonotope orders. The orders are defined by the inclusion of the column space, the conic span, and the zonotope of the matrices representing the experiments. For each order, I establish four equivalent characterizations – (i) inclusion of polyhedral sets of feasible state dependent utilities, (ii) matrix factorization, (iii) posterior belief distributions, and (iv) classes of moral hazard problems. The column space order characterizes the comparison of feasibility in all moral hazard problems. The conic span order characterizes the comparison of costs in all moral hazard problems with a risk neutral agent and limited liability. The zonotope order characterizes the comparison of costs in all moral hazard problems when the agent can have any utility exhibiting risk aversion.

*zizhe-xia@chicagobooth.edu I am indebted to Doron Ravid for his invaluable guidance and advices. I thank Alex Frankel, Emir Kamenica, Andrew McClellan, Daniel Rappoport, Lars Stole for their valuable comments and suggestions. All errors are my own.

1 Introduction

In a canonical moral hazard (MH) setting, the principal uses some noisy information to contract with the agent and align incentives. At its core lies a fundamental question of which information is better, that is, always yielding a lower agency cost. The classic informativeness principle by **?** gives one answer: If one information structure augments another with additional signals that contain new information about the agent's action, then it strictly reduces the agency cost.[1] **?** extends this result and provides a more general comparison based on the mean preserving spread order on the distribution of the likelihood ratios. More recently, **?** proposes using the linear convex order on the distribution of the likelihood ratios.

Although the literature offers a rich array of likelihood ratio-based comparisons, they do not directly tie information back to the agent's state dependent utility – the very object that governs incentives.[2] Given the state dependent utility, the agent chooses how to optimally affect the distribution of the states at some cost. On the other hand, the principal controls this state dependent utility by designing a contract which specifies the payment conditional on her information. The principal's information thus completely determines which utilities can be generated and at what cost. This insight motivates the comparisons of information based on the set of utilities they can generate.

This paper implements this geometric perspective to compare information in moral hazard problems. The moral hazard environment features a risk neutral principal (she) who hires an agent (he) to produce outcomes. The agent can produce any distribution over a finite state space at some smooth and convex cost. His objective is to maximize the expected utility minus the cost of production. The principal has access to some noisy information about the states. To provide incentives, she offers a contract that pays the agent based on her information. She has to minimize the cost of implementing a given state distribution. Depending on the applications, the agent may be risk neutral or risk averse, and the principal may face two additional constraints: limited liability, which requires non-negative payments, and ex post budget, which caps the payments from above.

My analysis is driven by the following question: Under what condition does one source of information yields a lower cost to the principal than another, irrespective of the agent's cost function and the target state distribution?[3] I answer this question across several classes of moral hazard problems, depending on whether the agent is risk neutral and whether the principal is subject to limited

[1] More precisely, the informativeness principle compares two information structures that have an inclusive relation, that is, one has additional signals than the other. In this case, the information structure with additional signals always yields a strictly lower agency cost if and only if the existing signals are not sufficient statistics of the additional signals.

[2] Formally, the state dependent utility is a vector that specifies the agent's expected utility across states.

[3] The total cost to the principal consists of two parts: the first best cost – the cost if the agent's action is directly contractible, which is the just the production cost – and the agency cost – the extra expenditure to the principal when the action is hidden. Since the principal's information cannot change the first best cost, comparing total costs is equivalent to comparing the agency costs alone.

Table 1: Comparisons of Information

Order	Set Inclusion	Matrix Factorization	Posterior Beliefs	Decision Problems
Column space $\geq_{\text{Col}}$	$\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M\}$	$\mathcal{E}' = \mathcal{E}G$ Any G	Affine span inclusion	MH implementability; MH with risk neutrality no limited liability
Conic span $\geq_{\text{Cone}}$	$\left\{ \begin{matrix} \mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, \\ \boldsymbol{v} \geq 0 \end{matrix} \right\}$	$\mathcal{E}' = \mathcal{E}G$ $G \geq 0$	Convex hull inclusion	MH with risk neutrality and limited liability
Zonotope $\geq_{\text{Zon}}$	$\left\{ \begin{matrix} \mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, \\ 0 \leq \boldsymbol{v} \leq 1 \end{matrix} \right\}$	$\mathcal{E}'B_N = \mathcal{E}G_N$ $0 \leq G_N \leq 1$	Linear convex order	MH with risk neutrality limited liability and ex post budget; MH with risk aversion
Blackwell $\geq_{\text{B}}$	$\bigcup_{K=1}^{\infty} \left\{ \begin{matrix} \mathcal{E}\Pi : \Pi \in \mathbb{R}^{M\times K}, \\ \Pi \geq 0, \Pi\mathbf{1} \leq \mathbf{1} \end{matrix} \right\}$	$\mathcal{E}' = \mathcal{E}G$ $G \geq 0, G\mathbf{1} = \mathbf{1}$,	Convex order	Any decision problem

Note: This table summarizes the results in the paper. $B_N := \begin{bmatrix} \boldsymbol{v}^{(1)} & \boldsymbol{v}^{(2)} & \dots & \boldsymbol{v}^{(2^N)} \end{bmatrix} \in \{0,1\}^{N\times 2^N}$ where $\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(2^N)}\} = \{0,1\}^N$ is the set of all binary vectors.

liability and ex post budget constraints.

The main contribution of the paper is the study of three geometric orders on information, namely the column space, the conic span, and the zonotope orders, and the demonstration of how each yields sharp comparisons of agency cost across different moral-hazard environments. To the author's knowledge, the column space and conic span orders are novel; the zonotope order has been studied previously,[4] and the contribution there is a new geometric perspective for moral hazard problems. To state the results precisely, I need to introduce the following notations. Let $\Omega = \{\omega_n\}_{n=1}^N$ be the finite state space. Information is modeled as finite (Blackwell) experiments.[5] A finite experiment $\mathcal{E}$ is an $N \times M$ row stochastic matrices where the n-th row represents the state ω_n, the m-th column represents the realization y_m, and the (n,m)-th entry is the conditional probability of realization y_m in state ω_n. Given the information $\mathcal{E}$, the principal's contract is described by a payment rule $\boldsymbol{t} \in \mathbb{R}^M$ that specifies the payment to the agent following each realization of $\mathcal{E}$. We are now ready to state the main results. They are summarized in Table 1.

As a prelude, let's recall the four equivalent ways to say $\mathcal{E}$ dominates $\mathcal{E}'$ in the Blackwell (**??**) order, denoted $\mathcal{E} \geq_{\text{B}} \mathcal{E}'$. The first is, for all decision problems, the expected payoff under $\mathcal{E}$ is no smaller than that under $\mathcal{E}'$. Equivalently, $\mathcal{E}'$ is a garbling of $\mathcal{E}$, meaning that there exists a garbling matrix G such that $\mathcal{E}' = \mathcal{E}G$.[6] Third, the posterior distribution induced by $\mathcal{E}$ dominates that by $\mathcal{E}'$ in the convex order for any prior.[7] Finally, for any action space A, the feasible set of joint distributions

[4]The zonotope order was first introduced as a comparison of inequality (**???**, see **?** for a textbook treatment). As far as the author is aware, **?** is the first to introduce it in information economics. They show that the zonotope order compares the value of information for all binary decision problems. **?** generalizes it to infinite experiments and illustrates its relation to the Lehmann order and moral hazard problems.

[5]All my results can be extended to infinite experiments.

[6]A matrix G is a garbling matrix if all entries are positive $G \geq 0$, and every row sums up to one $G\mathbf{1} = \mathbf{1}$.

[7]Recall that for distributions $F, G : \mathbb{R}^N \to [0,1]$, F dominates G in the convex order if for any convex function $\phi : \mathbb{R}^N \to \mathbb{R}$, $\mathbb{E}_{\boldsymbol{x}\sim F}[\phi(\boldsymbol{x})] \geq \mathbb{E}_{\boldsymbol{x}\sim F}[\phi(\boldsymbol{x})]$. The convex order compares the dispersion in the distributions.

over $\Omega \times A$ induced by $\mathcal{E}$ contains the corresponding set induced by $\mathcal{E}'$. While this feasible set can be rather intricate (see Table 1), in moral hazard problems they can be simplified as the principal only decides on how much to pay the agent.

The first order I study is the column space order, defined by the inclusion of the column spaces of experiments. The column space of experiment $\mathcal{E}$ is defined as $\operatorname{Col}\mathcal{E} := \{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M\}$. This is the set of all possible state dependent utilities that can be generated by experiment $\mathcal{E}$ with any payment rule. Say that $\mathcal{E}$ dominates $\mathcal{E}'$ in the column space order, denoted $\mathcal{E} \geq_{\text{Col}} \mathcal{E}'$, if $\operatorname{Col}\mathcal{E} \supseteq \operatorname{Col}\mathcal{E}'$. The column space order characterizes the comparisons of implementability in all moral hazard problems. More specifically, $\mathcal{E} \geq_{\text{Col}} \mathcal{E}'$ if and only if any implementable state distribution under $\mathcal{E}'$ is also implementable under $\mathcal{E}$ for any moral hazard problems. This is because a larger column space allows the principal to choose from a larger set of state dependent utilities, hence enlarging the implementable set.[8]

The column-space order also characterizes the cost comparisons with a risk-neutral agent and no further constraints.[9] That is, better information in the column space order is equivalent to lower cost in all such moral hazard problems. In this case, any implementable state distribution can be implemented at the first best cost without rents, reducing cost comparisons to implementability.[10]

The second order is the conic span order, defined by the inclusion of the conic spans of experiments. The conic span of experiment $\mathcal{E}$ is defined as $\operatorname{Cone}\mathcal{E} := \{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, \boldsymbol{v} \geq 0\}$, which is the set of all possible state dependent utilities that can be generated by experiment $\mathcal{E}$ with any non-negative payment rule.[11] Say that $\mathcal{E}$ dominates $\mathcal{E}'$ in the conic span order, denoted $\mathcal{E} \geq_{\text{Cone}} \mathcal{E}'$, if $\operatorname{Cone}\mathcal{E} \supseteq \operatorname{Cone}\mathcal{E}'$. The conic span order characterizes the cost comparisons moral hazard problems with a risk neutral agent and limited liability. Under risk neutrality, expected utility from money equals the expected payment. Hence, for cost minimization the principal minimizes the agent's utility from money. A larger conic span provides a larger feasible set of state dependent utilities, weakly reducing the cost. Risk neutrality is indispensible because a lower expected utility does not imply a lower expected payment under strict risk aversion.

Next, I study the zonotope order, defined by the inclusion of the zonotope of experiments. The zonotope of experiment $\mathcal{E}$ is defined as $\operatorname{Zon}\mathcal{E} := \{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, 0 \leq \boldsymbol{v} \leq 1\}$, which is the set of all possible state dependent utilities that can be generated by experiment $\mathcal{E}$ with any non-negative and bounded payment rule. Say that $\mathcal{E}$ dominates $\mathcal{E}'$ in the conic span order, denoted $\mathcal{E} \geq_{\text{Zon}} \mathcal{E}'$, if $\operatorname{Zon}\mathcal{E} \supseteq \operatorname{Zon}\mathcal{E}'$. This coincides with the linear Blackwell order studied in **?**; here I offer a geometric interpretation. Using the feasible state dependent utility interpretation, it is straightforward to see

[8] This simple argument works for the "only if" direction. The "if" direction requires a constructive proof.

[9] I focus on the principal's cost minimization problem to implement a given state distribution. For her profit maximization problem, a sell-the-firm contract always achieves the first best when the agent is risk neutral and not protected by limited liability, and the principal's information is irrelevant.

[10] The first best cost is the agent's production cost plus his outside option.

[11] For any vNM utility index $u : \mathbb{R} \to \mathbb{R}$ of the agent, I can normalize $u(0) = 0$ so that a non-negative utility corresponds to a non-negative payment.

that the zonotope order characterizes the cost comparison under risk neutrality, limited liability, and ex post budget.

The zonotope order also characterizes the cost comparison for all moral hazard problems when the agent can take any concave utility. This result is also proved in **?**. I provide a geometric proof to it. Specifically, I show that for any two experiments $\mathcal{E}$ and $\mathcal{E}'$, the inclusion of the zonotope is equivalent to the inclusion of the feasible set of state dependent utilities that can be generated with an *ex ante* budget constraint, for any state distribution and any concave utility. From that it is straightforward to see that the zonotope order implies the comparisons of costs.

