Knightian Uncertainty and Moral Hazard

Giuseppe Lopomo[a], Luca Rigotti[b,1], Chris Shannon[c]

[a] *Fuqua School of Business, Duke University*
[b] *Department of Economics, University of Pittsburgh*
[c] *Department of Economics, University of California Berkeley*

Abstract

This paper presents a principal-agent model in which the agent has imprecise beliefs. We model this situation formally by assuming the agent's preferences are incomplete as in [2]. In this setting, incentives must be robust to Knightian uncertainty. We study the implications of robustness for the form of the resulting optimal contracts. We give conditions under which there is a unique optimal contract, and show that it must have a simple flat payment plus bonus structure. That is, output levels are divided into two sets, and the optimal contract pays the same wage for all output levels in each set. We derive this result for the case in which the agent's utility function is linear and then show it also holds if this utility functions has some limited curvature.
JEL Codes: D82, D81, D23

Key words: Knightian uncertainty, moral hazard, contract theory, incomplete preferences

We are grateful to Truman Bewley, Ben Polak, the associate editor, and referees for thoughtful comments.

Email addresses: `glopomo@duke.edu` (Giuseppe Lopomo), `luca@pitt.edu` (Luca Rigotti), `cshannon@econ.berkeley.edu` (Chris Shannon)

[1] Corresponding author

1. Introduction

In this paper, we study a principal-agent model in which the agent has imprecise beliefs. Our model is motivated by situations in which the agent is less familiar with the details of the production process than the principal. For example, imagine an agent about to start a new job. Output depends on his effort as well as variables beyond his control, such as the work of others. Consequently he cannot precisely evaluate the stochastic relation between his effort and output. In contrast, the principal owns the production technology and can evaluate the relation between effort and output more precisely. In other words, the principal and the agent have different confidence in their beliefs.

The standard literature neglects this possibility by assuming that both principal and agent evaluate state-contingent contracts using expected utility, and hence both precisely evaluate the stochastic consequences of the agent's action. Asymmetric confidence in beliefs, on the other hand, is easily described using the Knightian uncertainty theory developed in [2], which provides foundations for a model in which beliefs are given by sets of probability distributions. Thus both the possibility of imprecise beliefs and a ranking in which the agent's beliefs are less precise than the principal's can be formalized in this framework.

Our main result characterizes optimal contracts in a principal-agent model with Knightian uncertainty where the parties have asymmetric confidence in beliefs and linear utility. We give conditions under which there is a unique optimal contract, and show that it has a simple flat payment plus bonus structure. In particular, a contract is optimal *only* if it takes two values across all states. The unique optimal incentive scheme divides all possible output levels into two groups, and pays a fixed wage within each group. This result extends to the case in which the agent's utility function has a limited amount of curvature. This feature of optimal contracts is striking because standard moral hazard models that rule out Knightian uncertainty typically generate complex incentive structures. Optimal contracts often involve as many different payments as there are possible levels of output.

Some authors have speculated that contracts are simple because they need to be robust. [11], for example, argue that real world incentives need to perform well across a wide range of circumstances. Once this need for robustness is considered, simple optimal schemes might obtain. Our results can be viewed through this lens since we introduce robustness to Knightian uncertainty. In our model this robustness, when paired with asymmetric confidence in beliefs, forces incentive schemes to be simple. Furthermore, these contracts have a shape we commonly observe since they can be written as a flat payment plus bonus.[2]

In the spirit of [2], we assume the agent's preferences are not necessarily complete; he may be unable to rank all pairs of contracts offered to him. As in the model axiomatized in [2], we assume that these preferences are represented by a utility function and a *set* of probability distributions, and that the agent prefers one contract to another if the former has higher expected utility for every probability distribution in this set. [2] argues that this approach formalizes the distinction between risk and uncertainty introduced in [15]. In this framework a unique probability is appropriate only

[2] [13] provide conditions under which linear incentive schemes are optimal. These conditions include constant relative risk aversion and a specific dynamic property of stochastic output. Neither of these requirements is related to the idea of robustness considered here. We allow the agent to consider many stochastic structures of output. We also adopt a different notion of simplicity. A two-wage scheme is simple because it can be thought of as contingent on only two events; a linear contract is simple because it is contingent on an intercept and a slope for all events.

when the decision maker regards all events as risky; the decision maker uses a set of distributions when he regards some events as uncertain.

With incomplete preference, choices may be indeterminate and incentive constraints hard to satisfy.[3] [2] proposes a behavioral assumption, inertia, that sometimes alleviates this problem. The inertia assumption states that, when faced with incomparable options, an individual chooses a status quo or reference point, often interpreted as his current behavior, unless there is an alternative that is strictly preferred. In the standard moral hazard setting, there is a natural candidate for the status quo: the agent's outside option or reservation utility. We consider two corresponding notions of implementation, depending on whether inertia is incorporated. Without inertia, an incentive scheme implements a particular action if the agent prefers that action to his reservation utility and to all other actions. With inertia, an incentive scheme implements an action if this action is preferred to the reservation utility and is preferred to all other actions that are comparable to the reservation utility. With inertia, implementing an action is easier because the desired action may be incomparable to some actions. We show that optimal contracts must take two values under either notion of implementation. Thus while inertia might change the cost of implementing an action, it does not change our main result regarding the nature of optimal contracts.

Our moral hazard model has standard features: the principal cannot observe the agent's action and looks for contracts that implement each action at the lowest possible expected cost; each action has a different disutility to the agent; each action induces different beliefs over output outcomes. We initially assume that principal and agent have linear utility functions over money, and then extend the main result to the case in which the agent's utility function is not necessarily linear. The agent perceives Knightian uncertainty, and has beliefs described by sets of probability distributions, one set for each action, while the principal's beliefs are a unique element of the relative interior of those sets. This formalizes the idea that the agent's beliefs are more imprecise than the principal's. Thus multiplicity of the agent's beliefs is the only formal difference between our model and a linear utility version of [10].

Our model shares important features with standard moral hazard models with both risk-averse and risk-neutral agents. As in models with risk-neutral agents, the agent is indifferent over purely objective mean-preserving randomizations that are risky but not uncertain. As in models with risk-averse agents, however, the agent cares about the distribution of payoffs across subjectively uncertain states. We show that optimal contracts with Knightian uncertainty can differ from those without. For example, unlike the standard risk-neutral model, it is no longer possible for the principal to do away with incentive concerns by selling the firm to the agent. Thus Knightian uncertainty provides an alternative to limited liability as an explanation for the importance of incentive provision in many settings where agents are plausibly indifferent to objective mean-preserving lotteries over money.[4] Unlike the standard risk-averse model, optimal incentives have a coarse structure regardless of other fine details of the model, including agents' beliefs.

[17] and [8] present moral hazard models similar in motivation to the one presented here. They use a different model of ambiguity, and in contrast show that incentive schemes are similar to those in a standard model. In both cases, ambiguity is modeled by assuming that the principal and the agent are Choquet expected utility maximizers with convex capacities (see [21]). Neither examines the role of asymmetric confidence, as in each model the principal and the agent share

[3] In a related paper, [16], we study general mechanism design problems. We show that in many standard mechanism design settings interim incentive compatibility is equivalent to ex post incentive compatibility.

[4] We thank a referee for suggesting and stressing the importance of this point.

the same belief set. On the other hand, [18] uses this framework to relate uncertainty to contract incompleteness.

The paper is organized as follows. The next section briefly describes the model of decision makers with incomplete preferences. Section 3 presents the basic framework and discusses the implementation rules. Section 4 characterizes optimal incentive schemes. Section 5 introduces a primitive model and establishes versions of our main results in this setting. Section 6 concludes.

2. Incomplete Preferences and Inertia

von Neumann and Morgenstern were the first to observe that completeness is not a satisfactory axiom for choice under uncertainty.[5] This idea was pursued by [1] who proposes a preference representation theorem for incomplete preferences and objective probabilities.[6] In a series of papers, [2], (1987), and (1989) develops Knightian decision theory, a model which allows for subjective probabilities and incompleteness.[7]

2.1. Incomplete Preferences

Under completeness, any pair of alternatives can be ranked. If preferences are not complete, some alternatives are incomparable. [2] axiomatizes a model allowing for incompleteness with subjective probabilities. To formalize this discussion, let the state space $\mathcal{N}$ be finite, and index the states by $i = 1, \ldots, N$. $\Delta(\mathcal{N}) := \{\pi \in \mathbf{R}^N : \pi_i \geq 0 \ \forall i, \sum_i \pi_i = 1\}$ denotes the set of probability distributions over $\mathcal{N}$, and $x, y \in X^N$ are random monetary payoffs where $X \subset \mathbf{R}$ is finite. In an Anscombe-Aumann setting, [2] characterizes incomplete preference relations represented by a unique nonempty, closed, convex set of probability distributions Π and a continuous, strictly increasing, concave function $u : X \to \mathbf{R}$, unique up to positive affine transformations, such that

$$x \succ y \qquad \text{if and only if} \qquad \sum_{i=1}^{N} \pi_i u(x_i) > \sum_{i=1}^{N} \pi_i u(y_i) \ \text{ for all } \pi \in \Pi.$$

With a small abuse of notation, we can rewrite this as

$$x \succ y \qquad \text{if and only if} \qquad E_\pi [u(x)] > E_\pi [u(y)] \ \text{ for all } \pi \in \Pi$$

where $E_\pi [\cdot]$ denotes the expected value with respect to the probability distribution π, and $u(x)$ denotes the vector $(u(x_1), \ldots, u(x_N))$.[8]

The set of probabilities Π reduces to a singleton whenever the preference ordering is complete, in which case the usual expected utility representation obtains. Without completeness, comparisons between alternatives are carried out "one probability distribution at a time", with one bundle

[5] They write:

> "It is conceivable - and may even in a way be more realistic - to allow for cases where the individual is neither able to state which of two alternatives he prefers nor that they are equally desirable...How real this possibility is, both for individuals and for organizations, seems to be an extremely interesting question...It certainly deserves further study." [23], Section 3.3.4, pg. 19.

[6] Aumann's work has been extended and clarified by [**?**] and [22].
[7] [2] has been published recently as [5].
[8] This representation has been extended by [9] in a Savage setting.

preferred to another if and only if it is preferred under every probability distribution considered by the agent.[9] [2] suggests that this approach formalizes the distinction between risk and uncertainty originating in [15]. Informally, the size of Π measures how much uncertainty the individual perceives, and can be thought of as reflecting confidence in beliefs.

Figure 1 illustrates Bewley's representation for the special case in which u is linear. The axes measure utility (or money) in each of the two possible states. Given a probability distribution, a line through the bundle x represents all the bundles that have the same expected value as x according to this distribution. As the probability distribution changes, we obtain a family of these indifference curves representing different expected utilities according to different probabilities. The thick curves represent the most extreme elements of this family, while thin curves represent other possible elements.

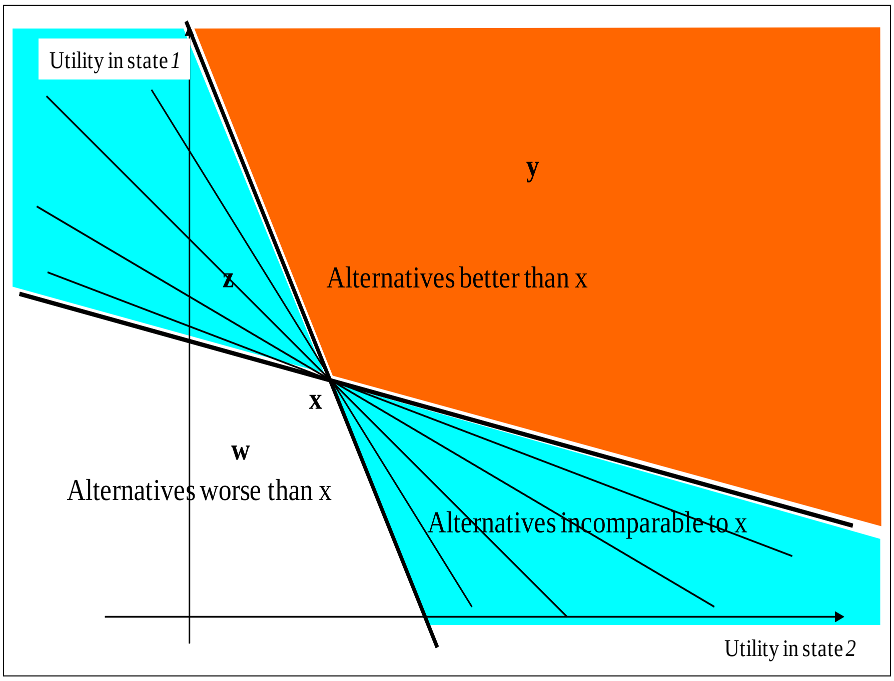

Figure 1: Incomplete Preferences

In Figure 1, y is preferred to x since it lies above all of the indifference curves corresponding to some expected utility of x. Also, x is preferred to w since w lies below all of the indifference curves through x. Finally, z is not comparable to x since it lies above some indifference curves through x and below others. Incompleteness induces three regions: bundles preferred to y, dominated by y, and incomparable to y. This last area is empty only if there is a unique probability distribution over the two states and the preferences are complete.

For any bundle x, the better-than-x set has a kink at x. This kink is a direct consequence of the multiplicity of probability distributions in Π, and vanishes only when Π is a singleton. As a consequence, the assumption of linear utility generates different behavior with multiple probabilities than in the standard expected utility model. Under expected utility, a linear utility index implies a

[9]The natural notion of indifference in this setting says two bundles are indifferent whenever they have the same expected utility for each probability distribution in Π.

constant marginal rate of substitution and indifference to mixtures of bundles in a given indifference class. Neither of these obtains when there is Knightian uncertainty. This can easily be seen in Figure 1, where convex combination of bundles that are not comparable to x can give rise to a bundle that is strictly preferred to x.

2.2. The Inertia Assumption

Revealed preference arguments must take incompleteness into account. If x is chosen when y is available, one cannot conclude that x is revealed preferred to y, but only that y is not revealed preferred to x. The concepts of status quo and inertia introduced in [2] can sharpen revealed preference inferences when preferences are incomplete. Bewley's inertia assumption posits the existence of planned behavior that is taken as a reference point, and assumes that this "status quo" is abandoned only for alternatives preferred to it. In Figure 1, for example, if x is the status quo and the inertia assumption holds, alternatives like z will not be chosen since they are incomparable to x.

In many economic contexts, there may not be a natural status quo. In the moral hazard model that follows, however, an obvious candidate for the status quo is the action yielding the agent's reservation utility. This corresponds to the payoff the agent receives if he rejects the contract offered by the principal.

3. The Moral Hazard Setup

The main features of our model are standard and follow the parametrized distribution approach: the principal owns resources that yield output, but needs the agent's input for production to take place; principal and agent observe realized output, but the principal does not observe the agent's action; each action imposes a cost on the agent and generates beliefs about output; these beliefs are common knowledge. The main innovation is that beliefs reflect Knightian uncertainty about output levels. The crucial assumption is that the agent perceives more uncertainty than the principal, and we simplify matters by assuming that the principal perceives no Knightian uncertainty.

Output is an N-vector $y = (y_1, \ldots, y_N)$, with states labeled so that $y_N > \cdots > y_1$ (throughout, we use subscripts to denote states). A contract is an N-vector $w = (w_1, \ldots, w_N)$, where w_j is the payment from the principal to the agent in state j. The agent's reservation utility is $\overline{w}$; we will interpret it as the status quo when we impose the inertia assumption. The agent chooses an action a from the finite set $\mathcal{A} := \{1, 2, \ldots, M\}$. The principal's beliefs are described by a function $\pi^P : \mathcal{A} \to \Delta$, while the agent's beliefs are described by a correspondence $\Pi : \mathcal{A} \to 2^{\Delta}$, where $\Pi(a)$ is nonempty, closed and convex for each a. In the parametrized distribution formulation of the principal-agent problem pioneered by [12], Π is a function and is identical to π^P.[10]

We assume that beliefs are consistent, but reflect different degrees of precision. Formally, we assume that for each a in $\mathcal{A}$, $\pi^P(a) \in \Pi(a)$. Uniqueness of the principal's beliefs is assumed for analytical tractability. Most of the analysis carries over if π^P is also a set, as long as asymmetric confidence holds (i.e. $\pi^P(a)$ is a proper subset of $\Pi(a)$ for each a).

The agent's disutility of action a is denoted $c(a)$. Actions are ordered such that whenever $a > a'$, $c(a) > c(a')$ and $\sum_{i=1}^N \pi_i^P(a) y_i \geq \sum_{i=1}^N \pi_i^P(a') y_i$. As usual, costlier actions increase the expected value of output to the principal.

[10] We discuss in section 5 some undesirable effects of this formulation with Knightian uncertainty.

Principal and agent have linear utility over money. The agent evaluates the difference between the expected value of the contract and the cost of his action for each probability distribution. For each a and each $\pi \in \Pi(a)$, this difference is given by

$$E_\pi[w] - c(a) = \sum_{j=1}^{N} \pi_j w_j - c(a).$$

Given a contract, the agent chooses an action by computing many expected values (one for each element of each belief set).

The principal evaluates the expected value of output minus the expected cost of the contract. For each a, the principal's expected utility is

$$E_{\pi^P(a)}[y - w] = \sum_{j=1}^{N} \pi_j^P(a) y_j - \sum_{j=1}^{N} \pi_j^P(a) w_j.$$

In the standard model, the principal is risk-neutral while the agent is either risk-neutral or risk-averse. Both parties have the same beliefs, and the same attitude toward uncertainty, but they might evaluate risk in different ways. In our model, both parties have consistent beliefs and evaluate risk in the same way, but might have different attitudes toward uncertainty. Notice that even though the agent's utility function is linear, the agent cares about the distribution of payoffs across states.

If the agent's action is observable (and/or verifiable), the contract can depend on it. In keeping with standard terminology, we denote the cost of this contract by $C^{FB}(a)$, where FB stands for first-best. One can immediately see that $C^{FB}(a) = \overline{w} + c(a)$ for each action a. In other words, the principal can implement a with a contract that pays $C^{FB}(a)$ in every state if action a is taken and $-\infty$ otherwise. A Pareto efficient contract must be constant across states: contracts that do not fully insure the agent are dominated by "flatter" contracts that do not make the agent worse off and lower the principal's expected costs.[11]

The principal and the agent do not value the firm equally. Therefore, one cannot appeal to the standard way of dealing with moral hazard when the parties have linear utility, which would have the principal sell the firm to the agent. For a given action, the highest price the agent is willing to pay for the firm is the lowest expected value of output minus the cost of taking that action. The highest price the agent is willing to pay is thus lower than what the firm is worth to the principal.

3.1. Two Implementation Rules

A contract implements an action a^* if it induces the agent to participate and choose that action. The contract must provide incentives that leave no doubt the agent's choice is a^* among all possible actions.[12] How this is done depends on whether or not the inertia assumption holds.

[11] [19] give a general characterization of Pareto optimal allocations with incomplete preferences.

[12] In [16] we distinguish two notions of incentive compatibility in a general mechanism design framework: optimal and maximal. A mechanism is optimal incentive compatible if truth telling is preferred to all other reports. A mechanism is maximal incentive compatible if no report is preferred to truth telling. The notion of incentive compatibility we consider here corresponds to what we call optimal incentive compatibility in [16].

3.1.1. Implementing without inertia

Without inertia, a contract must satisfy the standard participation and incentive compatibility conditions, modified to allow for Knightian uncertainty. Participation requires that a^* is preferred to the reservation utility, while incentive compatibility requires that a^* is preferred to all other actions.

Definition 1. *An incentive scheme* w implements a^* *if:*

$$\sum_{j=1}^{N} \pi_j(a^*)w_j - c(a^*) \geq \overline{w} \qquad \forall \pi(a^*) \in \Pi(a^*) \tag{P}$$

and for each $a \in \mathcal{A}$ *with* $a \neq a^*$

$$\sum_{j=1}^{N} \pi_j(a^*)w_j - c(a^*) \geq \sum_{j=1}^{N} \pi_j(a)w_j - c(a) \qquad \forall \pi(a^*) \in \Pi(a^*) \text{ and } \forall \pi(a) \in \Pi(a) \tag{IC}$$

For any probability distribution induced by a^*, the expected utility of the contract must be weakly higher than the reservation utility, and weakly higher than the expected utility calculated according to all probability distributions induced by any other action.

The implementation requirements imposed by Definition 1 are restrictive. In particular, there could be many actions that cannot be implemented.

Proposition 1. *If* $\Pi(a) \cap \Pi(a') \neq \emptyset$ *for some actions* $a \neq a'$*, then* a *and* a' *cannot both be implemented.*

Proof. Let the two actions be a and a' and, without loss of generality, assume $a > a'$. Suppose by way of contradiction that there exists $\pi \in \Pi(a) \cap \Pi(a')$, and there exists a contract w that implements a. Therefore, by (IC),

$$\sum_{j=1}^{N} \pi_j w_j - c(a) \geq \sum_{j=1}^{N} \pi_j w_j - c(a')$$

This implies $c(a) \leq c(a')$, and contradicts $a > a'$. □

A possible implication of Proposition 1 is that, when the amount of uncertainty is large, few actions could be implementable because no action for which the belief set intersects the belief set of a cheaper action can be implemented. For example, as an agent works harder he might gather valuable information regarding his influence on the productive process. Thus costlier actions might lead to a smaller set of beliefs. In such cases, the belief set for a cheaper action will always intersect the belief set of every more expensive action, and only the cheapest action could be implemented. The addition of inertia, which we discuss next, relaxes incentive constraints and can enlarge the set of implementable actions.