In addition to the above characterizations, I also provide equivalent conditions for the orders based on matrix factorization as well as posterior belief distributions. Similar to the Blackwell order, the matrix factorization requires $\mathcal{E}' = \mathcal{E}G$, but G may not be a garbling matrix. For the column space order, G can be any matrix. The conic span order requires $G \geq 0$. The zonotope order requires $0 \leq G \leq 1$. As for the posterior belief distributions, the column space order is equivalent to the containment of the affine span of the induced posteriors for any prior. The conic span order is characterized by the containment of the convex hull of the induced posteriors. The zonotope order requires the dominance in linear convex order between the induced posterior distributions.[12]

Lastly, I show that the orders are all strictly nested, with the only exception that when the state space is binary, the zonotope order coincides with the Blackwell order.[13] I also show that the zonotope order coincides with the Blackwell order when the experiments have full rank, and the conic span order coincides with the Blackwell order when the experiments have full column rank.[14]

The rest of the paper is organized as follows. Section 1.1 discusses the related literature. Section 2 introduces the moral hazard model. Section 3 studies the comparisons of information for different classes of moral hazard problems. Section 4 concludes.

1.1 Related Literature

This paper contributes to both the literature on information ordering and moral hazard. The comparisons of information start from Blackwell (**??**), who proposes garbling as a way to compare the value of information across all decision problems. Blackwell's comparison turns out to be very restrictive, especially beyond the case of a binary state space. Follow-up works explore ways to refine the Blackwell order by restricting attention to decision problems that are monotone (**??**), satisfy a single crossing property (**?**), have the interval dominance property (**?**), and binary decision

[12] For distributions $F, G : \mathbb{R}^N \to [0,1]$, F dominates G in the linear convex order, denoted $F \geq_{\text{lcx}} G$, if for any convex function $\phi : \mathbb{R} \to \mathbb{R}$ and any vector $\boldsymbol{\beta} \in \mathbb{R}^N$, $\mathbb{E}_{\boldsymbol{x}\sim F}\left[\phi(\boldsymbol{\beta}\cdot\boldsymbol{x})\right] \geq \mathbb{E}_{\boldsymbol{x}\sim F}\left[\phi(\boldsymbol{\beta}\cdot\boldsymbol{x})\right]$, where $\lambda \cdot x$ represents the dot product.

[13] The equivalence between the zonotope order and the Blackwell order when the state space is binary has been documented since Blackwell (**??**).

[14] For an $N \times M$ matrix $\mathcal{E}$, its rank, denoted $\operatorname{Rank}\mathcal{E}$, is defined as the maximal number of linearly independent columns. Say that $\mathcal{E}$ has full rank if $\operatorname{Rank}\mathcal{E} = \min\{M, N\}$. Say that $\mathcal{E}$ has full column rank if $\operatorname{Rank}\mathcal{E} = M$.

problems (?).[15] I follow the same route by looking at different classes of moral hazard problems.

Among these papers, (?) is the most relevant. He introduces two orders, the posterior mean order and the linear Blackwell order, and show how they sit between the Blackwell and the Lehmann orders. The linear Blackwell order coincides with the zonotope order studied in this paper. He also shows that the zonotope order characterizes the cost comparisons in moral hazard problems when the agent can be arbitrarily risk averse. My contribution, by contrast, is to shift to a geometric lens by pointing out the state dependent utility is the only object that matters for incentives.

This paper also relates to the literature on moral hazard starting from ?, ?, and ?. Since the celebrated informativeness principle by ?, economists have been studying what information is better at providing incentives. ? considers a more general model with an unknown state that affects the utilities in addition to a hidden action. ? extends the informativeness principle to a mean preserving spread order. ? provide an equivalent integral condition. Whereas prior work focuses on likelihood ratio properties, I instead develops the orders with the geometry of state dependent utilities. I also consider different classes of moral hazard problems and provide multiple equivalent characterizations.

2 Moral Hazard Problems

Throughout the paper, I adopt the following notations: Matrices are denoted by uppercase letters, e.g. $\mathcal{E}$; vectors are denoted by boldface lowercase letters, e.g. $\boldsymbol{x} = (x_1, x_2, ..., x_M)$; scalars are denoted by plain lowercase letters, e.g. t; inequality $\boldsymbol{t} \geq a$ means every entry of $\boldsymbol{t}$ has to be above scalar a; boldface $\mathbf{1}$ and $\mathbf{0}$ are vectors of ones and zeros of conformable shapes.

States and Information

Fix a finite set of N states $\Omega = \{\omega_n\}_{n=1}^N$. Information is modeled as finite (Blackwell) experiments. An experiment $\mathcal{E}$ with M realizations is an $N \times M$ row stochastic matrix, that is, every entry is weakly positive $\mathcal{E} \geq 0$, and every row of $\mathcal{E}$ sums to one $\mathcal{E}\mathbf{1} = \mathbf{1}$. Let $Y := \{y_m\}_{m=1}^M$ denote the set of realizations of $\mathcal{E}$. The n-th row of $\mathcal{E}$ represents the state ω_n, the m-th column of $\mathcal{E}$ represents the realization y_m, and the (n, m)-th entry $\mathcal{E}_{n,m} := \mathcal{E}(y_m \mid \omega_n)$ is the conditional probability of realization y_m in state ω_n. Let E^M be the set of all experiments with M realizations. Let $E = \cup_{M=1}^{\infty} E^M$ be the set of all finite experiments. Let $\langle \mathcal{E} \mid \boldsymbol{\mu} \rangle_Y$ denote the distribution of realizations of $\mathcal{E}$ when the prior is $\boldsymbol{\mu} := (\mu_1, \mu_2, ..., \mu_N)$, and let $\langle \mathcal{E} \mid \boldsymbol{\mu} \rangle_\Delta$ represent the distribution of posteriors

[15] Apart from restricting the set of decision problems, there are other ways to refine the Blackwell order. For example, ? consider robustness to additional information that can be arbitrarily correlated with the current information. ? consider taking mulitiple i.i.d. draws from the same experiments.

induced by $\mathcal{E}$ when the prior is $\boldsymbol{\mu}$.[16]

$$\mathcal{E} := \begin{array}{c} \\ \omega_1 \\ \vdots \\ \omega_N \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & \dots & y_M \end{array} \\ \begin{bmatrix} \mathcal{E}_{1,1} & \dots & \mathcal{E}_{1,M} \\ \vdots & \ddots & \vdots \\ \mathcal{E}_{N,1} & \dots & \mathcal{E}_{N,M} \end{bmatrix} \end{array}$$

The Principal

A principal (she) hires an agent (he) to produce certain distribution over the finite state space $\Omega = \{\omega_n\}_{n=1}^N$. The principal wants the agent to produce some (exogenously given) state distribution $\boldsymbol{\mu}_0 \in \Delta\Omega$. The principal does not observe and cannot contract directly on the agent's action. She only observes a noisy experiment $\mathcal{E} \in E^M$ about the states with realizations $Y := \{y_m\}_{m=1}^M$. $\mathcal{E}$ is publicly known and its realization is contractible.

To provide incentives, the principal can offer a contract to the agent. Specifically, a contract under $\mathcal{E}$ specifies a payment rule $\boldsymbol{t} : Y \to \mathbb{R}$ which maps realizations of $\mathcal{E}_P$ to payments.[17] Since the experiment is finite, the contract $\boldsymbol{t}$ can also be viewed as a vector in $\mathbb{R}^M$, each component of which specifies the payment to the agent following a realization of $\mathcal{E}$.

The Agent

The agent can incur a cost to produce any state distribution $\boldsymbol{\mu} \in \Delta\Omega$. His production cost is commonly known and is described by some function $C : \Delta\Omega \to \mathbb{R}_+ \cup \{\infty\}$, where an infinite cost means the state distribution is not feasible. The agent's payoff is additively separable in his utility from money and the production cost. His utility from money $u : \mathbb{R} \to \mathbb{R}$ is strictly increasing, continuous, concave, unbounded, and normalized so that $u(0) = 0$. Let $\mathcal{U}$ be the set of utilities that satisfy the above assumptions. Later I will also consider the special case where the agent is risk neutral with $u(t) = t$.[18] His payoff when producing $\boldsymbol{\mu}$ and receiving payment t is $u(t) - C(\boldsymbol{\mu})$. He maximizes the expected payoff and has an outside option $\underline{u}$.

Given a contract $\boldsymbol{t}$, the agent decides whether to accept it, and if so, what distribution to produce. Formally, the agent's expected payoff from producing $\boldsymbol{\mu} \in \Delta\Omega$ is given by

$$U(\boldsymbol{\mu}; \mathcal{E}, \boldsymbol{t}) := \mathbb{E}_{y_m \sim \langle \mathcal{E} | \boldsymbol{\mu} \rangle_Y} [u(\boldsymbol{t}(y_m))] - C(\boldsymbol{\mu}). \tag{1}$$

[16]While one often writes $\langle \mathcal{E} \mid \boldsymbol{\mu} \rangle$ without distinguishing the distribution of realizations from the distribution of induced posteriors, I need both here for notational clarity. Since we may implement different target state distributions and hence change the prior $\boldsymbol{\mu}$, the posterior attached to each y_m moves as $\boldsymbol{\mu}$ varies. The notation $\langle \mathcal{E} \mid \boldsymbol{\mu} \rangle_Y$ allows me to speak only to the distribution of realizations without worrying about the meaning of each y_m.

[17]It is without loss to assume that the contract is deterministic if the agent is risk neutral or risk averse. In this case, there is no benefit to offer lotteries following any realization of $\mathcal{E}$.

[18]The plain lowercase t represents one payment the agent may receive, and the boldface $\boldsymbol{t}$ represents a payment rule which is a vector in $\mathbb{R}^M$.

The agent then optimally chooses $\boldsymbol{\mu}$ to maximize his expected payoff, provided that this payoff is above his outside option $\underline{u}$.

The production cost C is lower semi-continuous, convex, Gateaux differentiable, and has a free option $\underline{\boldsymbol{\mu}}$ with $C(\underline{\boldsymbol{\mu}}) = 0$. Lower semi-continuity guarantees the existence of a solution in the agent's problem. Convexity and the existence of the free option are without loss.[19] Gateaux differentiability requires more explanation. C being Gateaux differentiable means that it admits a function $\boldsymbol{\nabla} C : \Delta\Omega \to \bar{\mathbb{R}}^N$ where $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ such that, for any $\boldsymbol{\mu}, \boldsymbol{\mu}' \in \Delta\Omega$,

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \left[C\left(\boldsymbol{\mu} + \epsilon(\boldsymbol{\mu}' - \boldsymbol{\mu})\right) - C\left(\boldsymbol{\mu}\right) \right] = (\boldsymbol{\mu}' - \boldsymbol{\mu}) \cdot \boldsymbol{\nabla} C(\boldsymbol{\mu}). \tag{2}$$

The function $\boldsymbol{\nabla} C$ is called a (Gateaux) derivative of C. Equation (2) says that the cost of any marginal change in production can be priced linearly with weights $\boldsymbol{\nabla} C(\boldsymbol{\mu})$. The n-th component of $\boldsymbol{\nabla} C(\boldsymbol{\mu})$ has the interpretation of the marginal cost to increase the probability of state ω_n. This is the standard notion of differentiability if we view C as a function from $\mathbb{R}^{N-1}$ to $\mathbb{R}_+ \cup \{\infty\}$.[20]

Lastly, let $\mathcal{C}$ be the set of all cost functions that satisfy the assumptions above.