3.1.2. Implementing with inertia

Under the inertia assumption, an alternative is chosen only if it is preferred to the status quo. Hence, actions not comparable to the reservation utility are not chosen. If a^* and a' are not comparable, but a^* is preferred to the status quo while a' is not comparable to it, then the inertia assumption implies a' is not chosen. Therefore, with inertia a contract does not need to make a^* preferred to all other actions, but just to those that are comparable to the status quo.

Definition 2. *An incentive scheme w* implements a^* with inertia *in $\mathcal{A}$ if*

$$\sum_{j=1}^{N} \pi_j(a^*)w_j - c(a^*) \geq \overline{w} \qquad \forall \pi(a^*) \in \Pi(a^*) \tag{P}$$

and for each $a \in \mathcal{A}$ with $a \neq a^*$, either

$$\sum_{j=1}^{N} \pi_j(a^*)w_j - c(a^*) \geq \sum_{j=1}^{N} \pi_j(a)w_j - c(a) \qquad \forall \pi(a^*) \in \Pi(a^*) \quad \textit{and } \forall \pi(a) \in \Pi(a) \tag{IC}$$

or

$$\sum_{j=1}^{N} \pi_j(a)w_j - c(a) \leq \overline{w} \textit{ for some } \pi(a) \in \Pi(a) \tag{NC}$$

NC is a non-comparability constraint. It says there exists at least one probability distribution induced by a such that the corresponding expected utility of the contract is weakly lower than the reservation utility. Inertia weakens the incentive constraints since actions that are not comparable to the outside option are not chosen. Obviously, contracts that satisfy Definition 1 also satisfy Definition 2. On the other hand, there could be contracts that implement a costlier action with inertia even if belief sets intersect, and therefore inertia enlarges the set of implementable actions.

3.2. The principal's problem

Given an action, the principal looks for the cheapest contract that implements it. She then decides which action to implement. As usual, we focus only on the first of these problems. For a given action a, let $\mathcal{H}(a)$ be the (possibly empty) set of incentive schemes that implement a, and $\mathcal{H}^I(a)$ be the (also possibly empty) set of incentive schemes that implement a with inertia. We say $\widehat{w}(a)$ is an *optimal incentive scheme to implement a* if it is a solution to

$$\min_{w \in \mathcal{H}(a)} \sum_{j=1}^{N} \pi_j^P(a)w_j \tag{1}$$

Let $\widehat{W}(a)$ denote the set of solutions to (1). Similarly, we say that $\widehat{w}(a)$ is an *optimal incentive scheme to implement a with inertia* if it is a solution to

$$\min_{w \in \mathcal{H}^I(a)} \sum_{j=1}^{N} \pi_j^P(a)w_j \tag{2}$$

Let $\widehat{W^I}(a)$ denote the set of solutions to (2).

3.3. Some characteristics of the optimal incentive scheme

Although the agent's behavior depends on all the probability distributions in his belief set, some are particularly relevant because they determine whether incentive and participation constraints are satisfied.

For any action a and contract w, let

$$\underline{\Pi}(a;w) := \arg \min_{\pi(a)\in\Pi(a)} \sum_{j=1}^{N} \pi_j(a)w_j$$

Given w, any $\underline{\pi} \in \underline{\Pi}(a;w)$ is a probability distribution yielding the minimum expected value for the agent when he chooses action a. A contract w satisfies the participation constraint for action a if and only if $E_\pi[w] - c(a) \geq \bar{w}$ for every $\pi \in \underline{\Pi}(a;w)$. Similarly, given a contract w, action a satisfies the non-comparability constraint if and only if $E_\pi[w] - c(a) < \bar{w}$ for some $\pi \in \underline{\Pi}(a;w)$. For any fixed action a and contract w, let

$$\bar{\Pi}(a;w) := \arg \max_{\pi(a)\in\Pi(a)} \sum_{j=1}^{N} \pi_j(a)w_j$$

Then $\bar{\pi} \in \bar{\Pi}(a;w)$ is a probability yielding the maximum expected value of the contract w for the agent when he chooses action a. A contract w satisfies the incentive compatibility constraint for action a versus a' if and only if $E_{\underline{\pi}}[w]-c(a) \geq E_{\bar{\pi}}[w]-c(a')$ for all $\underline{\pi} \in \underline{\Pi}(a;w)$ and all $\bar{\pi} \in \bar{\Pi}(a';w)$.

The following proposition extends some standard results about optimal contracts (all proofs are in the appendix).

Proposition 2. *Let a be the action to be implemented.*

(i) For any $\widehat{w}$ in $\widehat{W}(a)$ or $\widehat{W}^I(a)$, the participation constraint binds when computed according to $\underline{\pi}(a) \in \underline{\Pi}(a;\widehat{w})$; formally,

$$\sum_{j=1}^{N} \underline{\pi}_j(a)\widehat{w}_j - c(a) = \overline{w}$$

for any $\widehat{w} \in \widehat{W}(a) \cup \widehat{W}^I(a)$ and $\underline{\pi}(a) \in \underline{\Pi}(a;\widehat{w})$.

(ii) If a is the least costly action, the scheme that pays $\overline{w} + c(1)$ in all states implements a (with inertia); that is, $\overline{w} + c(1) \in \widehat{W}(1)$.

(iii) If $\widehat{w} \in \widehat{W}^I(a)$ and a is not the least costly action, then there exists an action less costly than a such that either (IC) or (NC) binds for this action; formally, if $a > 1$ there exists an $a' < a$ such that either

$$\sum_{j=1}^{N} \underline{\pi}_j(a)\widehat{w}_j - c(a) = \sum_{j=1}^{N} \bar{\pi}_j(a')\widehat{w}_j - c(a') \qquad \forall \underline{\pi}(a) \in \underline{\Pi}(a;\widehat{w}) \text{ and } \forall \bar{\pi}(a') \in \bar{\Pi}(a';\widehat{w})$$

or

$$\sum_{j=1}^{N} \underline{\pi}_j(a')\widehat{w}_j - c(a') = \overline{w} \qquad \forall \underline{\pi}(a') \in \underline{\Pi}(a';\widehat{w})$$

for any $\widehat{w} \in \widehat{W}^I(a)$.

Similarly, if $\widehat{w} \in \widehat{W}(a)$ and a is not the least costly action, then there exists an action less costly than a such that (IC) binds for this action.

4. Optimal Incentive Schemes

In this section, we first consider the case of two actions. Then we generalize the result to many actions.

The main results depend on two properties of beliefs.

Assumption A-1: For each action $a \in \mathcal{A}$, $\Pi(a)$ is the core of a convex capacity v^a on $\mathcal{N}$.[13] That is,

$$\Pi(a) = \{\pi \in \Delta(\mathcal{N}) : \pi(E) \geq \nu^a(E) \text{ for each } E \subset \mathcal{N}\} \tag{3}$$

Geometrically, assumption A-1 requires that $\Pi(a)$ is a polyhedron with boundaries determined by the linear inequalities in (3). Figure 2 displays some examples with three states: Π^A and Π^B are cores of convex capacities while Π^C is not.

Some intuition for the content of this assumption might be gleaned from a slightly stronger condition that makes use of probability bounds. That is, for each state $i = 1, \ldots, N$, take $q_i, r_i \in [0, 1]$ with $q_i \leq r_i$; the interval $[q_i, r_i]$ will be interpreted as bounds on the probability of state i occurring. Under several consistency conditions,[14]

$$\Pi := \{\pi \in \Delta(\mathcal{N}) : \pi_i \in [q_i, r_i] \ \forall \ i\}$$

is the core of the convex capacity ν defined by

$$\nu(E) = \min_{\pi \in \Pi} \pi(E) \quad \forall E \subset \mathcal{N}$$

A similar restriction is found in much of the work on applications of ambiguity, which frequently uses a version of Choquet expected utility ([21]). In these models uncertainty aversion and a limited notion of independence generate beliefs represented by the core of a convex capacity. For example, both [17]and [8] assume the beliefs of both parties involved in a moral hazard model are represented by the cores of convex capacities.

[13] A convex capacity on $\mathcal{N}$ is a function $\nu : 2^{\mathcal{N}} \rightarrow [0, 1]$ such that (i) $\nu(\emptyset) = 0$, (ii) $\nu(\mathcal{N}) = 1$, (iii) $\forall E, E' \subset \mathcal{N}$, $E \subseteq E'$ implies $\nu(E) \leq \nu(E')$, and (iv) $\forall E, E' \subset S$, $\nu(E \cup E') \geq \nu(E) + \nu(E') - \nu(E \cap E')$.

[14] Specifically, the collection $\{q_i, r_i : i = 1, \ldots, N\}$ is *proper* if

$$\sum_i q_i \leq 1 \leq \sum_i r_i$$

For a proper collection, the set of probabilities consistent with the associated intervals state by state is non-empty. Thus if $\{q_i, r_i : i = 1, \ldots, N\}$ is proper,

$$\{\pi \in \Delta(\mathcal{N}) : \pi_i \in [q_i, r_i] \ \forall \ i\}$$

is non-empty. The collection $\{q_i, r_i : i = 1, \ldots, N\}$ is *reachable* if for each i,

$$\sum_{j \neq i} q_j + r_i \leq 1 \text{ and } \sum_{j \neq i} r_j + q_i \geq 1$$

This guarantees that

$$q_i = \inf\{\pi_i : \pi \in \Delta(\mathcal{N}), \pi_j \in [q_j, r_j] \ \forall j\} \text{ and } r_i = \sup\{\pi_i : \pi \in \Delta(\mathcal{N}), \pi_j \in [q_j, r_j] \ \forall j\}$$

for each state i. If $\{q_i, r_i : i = 1, \ldots, N\}$ is proper and reachable, then the set Π consistent with these bounds is the core of a convex capacity, in particular $\nu(E) = \min_{\pi \in \Pi} \pi(E) \ \forall E \subset \mathcal{N}$. For example, see [6].

For simplicity we have chosen to state assumption A-1 directly as a technical condition on the belief sets. We could give an alternative formulation based on conditions on the agent's underlying preferences, however. That is, for each $a \in \mathcal{A}$, let $\succ^a$ denote the agent's strict preference order over state-contingent monetary payoffs $\mathbf{R}^N$, conditional on choosing the action a. Given our assumptions, for any pair $w, y \in \mathbf{R}^N$,

$$w \succ^a y \iff E_\pi[w] - c(a) > E_\pi[y] - c(a) \;\; \forall\, \pi \in \Pi(a)$$

Assumption A-1 can be rephrased as a condition on $\succ^a$ by making use of a natural extension of the notion of subjective beliefs in [20].