The Contracting Problem

A moral hazard environment is a tuple $P := (\boldsymbol{\mu}_0, u, C)$ where $\boldsymbol{\mu}_0 \in \Delta\Omega$ is the target state distribution, $u \in \mathcal{U}$ is the agent's utility for money, and $C \in \mathcal{C}$ is the production cost.[21] Given an enviroment P and information $\mathcal{E}$, the principal chooses a contract to minimize the expected cost subject to the agent's incentive constraint (IC) and participation constraint (PC),[22]

$$\min_{\boldsymbol{t}} \mathbb{E}_{y_m \sim \langle \mathcal{E}_P | \boldsymbol{\mu} \rangle_Y} \left[\boldsymbol{t}(y_m) \right], \tag{3}$$

$$\text{s.t.(IC)}\ \boldsymbol{\mu}_0 \in \underset{\boldsymbol{\mu} \in \Delta\Omega}{\operatorname{argmax}}\, U\left(\boldsymbol{\mu}; \mathcal{E}, \boldsymbol{t}\right), \tag{4}$$

$$\text{(PC)}\ \max_{\boldsymbol{\mu} \in \Delta\Omega} U\left(\boldsymbol{\mu}; \mathcal{E}, \boldsymbol{t}\right) \geq \underline{u}, \tag{5}$$

[19] Convexity is without loss because the agent may randomize. The cost of some $\boldsymbol{\mu}$ is the cheapest expected cost from randomizing across all state distributions to generate $\boldsymbol{\mu}$, and the resulting cost function must be convex. The existence of free option is a normalization because, if the cheapest option $\underline{\boldsymbol{\mu}}$ (which must exist due to lower semi-continuity) has a strictly positive cost $\underline{c}$, we can redefine the cost function as $C(\boldsymbol{\mu}) - \underline{c}$ and change the outside option to $\underline{u} + \underline{c}$.

[20] The derivative $\boldsymbol{\nabla} C(\boldsymbol{\mu})$ differs from the usual partial derivatives only by a normalizing constant. Equation (2) only defines the derivative up to a constant. If $\boldsymbol{\nabla} C(\boldsymbol{\mu})$ satisfies Equation (2), then so does $\boldsymbol{\nabla} C(\boldsymbol{\mu}) + k$ for any $k \in \mathbb{R}$, because the constant k cancels out as probabilities must sum to one. If we view C as a function defined over $\mathbb{R}^{N-1}$, its partial derivative with respect to the n-th component is the marginal cost of increasing the probability of ω_n while decreasing the probability of ω_N. The partial derivatives are indeed $\boldsymbol{\nabla} C(\boldsymbol{\mu}) + k$ where its N-th component is normalized to zero.

[21] The agent's outside option $\underline{u}$ is not included for brevity. The principal can always solve the moral hazard problem assuming $\underline{u} = 0$, and then provide the agent with a lump-sum payment that equals his outside option. Since $\mathcal{E}$ does not affect $\underline{u}$, it does not affect the comparisons of information.

[22] A solution always exists as long as the problem is feasible. The existence does not require the limited liability constraint due to the finiteness of the state space and the principal's experiment.

where the incentive constraint requires that $\boldsymbol{\mu}_0$ is an optimal choice for the agent, and the participation constraint requires that the agent does at least as good as his outside option.[23] I also consider two additional constraints faced by the principal: limited liability (LL), which requires the payments to be non-negative, and ex post budget (B), which bounds the payments from above by some constant $B > 0$,

$$\text{(LL) } \boldsymbol{t} \geq 0, \tag{6}$$

$$\text{(B) } \boldsymbol{t} \leq B. \tag{7}$$

The main exercise of the paper is to compare experiments across different classes of moral hazard environments in terms of the cost to the principal. To that end, I introduce the following definitions. Let $K(\mathcal{E}; P)$ be the indirect cost from problem (3) where the principal has experiment $\mathcal{E}$ and faces environment P. In addition, let $K^{\text{LL}}(\mathcal{E}; P)$ be the indirect cost from problem (3) with the limited liability constraint (6), and $K^{\text{LL,B}}(\mathcal{E}; P)$ be the indirect cost from problem (3) with both limited liability (6) and ex post budget (7). Finally, let $\mathcal{P} := \{(\boldsymbol{\mu}_0, u, C) : \boldsymbol{\mu}_0 \in \Delta\Omega, u \in \mathcal{U}, C \in \mathcal{C}\}$ be the collection of all moral hazard environments, and $\mathcal{P}^{\text{RN}} := \{P \in \mathcal{P} : u(t) = t\}$ be the environments where the agent is risk neutral.

As a final note, the problem (3) is not always feasible because experiment $\mathcal{E}$ may not contain the correct information to satisfy the agent's incentive constraint. A simple example is when $\mathcal{E}$ contains no information, in which case only the free option $\underline{\boldsymbol{\mu}}$ can be implemented. When the problem is infeasible, I adopt the convention to write $K(\mathcal{E}; P) = \infty$. The theorems in the next section involve comparing $K(\mathcal{E}; P)$ against $K(\mathcal{E}'; P)$ for a class of environments P. When both $K(\mathcal{E}; P)$ and $K(\mathcal{E}'; P)$ are infinity, I adopt the convention that $K(\mathcal{E}; P) \leq K(\mathcal{E}'; P)$ holds as it simplifies the notation.

3 Comparisons of Experiments

In this section I study three orders on experiments to compare information in moral hazard problems, each defined by the inclusion of polyhedral sets of state dependent utilities that can be generated by the experiment using different sets of payment rule. I will also provide equivalent characterizations using matrix factorization and posterior beliefs.

The reason to focus on the state dependent utility is that it completely determines the agent's incentives. To see this, recall that the agent chooses a state distribution to maximize (1), trading off his expected utility from payments against the production cost. The production cost is out of the principal's control. She can only affect the agent's expected utility through her contract

[23] In case of indifference, I assume the agent always breaks ties in favor of the principal. The incentive constraint is hence weak, in the sense that it suffices for $\boldsymbol{\mu}_0$ to be an optimal choice, rather than the unique optimal choice. For unique optimality, I need the cost function C to be strictly convex in a neighborhood of $\boldsymbol{\mu}_0$. In this case, $\boldsymbol{\mu}_0$ is an optimal choice if and only if it is the unique optimal choice in the agent's problem.

t coupled with her information $\mathcal{E}$. The marginal benefit of making ω_n more likely is the agent's expected utility in state ω_n. Given his expected utility in every state, the agent then chooses a state distribution to produce. The moral hazard problem thus reduces to choosing a state dependent utility feasible under $\mathcal{E}$.

Formally, let $v_m := u(t_m)$ be the utility from t_m, and $\boldsymbol{v} := [v_m]_{m=1}^M \in \mathbb{R}^M$ be the vector of utilities across all possible payments in $\boldsymbol{t}$. The agent's state dependent utility given experiment $\mathcal{E}$ and contract $\boldsymbol{t}$ is a vector $\mathcal{E}v$, where its n-th component $\sum_{m=1}^M \mathcal{E}_{n,m}v_m$ is the expected utility in state ω_n.[24]

I will define the orders based on the inclusion of the convex polyhedral sets of feasible state dependent utility vectors $\mathcal{E}v$.[25]

3.1 Column Space Order

Recall that an experiment $\mathcal{E}$ is $N \times M$ row stochastic matrix. Its column space is defined as $\operatorname{Col}\mathcal{E} := \left\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M\right\}$. For any utility $u \in \mathcal{U}$, $\operatorname{Col}\mathcal{E}$ is the set of all state dependent utilities that can be generated with experiment $\mathcal{E}$ using any payment rule.[26] $\operatorname{Col}\mathcal{E}$ captures the full set of state dependent utilities the principal can choose from.

The column space order is defined as the inclusion of the column space. Formally, say that $\mathcal{E}$ dominates $\mathcal{E}'$ in the column space order, denoted $\mathcal{E} \geq_{\text{Col}} \mathcal{E}'$, if $\operatorname{Col}\mathcal{E} \supseteq \operatorname{Col}\mathcal{E}'$.

The next theorem summarizes the results on the column space order. In addition to the characterization with moral hazard problems, I also provide equivalent characterizations with matrix factorization as well as posterior beliefs. To state the theorem, I need to introduce some notations. Recall that the affine span of a set $A \subseteq \mathbb{R}^N$ is defined as $\operatorname{Aff} A = \{\sum w_i \boldsymbol{a}_i : w_i \in \mathbb{R}, \sum w_i = 1, \boldsymbol{a}_i \in A, \forall i\}$, which is the set of all affine combinations of elements in A. For any experiment $\mathcal{E}$, utility function $u \in \mathcal{U}$, and cost function $C \in \mathcal{C}$, define the implementable set of state distributions as

$$\mathcal{F}(\mathcal{E}; u, C) := \left\{\boldsymbol{\mu} \in \Delta\Omega : K\left(\mathcal{E}', (\boldsymbol{\mu}, u, C)\right) < \infty\right\}.$$

This is the set of distributions the principal is able to implement with $\mathcal{E}$. Consider any experiments $\mathcal{E} \in E^M$ and $\mathcal{E}' \in E^{M'}$.

Theorem 1. *The following are equivalent:*

(1) $\mathcal{E} \geq_{Col} \mathcal{E}'$,

[24] Recall that $\mathcal{E}$ is an $N \times M$ row stochastic matrix, and v is a vector of size M. The resulting vector $\mathcal{E}v$ is of size N, and the n-th component is the expected utility $\sum_{m=1}^M \mathcal{E}_{n,m}v_m$ in state ω_n.

[25] A subset of $\mathbb{R}^N$ is polyhedral if it can be expressed as the intersection of finitely many closed half spaces. In particular, for any matrix $\mathcal{E}$, its column space, conic span, and zonotope are all polyhedral sets (see definitions in Sections 3.1-3.3).

[26] The utility function u is assumed to be unbounded. Therefore, any $\boldsymbol{v} \in \mathbb{R}^M$ can be generated with some payment rule $\boldsymbol{t} \in \mathbb{R}^M$.

(2) $\mathcal{E}' = \mathcal{E}G$ *for some matrix* G*,*

(3) $\text{Aff Supp}\,\langle \mathcal{E} \mid \boldsymbol{\mu}_0 \rangle_\Delta \supseteq \text{Aff Supp}\,\langle \mathcal{E}' \mid \boldsymbol{\mu}_0 \rangle_\Delta$ *for any prior* $\boldsymbol{\mu}_0 \in \Delta\Omega$*,*[27]

(4) $K(\mathcal{E}; P) \leq K(\mathcal{E}'; P)$ *for any* $P \in \mathcal{P}^{RN}$*,*

(5) $\mathcal{F}(\mathcal{E}'; u, C) \subseteq \mathcal{F}(\mathcal{E}; u, C)$ *for any* $u \in \mathcal{U}, C \in \mathcal{C}$*.*

I now explain the theorem and provide a brief proof sketch.

(2) characterizes the column space order in terms of matrix factorization. Unlike Blackwell, the column space order only requires the existence of some matrix factor G with no additional requirements.[28] In some sense, the column space order compares the qualitative information contained in experiments. (1) $\Rightarrow$ (2) follows from the fact that each column of $\mathcal{E}'$ is in $\operatorname{Col} \mathcal{E}' \subseteq \operatorname{Col} \mathcal{E}$. (2) $\Rightarrow$ (1) follows from checking the definition.

(3) characterizes the column space order in terms of posterior beliefs. Instead of the posterior distributions, it only compares the affine span of the induced posteriors for any prior. For (2) $\Rightarrow$ (3), use Bayes' rule to express the induced posteriors and plug in $\mathcal{E}' = \mathcal{E}G$. For (3) $\Rightarrow$ (2), pick μ_0 to be the uniform prior and the induced posteriors are proportional to columns of $\mathcal{E}$. The affine span inclusion then gives (2).

(4) and (5) are characterizations based on moral hazard problems. (4) says the order characterizes the comparisons of implementability in all moral hazard problems. Intuitively, for every state dependent utility, there is an optimal state distribution for the agent. Therefore, a larger set of state dependent utilities makes more state distributions implementable. (5) says the order also characterizes the comparisons of costs in all moral hazard problems with a risk neutral agent but without limited liability.[29] That is, $\mathcal{E} \geq_{\text{Col}} \mathcal{E}'$ if and only if the cost is always lower under $\mathcal{E}$ in all such problems. This is again because the principal can choose from a larger set of state dependent utilities under $\mathcal{E}$ than $\mathcal{E}'$ for her cost minimization problem.[30]

The proof for (1) $\Rightarrow$ (5) simply follows from an argument of expanding the principal's choice set. (5) $\Rightarrow$ (4) is obvious. The converse (4) $\Rightarrow$ (1) can be proved with a constructive approach. Roughly speaking, if $\mathcal{E} \not\geq_{\text{Col}} \mathcal{E}'$, there exists some vector $\boldsymbol{x} \in \operatorname{Col} \mathcal{E}'_P \setminus \operatorname{Col} \mathcal{E}_P$. I can then use $\boldsymbol{x}$ to construct a state distribution $\boldsymbol{\mu}_0 \in \Delta\Omega$, a utility $u \in \mathcal{U}$, and a cost function $C \in \mathcal{C}$ so that $\boldsymbol{\mu}_0$ is implementable under $\mathcal{E}'$ but not $\mathcal{E}$.