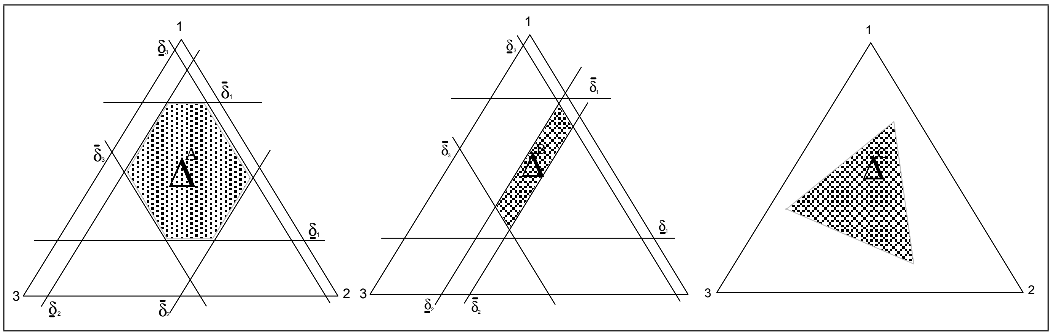

Figure 2: Agent's Beliefs with 3 Output States

Assumption A-2: For each action $a \in \mathcal{A}$, $\pi^P(a)$ is an element of the relative interior of $\Pi(a)$.[15] In addition to the relative imprecision condition we maintain by assuming $\pi^P(a) \in \Pi(a)$ for each a, assumption A-2 requires that the agent perceives sufficiently rich uncertainty so that $\Pi(a)$ has a non empty relative interior for each action a, and refines the restriction on principal's beliefs to rule out cases where the principal's belief is an extreme element of the agent's set of beliefs.

4.1. Optimal incentive schemes with two actions

Assume there are only two actions, H and L, with $c(H) > c(L)$. We interpret these as high and low effort respectively. By Proposition 2, the principal can implement low effort at the first best cost with a constant contract. Our main result shows that a contract implementing high effort can be optimal only if it takes two values.

Proposition 3. *Suppose Assumptions A-1 and A-2 hold, and the high action H is implementable (implementable with inertia). Then, the unique optimal incentive scheme to implement H (with inertia) divides the N states into two subsets and is constant on each subset.*

This result suggests Knightian uncertainty could explain the emergence of simple contracts. Without it, our setup reduces to the standard principal agent model with linear utilities. In that case, both parties are risk neutral and a continuum of contracts is optimal. These include simple two-valued contracts and many complicated ones. The simplest in this class is the contract that corresponds to selling the firm to the agent, hence avoiding incentive problems altogether. In our

[15] Here we mean interior relative to $\Delta(\mathcal{N})$, that is, the points $\pi \in \Pi(a)$ such that there exists an open set $O \subset \mathbf{R}^N$ with $\pi \in O \cap \Delta(\mathcal{N}) \subset \Pi(a)$.

setup, because of Knightian uncertainty the parties cannot avoid the incentive problem even when utility functions are linear. Furthermore, a two-tier contract is the *unique* optimal contract with Knightian uncertainty. Adding Knightian uncertainty to the standard model selects a unique, simple contract.

Our model predicts simple contracts when the standard model with risk averse agent does not. Although the standard model makes a different assumption on the curvature of the agent's utility function, it shares some behavioral characteristics with the particular form of Knightian uncertainty we assume, since the agent's better-than sets are not half-spaces but proper cones. With Knightian uncertainty, thinking of an agent as risk neutral when his utility function is linear seems inappropriate since he cares about the distribution of payoffs across states.

Notice that the inertia assumption plays no role in Proposition 3. In the Appendix, we provide a detailed proof for the inertia case, but the other case can be proved very similarly. Intuitively, the result follows from a simple consequence our assumptions. For any state, there is at least one belief of the agent that assigns lower probability than the principal to that state. Therefore, for any state there is a belief that makes the principal more optimistic about the likelihood of that state. Because all beliefs matter, the optimal contract must "accommodate" this probability and doing so is costly to the principal. Since this is true for all states, and everyone utility is linear, the principal can always benefit by using a "flatter" contract. Eventually, this process must stop because a contract must have enough variation to provide incentives.

Next we show that Proposition 3 is robust to the introduction of a small degree of risk aversion. Consider the generalization of our basic model in which the agent's utility function over certain monetary outcomes is a strictly increasing and strictly concave function $u : \mathbf{R} \to \mathbf{R}$ that is twice continuously differentiable, with $u'(x) > 0, u''(x) < 0$ for all $x \in \mathbf{R}$. Let v_M denote the highest utility level that the agent can receive, i.e. $v_M = u(V_*)$, where V_* denotes the highest profit generated by any action. Let $h : \mathbf{R} \to \mathbf{R}$ denote the inverse of u. Since h is strictly increasing and convex, we can represent it as the sum $h(u) = u + g(u)$, where $g : \mathbf{R} \to \mathbf{R}$ is convex and continuously differentiable. Our main result remains valid provided g' is sufficiently small. Because the derivation of the requisite bound is somewhat complicated, we defer the proof to the appendix.

Proposition 4. *Suppose Assumptions A-1 and A-2 hold, and the high action H is implementable (implementable with inertia). Then there exists $b > 0$ such that if $g'(v_M) < b$, then the unique optimal incentive scheme to implement H (with inertia) divides the N states into two subsets and is constant on each subset.*

The idea is that when the curvature of the utility function is small, Knightian uncertainty dominates, and the intuition behind Proposition 3 is powerful enough to overcome the usual arguments that make the optimal contract depend on the likelihood ratios of different states across different actions.

4.2. Optimal incentive schemes with many actions

We now provide conditions to extend the previous results to many actions. Simple reasoning shows that the main result is easily extended to more actions without additional restrictions; it implies that an optimal contract takes as many values as there are actions. The proof of Proposition 3 shows that reducing the number of states on which a contract depends is profitable for the principal. Clearly, this becomes impossible when all constraints are binding. In the case of two actions, this implies the contract has two values. If there are M possible actions, there will be at most M binding constraints, and the optimal contract must be at most M-valued.

Under additional restrictions, we can show that the optimal contract must be two-valued even when there are many actions. These conditions reduce the many-action case to the two-action case. The exercise parallels what is done to obtain monotonicity (in output) of the optimal contract in the standard model, and uses analogous assumptions.
Assumption A-3: For each action a, every selection in

$$\{\pi^e : \mathcal{A} \to \Delta(\mathcal{N}) | \pi^e(a) \text{ is an extreme point of } \Pi(a) \text{ for each } a\}$$

satisfies the monotone likelihood ratio property and concavity of the distribution function. That is, if $\pi^e : \mathcal{A} \to \Delta(\mathcal{N})$ is a selection such that $\pi^e(a)$ is an extreme point of $\Pi(a)$ for each a, then

(i) (MLR) for any two actions a', a, $c(a) > c(a')$ implies that $\frac{\pi_i^e(a')}{\pi_i^e(a)}$ is decreasing in i

(ii) (CDFC) for any three actions a'', a', a such that $c(a) = \lambda c(a') + (1-\lambda) c(a'')$ for some $\lambda \in [0,1]$, $\sum_{i=1}^{j} \pi_i^e(a) \leq \lambda \sum_{i=1}^{j} \pi_i^e(a') + (1-\lambda) \sum_{i=1}^{j} \pi_i^e(a'')$ for all j.

The first part of this assumption says that the standard monotone likelihood ratio assumption holds for any extreme point of the set of probability distribution induced by each action. The second part extends concavity of the distribution function along similar lines by also requiring that it holds for any extreme point of the set of probability distribution induced by each action. Without uncertainty, these are pretty standard assumptions in the moral hazard literature. With uncertainty, all we do is extend them in the most obvious fashion so that they hold for the sets of probability distribution of our previous assumptions.

Proposition 5. *Suppose Assumptions A-1, A-2 and A-3 hold. Then for any action $a^* > 1$, the optimal contract to implement a^* (with inertia) has a two-wage structure.*

Interestingly, a sufficient condition to reduce the multi-action case to the two-action case is a generalized version of the requirement one needs in the standard model to obtain monotonicity in output of the optimal contract. Notice that, a fortiori, this same condition would induce monotonicity of the contract in our model.

5. The Primitive Model

The standard principal-agent model is formalized by using probability distributions that depend on actions. This is a shortcut for the formal model considered by the early literature in which actions determine output jointly with the state of the world while probabilities over states are given. As noted in [14], the parameterized distribution model lacks an axiomatic foundation since in both Savage and Anscombe-Aumann probability distributions over states do not depend on choices. Without Knightian uncertainty, any primitive model can be written in the parameterized distribution formulation but not vice-versa (see [12] for a discussion). Hence assumptions made directly on the parameterized distribution model may not have a counterpart in any primitive model and thus may not be consistent with first principles. In this section, we formalize the primitive model with Knightian uncertainty, and show that our main assumptions can be deduced from similar assumptions on preferences in that model. In the process, we show how action-based beliefs over outcomes are derived from a fixed set of beliefs over states and therefore establish that the parameterized distribution formulation can also be thought of as a shortcut for the primitive model in the presence of Knightian uncertainty.

Let $\mathcal{S}$ be the state space, with elements $s = 1, \ldots, S$. Output is determined by a production function that depends on effort as well as stochastic factors. For each action $a \in \mathcal{A}$, $y(a) : S \to \mathbf{R}$ denotes the output that results from the action a. For each action a, a', we assume $\operatorname{supp} y(a) = \operatorname{supp} y(a') := \{y_1, \ldots, y_N\}$; different output levels across actions correspond to different permutations of the state space, so that observing a particular output level does not reveal the agent's action.

If the output is y, the principal's payoff is $y - w$. The agent's payoff is $w - c(a)$, where as above $c : \mathcal{A} \to \mathbf{R}$ denotes the agent's disutility of effort. The wage depends on output alone and not on the state. Since the agent knows which action he has taken, he also knows which state has occurred. The principal, on the other hand, does not know the realized state of the world since he can only observe realized output. We assume that the agent's beliefs over S are given by a closed and convex set $\Pi \subset \Delta(\mathcal{S})$, and that the principal's belief is $\pi^P \in \Delta(\mathcal{S})$.

Our notions of implementation can be reformulated as follows.

Definition 3. *An incentive scheme w* implements a^* *if:*

$$\sum_{s=1}^{S} \pi_s w(y_s(a^*)) - c(a^*) \geq \overline{w} \qquad \forall \pi \in \Pi \tag{P}$$

and for each $a \in \mathcal{A}$ with $a \neq a^*$

$$\sum_{s=1}^{S} \pi_s w(y_s(a^*)) - c(a^*) \geq \sum_{s=1}^{S} \pi_s w(y_s(a)) - c(a) \qquad \forall \pi \in \Pi \tag{IC}$$

Definition 4. *An incentive scheme w* implements a^* with inertia *in $\mathcal{A}$ if*

$$\sum_{s=1}^{S} \pi_s w(y_s(a^*)) - c(a^*) \geq \overline{w} \qquad \forall \pi \in \Pi \tag{P}$$

and for each $a \in \mathcal{A}$ with $a \neq a^*$, either

$$\sum_{s=1}^{S} \pi_s w(y_s((a^*)) - c(a^*) \geq \sum_{s=1}^{S} \pi_s w(y_s(a)) - c(a) \qquad \forall \pi \in \Pi \tag{IC}$$

or

$$\sum_{s=1}^{S} \pi_s w(y_s(a)) - c(a) \leq \overline{w} \textit{ for some } \pi \in \Pi \tag{NC}$$

To relate this model to the parameterized distribution model, given $\pi \in \Pi$ and $a \in \mathcal{A}$, define $\pi(a) \in \Delta(\mathcal{N})$ by

$$\pi_i(a) = \sum_{\{s \in \mathcal{S} : y_s(a) = y_i\}} \pi_s$$

For each a, set $\Pi(a) := \{\hat{\pi} \in \Delta(\mathcal{N}) : \hat{\pi} = \pi(a) \text{ for some } \pi \in \Pi\}$. Notice that $\Pi(a)$ is closed and convex for each a. This means assumptions about the set Π can easily be translated into assumptions about $\Pi(a)$. In particular, we show that Assumptions A-1 and A-2 can be derived from analogous assumptions on Π (the beliefs over the primitive state space).

For each $E \subset \{1, \ldots, N\}$ and each a, let $E(a) := \{s \in \mathcal{S} : y_s(a) \in E\}$. Then notice that $\hat{\pi} \in \Pi(a) \iff \exists \pi \in \Pi$ such that $\hat{\pi}(E) = \pi(E(a))$ for each $E \subset \{1, \ldots, N\}$.

Proposition 6. *Let $\nu : 2^{\mathcal{S}} \to [0,1]$ be a convex capacity on $\mathcal{S}$ and let Π be the core of ν. For each $a \in \mathcal{A}$ define $\nu^a : 2^{\mathcal{N}} \to [0,1]$ by $\nu^a(E) = \nu(E(a))$ for each $E \subset \{1, \ldots, N\}$. Then for each $a \in A$, ν^a is a convex capacity on $\{1, \ldots, N\}$ with core $\Pi(a)$.*

We say that a subset $B \subset \mathcal{S}$ is *unambiguous* if $\pi(B) = \pi'(B)$ for all $\pi, \pi' \in \Pi$. While the empty set and the entire state space $\mathcal{S}$ are always unambiguous, every non empty, proper subset of $\mathcal{S}$ is ambiguous under the assumption that Π has non empty relative interior. Here we assume that Π has non empty relative interior, and that the principal's unique prior $\pi^P \in \Delta(\mathcal{S})$ is an element of $\text{rel int } \Pi$. Under these assumptions, together with the assumption that the support of stochastic output $y(a)$ is the same for all actions a, no action eliminates ambiguity. That is, $\Pi(a)$ also has a non empty relative interior, and $\pi^P(a) \in \text{ rel int } \Pi(a)$ for each action a.

Proposition 7. *If $\pi^P \in$ rel int Π, then $\pi^P(a) \in$ rel int $\Pi(a)$ for each action a.*

These two results allow us to identify assumptions in this primitive model that will suffice to ensure that optimal contracts robust to uncertainty will be simple, as in the previous section.
Assumption A-4: Π is the core of a convex capacity ν on $\mathcal{S}$.
Assumption A-5: π^P is an element of the relative interior of Π.

Proposition 8. *Suppose Assumptions A-4 and A-5 hold, and the high action H is implementable (implementable with inertia) in $\{L, H\}$. Then, the unique optimal incentive scheme to implement H (with inertia) divides the S states into two subsets and is constant on each subset.*

6. Conclusions

We study a principal-agent model with Knightian uncertainty and asymmetric confidence in beliefs. In this model, optimal contracts must be robust to the agent's imprecise beliefs about the stochastic relationship between effort and output, as well as reflect the fact that the principal is more confident than the agent in evaluating that relationship. Our main result shows that optimal contracts must be two-valued, and are therefore simpler than the contracts that frequently obtain in the standard principal-agent model. This result also allows for the presence of a small amount of curvature in the agent's utility function. Knightian uncertainty thus introduces a force that pushes incentive schemes to be as flat as possible, while still providing incentives. This force can be sufficient to prevail over the usual motives that make contracts highly dependent on fine details of beliefs.

Appendix

Proof of Proposition 2

We prove each statement separately.
Proof of (i): Suppose not. Then $\widehat{w}(a)$ is optimal and $\sum_{j=1}^{N} \underline{\pi}_j(a)\widehat{w}_j(a) - c(a) > \overline{w}$. Reduce the payment in each state by $\varepsilon > 0$; that is, for all j, let $\widetilde{w}_j = \widehat{w}_j^a - \varepsilon$. For ε small enough, $\widetilde{w}_j$ implements a according to both definitions because it satisfies (P), (IC), and (NC). Furthermore, $\sum_{j=1}^{N} \pi_j^P(a)\widetilde{w}_j < \sum_{j=1}^{N} \pi_j^P(a)\widehat{w}_j(a) - \varepsilon$, contradicting the optimality of $\widehat{w}(a)$.
Proof of (ii): Because $\pi^P(1) \in \Pi(1)$, for any payment scheme w and any $\underline{\pi}(1) \in \underline{\Pi}(1; w)$, by definition $\sum_{j=1}^{N} \underline{\pi}_j(1)w_j \leq \sum_{j=1}^{N} \pi_j^P(1)w_j$. The constant contract $\widehat{w}_j = \overline{w} + c^1$ for all j is feasible:

it satisfies (P), and it satisfies (IC) because alternative actions are more costly for the agent. It is a solution to (1) because $\overline{w} + c^1 = \sum_{j=1}^N \underline{\pi}_j(1)\widehat{w}_j(1) = \sum_{j=1}^N \pi_j^P(1)\widehat{w}_j(1)$.

Proof of (iii): Suppose the claim does not hold. That is, $\widehat{w}(a)$ is optimal and for each $a' < a$, either

$$\sum_{j=1}^{N} \underline{\pi}_j(a)\widehat{w}_j(a) - c(a) > \sum_{j=1}^{N} \bar{\pi}_j(a')\widehat{w}_j(a) - c(a')$$

for all $\underline{\pi}(a) \in \underline{\Pi}(a;\widehat{w})$ and $\bar{\pi}(a') \in \bar{\Pi}(a';\widehat{w})$, or

$$\sum_{j=1}^{N} \pi_j(a')\widehat{w}_j(a) - c(a') < \overline{w}$$

for some $\pi(a') \in \Pi(a')$. The latter inequality implies $\sum_{j=1}^N \underline{\pi}_j(a')\widehat{w}_j(a) - c(a') < \overline{w}$ for any $\underline{\pi}(a') \in \underline{\Pi}(a';\widehat{w})$. Because none of the respective constraints binds, $\widehat{w}(a)$ is a solution for a problem like (1) where all actions like a' have been dropped from the constraints. In this new problem, a is the least costly action and a contract that pays $\overline{w} + c(a)$ in all states is optimal. That is, $\widehat{w}_j = \overline{w} + c(a)$ for each j. Thus, $\overline{w} + c(a) = \sum_{j=1}^N \underline{\pi}_j(a)\widehat{w}_j(a) = \sum_{j=1}^N \bar{\pi}_j(a')\widehat{w}_j(a) = \sum_{j=1}^N \underline{\pi}_j(a')\widehat{w}_j(a)$. Hence $c(a') > c(a)$, contradicting $a' < a$.

Proofs for Section 4

We will use the following results adapted from [7].

Lemma 1. *Let v be a convex capacity on S with core Π, and let Π^e be the set of extreme points of Π. Then, for any N-vector z such that $z_1 \leq z_2 \leq \cdots \leq z_N$,*

$$\min_{\pi\in\Pi} \sum_{j=1}^{N} \pi_j z_j$$

is attained at $\pi \in \Pi^e$ such that $\sum_{s=j}^N \pi_s = \sum_{s=j}^N v(\{s\})$ for each $j = 2, \dots, N$. In particular, if z and z' are two N-vectors such that $z_1 \leq \cdots \leq z_N$ and $z'_1 \leq \cdots \leq z'_N$, then

$$\arg\min_{\pi\in\Pi} \sum_{j=1}^{N} \pi_j z_j = \arg\min_{\pi\in\Pi} \sum_{j=1}^{N} \pi_j z'_j$$

The first result follows from [7], Propositions 10 and 13. The second follows from the first, which shows that the set of minimizing distributions depends only on the order of the elements of z and not on the values of the components.

When Assumption A-1 holds, the lemma provides a characterization of the probability distributions that, among a set, yield the smallest and largest expected values of a fixed contract. In particular, this lemma can be used to show that for each action a, the smallest expected value for a given z is attained at an extreme point of $\Pi(a)$ such that $\pi_K(a) < \pi_K^P(a)$ and $\pi_1(a) \geq \pi_1^P(a)$, where K is an index at which the maximum value of z_k is attained. Moreover, when two incentive schemes are "ordered" in the same way, their minimum or maximum expectations are attained using the same distributions.

Proof of Proposition 3

The main step is to show, by contradiction, than an optimal contract cannot be contingent on more than two subsets of output levels. Define a partition of the state space such that each element in this partition corresponds to different payments in the optimal contract $\widehat{w}$. Let k be an element of this partition; we say k is an event. Any probability distribution over the original space defines a probability distribution over this partition. Label events so that K corresponds to the largest $\widehat{w}_k$, $K-1$ to the second largest, and so on until event 1 corresponds to the smallest value of $\widehat{w}_k$. By construction, K is the number of events the optimal contract is contingent upon, and $\widehat{w}_1 < \widehat{w}_2 < \cdots < \widehat{w}_K$. We need to show K equals 2.