[27] An equivalent way to state (3) is that the same inclusion holds for some interior prior μ_0. In the appendix, I provide the proof for both statements.

[28] Recall that the Blackwell order requires G to be a garbling matrix, that is, $G \geq 0$ and $G\mathbf{1} = \mathbf{1}$.

[29] I focus on the principal's cost minimization problem to implement a given state distribution. For her profit maximization problem, a sell-the-firm contract always achieves the first best when the agent is risk neutral and not protected by limited liability, and the principal's information is irrelevant.

[30] Alternatively, this also follows immediately from the comparison of implementability. In this case, if the target distribution $\boldsymbol{\mu}_0$ is implementable, the principal can always implement it at the first best cost, leaving zero rent to the agent. This means the cost to implement any μ_0 is its first best cost, independent of the principal's information $\mathcal{E}$. Therefore, the comparison of costs in this case reduces to the comparison of implementability. The other direction is not trivial and requires a constructive proof through the column space.

3.2 Conic Span Order

The conic span of an experiment $\mathcal{E}$ is defined as $\operatorname{Cone}\mathcal{E} := \left\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, \boldsymbol{v} \geq 0\right\}$. For any utility $u \in \mathcal{U}$, $\operatorname{Cone}\mathcal{E}$ is the set of all state dependent utilities that can be generated with experiment $\mathcal{E}$ using any non-negative payment rule.[31] This is the full set of state dependent utilities the principal can choose from when there is a limited liability constraint.

The conic span order is defined as the inclusion of the conic span. Formally, say that $\mathcal{E}$ dominates $\mathcal{E}'$ in the conic span order, denoted $\mathcal{E} \geq_{\text{Cone}} \mathcal{E}'$, if $\operatorname{Cone}\mathcal{E} \supseteq \operatorname{Cone}\mathcal{E}'$.

The next theorem summarizes the results on the conic span order. Recall that the convex hull of a set $A \subseteq \mathbb{R}^N$ is defined as $\operatorname{Co} A = \{\sum w_i \boldsymbol{a}_i : w_i \geq 0, \sum w_i = 1, \boldsymbol{a}_i \in A, \forall i\}$, which is the set of all convex combinations of elements in A. Consider any experiments $\mathcal{E} \in E^M$ and $\mathcal{E}' \in E^{M'}$.

Theorem 2. *The following are equivalent:*

(1) $\mathcal{E} \geq_{Cone} \mathcal{E}'$,

(2) $\mathcal{E}' = \mathcal{E}G$ *for some matrix* $G \geq 0$,

(3) $\operatorname{Co}\operatorname{Supp}\langle \mathcal{E} \mid \boldsymbol{\mu}_0\rangle_\Delta \supseteq \operatorname{Co}\operatorname{Supp}\langle \mathcal{E}' \mid \boldsymbol{\mu}_0\rangle_\Delta$ *for any prior* $\boldsymbol{\mu}_0 \in \Delta\Omega$,

(4) $K^{LL}(\mathcal{E}; P) \leq K^{LL}(\mathcal{E}'; P)$ *for any* $P \in \mathcal{P}^{RN}$.

(2) says that, for the conic span order, the matrix factorization requires a non-negative factor G.[32] Observing that each column of $\mathcal{E}'$ is in the conic span $\operatorname{Cone}\mathcal{E}'$, (1) $\Leftrightarrow$ (2) follows from checking the definition.

(3) says that the conic span order is characterized by the containment of the convex hull of the induced posteriors for any prior. In words, the conic span order only compares how extreme the extremal beliefs induced by the experiments are, without referring to how often these beliefs are induced. To prove (2) $\Rightarrow$ (3), write down the induced posteriors using Bayes' rule and apply $\mathcal{E}' = \mathcal{E}G$. For (3) $\Rightarrow$ (2), simply let $\boldsymbol{\mu}_0$ be the uniform prior, and the inclusion of the convex hull implies the existence of matrix factor $G \geq 0$.

(4) says that the conic span order characterizes the comparisons of costs in all moral hazard problems with risk neutrality and limited liability. Limited liability is built into the definition of the conic span. Risk neutrality makes the agent's expected utility from money equal his expected payment. Hence, to minimize the cost the principal simply minimizes the agent's utility from money. A larger conic span provides a larger set of state dependent utilities to choose from, weakly reducing the cost. This argument relies crucially on risk neutrality: With a risk averse agent, lowering expected

[31] Recall that u is strictly increasing and is normalized so that $u(0) = 0$. Therefore any $\boldsymbol{v} \geq 0$ can be generated with some non-negative payment rule $\boldsymbol{t} \geq 0$.

[32] (2) requires that there exists a matrix $G \geq 0$ such that $\mathcal{E}' = \mathcal{E}G$. In particular, it does not require $G \geq 0$ for any G with $\mathcal{E}' = \mathcal{E}G$. When $\mathcal{E}$ has deficient rank, there can be many factors G that satisfy $\mathcal{E}' = \mathcal{E}G$. This condition only requires one of them to be non-negative.

utility does not guarantee a lower expected payment.[33] For the proof, (1) $\Rightarrow$ (4) follows from a larger choice set. (4) $\Rightarrow$ (1) again comes from a constructive approach.

As a final remark, I also show in the appendix that the conic span order characterizes the comparisons of costs in all moral hazard problems with risk neutrality and the ex post budget constraint. Intuitively, the set of state dependent utilities in this case is a constant minus the conic span of $\mathcal{E}$, so the same proof applies.[34]

3.3 Zonotope Order

The zonotope of an experiment $\mathcal{E}$ is defined as $\operatorname{Zon}\mathcal{E} := \left\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, 0 \le \boldsymbol{v} \le 1\right\}$. For any utility $u \in \mathcal{U}$, $\operatorname{Zon}\mathcal{E}$ is the set of all state dependent utilities that can be generated with experiment $\mathcal{E}$ using any non-negative and bounded payment rule.[35] This is the full set of state dependent utilities the principal can choose from when she is constraint both by limited liability and ex post budget.

The zonotope order is defined as the inclusion of the zonotope. Formally, say that $\mathcal{E}$ dominates $\mathcal{E}'$ in the zonotope order, denoted $\mathcal{E} \geq_{\text{Zon}} \mathcal{E}'$, if $\operatorname{Zon}\mathcal{E} \supseteq \operatorname{Zon}\mathcal{E}'$.

The next theorem summarizes the results on the zonotope order. The equivalence between (3) and (5) also appear in **?**. Here I provide a geometric argument in the proof. To state the theorem, recall that for distributions $F, G : \mathbb{R}^N \to [0,1]$, F dominates G in the linear convex order, denoted $F \geq_{\text{lcx}} G$, if for any convex function $\phi : \mathbb{R} \to \mathbb{R}$ and any vector $\boldsymbol{\beta} \in \mathbb{R}^N$, $\mathbb{E}_{\boldsymbol{x}\sim F}\left[\phi(\boldsymbol{\beta}\cdot\boldsymbol{x})\right] \geq \mathbb{E}_{\boldsymbol{x}\sim F}\left[\phi(\boldsymbol{\beta}\cdot\boldsymbol{x})\right]$. Consider any experiments $\mathcal{E} \in E^M$ and $\mathcal{E}' \in E^{M'}$.

Theorem 3. *The following are equivalent:*

(1) $\mathcal{E} \geq_{Zon} \mathcal{E}'$,

(2) $\mathcal{E}' B_N = \mathcal{E} G_N$ *for some matrix* $G_N \in \mathbb{R}^{N\times 2^N}$ *with* $0 \le G_N \le 1$, *where* $B_N := \begin{bmatrix} \boldsymbol{v}^{(1)} & \boldsymbol{v}^{(2)} & \dots & \boldsymbol{v}^{(2^N)} \end{bmatrix}$ *with* $\{\boldsymbol{v}^{(1)}, \dots, \boldsymbol{v}^{(2^N)}\} = \{0,1\}^N$ *being the set of all binary vectors.*

(3) $\langle \mathcal{E} \mid \boldsymbol{\mu}_0 \rangle_\Delta \geq_{lcx} \langle \mathcal{E}' \mid \boldsymbol{\mu}_0 \rangle_\Delta$ *for any prior* $\boldsymbol{\mu}_0 \in \Delta\Omega$,

(4) $K^{LL,B}(\mathcal{E}; P) \le K^{LL,B}(\mathcal{E}'; P)$ *for any* $P \in \mathcal{P}^{RN}$.

(5) $K^{LL}(\mathcal{E}; P) \le K^{LL}(\mathcal{E}'; P)$ *for any* $P \in \mathcal{P}$.

(2) requires that partial sum of columns in $\mathcal{E}'$ must be a linear combination of $\mathcal{E}$ with coefficients between 0 and 1. This necessarily implies that $\mathcal{E}' = \mathcal{E}G$ for some $0 \le G \le 1$. (3) says the zonotope order is equivalent to dominance in the linear convex order of the induced posterior distributions.

[33] To see this, suppose the agent has utility $u(t) = \sqrt{t}$. If the agent gets paid 1 with probability 0.6, his utility is 0.6 and the expected payment is 0.6. If the agent gets paid 4 with probability 0.25, his utility is $0.5 < 0.6$, but the expected payment is $1 > 0.6$.

[34] Specifically, if the principal's payment is bounded from above, she can only generate state dependent utilities in $\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \le 1\}$, which reduces to $\mathbf{1} - \operatorname{Cone}\mathcal{E}$, observing that $\mathcal{E}\mathbf{1} = \mathbf{1}$.

[35] For any upper bound $B > 0$, the set of state dependent utilities from any $0 \le \boldsymbol{t} \le B$ is given by $\left\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^M, 0 \le \boldsymbol{v} \le u(B)\right\} = u(B) \cdot \operatorname{Zon}\mathcal{E}$, which is the zonotope scaled by a constant $u(B)$.

(4) says the zonotope order characterizes the comparisons of costs in all moral hazard problems where the agent is risk neutral and protected by limited liability, and the principal is subject to an ex post budget constraint. Again, due to risk neutrality, the principal can focus on minimizing the agent's expected utility from money. (1) $\Rightarrow$ (4) because a larger zonotope means the principal has a larger set of feasible state dependent utilities, hence lowering the cost. For (4) $\Rightarrow$ (1) I have another constructive proof in the appendix.

(5) says the zonotope order also characterizes the cost comparisons in moral hazard problems with arbitrary risk aversion and limited liability. In fact, if the agent can take any utility $u \in \mathcal{U}$, whether the principal is subject to limited liability or ex post budget is inconsequential. I show in the appendix that zonotope dominance is equivalent to lower costs in all moral hazard problems without limited liability, or with ex post budget, or with both constraints. I present the result with limited liability for ease of exposition.

I will use a geometric approach to illustrate the intuition behind (1) $\Leftrightarrow$ (5) and sketch the proof. Risk aversion breaks the previous argument: Although a larger set of feasible state dependent utilities still lets the principal drive down the agent's expected utility, the expected utility no longer equals expected payment, and reducing utility need not lower cost.

To tether utilities back to expected payments, abusing the notation to denote $u(\boldsymbol{t}) := [u(t_m)]_{m=1}^M$ define

$$\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}) := \left\{\mathcal{E}\boldsymbol{v} : \exists \boldsymbol{t} \in \mathbb{R}^M \text{ such that } \boldsymbol{v} = u(\boldsymbol{t}), \boldsymbol{t} \geq 0, \boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \leq 1\right\},$$

where $u \in \mathcal{U}$ is a utility for money, and $\boldsymbol{\mu}_0 \in \Delta\Omega$ is the reference state distribution at which we compute the expected payment $\boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t}$. In words, $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$ is the set of state dependent utilities that can be generated with an *ex ante* budget normalized to one. Due to the concavity of u, $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$ may no longer be polyhedral, but it is still convex.[36]

(1) $\Leftrightarrow$ (5) will be established via two intermediate equivalences, stated in the following two lemmas.

Lemma 1. *$K^{LL}(\mathcal{E}; P) \leq K^{LL}(\mathcal{E}'; P)$ for any $P \in \mathcal{P}$ if and only if $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}) \supseteq \mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}')$ for any $\boldsymbol{\mu}_0 \in \Delta\Omega$ and $u \in \mathcal{U}$.*

Lemma 2. *$\mathcal{E} \geq_{Zon} \mathcal{E}'$ if and only if $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}) \supseteq \mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}')$ for any $\boldsymbol{\mu}_0 \in \Delta\Omega$ and $u \in \mathcal{U}$.*

Lemma 1 says that experiment $\mathcal{E}$ leads to lower costs in every moral hazard problem than $\mathcal{E}'$ if and only if it generates a larger set of feasible state dependent utilities with an ex ante budget. Thus, we can dispense with the agent's cost function C, which only appears as part of an environment $P := (\boldsymbol{\mu}_0, u, C)$, and direct our attention to the feasible state-dependent utilities.

Lemma 1 is obtained by turning the cost minimization problem (3) into a feasibility problem. Take any environment P. The value of the principal's problem is the minimum value of $B \geq 0$ under which

[36] Convexity is proved by checking the definition and realizing that $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$ satisfies a free disposal above zero property, that is, if some $\mathcal{E}\boldsymbol{v}$ is contained in $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$, then so is any $\mathcal{E}\boldsymbol{v}'$ with $0 \leq \boldsymbol{v}' \leq \boldsymbol{v}$. The detail is in the appendix.

the principal's problem is still feasible under an additional ex ante budget constraint $\boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \leq B$. If $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}) \supseteq \mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}')$, then the feasible set is always larger under $\mathcal{E}$ than $\mathcal{E}'$ in this new problem, which implies $K^{\mathrm{LL}}(\mathcal{E}; P) \leq K^{\mathrm{LL}}(\mathcal{E}'; P)$.[37] For the "only if" direction, consider the contrapositive. If the inclusion of $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}$ fails, one can find some state dependent utility that is cheaper under $\mathcal{E}'$ than $\mathcal{E}$, and a moral hazard problem can be constructed to violate the cost comparison.

Lemma 2 says that the containment of $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}$ for any $\boldsymbol{\mu}_0 \in \Delta\Omega$ and $u \in \mathcal{U}$ is equivalent to the zonotope order. The "if" direction follows from taking a specific utility $\bar{u}(t) = \min\{t, 1\}$ and payment rules $\boldsymbol{t} = \boldsymbol{e}_m$ which pays 1 when y_m realizes and 0 otherwise.[38] The "only if" direction uses two additional math facts. Say that a vector $\boldsymbol{x} \in \mathbb{R}^M$ majorizes $\boldsymbol{z} \in \mathbb{R}^{M'}$ if[39]

$$\max_{\boldsymbol{v} \in \mathbb{R}^M, 0 \leq \boldsymbol{v} \leq 1} \boldsymbol{x} \cdot \boldsymbol{v} \geq \max_{\boldsymbol{v}' \in \mathbb{R}^{M'}, 0 \leq \boldsymbol{v}' \leq 1} \boldsymbol{z} \cdot \boldsymbol{v}'.$$

The first fact is the Karamata's majorization inequality (see **?** Theorem 1.A.3), and the second fact is an application of supporting hyperplane theorem (**?**; see **?** Corollary 2.B.3).

Fact 1. *$\boldsymbol{x}$ majorizes $\boldsymbol{z}$ if and only if for any convex function $\phi : \mathbb{R} \to \mathbb{R}$,*

$$\sum_{m=1}^{M} \phi(x_m) \geq \sum_{m'=1}^{M'} \phi(z_{m'}). \tag{8}$$

Fact 2. *A vector $\boldsymbol{x} \in \mathbb{R}^N$ is in a convex set $\mathcal{V}$ if and only if, for every $\boldsymbol{\beta} \in \mathbb{R}^N$,*

$$\boldsymbol{\beta} \cdot \boldsymbol{x} \leq \max_{\boldsymbol{v} \in \mathcal{V}} \boldsymbol{\beta} \cdot \boldsymbol{v}. \tag{9}$$

Apply Fact 2 to the inclusion of the zonotope, we have for any $\boldsymbol{\beta} \in \mathbb{R}^N$,

$$\max_{0 \leq \boldsymbol{v} \leq 1} \boldsymbol{\beta} \cdot \mathcal{E}\boldsymbol{v} \geq \max_{0 \leq \boldsymbol{v} \leq 1} \boldsymbol{\beta} \cdot \mathcal{E}'\boldsymbol{v}. \tag{10}$$

The above inequality simply says that $\boldsymbol{\beta} \cdot \mathcal{E}$ majorizes $\boldsymbol{\beta} \cdot \mathcal{E}'$ for any $\boldsymbol{\beta} \in \mathbb{R}^N$.[40]

[37] The ex ante budget is normalized to one in the definition of $\mathcal{V}_{\boldsymbol{\mu}_0,u}$, but one can show that (see appendix) $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}) \supseteq \mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E}')$ if and only if $\mathcal{V}^+_{\boldsymbol{\mu}_0,u,B}(\mathcal{E}) \supseteq \mathcal{V}^+_{\boldsymbol{\mu}_0,u,B}(\mathcal{E}')$ for any $B \geq 0$ where

$$\mathcal{V}^+_{\boldsymbol{\mu}_0,u,B}(\mathcal{E}) := \left\{ \mathcal{E}\boldsymbol{v} : \exists \boldsymbol{t} \in \mathbb{R}^M \text{ such that } \boldsymbol{v} = u(\boldsymbol{t}), \boldsymbol{t} \geq 0, \boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \leq B \right\}.$$

[38] Technically speaking, $u \in \mathcal{U}$ is required to be strictly increasing. I need to use a sequence $\bar{u}_n(t) = \min\left\{t, \frac{1}{n}(t-1) + 1\right\}$ which is strictly increasing and converges pointwise to $\bar{u}$.

[39] Equivalently (and more commonly), majorization is defined as follows. Order the entries of x in descending order as $x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[M]}$ where $x_{[m]}$ denotes the m-th largest component of $\boldsymbol{x}$. Say that $\boldsymbol{x}$ majorizes $\boldsymbol{z}$ if $\sum_{i=1}^{k} x_{[i]} \geq \sum_{i=1}^{k} z_{[i]}, \forall 1 \leq k \leq \min\{M, M'\}$ with equality at $k = \min\{M, M'\}$. From **?**, this is equivalent to the definition given above.

[40] $\boldsymbol{\beta} \cdot \mathcal{E}$ is a vector given by $(\boldsymbol{\beta} \cdot \mathcal{E})_n := \sum_{m=1}^{M} \beta_m \mathcal{E}_{n,m}$. It is the usual product $\boldsymbol{\beta}^\intercal \mathcal{E}$ written without a transpose symbol.

Lemma 2 translates the inclusion of $\mathcal{V}^{+}_{\boldsymbol{\mu}_0,u}$ into the following comparisons of values of the following optimization problem: for every $\boldsymbol{\beta} \in \mathbb{R}^N$, $u \in \mathcal{U}$, and $\boldsymbol{\mu}_0 \in \Delta\Omega$,

$$\max_{\boldsymbol{t}\in\mathbb{R}^M_+, \boldsymbol{\mu}_0\cdot\mathcal{E}\boldsymbol{t}\leq 1} \boldsymbol{\beta}\cdot\mathcal{E}u(\boldsymbol{t}) \geq \max_{\boldsymbol{t}\in\mathbb{R}^{M'}_+, \boldsymbol{\mu}_0\cdot\mathcal{E}'\boldsymbol{t}\leq 1} \boldsymbol{\beta}\cdot\mathcal{E}'u(\boldsymbol{t}). \tag{11}$$

(11) holds if

$$\max_{\boldsymbol{t}\geq 0} \boldsymbol{\beta}\cdot\mathcal{E}u(\boldsymbol{t}) - \boldsymbol{\mu}_0\cdot\mathcal{E}\boldsymbol{t} \geq \max_{\boldsymbol{t}\geq 0} \boldsymbol{\beta}\cdot\mathcal{E}'u(\boldsymbol{t}) - \boldsymbol{\mu}_0\cdot\mathcal{E}'\boldsymbol{t}, \tag{12}$$

where we incorporate the ex ante budget constraint with a multiplier.[41] The expression in (12) is a convex function of $(\boldsymbol{\beta}\cdot\mathcal{E}, \boldsymbol{\mu}_0\cdot\mathcal{E})$ because it is the maximum over linear functions. (12) then follows from Fact 1.

3.4 Relations Between the Orders

I now discuss the relations between the orders.

First of all, they are nested.

Proposition 1. $\geq_{Col} \supsetneq \geq_{Cone} \supsetneq \geq_{Zon} \supseteq \geq_B$.

The inclusion is the easiest to see using the matrix factorization condition. The strictness can be shown with the following examples.

Example 1. Suppose $N = 2$. Consider the following experiments.

$$\mathcal{E} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix};\ \mathcal{E}' = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.1 & 0.4 & 0.5 \end{bmatrix};\ \mathcal{E}'' = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}.$$

We have $\mathcal{E} \geq_{\text{Col}} \mathcal{E}'$ but $\mathcal{E} \not\geq_{\text{Cone}} \mathcal{E}'$, and $\mathcal{E}' \geq_{\text{Cone}} \mathcal{E}''$ but $\mathcal{E}' \not\geq_{\text{Zon}} \mathcal{E}''$.

The zonotope order coincides with Blackwell when the state space is binary. This is a known result due to **?**.

Proposition 2. *When $N = 2$, $\geq_{Zon} = \geq_B$. When $N > 2$, $\geq_{Zon} \supsetneq \geq_B$.*

The conic span order may also coincide with Blackwell if the experiments have full column rank. Full column rank implies that the induced posteriors are linearly independent, assuming any interior prior. This means there is no redundant realizations. Blackwell can be restored because experiments are required to be row stochastic, which places some implicit requirements on the matrix factor G.[42]

Proposition 3. *Suppose $\mathcal{E}$ has full column rank. $\mathcal{E} \geq_{Cone} \mathcal{E}'$ if and only if $\mathcal{E} \geq_B \mathcal{E}'$.*

[41] Specifically, let $\eta \in \mathbb{R}$ be the multiplier on the ex ante budget constraint. Fix any $\beta, u, \boldsymbol{\mu}_0$, the Lagrangian is $\mathcal{L}(\boldsymbol{t}, \eta; \mathcal{E}) := \boldsymbol{\beta}\cdot\mathcal{E}u(\boldsymbol{t}) - \eta\boldsymbol{\mu}_0\cdot\mathcal{E}\boldsymbol{t} + \eta$. To show (11), it suffices to show that $\mathcal{L}(\boldsymbol{t}, \eta; \mathcal{E}) \geq \mathcal{L}(\boldsymbol{t}, \eta; \mathcal{E}')$ for any η. To obtain the above expression, we divide η on both sides sides and redefine $\boldsymbol{\beta}$ as $\boldsymbol{\beta}/\eta$ since $\boldsymbol{\beta}$ can take any value.

[42] Suppose $\mathcal{E} \geq_{\text{Cone}} \mathcal{E}'$. $\mathcal{E}' = \mathcal{E}G$ for some $G \geq 0$. Due to row stochasticity of $\mathcal{E}$ and $\mathcal{E}'$, we have $\mathcal{E}\mathbf{1} = 1 = \mathcal{E}'\mathbf{1} = \mathcal{E}G\mathbf{1}$, which implies $\mathcal{E}(G\mathbf{1} - \mathbf{1}) = \mathbf{0}$. When $\mathcal{E}$ has full column rank, this necessarily implies $G\mathbf{1} = \mathbf{1}$.

Lastly, I briefly discuss why the Blackwell order is not the correct order to compare experiments in moral hazard problems. One way to characterize Blackwell is to use the containment of the set of feasible joint distribution of state-action pairs. Specifically, for any action space $A = \{a_k\}_{k=1}^{K+1}$, the feasible set of joint distributions over $\Omega \times A$ induced by $\mathcal{E}$ can be described by

$$\mathcal{S}_K(\mathcal{E}) := \left\{\mathcal{E}\Pi : \Pi \in \mathbb{R}^{M\times K}, \Pi \geq 0, \Pi\mathbf{1} \leq \mathbf{1}\right\},$$

where $\Pi_{m,k}$ is the probability of taking action a_k after realization y_m, and $1 - \sum_{k=1}^{K} \Pi_{m,k}$ is the probability of taking action a_{K+1} after realization y_m. $\mathcal{E} \geq_{\mathrm{B}} \mathcal{E}'$ if and only if $\bigcup_{K=1}^{\infty} \mathcal{S}_K(\mathcal{E}) \supseteq \bigcup_{K=1}^{\infty} \mathcal{S}_K(\mathcal{E}')$. This works well for general decision problems.