By Proposition 2, $\widehat{w}$ must satisfy

$$\sum_{k=1}^{K} \underline{\pi}_k(H)\widehat{w}_k = \overline{w} + c(H) \tag{4}$$

for any $\underline{\pi}(H) \in \underline{\Pi}(H;\widehat{w})$. We claim $\widehat{w}$ must also satisfy

$$\sum_{k=1}^{K} \underline{\pi}_k(L)\widehat{w}_k = \overline{w} + c(L) \tag{5}$$

for any $\underline{\pi}(L) \in \underline{\Pi}(L;\widehat{w})$. Suppose not. Then, $\sum_{k=1}^{K} \underline{\pi}_k(L)\widehat{w}_k < \overline{w} + c(L)$. Let

$$\widetilde{w} = \left(\widehat{w}_1 - \varepsilon\frac{\underline{\pi}_K(H)}{\underline{\pi}_1(H)}, \widehat{w}_2, \ldots, \widehat{w}_{K-1}, \widehat{w}_K + \varepsilon\right)$$

where ε is positive and small enough so that the ranking of the payments for $\widetilde{w}$ and $\widehat{w}$ is the same. By Lemma 1, $\underline{\pi}(H)$ and $\underline{\pi}(L)$ minimize the expected value of $\widetilde{w}$ for the agent. For any distribution π, the expected values of $\widetilde{w}$ and $\widehat{w}$ are related by the following:

$$\sum_{k=1}^{K} \pi_k\widetilde{w}_k - \sum_{k=1}^{K} \pi_k\widehat{w}_k = \frac{\pi_1\underline{\pi}_K(H) - \pi_K\underline{\pi}_1(H)}{\underline{\pi}_1(H)}\varepsilon \tag{6}$$

$\pi = \underline{\pi}(H)$ implies the right hand side of (6) is equal to 0; thus, $\widetilde{w}$ satisfies (4). For some ε close enough to zero and $\pi = \underline{\pi}(L)$ the right hand side of (6) is very small; thus $\widetilde{w}$ satisfies (NC) because $\widehat{w}$ satisfies it strictly. By Lemma 1 and Assumption A-1, $\underline{\pi}_K(H) < \pi^P_K(H)$ and $\underline{\pi}_1(H) \geq \pi^P_1(H)$; thus, the right hand side of (6) is negative when $\pi = \pi^P(H)$. Summarizing, $\widetilde{w}$ is feasible and cheaper than $\widehat{w}$, contradicting the optimality of $\widehat{w}$. Hence (5) must hold for $\widehat{w}$ to be an optimum.

We claim K must be strictly larger than 1. Suppose not, i.e., $K = 1$. If this is the case, all payments are the same, and the left hand sides of equations (4) and (5) are the same. Thus, we have $\overline{w} + c(L) = \overline{w} + c(H)$, contradicting $c(H) > c(L)$.

We claim K is not larger than 2. Suppose not. Then $\widehat{w}$ is optimal, so satisfies equations (4) and (5), and $K > 2$. Equations (4) and (5) constitute a system of two equations which can be solved for $\widehat{w}_K$ and some $\widehat{w}_{k'}$, yielding:

$$\widehat{w}_K = \frac{\underline{\pi}_{k'}(L)\left(\overline{w} + c(H)\right) - \underline{\pi}_{k'}(H)\left(\overline{w} + c(L)\right)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} + \sum_{k \neq K,k'} \frac{\underline{\pi}_{k'}(H)\underline{\pi}_k(L) - \underline{\pi}_k(H)\underline{\pi}_{k'}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}\widehat{w}_k \tag{7}$$

$$\widehat{w}_{k'} = \frac{\underline{\pi}_K(H)\left(\overline{w}+c(L)\right) - \underline{\pi}_K(L)\left(\overline{w}+c(H)\right)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} + \sum_{k\neq K,k'} \frac{\underline{\pi}_K(L)\underline{\pi}_k(H) - \underline{\pi}_K(H)\underline{\pi}_k(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}\widehat{w}_k \tag{8}$$

These are well defined unless

$$\underline{\pi}_K(L)\underline{\pi}_{k'}(H) - \underline{\pi}_{k'}(L)\underline{\pi}_K(H) = 0 \text{ for all } k' \neq K \tag{9}$$

In that case:

$$\begin{aligned}
0 &= \underline{\pi}_K(L)\sum_{k=1}^{K-1}\underline{\pi}_k(H) - \underline{\pi}_K(H)\sum_{k=1}^{K-1}\underline{\pi}_k(L) \\
&= \underline{\pi}_K(L)\left(1-\underline{\pi}_K(H)\right) - \underline{\pi}_K(H)\left(1-\underline{\pi}_K(L)\right) \\
&= \underline{\pi}_K(L) - \underline{\pi}_K(H) \\
&\Rightarrow \underline{\pi}_K(H) = \underline{\pi}_K(L)
\end{aligned}$$

Using this result in (9):

$$\underline{\pi}_K(H)\left(\underline{\pi}_{k'}(H) - \underline{\pi}_{k'}(L)\right) = 0$$

Thus, either $\underline{\pi}_K(H) = \underline{\pi}_K(L) = 0$, or $\underline{\pi}_{k'}(H) = \underline{\pi}_{k'}(L)$ for all $k' \neq K$. If $\underline{\pi}_K(H) = \underline{\pi}_K(L) = 0$, $\widehat{w}$ cannot be optimal because it makes the largest payment in a state that does not affect the constraints and, by Assumption A-2, has positive probability for the principal. If $\underline{\pi}_{k'}(H) = \underline{\pi}_{k'}(L)$ for all $k' \neq K$, then $\underline{\pi}(L) = \underline{\pi}(H)$. In this case, $\sum_{k=1}^{K}\underline{\pi}_k(H)\widehat{w}_k = \sum_{k=1}^{K}\underline{\pi}_k(L)\widehat{w}_k$ and $c(H) = c(L)$, a contradiction.

Using equations (7) and (8), we can write the expected cost of the optimal incentive scheme as follows:

$$\sum_{k=1}^{K}\pi_k^P(H)\widehat{w}_k = \widehat{\alpha} + \sum_{k\neq K,k'}\widehat{\beta}_k\widehat{w}_k$$

where

$$\begin{aligned}
\widehat{\alpha} \quad = \quad & \frac{\pi_K^P(H)\underline{\pi}_{k'}(L) - \pi_{k'}^P(H)\underline{\pi}_K(L) + \pi_{k'}^P(H)\underline{\pi}_K(H) - \pi_K^P(H)\underline{\pi}_{k'}(H)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}\overline{w} \\
& + \frac{\pi_K^P(H)\underline{\pi}_{k'}(L) - \pi_{k'}^P(H)\underline{\pi}_K(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}c(H) \\
& + \frac{\pi_{k'}^P(H)\underline{\pi}_K(H) - \pi_K^P(H)\underline{\pi}_{k'}(H)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}c(L)
\end{aligned}$$

and

$$\widehat{\beta}_k = \pi_k^P(H) - \underline{\pi}_k(H)\frac{\pi_K^P(H)\underline{\pi}_{k'}(L) - \pi_{k'}^P(H)\underline{\pi}_K(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} - \underline{\pi}_k(L)\frac{\pi_{k'}^P(H)\underline{\pi}_K(H) - \pi_K^P(H)\underline{\pi}_{k'}(H)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} \tag{10}$$

We claim that $\widehat{\beta}_{k''} \neq 0$ for some $k'' \neq K, k'$. Suppose not. Then, $\widehat{\beta}_k = 0$ for each $k \neq K, k'$. Using equation (10),

$$\begin{aligned}
\pi_k^P(H)\left(\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)\right) \quad = \quad & \underline{\pi}_k(H)\left(\pi_K^P(H)\underline{\pi}_{k'}(L) - \pi_{k'}^P(H)\underline{\pi}_K(L)\right) \\
& + \underline{\pi}_k(L)\left(\pi_{k'}^P(H)\underline{\pi}_K(H) - \pi_K^P(H)\underline{\pi}_{k'}(H)\right)
\end{aligned}$$

Summing over k, rearranging, and solving for $\pi_{k'}^P(H)$:

$$\pi_{k'}^P(H) = -\frac{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H) + \pi_K^P(H)\left(\underline{\pi}_{k'}(H) - \underline{\pi}_{k'}(L)\right)}{\underline{\pi}_K(L) - \underline{\pi}_K(H)}$$

This implies:

$$\frac{\pi_K^P(H)\underline{\pi}_{k'}(L) - \pi_{k'}^P(H)\underline{\pi}_K(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} = \frac{\underline{\pi}_K(L) - \pi_K^P(H)}{\underline{\pi}_K(L) - \underline{\pi}_K(H)}$$

and

$$\frac{\pi_{k'}^P(H)\underline{\pi}_K(H) - \pi_K^P(H)\underline{\pi}_{k'}(H)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} = \frac{\pi_K^P(H) - \underline{\pi}_K(H)}{\underline{\pi}_K(L) - \underline{\pi}_K(H)}$$

We know that $\underline{\pi}_K(H) < \pi_K^P(H)$. Thus, $\frac{\underline{\pi}_K(L)-\pi_K^P(H)}{\underline{\pi}_K(L)-\underline{\pi}_K(H)} < 1$ and $\frac{\pi_K^P(H)-\underline{\pi}_K(H)}{\underline{\pi}_K(L)-\underline{\pi}_K(H)} > 0$. Hence:

$$\begin{aligned}\sum_{k=1}^{K}\pi_k^P(H)\widehat{w}_k &= \widehat{\alpha} = \overline{w} + c(H) + \frac{\pi_K^P(H) - \underline{\pi}_K(H)}{\underline{\pi}_K(L) - \underline{\pi}_K(H)}\left(c(L) - c(H)\right) \\ &< \overline{w} + c(H)\end{aligned}$$

a contradiction.

Because $\widehat{\beta}_{k''} \neq 0$ for some $k'' \neq K, k'$, we find a feasible contract which is cheaper than $\widehat{w}$. Let $\widetilde{w}$ be defined as follows:

$$\begin{aligned}&\widetilde{w}_k = \widehat{w}_k && \text{when} \quad k \neq K, k', k'' \\ &\widetilde{w}_K = \widehat{w}_K + \tfrac{\underline{\pi}_{k'}(H)\underline{\pi}_{k''}(L) - \underline{\pi}_{k''}(H)\underline{\pi}_{k'}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}\varepsilon \\ &\widetilde{w}_{k'} = \widehat{w}_{k'} + \tfrac{\underline{\pi}_K(L)\underline{\pi}_{k''}(H) - \underline{\pi}_K(H)\underline{\pi}_{k''}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)}\varepsilon \\ &\widetilde{w}_{k''} = \widehat{w}_{k''} + \varepsilon\end{aligned}$$

where

$$|\varepsilon| < \min\left\{\widehat{w}_{k''} - \widehat{w}_{k''-1}, \widehat{w}_{k''+1} - \widehat{w}_{k''}, \widehat{w}_K - \widehat{w}_{K-1}, \widehat{w}_{k'} - \widehat{w}_{k'-1}, \widehat{w}_{k'+1} - \widehat{w}_{k'}\right\}$$

By construction, the payments in $\widetilde{w}$ and $\widehat{w}$ are ranked in the same order. Lemma 1 applies, and $\underline{\pi}(H)$ and $\underline{\pi}(L)$ yield the minimum expected values of $\widetilde{w}$ under actions H and L. Moreover,

$$\begin{aligned}0 &= \underline{\pi}_K(L)\frac{\underline{\pi}_{k'}(H)\underline{\pi}_{k''}(L) - \underline{\pi}_{k''}(H)\underline{\pi}_{k'}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} + \underline{\pi}_{k'}(L)\frac{\underline{\pi}_K(L)\underline{\pi}_{k''}(H) - \underline{\pi}_K(H)\underline{\pi}_{k''}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} + \underline{\pi}_{k''}(L) \\ 0 &= \underline{\pi}_K(L)\frac{\underline{\pi}_{k'}(H)\underline{\pi}_{k''}(L) - \underline{\pi}_{k''}(H)\underline{\pi}_{k'}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} + \underline{\pi}_{k'}(L)\frac{\underline{\pi}_K(L)\underline{\pi}_{k''}(H) - \underline{\pi}_K(H)\underline{\pi}_{k''}(L)}{\underline{\pi}_{k'}(L)\underline{\pi}_K(H) - \underline{\pi}_K(L)\underline{\pi}_{k'}(H)} + \underline{\pi}_{k''}(L)\end{aligned}$$

Hence, $\widetilde{w}$ is feasible because $\widehat{w}$ is. The expected cost of $\widetilde{w}$ is given by

$$\begin{aligned}\sum_{k=1}^{K}\pi_k^P(H)\widetilde{w}_k &= \widehat{\alpha} + \sum_{k\neq K,k'}\widehat{\beta}_k\widetilde{w}_k \\ &= \widehat{\alpha} + \sum_{k\neq K,k'}\widehat{\beta}_k\widehat{w}_k + \widehat{\beta}_{k''}\varepsilon\end{aligned}$$

Thus, we can choose $\varepsilon > 0$ whenever $\widehat{\beta}_{k''} < 0$ and $\varepsilon < 0$ whenever $\widehat{\beta}_{k''} > 0$. In either case, $\widetilde{w}$ is feasible and cheaper than $\widehat{w}$, contradicting the optimality of $\widehat{w}$. Summarizing, if an optimal contract is contingent on $K > 2$ events, we can find a feasible contract which is cheaper. Therefore, because we already proved $K < 2$ is impossible, a contract can be optimal only if $K = 2$.

Proof of Proposition 4

To simplify notation, throughout we use π^P in place of $\pi^P(H)$.

First, set

$$\begin{aligned} S := \{(A,B,C,\pi^1,\pi^2) : A,B,C \subset \mathcal{N} \text{ are pairwise disjoint,}& \\ \pi^1 \in \Pi(H), \pi^2 \in \Pi(L) \text{ are extreme points,}& \\ \pi^1(A)\pi^2(B) - \pi^1(B)\pi^2(A) &\neq 0 \\ \left|\pi^P(B)\frac{\pi^1(C)\pi^2(A)-\pi^1(A)\pi^2(C)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)}\right| + \left|\pi^P(A)\frac{\pi^1(B)\pi^2(C)-\pi^1(C)\pi^2(B)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)}\right| &\neq 0 \\ \left|\pi^P(C) + \pi^P(B)\frac{\pi^1(C)\pi^2(A)-\pi^1(A)\pi^2(C)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)} + \pi^P(A)\frac{\pi^1(B)\pi^2(C)-\pi^1(C)\pi^2(B)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)}\right| &\neq 0\} \end{aligned}$$

Notice that S is finite; we argue below that S is also nonempty.

For any pairwise disjoint events $A,B,C \subset \mathcal{N}$ and any extreme points $\pi^1 \in \Pi(H)$ and $\pi^2 \in \Pi(L)$ with $(A,B,C,\pi^1,\pi^2) \in S$, set

$$R(A,B,C,\pi^1,\pi^2) := \frac{\left|\pi^P(C) + \pi^P(B)\frac{\pi^1(C)\pi^2(A)-\pi^1(A)\pi^2(C)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)} + \pi^P(A)\frac{\pi^1(B)\pi^2(C)-\pi^1(C)\pi^2(B)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)}\right|}{\left|\pi^P(B)\frac{\pi^1(C)\pi^2(A)-\pi^1(A)\pi^2(C)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)}\right| + \left|\pi^P(A)\frac{\pi^1(B)\pi^2(C)-\pi^1(C)\pi^2(B)}{\pi^1(A)\pi^2(B)-\pi^1(B)\pi^2(A)}\right|}$$

By definition, R is well-defined and strictly positive.

Finally, set

$$b := \min_{(A,B,C,\pi^1,\pi^2)\in S} R(A,B,C,\pi^1,\pi^2)$$

Since S is finite, b is positive.

Arguments similar to the ones used in the first part of the proof of Proposition 2, with each payment variable w_k replaced by the utility variable $v_k := u(w_k)$, can be used to establish that: i) (P) must hold with equality for all $\underline{\pi} \in \underline{\Pi}(H;v)$, and ii) (NC) must hold with equality for all $\underline{\pi} \in \underline{\Pi}(H;v)$ and $\overline{\pi} \in \overline{\Pi}(L;v)$. Thus we can restrict the search for optimal contracts in the subset F of R^m defined by the set of vectors v for which (P) and (NC) hold with equality.

Suppose that $v_1,\dots,v_K$ are the agent's utility levels corresponding to an *optimal* contract with K levels $w_1 < \cdots < w_K$, i.e. $v_k := u(w_k)$, and suppose by way of contradiction that $K > 2$. Thus $v_1,\dots,v_K$ solves the following program:

$$\min_{v_1,\dots,v_K} \sum_{k=1}^{K} \pi_k^P h(v_k)$$

subject to (P) and (NC).

Claim 1. *For any $v \in F$ there exists $j \in \{1,\dots,K\}$ such that (NC) and (P) can be solved for (v_j, v_K), i.e.*

$$\begin{aligned} v_j &= \frac{\underline{\pi}_K(w_0+c_L) - \overline{\pi}_K(w_0+c_H)}{\underline{\pi}_K\overline{\pi}_j - \underline{\pi}_j\overline{\pi}_K} + \sum_{k\neq j,K} \frac{\underline{\pi}_k\overline{\pi}_K - \underline{\pi}_K\overline{\pi}_k}{\underline{\pi}_K\overline{\pi}_j - \underline{\pi}_j\overline{\pi}_K} v_k = \alpha_j + \sum_{k\neq j,K}\sigma_{jk}v_k \\ v_K &= \frac{\underline{\pi}_j(w_0+c_L) - \overline{\pi}_j(w_0+c_H)}{\underline{\pi}_j\overline{\pi}_K - \underline{\pi}_K\overline{\pi}_j} + \sum_{k\neq j,K} \frac{\underline{\pi}_j\overline{\pi}_k - \underline{\pi}_k\overline{\pi}_j}{\underline{\pi}_K\overline{\pi}_j - \underline{\pi}_j\overline{\pi}_K} v_k = \alpha_K + \sum_{k\neq j,K}\sigma_{Kk}v_k \end{aligned} \tag{11}$$

where $\underline{\pi}_K \overline{\pi}_j - \underline{\pi}_j \overline{\pi}_K \neq 0$, and $\underline{\pi} \in \underline{\Pi}(L;v), \overline{\pi} \in \overline{\Pi}(H;v)$ are extreme points.

Proof of Claim 1. See the part of the proof of Proposition 2 which begins with "These are well defined unless" immediately before formula (9) (here we have replaced k' with j).

Substituting the expressions in (11) into the objective yields

$$\sum_{i=1}^{K} \pi_k^P \ h(v_i) = \sum_{k \neq j,K} \pi_k^P \ h\left(v_k\right) + \pi_j^P \ \ h(\underbrace{\alpha_j + \sum_{k \neq j,K} \sigma_{jk} v_k}_{v_j}) + \pi_K^P \ \ h(\underbrace{\alpha_K + \sum_{k \neq j,K} \sigma_{Kk} v_k}_{v_K}).$$

Since the given contract $v_1, \ldots, v_K$ is optimal it must minimize the last expression, hence satisfy the first order condition (FOC):

$$\forall k \neq j, K:$$

$$\begin{aligned} 0 &= \pi_k^P \ h'\left(v_k\right) + \pi_j^P \ \sigma_{jk} \ \ h'(v_j) + \pi_K^P \ \sigma_{Kk} \ \ h'(v_K) \\ &= \underbrace{\left[\pi_k^P + \pi_j^P \ \sigma_{jk} \ + \pi_K^P \ \sigma_{Kk}\right] h'\left(v_k\right)}_{T_1} \ + \ \underbrace{\pi_j^P \ \sigma_{jk} \ \left[h'(v_j) - h'(v_k)\right] + \pi_K^P \ \sigma_{Kk} \ \left[h'(v_K) - h'(v_k)\right]}_{T_2} \end{aligned}$$

Claim 2. *Since* $\pi^P \neq \underline{\pi}$*, for any* $v \in F$ *there exists* $k_* \neq j, K$ *such that the first term in square brackets*

$$T_1 = \pi_k^P + \pi_j^P \ \sigma_{jk} \ + \pi_K^P \ \sigma_{Kk}$$

is not zero.

Proof of Claim 2. See the part of the proof of Proposition 2 which establishes that $\widehat{\beta}_{kk} \neq 0$ for some $k'' \neq k, K$.

Remark. Given Claim 2, the FOC is violated (for v_{k_*}) if h' is constant (i.e. the agent's utility function is linear); or if h' does not vary too much, so that

$$\left|T_1\right| > \left|T_2\right|.$$

The remainder of the argument verifies that this holds provided $g'(v_M) < b$.

Note that by Lemma 1, the two sets of critical probabilities $\underline{\Pi}\left(H;w\right)$ and $\overline{\Pi}\left(L;w\right)$ do not change if the order of the payments in the contract does not change.

By Claims 1 and 2 we can conclude that for any $v \in F$ there exists $k_*\left(v\right)$ such that the first term in square brackets (T_1) is not zero.

Then the FOC can be written as

$$\begin{aligned} \forall k \neq j, K: \qquad\qquad 0 \ &= \ \left[\pi_k^P + \pi_j^P \ \sigma_{jk} \ + \pi_K^P \ \sigma_{Kk}\right] \left[1 + g'\left(v_k\right)\right] \\ &\quad + \pi_j^P \ \sigma_{jk} \ \left[g'(v_j) - g'(v_k)\right] + \pi_K^P \ \sigma_{Kk} \ \left[g'(v_K) - g'(v_k)\right] \end{aligned}$$

Recall that h is unique up to affine transformations. Thus without loss of generality we can set $g'\left(\overline{u}\right) = 0$, where $\overline{u}$ denotes the agent's reservation utility, i.e. $\overline{u} := u\left(\overline{w}\right)$. This together with the convexity of g implies $g'\left(u\right) \geq 0$ for all $u > u\left(\overline{w}\right)$; and in particular $g'\left(v_k\right) \geq 0$. Thus it suffices to show that the first term in square brackets is larger in absolute value than the term on the last line (T_2) in absolute value.