In moral hazard problems, however, the principal only has to decide whether and how much to pay the agent following each realization. The principal's choice can be summarized in a $M \times 1$ vector of payments t. As a result, only $\mathcal{S}_1(\mathcal{E})$ or its subsets matter, and $\mathcal{S}_K(\mathcal{E})$ for $K \geq 2$ are irrelevant. In fact, $\mathcal{S}_1(\mathcal{E}) = \mathrm{Zon}(\mathcal{E})$, and the zonotope order characterizes the value of information comparisons for all binary decision problems where the decision maker has only two actions to choose from.

4 Concluding Remarks

I conclude by discussing several extensions of my results.

Infinite State Space and Infinite Experiments While I focus on finite state space and finite experiments, the geometric method can also be extended to infinite experiments and infinite state space. In this case, one has to be careful about the existence of a solution due to the **?** example.[43] One has to either assume the agent's utility is bounded, or the likelihood ratios of the principal's experiments are bounded.

Non-Flexible Production I assume the agent can flexibly produce any state distribution with a smooth cost function. This convexifies the agent's problem and renders the first order condition both necessary and sufficient. Without flexibility, there are less deviations available to the agent. Stochastic orders in Section 3 are still sufficient for lower costs in the corresponding classes of moral hazard problems. However, they cease to be necessary because the principal only has to guard against a smaller set of deviations.

Strict Orders Strict orders can also be defined in the usual way. That is, say that $\mathcal{E}$ dominates $\mathcal{E}'$ strictly in the column space (or, conic span, zonotope) order, denoted $\mathcal{E} >_{\mathrm{Col}} \mathcal{E}'$ (or $\mathcal{E} >_{\mathrm{Cone}} \mathcal{E}'$, $\mathcal{E} >_{\mathrm{Zone}} \mathcal{E}'$), if $\mathcal{E}$ dominates $\mathcal{E}'$ but $\mathcal{E}'$ does not dominate $\mathcal{E}$ in the weak order. One can show that the dominance in the strict order is characterized by the weak comparisons of costs as in Theorems 1-3 and the existence of a strict cost comparison in some moral hazard problem.

[43] **?** constructs an example under which the principal can get arbitrarily close to the first best but an optimal solution does not exist. The idea is to impose an arbitrarily large punishment with an arbitrarily small probability.

References

Bertschinger, Nils, and Johannes Rauh. 2014. "The Blackwell relation defines no lattice." 2479–2483, IEEE.

Blackwell, David. 1951. "Comparison of experiments." Vol. 1, 26.

Blackwell, David. 1953. "Equivalent comparisons of experiments." *The annals of mathematical statistics*, 265–272.

Brooks, Benjamin, Alexander Frankel, and Emir Kamenica. 2024. "Comparisons of signals." *American Economic Review*, 114(9): 2981–3006.

Chen, Kailin. 2025. "Experiments in the Linear Convex Order." *arXiv preprint arXiv:2502.06530.*

Demougin, Dominique, and Claude Fluet. 2001. "Ranking of information systems in agency models: an integral condition." *Economic Theory*, 17: 489–496.

Gjesdal, Frøystein. 1982. "Information and incentives: The agency information problem." *The Review of Economic Studies*, 49(3): 373–390.

Holmström, Bengt. 1979. "Moral hazard and observability." *The Bell journal of economics*, 74–91.

Kim, Son Ku. 1995. "Efficiency of an information system in an agency model." *Econometrica: Journal of the Econometric Society*, 89–102.

Kim, Yonggyun. 2023. "Comparing information in general monotone decision problems." *Journal of Economic Theory*, 211: 105679.

Koshevoy, Gleb. 1995. "Multivariate lorenz majorization." *Social Choice and Welfare*, 93–102.

Koshevoy, Gleb. 1997. "The Lorenz zonotope and multivariate majorizations." *Social Choice and Welfare*, 15(1): 1–14.

Koshevoy, Gleb, and Karl Mosler. 1996. "The Lorenz zonoid of a multivariate distribution." *Journal of the American Statistical Association*, 91(434): 873–882.

Lehmann, E. L. 1988. "Comparing Location Experiments." *The Annals of Statistics*, 16(2): 521–533.

Marshall, Albert W, Ingram Olkin, and Barry C Arnold. 2009. "Inequalities: theory of majorization and its applications."

Mirrlees, James A. 1999. "The theory of moral hazard and unobservable behaviour: Part I." *The Review of Economic Studies*, 66(1): 3–21.

Mosler, Karl. 2002. *Multivariate dispersion, central regions, and depth: the lift zonoid approach.* Vol. 165, Springer Science & Business Media.

Mu, Xiaosheng, Luciano Pomatto, Philipp Strack, and Omer Tamuz. 2021. "From Blackwell dominance in large samples to Rényi divergences and back again." *Econometrica*, 89(1): 475–506.

Persico, Nicola. 2000. "Information acquisition in auctions." *Econometrica*, 68(1): 135–148.

Quah, John K-H, and Bruno Strulovici. 2009. "Comparative statics, informativeness, and the interval dominance order." *Econometrica*, 77(6): 1949–1992.

Rado, Richard. 1952. "An inequality." *Journal of the London Mathematical Society*, 1(1): 1–6.

Ross, Stephen A. 1973. "The economic theory of agency: The principal's problem." *The American economic review*, 63(2): 134–139.

A Omitted Proofs

This appendix collects the omitted proofs from the main text. I provide some preliminary results in Section A.1. Section A.2 proves the main results.

A.1 Agent's Incentive Constraint

I first present a useful lemma on the agent's incentive constraint. It illustrates how the state dependent utility completely determines the agent's incentives. Given the principal's information $\mathcal{E} \in E^M$ and contract $\boldsymbol{t} \in \mathbb{R}^M$, it provides a necessary and sufficient condition for $\boldsymbol{\mu}_0$ to maximize the agent's objective (1). Note that the agent's objective can be rewritten as $U(\boldsymbol{\mu}; \mathcal{E}, \boldsymbol{t}) = \boldsymbol{\mu} \cdot \mathcal{E}u(\boldsymbol{t}) - C(\boldsymbol{\mu})$.

Lemma 3. *$\boldsymbol{\mu}_0 \in \operatorname*{argmax}_{\boldsymbol{\mu} \in \Delta\Omega} U(\boldsymbol{\mu}; \mathcal{E}, \boldsymbol{t})$ if and only if there exists some $\lambda \in \mathbb{R}$ and $\boldsymbol{\eta} \in \mathbb{R}^N_+$ such that*

$$\mathcal{E}u(\boldsymbol{t}) = \boldsymbol{\nabla} C(\boldsymbol{\mu}_0) + \lambda \mathbf{1} + \boldsymbol{\eta}, \tag{13}$$

$$\eta_n \boldsymbol{\mu}_0(\omega_n) = 0, \forall 1 \leq n \leq N, \tag{14}$$

where $u(\boldsymbol{t}) := [u(t_m)]_{m=1}^M \in \mathbb{R}^M$ and $\boldsymbol{\mu}_0(\omega_n)$ denotes the probability of ω_n under $\boldsymbol{\mu}_0$.

Proof. I first show the necessity. Indeed, this is a first order condition. The additional constant λ is the multiplier on the constraint that the probabilities must add up to one, and $\boldsymbol{\eta}$ is the multipliers on probabilities being non-negative. Suppose $\boldsymbol{\mu}_0 \in \operatorname*{argmax}_{\boldsymbol{\mu} \in \Delta\Omega} U(\boldsymbol{\mu}; \mathcal{E}, \boldsymbol{t})$. View the agent's problem as choosing some $\boldsymbol{\mu} \in \mathbb{R}^N$ subject to the constraint that $\boldsymbol{\mu} \cdot \mathbf{1} = 1$ and $\boldsymbol{\mu} \geq 0$ because the probabilities must add up to one and be non-negative.[44] Let $\lambda \in \mathbb{R}$ be the multiplier on the adding up constraint, and $\boldsymbol{\eta} \in \mathbb{R}^N_+$ be the multipliers on the non-negativity constraint. The Lagrangian of the problem is $\mathcal{L}(\boldsymbol{\mu}, \lambda, \boldsymbol{\eta}) := \boldsymbol{\mu} \cdot \mathcal{E}u(\boldsymbol{t}) - C(\boldsymbol{\mu}) + \lambda(\boldsymbol{\mu} \cdot \mathbf{1} - 1) + \boldsymbol{\mu} \cdot \boldsymbol{\eta}$. Optimality of $\boldsymbol{\mu}_0$ implies that no perturbation in $\boldsymbol{\mu}_0$ can improve the value of $\mathcal{L}(\boldsymbol{\mu}, \lambda, \boldsymbol{\eta})$. Specifically, perturb $\boldsymbol{\mu}_0$ to any $\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \epsilon \boldsymbol{\nu}$ for some $\epsilon > 0$ and $\boldsymbol{\nu} \in \mathbb{R}^N$, we must have

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \left[C(\boldsymbol{\mu}_0 + \epsilon \boldsymbol{\nu}) - C(\boldsymbol{\mu}_0) \right] - \boldsymbol{\nu} \cdot \mathcal{E}u(\boldsymbol{t}) - \lambda \boldsymbol{\nu} \cdot \mathbf{1} - \boldsymbol{\nu} \cdot \boldsymbol{\eta} \geq 0.$$

Apply the definition of $\boldsymbol{\nabla} C$ and notice $\boldsymbol{\nu} \cdot \mathbf{1} = 1$,

$$\lim_{\epsilon \downarrow 0} \boldsymbol{\nu} \cdot \left[\boldsymbol{\nabla} C(\boldsymbol{\mu}) - \mathcal{E}u(\boldsymbol{t}) - \lambda \mathbf{1} - \boldsymbol{\eta} \right] \geq 0.$$

This must hold for any $\boldsymbol{\nu} \in \mathbb{R}^N$. Therefore, we must have (13), and (14) is the corresponding complementary slackness condition. Note that if $\boldsymbol{\mu}_0$ is interior, then the condition reduces to $\mathcal{E}u(\boldsymbol{t}) = \boldsymbol{\nabla} C(\boldsymbol{\mu}_0) + \lambda \mathbf{1}$ for some $\lambda \in \mathbb{R}$. This completes the proof for necessity.

[44] The probabilities has to also be less than one. But the less than one constraint is implied by non-negativity and adding up.

For sufficiency, note that the agent's problem is convex, because in his objective (1) the expected utility from money is linear in his choice $\boldsymbol{\mu}$, and the production cost is smooth and convex. As a result, the first order condition is sufficient for global optimality. $\square$

From Lemma 3, the principal's problem can be viewed as choosing a state dependent utility $\mathcal{E}u(\boldsymbol{t})$ for the agent to minimize the expected cost subject to the agent's first order condition (13) and (14), as well as his participation constraint (5) and potentially limited liability (6) and ex post budget (7). To compare experiments, it is then clear that one only has to compare the set of feasible state dependent utilities. The proofs in the next section repeatedly employ this idea.

A.2 Proofs for Main Results

I now prove a stronger version of Theorems 1-3.

Proof of Theorem 1. In addition to the equivalences in Theorem 1, I also prove their equivalence to

(3') $\operatorname{Aff}\operatorname{Supp}\langle\mathcal{E}\mid\boldsymbol{\mu}_0\rangle_\Delta \supseteq \operatorname{Aff}\operatorname{Supp}\langle\mathcal{E}'\mid\boldsymbol{\mu}_0\rangle_\Delta$ for some interior prior $\boldsymbol{\mu}_0\in\Delta\Omega$.

(1) $\Rightarrow$ (2). Consider the m-th column of $\mathcal{E}'$, denoted $\mathcal{E}'_m$. Due to the column space inclusion, $\mathcal{E}'_m\in\operatorname{Col}\mathcal{E}'\subseteq\operatorname{Col}\mathcal{E}$. Therefore, for each $1\le m\le M'$, $\mathcal{E}'_m=\mathcal{E}\boldsymbol{v}_m$ for some $\boldsymbol{v}_m$. We have $\mathcal{E}'=\mathcal{E}G$ for $G=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_{M'}\end{bmatrix}$.

(2) $\Rightarrow$ (1). Take any $\boldsymbol{x}\in\operatorname{Col}\mathcal{E}'$. $\boldsymbol{x}=\mathcal{E}'\boldsymbol{v}$ for some $\boldsymbol{v}\in\mathbb{R}^{M'}$. Thus, $\boldsymbol{x}=\mathcal{E}'\boldsymbol{v}=\mathcal{E}G\boldsymbol{v}\in\operatorname{Col}\mathcal{E}$.

(2) $\Rightarrow$ (3). Take any prior $\boldsymbol{\mu}_0$. Consider $\mathcal{E}$ first. Apply the Bayes' rule, and the posterior $\boldsymbol{\mu}_m$ induced by realization y_m is given by

$$\boldsymbol{\mu}_m(\omega_n)=\frac{\boldsymbol{\mu}_0(\omega_n)\mathcal{E}_{n,m}}{\sum_{n'=1}^N\boldsymbol{\mu}_0(\omega_{n'})\mathcal{E}_{n',m}}, \tag{15}$$

where $\mathcal{E}_{n,m}$ is the probability of y_m in state ω_n. The same formula applies to $\mathcal{E}'$ as well. Let $\boldsymbol{\mu}'_m$ be the posteriors induced by $\mathcal{E}'$. We have

$$\boldsymbol{\mu}'_m=\boldsymbol{\mu}_mG\cdot\frac{\sum_{n=1}^N\boldsymbol{\mu}_0(\omega_n)\mathcal{E}_{n,m}}{\sum_{n=1}^N\boldsymbol{\mu}_0(\omega_n)\mathcal{E}'_{n,m}}, \tag{16}$$

where the last term is a constant. To check for the inclusion of the affine span, for any $\boldsymbol{x}\in\operatorname{Aff}\operatorname{Supp}\langle\mathcal{E}'\mid\boldsymbol{\mu}_0\rangle_\Delta$, we must show that $\boldsymbol{x}$ is also an affine combination of $\operatorname{Supp}\langle\mathcal{E}\mid\boldsymbol{\mu}_0\rangle_\Delta$. Equation (16) tells us that $\boldsymbol{x}$ is a linear combination of $\operatorname{Supp}\langle\mathcal{E}\mid\boldsymbol{\mu}_0\rangle_\Delta$. It is easy to see that it is also an affine combination (i.e., the coefficients sum up to one) because the posterior beliefs must all sum up to one. This means $\boldsymbol{x}\cdot\mathbf{1}=1$ because $\boldsymbol{x}\in\operatorname{Aff}\operatorname{Supp}\langle\mathcal{E}'\mid\boldsymbol{\mu}_0\rangle_\Delta$, which forces the coefficients on $\operatorname{Supp}\langle\mathcal{E}\mid\boldsymbol{\mu}_0\rangle_\Delta$ to also sum up to one.

(3) $\Rightarrow$ (2). Take $\boldsymbol{\mu}_0$ to be the uniform prior. In this case,

$$\boldsymbol{\mu}_m(\omega_n) = \frac{\mathcal{E}_{n,m}}{\sum_{n'=1}^N \mathcal{E}_{n',m}}.$$

The affine span inclusion thus implies that for any column $\mathcal{E}'_m$ of $\mathcal{E}'$, we have $\mathcal{E}'_m = \mathcal{E}\boldsymbol{v}_m$ for some $\boldsymbol{v}_m$, which implies the matrix factorization condition.

(3) $\Leftrightarrow$ (3′). The $\Rightarrow$ direction is trivial. For $\Leftarrow$, take any interior prior $\boldsymbol{\mu}_0$ such that (3′) holds. Consider any other prior $\boldsymbol{\nu}_0$. We can always express the posteriors in $\operatorname{Supp}\langle\mathcal{E} \mid \boldsymbol{\nu}_0\rangle_\Delta$ as some linear transformation of posteriors in $\operatorname{Supp}\langle\mathcal{E} \mid \boldsymbol{\mu}_0\rangle_\Delta$. Specifically, let $\boldsymbol{\mu}_m$ and $\boldsymbol{\nu}_m$ be the posterior induced by y_m when the prior is $\boldsymbol{\mu}_0$ and $\boldsymbol{\nu}_0$. From (15), we have

$$\boldsymbol{\nu}_m(\omega_n) = \frac{\boldsymbol{\nu}_0(\omega_n)\mathcal{E}_{n,m}}{\sum_{n'=1}^N \boldsymbol{\nu}_0(\omega_{n'})\mathcal{E}_{n',m}} = \boldsymbol{\mu}_m(\omega_n) \cdot \frac{\boldsymbol{\nu}_0(\omega_n)\sum_{n'=1}^N \boldsymbol{\mu}_0(\omega_{n'})\mathcal{E}_{n',m}}{\boldsymbol{\mu}_0(\omega_n)\sum_{n'=1}^N \boldsymbol{\nu}_0(\omega_{n'})\mathcal{E}_{n',m}}.$$

Therefore, every $\boldsymbol{\nu}_m$ is $\boldsymbol{\mu}_m$ multiplied by some state specific constants that do not depend on m. Any linear combination of $\boldsymbol{\nu}_m$ is the same linear combination of $\boldsymbol{\mu}_m$ multiplied by the state specific constants, which gives (3).

(1) $\Rightarrow$ (5). From Lemma 3, to implement some $\boldsymbol{\mu}_0$ under some $u \in \mathcal{U}, C \in \mathcal{C}$, it suffices to find some contract $\boldsymbol{t}$ to generate a state dependent utility that satisfies the first order condition (13) for some multipliers $\lambda \in \mathbb{R}$ and $\boldsymbol{\eta} \in \mathbb{R}_+^N$. If $\mathcal{E} \geq_{\text{Col}} \mathcal{E}'$, then any state dependent utility that can be generated under $\mathcal{E}'$ can also be generated under $\mathcal{E}$: For any $\mathcal{E}'u(\boldsymbol{t}')$ for some $\boldsymbol{t}' \in \mathbb{R}^{M'}$, one can always find $\boldsymbol{t} \in \mathbb{R}^M$ so that $u(\boldsymbol{t}) = Gu(\boldsymbol{t}')$, and $\boldsymbol{t}$ generates the same state dependent utility $\mathcal{E}u(\boldsymbol{t}) = \mathcal{E}Gu(\boldsymbol{t}') = \mathcal{E}'u(\boldsymbol{t}')$. Therefore, if the first order condition (13) can be satisfied for some $\lambda \in \mathbb{R}$ and $\boldsymbol{\eta} \in \mathbb{R}_+^N$ under $\mathcal{E}'$, it can also be satisfied under $\mathcal{E}$ for the same multipliers. The argument applies even if $\boldsymbol{\mu}_0$ is not interior and additional multipliers $\boldsymbol{\eta} \in \mathbb{R}_+^N$ kick in.

(5) $\Rightarrow$ (4). It suffices to observe that when the agent is risk neutral without limited liability, any $\boldsymbol{\mu}_0 \in \mathcal{F}(\mathcal{E}; u, C)$ can be implemented at the first best cost. To see this, suppose $\boldsymbol{\mu}_0$ can be implemented by some contract $\boldsymbol{t} \in \mathbb{R}^M$. The principal can simply take away a constant bonus from $\boldsymbol{t}$ until the agent earns no rent. The cost to the principal is always $C(\boldsymbol{\mu}_0) + \underline{u}$, regardless of her information. The only place information matters is to determine $\mathcal{F}(\mathcal{E}; u, C)$ because if $\boldsymbol{\mu}_0 \notin \mathcal{F}(\mathcal{E}; u, C)$, the cost is infinity. Therefore, the comparisons of cost are implied by the comparisons of implementability.

(4) $\Rightarrow$ (1), (5) $\Rightarrow$ (1). I consider the contrapositive. Suppose $\mathcal{E} \ngeq_{\text{Col}} \mathcal{E}'$. I will find a moral hazard problem with infinite cost under $\mathcal{E}$ but feasible under $\mathcal{E}'$. This proves (4) $\Rightarrow$ (1) and (5) $\Rightarrow$ (1) at the same time. There exists some $\boldsymbol{x} \in \operatorname{Col}\mathcal{E}' \setminus \operatorname{Col}\mathcal{E}$. Consider a risk neutral agent with $u(t) = t$, and pick a cost function C such that $\lim_{\boldsymbol{\mu}\to\boldsymbol{\mu}'} \|\boldsymbol{\nabla} C(\boldsymbol{\mu})\| = +\infty$ for all $\boldsymbol{\mu}' \in \partial\Delta\Omega$ where $\partial\Delta\Omega$ is the boundary of $\Delta\Omega$. Such cost functions make sure that the first order condition (13) does not involve multipliers $\boldsymbol{\eta} \in \mathbb{R}_+^N$ since we must have $\boldsymbol{\eta} = \mathbf{0}$. I can choose some C and $\boldsymbol{\mu}_0$ so that the marginal cost of C at $\boldsymbol{\mu}_0$ is $\boldsymbol{x}$, which makes it implementable under $\mathcal{E}'$ but not $\mathcal{E}$. Note that to implement $\boldsymbol{\mu}_0$,

the principal can pick any multiplier $\lambda \in \mathbb{R}$ to satisfy (13). But if $\boldsymbol{x} \notin \operatorname{Col} \mathcal{E}$, the principal cannot satisfy (13) for any value of λ. This is because, if so, $\boldsymbol{x} + \lambda \mathbf{1} \in \operatorname{Col} \mathcal{E}$ which is a contradiction since $\mathbf{1} \in \operatorname{Col} \mathcal{E}$. $\square$

Proof of Theorem 1. In addition to the equivalences in Theorem 2, I also prove their equivalence to

(3') $\operatorname{Co}\operatorname{Supp} \langle \mathcal{E} \mid \boldsymbol{\mu}_0 \rangle_{\Delta} \supseteq \operatorname{Co}\operatorname{Supp} \langle \mathcal{E}' \mid \boldsymbol{\mu}_0 \rangle_{\Delta}$ for some interior prior $\boldsymbol{\mu}_0 \in \Delta\Omega$.

(4') $K^{\mathrm{B}}(\mathcal{E}; P) \leq K^{\mathrm{B}}(\mathcal{E}'; P)$ for any $P \in \mathcal{P}^{\mathrm{RN}}$.

(1) $\Rightarrow$ (2). The m-th column of $\mathcal{E}'$ is in the conic span, $\mathcal{E}'_m \in \operatorname{Cone} \mathcal{E}' \subseteq \operatorname{Cone} \mathcal{E}$. Therefore, for each $1 \leq m \leq M'$, $\mathcal{E}'_m = \mathcal{E}\boldsymbol{v}_m$ for some $\boldsymbol{v}_m \geq 0$. We have $\mathcal{E}' = \mathcal{E}G$ for $G = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_{M'} \end{bmatrix} \geq 0$.

(2) $\Rightarrow$ (1). Take any $\boldsymbol{x} \in \operatorname{Cone} \mathcal{E}'$. $\boldsymbol{x} = \mathcal{E}'\boldsymbol{v}$ for some $\boldsymbol{v} \geq 0$. Thus, $\boldsymbol{x} = \mathcal{E}'\boldsymbol{v} = \mathcal{E}G\boldsymbol{v} \in \operatorname{Cone} \mathcal{E}$ since $G\boldsymbol{v} \geq 0$.

(2) $\Rightarrow$ (3). Take any prior $\boldsymbol{\mu}_0$. From (15), we know that $\operatorname{Co}\operatorname{Supp} \langle \mathcal{E} \mid \boldsymbol{\mu}_0 \rangle_{\Delta}$ is simply the intersection of $\operatorname{Cone} \boldsymbol{\mu}_0 \odot \mathcal{E}$ and the hyperplane $\boldsymbol{\mu} \cdot \mathbf{1} = 1$, where $\boldsymbol{\mu}_0 \odot \mathcal{E}$ is the matrix $\mathcal{E}$ weighted by the prior with $(\boldsymbol{\mu}_0 \odot \mathcal{E})_{n,m} := \boldsymbol{\mu}_0(\omega_n)\mathcal{E}_{n,m}$. The inclusion of the conic span immediately implies $\operatorname{Cone} \boldsymbol{\mu}_0 \odot \mathcal{E} \supseteq \operatorname{Cone} \boldsymbol{\mu}_0 \odot \mathcal{E}'$ for any prior $\boldsymbol{\mu}_0$, which implies (3).