Since by assumption $g'(v_M) < b$, we have

$$\left|\pi_j^P\ \sigma_{jk}\ [g'(v_j) - g'(v_k)]\ +\ \pi_K^P\ \sigma_{Kk}\ [g'(v_K) - g'(v_k)]\right|$$

$$\begin{aligned}
&\leq \left|\pi_j^P\ \sigma_{jk}\ [g'(v_j) - g'(v_k)]\right|\ +\ \left|\pi_K^P\ \sigma_{Kk}\ [g'(v_K) - g'(v_k)]\right| && \text{(triangle inequality)}\\
&= \left|\pi_j^P\ \sigma_{jk}\right| |g'(v_j) - g'(v_k)|\ +\ \left|\pi_K^P\ \sigma_{Kk}\right| |g'(v_K) - g'(v_k)| && |xy| = |x|\,|y|\\
&\leq \left[\left|\pi_j^P\ \sigma_{jk}\right|\ +\ \left|\pi_K^P\ \sigma_{Kk}\right|\right] [g'(v_M) - g'(\overline{u})] && \text{(convexity of } g)\\
&= \left[\left|\pi_j^P\ \sigma_{jk}\right|\ +\ \left|\pi_K^P\ \sigma_{Kk}\right|\right] g'(v_M) && (g'(\overline{u}) = 0)\\
&< \left[\left|\pi_j^P\ \sigma_{jk}\right|\ +\ \left|\pi_K^P\ \sigma_{Kk}\right|\right] \frac{\left|\pi_k^P + \pi_j^P\ \sigma_{jk}\ + \pi_K^P\ \sigma_{Kk}\right|}{\left|\pi_j^P\ \sigma_{jk}\right|\ +\ \left|\pi_K^P\ \sigma_{Kk}\right|} && (g'(v_M) < b)\\
&= \left|\pi_k^P + \pi_j^P\ \sigma_{jk}\ + \pi_K^P\ \sigma_{Kk}\right|
\end{aligned}$$

Since this is a contradiction, the contract could not be optimal and the result is established.

Proof of Proposition 5

Again we prove the result for the case of implementation with inertia; the other case is analogous. Let $\widehat{w}$ be an optimal contract that implements a^* with inertia. Without loss of generality, label payments so that $\widehat{w}_N$ corresponds to the highest, $\widehat{w}_{N-1}$ the second highest, and so on. We claim there exists only one action $a' \neq a^*$ for which the incentive compatibility constraint binds, and $a' < a^*$. Suppose not. Then, there exist two actions a' and a'' different from a^* such that

$$\sum_{j=1}^{N} \underline{\pi}_j(a'')\widehat{w}_j - c(a'') = \sum_{j=1}^{N} \underline{\pi}_j(a')\widehat{w}_j - c(a') = \sum_{j=1}^{N} \underline{\pi}_j(a^*)\widehat{w}_j - c(a^*) = \overline{w}$$

for any $\underline{\pi}(a^*) \in \underline{\Pi}(a^*;\widehat{w}), \underline{\pi}(a') \in \underline{\Pi}(a';\widehat{w})$, and $\underline{\pi}(a'') \in \underline{\Pi}(a'';\widehat{w})$. By construction $\underline{\pi}(a^*)$, $\underline{\pi}(a')$, and $\underline{\pi}(a'')$ can be taken to be extreme points of the corresponding belief sets. From here on, following the argument in [10] establishes the claim.

If only the constraint relative to one action a'^* binds, $\widehat{w}$ must also be optimal in a problem where all other actions $a \neq a'^*$ are dropped from the constraints. Therefore, Proposition 3 applies to that problem and the optimal contract has a two-wage structure.

Proofs for Section 5

Proof of Proposition 6

Fix $a \in \mathcal{A}$. Let $E, F \subset \{1, \ldots, N\}$. Then

$$\begin{aligned}
\nu^a(E) + \nu^a(F) &= \nu(E(a)) + \nu(F(a))\\
&\leq \nu(E(a) \cap F(a)) + \nu(E(a) \cup F(a))\\
&= \nu((E \cap F)(a)) + \nu((E \cup F)(a)\\
&= \nu^a(E \cap F) + \nu^a(E \cup F)
\end{aligned}$$

Thus ν^a is a convex capacity. Moreover, if $\hat{\pi} \in \Pi(a)$, then $\hat{\pi} = \pi(a)$ for some $\pi \in \Pi$. For each $\hat{\pi} \in \Pi(a)$, $\hat{\pi}(E) = \pi(E(a)) \geq \nu(E(a)) = \nu^a(E)$. Thus $\Pi(a)$ is a subset of the core of ν^a. Since $\Pi(a)$ is closed and convex, it suffices to show that $\Pi(a)$ contains all of the "marginal contribution" vectors for ν^a, that is, any vector π of the form $\pi_j = \nu^a(\{\sigma(1), \ldots, \sigma(j)\}) - \nu^a(\{\sigma(1), \ldots, \sigma(j-1)\})$ where σ is a permutation on $\{1, \ldots, N\}$. Without loss of generality, consider the identity permutation and corresponding vector π in which $\pi_j = \nu^a(\{1, \ldots, j\}) - \nu^a(\{1, \ldots, j-1\})$ for each j. Thus for each j, $\pi_j = \nu(\{y_1, \ldots, y_j\}(a)) - \nu(\{y_1, \ldots, y_{j-1}\}(a))$. For each j, set $\{s_1^j, \ldots, s_{k_j}^j\} := \{y_j\}(a) = \{y_1, \ldots, y_j\}(a) \setminus \{y_1, \ldots, y_{j-1}\}(a)$. Define $\bar{\pi} \in \Delta(\mathcal{S})$ as follows. Set

$$\bar{\pi}(s_1^1) = \nu(\{s_1^1\})$$

and for each $k \neq 2, \ldots, k_1$, set

$$\bar{\pi}(s_k^1) = \nu(\{s_1^1, \ldots, s_k^1\}) - \nu(\{s_1^1, \ldots, s_{k-1}^1\})$$

For $j = 2, \ldots, N$, similarly define

$$\pi(s_1^j) = \nu(\{y_1, \ldots, y_{j-1}\}(a) \cup \{s_1^j\}) - \nu(\{y_1, \ldots, y_{j-1}\}(a))$$

and for $k = 2, \ldots, k_j$,

$$\bar{\pi}(s_k^j) = \nu(\{y_1, \ldots, y_{j-1}\}(a) \cup \{s_1^j, \ldots, s_k^j\}) - \nu(\{y_1, \ldots, y_{j-1}\}(a) \cup \{s_1^j, \ldots, s_{k-1}^j\})$$

Then $\bar{\pi}$ is an element of the core of ν, since it is the marginal contribution vector corresponding to the permutation $(s_1^1, \ldots, s_{k_1}^1, \ldots, s_1^N, \ldots, s_{k_N}^N)$. Thus $\bar{\pi} \in \Pi$. By construction, for each j,

$$\bar{\pi}(\{y_j\}(a)) = \sum_{k=1}^{k_j} \bar{\pi}(s_k^j) = \nu(\{y_1, \ldots, y_j\}(a)) - \nu(\{y_1, \ldots, y_{j-1}\}(a)) = \pi_j$$

Thus $\pi \in \Pi(a)$, and the claim is established.

Proof of Proposition 7

Fix an action a. For each $j = 1, \ldots, N$, $\{y_j\}(a) \subset \mathcal{S}$ is a non empty and proper subset of $\mathcal{S}$, since $y(a)$ has support $\{y_1, \ldots, y_N\}$. Thus for a non empty, proper subset $E \subset \{1, \ldots, N\}$, $E(a) \subset \mathcal{S}$ must also be non empty and proper. By assumption,

$$\min_{\pi \in \Pi} \pi(E(a)) < \pi^P(E(a)) < \max_{\pi \in \Pi} \pi(E(a))$$

>From this we conclude

$$\min_{\pi \in \Pi(a)} \pi(E) < \pi_j^P(a)(E) < \max_{\pi \in \Pi(a)} \pi(E)$$

Thus the claim is established.

References

[1] Robert J. Aumann. Utility theory without the completeness axiom. *Econometrica*, 30:445–462, 1962.

[2] Truman F. Bewley. Knightian decision theory: Part i. Technical report, Cowles Foundation, 1986.

[3] Truman F. Bewley. Knightian decision theory: Part ii. Technical report, Cowles Foundation, 1987.

[4] Truman F. Bewley. Market innovation and entrepreneurship: A knightian view. Technical report, Cowles Foundation, 1989.

[5] Truman F. Bewley. Knightian decision theory: Part i. *Decisions in Economics and Finance*, 2:79–110, 2002.

[6] L.M. de Campos, J.F. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2(5):167–196, 1994.

[7] Alain Chateauneuf and Jean-Yves Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.

[8] Paolo Ghirardato. Agency theory with uncertainty aversion. Technical report, Caltech, 1994.

[9] Paolo Ghirardato, Fabio Maccheroni, Massimo Marinacci, and Marciano Siniscalchi. A subjective spin on roulette wheels. *Econometrica*, 71:1897–1908, 2003.

[10] Sanford J. Grossman and Oliver D. Hart. An analysis of the principal-agent problem. *Econometrica*, 51:7–45, 1983.

[11] Oliver D. Hart and Bengt Holmstrom. The theory of contracts. In Truman F. Bewley, editor, *Advances in Economic Theory*. 1987.

[12] Bengt Holmstrom. Moral hazard and observability. *Bell Journal of Economics*, 10:74–91, 1987.

[13] Bengt Holmstrom and Paul Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55:303–328, 1987.

[14] Edi Karni. Agency theory: Choice-based behavioral foundations of the parametrized distribution formulation. *Economic Theory*, 36:337–351, 2008.

[15] Frank H. Knight. *Uncertainty and Profit*. Houghton Mifflin, Boston, 1921.

[16] Giuseppe Lopomo, Luca Rigotti, and Chris Shannon. Uncertainty in mechanism design. *mimeo*, 2009.

[17] Sujoy Mukerji. Ambiguity and contractual forms. Technical report, Yale University, 1998.

[18] Sujoy Mukerji. Ambiguity aversion and incompleteness of contractual form. *American Economic Review*, 88:1207–1231, 1998.

[19] Luca Rigotti and Chris Shannon. Uncertainty and risk in financial markets. *Econometrica*, 73:203–243, 2005.

[20] Luca Rigotti, Chris Shannon, and Tomasz Strzalecki. Subjective beliefs and ex-ante trade. *Econometrica*, 76(5):1167–1190, 2008.

[21] David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.

[22] Lloyd S. Shapley and Manel Baucells. Multiperson utility. Technical report, University of California, Los Angeles, 1998.

[23] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 1953.