(3) $\Rightarrow$ (2). Take $\boldsymbol{\mu}_0$ to be the uniform prior. (3) implies that any posterior induced by $\mathcal{E}'$ is a convex combination of posteriors induced by $\mathcal{E}$. From (15), the posteriors induced by $\mathcal{E}$ and $\mathcal{E}'$ differ from the columns of $\mathcal{E}$ and $\mathcal{E}'$ only by some positive multiplicative constant. Therefore, columns of $\mathcal{E}'$ are positive combinations of columns of $\mathcal{E}$, that is, (2) holds.

(3) $\Leftrightarrow$ (3$'$). The proof is exactly the same as that in Theorem 1, except that the linear combination is replaced by convex combination.

(1) $\Rightarrow$ (4). To implement $\boldsymbol{\mu}_0$, the principal has a larger set of state dependent utilities to choose from under $\mathcal{E}$. Moreover, due to risk neutrality, the principal's expected cost equals the agent's expected utility. Therefore, a larger choice set always decreases the cost.

(4) $\Rightarrow$ (1). Suppose $\operatorname{Cone} \mathcal{E} \not\supseteq \operatorname{Cone} \mathcal{E}'$, then there exists some $\boldsymbol{x} \in \operatorname{Cone} \mathcal{E}' \setminus \operatorname{Cone} \mathcal{E}$. Focus on cost functions C such that $\lim_{\boldsymbol{\mu} \to \boldsymbol{\mu}'} \|\boldsymbol{\nabla} C(\boldsymbol{\mu})\| = +\infty$ for all $\boldsymbol{\mu}' \in \partial\Delta\Omega$. This condition ensures that $\boldsymbol{\eta}$ does not show up in the first order condition. Pick a cost function C and $\boldsymbol{\mu}_0$ so that the marginal cost of C at $\boldsymbol{\mu}_0$ is $\boldsymbol{x}$. To optimally implement $\boldsymbol{\mu}_0$ under $\mathcal{E}'$, the principal already chooses the optimal state dependent utility $\boldsymbol{x} + \lambda \mathbf{1} \in \operatorname{Cone} \mathcal{E}'$ for some $\lambda \in \mathbb{R}$. Under $\mathcal{E}$, the principal either cannot implement $\boldsymbol{\mu}_0$, or she can at best pick some $\lambda' > \lambda$ because $\lambda' \leq \lambda$ would imply $\boldsymbol{x} \in \operatorname{Cone} \mathcal{E}$, a contradiction. As a result, the cost is strictly higher under $\mathcal{E}$ than $\mathcal{E}'$.

(1) $\Leftrightarrow$ (4$'$). It suffices to realize, if the principal is subject to only the ex post budget constraint, the set of feasible state dependent utilities is $\{\mathcal{E}\boldsymbol{v} : \boldsymbol{v} \leq 1\}$, which reduces to $\mathbf{1} - \operatorname{Cone} \mathcal{E}$, observing that $\mathcal{E}\mathbf{1} = \mathbf{1}$. The rest of the proof follows the steps for (1) $\Leftrightarrow$ (4). $\square$

Proof of Theorem 3. In addition to the equivalences in Theorem 2, I also prove their equivalence to

(3') $\langle \mathcal{E} \mid \boldsymbol{\mu}_0 \rangle_\Delta \geq_{\text{lcx}} \langle \mathcal{E}' \mid \boldsymbol{\mu}_0 \rangle_\Delta$ for some interior prior $\boldsymbol{\mu}_0 \in \Delta\Omega$,

(5') $K(\mathcal{E}; P) \leq K(\mathcal{E}'; P)$ for any $P \in \mathcal{P}$,

(5") $K^{\mathrm{B}}(\mathcal{E}; P) \leq K^{\mathrm{B}}(\mathcal{E}'; P)$ for any $P \in \mathcal{P}$,

(5"') $K^{\mathrm{LL,B}}(\mathcal{E}; P) \leq K^{\mathrm{LL,B}}(\mathcal{E}'; P)$ for any $P \in \mathcal{P}$.

(1) $\Rightarrow$ (2). Any partial sum of the columns in $\mathcal{E}'$ is still in the zonotope. Therefore, $\mathcal{E}' B_N = \mathcal{E} G_N$.

(2) $\Rightarrow$ (1). To show that $\operatorname{Zon} \mathcal{E}' \subseteq \operatorname{Zon} \mathcal{E}$, it suffices to show that all the extreme points of $\operatorname{Zon} \mathcal{E}'$ is included in $\operatorname{Zon} \mathcal{E}$. This is exactly what $\mathcal{E}' B_N = \mathcal{E} G_N$ says.

(1) $\Leftrightarrow$ (3), (3'). See **?**.

(1) $\Rightarrow$ (4). This simply follows from the fact that the principal can choose from a larger set of utilities.

(4) $\Rightarrow$ (1). Suppose $\operatorname{Zon} \mathcal{E} \not\supseteq \operatorname{Zon} \mathcal{E}'$, then there exists some $\boldsymbol{x} \in \operatorname{Zon} \mathcal{E}' \setminus \operatorname{Zon} \mathcal{E}$. Focus on cost functions C such that $\lim_{\boldsymbol{\mu} \to \boldsymbol{\mu}'} \|\boldsymbol{\nabla} C(\boldsymbol{\mu})\| = +\infty$ for all $\boldsymbol{\mu}' \in \partial\Delta\Omega$. This condition ensures that $\boldsymbol{\eta}$ does not show up in the first order condition. Pick a cost function C and $\boldsymbol{\mu}_0$ so that the marginal cost of C at $\boldsymbol{\mu}_0$ is $\boldsymbol{x}$. To optimally implement $\boldsymbol{\mu}_0$ under $\mathcal{E}'$, the principal chooses an optimal state dependent utility $\boldsymbol{x} + \lambda \mathbf{1} \in \operatorname{Zon} \mathcal{E}'$ for some $\lambda \in \mathbb{R}$ so that is on the boundary of $\operatorname{Zon} \mathcal{E}'$. If $\boldsymbol{\mu}_0$ is not implementable under $\mathcal{E}$, then we are done. Otherwise, the principal optimally implements it at some $\boldsymbol{x} + \lambda' \mathbf{1} \in \operatorname{Zon} \mathcal{E}$ on the boundary of $\operatorname{Zon} \mathcal{E}$. If the cost is lower under $\boldsymbol{x} + \lambda \mathbf{1}$, we are done. Otherwise, since the zonotope is centrally symmetric, we can find a symmetric $\boldsymbol{x}'$ so that the cost is lower under $\mathcal{E}'$ than $\mathcal{E}$.

(1) $\Leftrightarrow$ (5). This follows from Lemma 1 and Lemma 2. I fill in the gaps in the proof for the lemmas at the end.

(1) $\Leftrightarrow$ (5'), (5''), (5'''). Note that in the proof for (1) $\Leftrightarrow$ (5), I do not explicitly use the property of $\boldsymbol{t} \geq 0$. In fact, such restrictions on $\boldsymbol{t}$ does not play a role. Therefore, we can similarly define

$$
\begin{aligned}
\mathcal{V}_{\boldsymbol{\mu}_0,u}(\mathcal{E}) &:= \left\{ \mathcal{E}\boldsymbol{v} : \exists \boldsymbol{t} \in \mathbb{R}^M \text{ such that } \boldsymbol{v} = u(\boldsymbol{t}), \boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \leq 1 \right\}, \\
\mathcal{V}^-_{\boldsymbol{\mu}_0,u}(\mathcal{E}) &:= \left\{ \mathcal{E}\boldsymbol{v} : \exists \boldsymbol{t} \in \mathbb{R}^M \text{ such that } \boldsymbol{v} = u(\boldsymbol{t}), \boldsymbol{t} \leq 1, \boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \leq 1 \right\}, \\
\mathcal{V}^\pm_{\boldsymbol{\mu}_0,u}(\mathcal{E}) &:= \left\{ \mathcal{E}\boldsymbol{v} : \exists \boldsymbol{t} \in \mathbb{R}^M \text{ such that } \boldsymbol{v} = u(\boldsymbol{t}), 0 \leq \boldsymbol{t} \leq 1, \boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \leq 1 \right\},
\end{aligned}
$$

and similarly write down the lemmas 1 and Lemma 2 with the above set of feasible utilities. The proof follows the same steps. □

Proof. Proof of Lemma 1 The main text has outlined the proof. It suffices to check the convexity of $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$. To see that $\mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$ is convex, take any $\boldsymbol{v}_1, \boldsymbol{v}_2$ such that $\mathcal{E}\boldsymbol{v}_1, \mathcal{E}\boldsymbol{v}_2 \in \mathcal{V}^+_{\boldsymbol{\mu}_0,u}(\mathcal{E})$. Let $\boldsymbol{t}_1, \boldsymbol{t}_2 \geq 0$ be the corresponding payment rules that generates $\boldsymbol{v}_1, \boldsymbol{v}_2$ with $\boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t}_1 \leq 1$ and $\boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t}_2 \leq 1$. Take any $\alpha \in [0, 1]$. Let $\boldsymbol{t} = \alpha \boldsymbol{t}_1 + (1 - \alpha)\boldsymbol{t}_2$ and $\boldsymbol{v} := u(\boldsymbol{t})$. Since the budget constraint is linear, we have $\boldsymbol{\mathcal{E}v} \in \mathcal{V}^{+}_{\boldsymbol{\mu}_0,u}(\mathcal{E})$. Due to concavity of u, $\boldsymbol{v} \geq \alpha\boldsymbol{v}_1 + (1-\alpha)\boldsymbol{v}_2 \geq 0$. We can lower the payment to pick some $0 \leq \boldsymbol{t}' \leq \boldsymbol{t}$ so that $\boldsymbol{v}' := u(\boldsymbol{t}')$ satisfies $\boldsymbol{v}' = \alpha\boldsymbol{v}_1 + (1-\alpha)\boldsymbol{v}_2$. It suffices to show $\boldsymbol{v}' \in \mathcal{V}^{+}_{\boldsymbol{\mu}_0,u}(\mathcal{E})$, which follows from $\boldsymbol{t}' \leq \boldsymbol{t}$ because $\boldsymbol{t}'$ pays less and must also satisfy the budget constraint. $\square$

Proof. Proof of Lemma 2 We have the following equivalences.

$$
\begin{aligned}
&\mathcal{E} \geq_{\text{Zon}} \mathcal{E}', \\
\Leftrightarrow &\max_{0 \leq \boldsymbol{v} \leq 1} \boldsymbol{\beta} \cdot \mathcal{E}\boldsymbol{v} \geq \max_{0 \leq \boldsymbol{v} \leq 1} \boldsymbol{\beta} \cdot \mathcal{E}'\boldsymbol{v}, \forall \boldsymbol{\beta} \in \mathbb{R}^N, \\
\Leftrightarrow &\boldsymbol{\beta} \cdot \mathcal{E} \text{ majorizes } \boldsymbol{\beta} \cdot \mathcal{E}', \forall \boldsymbol{\beta} \in \mathbb{R}^N, \\
\Leftrightarrow &\max_{\boldsymbol{t} \geq 0} \boldsymbol{\beta} \cdot \mathcal{E}u(\boldsymbol{t}) - \boldsymbol{\mu}_0 \cdot \mathcal{E}\boldsymbol{t} \geq \max_{\boldsymbol{t} \geq 0} \boldsymbol{\beta} \cdot \mathcal{E}'u(\boldsymbol{t}) - \boldsymbol{\mu}_0 \cdot \mathcal{E}'\boldsymbol{t}, \forall \boldsymbol{\beta} \in \mathbb{R}^N, u \in \mathcal{U}, \boldsymbol{\mu}_0 \in \Delta\Omega, \\
\Leftrightarrow &\mathcal{V}^{+}_{\boldsymbol{\mu}_0,u}(\mathcal{E}) \supseteq \mathcal{V}^{+}_{\boldsymbol{\mu}_0,u}(\mathcal{E}').
\end{aligned}
$$

where the second line uses Fact 2, the third line uses the definition of majorization, the last line uses Fact 1 and realizes that the expression is a convex function in $\boldsymbol{\beta} \cdot \mathcal{E}$, and the last line follows from the argument in the main text. $\square$