

# Knightian Uncertainty and Moral Hazard\*

Giuseppe Lopomo<sup>†</sup>  
Fuqua School of Business  
Duke University

Luca Rigotti<sup>‡</sup>  
Department of Economics  
University of Pittsburgh

Chris Shannon<sup>§</sup>  
Department of Economics  
UC Berkeley

August 2010

## 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 Bewley (1986). In this setting, incentives must be robust to Knightian uncertainty. We study whether robustness leads to simplicity of the resulting optimal contracts. Under mild conditions, we show that optimal contracts have a two-wage structure. That is, all output levels are divided into two groups, and the optimal wage contract pays the same amount for all output levels in each group. This can be interpreted as a flat payment plus bonus contract.

*JEL Codes:* D82, D81, D23

*Keywords:* Knightian uncertainty, moral hazard, contract theory, incomplete preferences.

---

\*A previous version of this paper was titled Imprecise Beliefs in a Principal Agent Model

<sup>†</sup>E-mail: [glopomo@duke.edu](mailto:glopomo@duke.edu)

<sup>‡</sup>E-mail: [luca@pitt.edu](mailto:luca@pitt.edu)

<sup>§</sup>E-mail: [cshannon@econ.berkeley.edu](mailto:cshannon@econ.berkeley.edu)

# 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, the agent is entering a new environment and cannot evaluate precisely the consequences of his effort because output depends on his effort as well as variables beyond his control like other workers’s input, pure luck, and so on, while the principal owns the production technology and can evaluate the relation between effort and output more precisely. In these cases, the agent may not be able to assess precisely the stochastic relation between effort and output. The standard literature neglects this possibility by assuming that both principal and agent evaluate state-contingent contracts using expected utility, and hence can precisely evaluate the stochastic consequences of the agent’s action. This contrast is easily modeled using the model of Knightian uncertainty developed in Bewley (1986), however, which provides foundations for a model in which beliefs are given by a set of probability distributions over output. 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 shows that in an otherwise standard principal-agent model with Knightian uncertainty, optimal contracts are simple under relatively mild conditions. In particular, we show that a contract can be optimal *only* if it takes two values across all states. The optimal incentive scheme divides all possible output levels into two groups, and pays a fixed wage within each group. This result is striking because the standard moral hazard model typically generates complex incentive structures. Optimal contracts often involve as many different payments as there are possible levels of output. Casual empiricism, on the other hand, suggests that many contracts are quite simple. For example, many labor contracts have a simple two-wage structure: a flat payment plus an “incentive bonus” at the end of the year.

Some authors have speculated that contracts are simple because they need to be robust. Hart and Holmstrom (1987), for example, argue that real world incentives need to perform well across a wider range of circumstances. Once this need for robustness is considered, simple optimal schemes might obtain. We introduce a particular form of robustness: Knightian uncertainty. Thus, an incentive scheme must be robust to the agent’s multiple beliefs. We show that, under mild conditions, the optimal scheme must have a two-valued structure. In this sense, Knightian uncertainty generates simple contracts. Furthermore, these contracts have a shape we commonly observe in practice.<sup>1</sup>

In the spirit of Bewley (1986), the agent’s preferences are not necessarily complete; he may be unable to rank all pairs of alternatives offered to him. As in the model axiomatized in Bewley (1986), 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. Bewley (1986) argues that this approach formalizes the distinction between risk and uncertainty introduced in Knight (1921). In this framework a unique

---

<sup>1</sup>Holmstrom and Milgrom (1987) 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.

probability is appropriate only 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.<sup>2</sup> Bewley (1986) 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 abstract settings, identifying a plausible status quo element can be hard. In the standard moral hazard setting, however, 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 that optimal contracts must be simple.

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 assume that principal and agent have linear utility functions over money. 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. Thus multiplicity of the agent's beliefs is the only formal difference between our model and a linear utility version of Grossman and Hart (1983). It is important to realize, however, that with Knightian uncertainty linear utility does not imply risk neutrality. In the expected utility model, linear utility rules out preference for mixing and is equivalent to risk neutrality. With Knightian uncertainty and linear utility, on the other hand, one sees the preference for mixing that is usually associated with risk aversion.

Mukerji (1998a) and Ghirardato (1994) 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, the principal and the agent are Choquet expected utility maximizers (see Schmeidler (1989)). Optimal incentive schemes need not be simple in those settings. Neither examines the role of asymmetric confidence, as in each model the principal and the agent share the same belief set. On the other hand, Mukerji (1998b) 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 shows that optimal incentive schemes are simple. Section 5 introduces a primitive model and establishes versions of our main results in this setting.

---

<sup>2</sup>In a related paper, Lopomo, Rigotti, and Shannon (2009), 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.

## 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.<sup>3</sup> This idea was pursued by Aumann (1962) who proposes a preference representation theorem for incomplete preferences and objective probabilities.<sup>4</sup> In a series of papers, Bewley (1986), (1987), and (1989) develops Knightian decision theory, a model which allows for subjective probabilities and incompleteness.<sup>5</sup>

### 2.1 Preference Representation without Completeness

Under completeness, any pair of alternatives can be ranked. If preferences are not complete, some alternatives are incomparable. Bewley (1986) 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  $i = 1, \dots, N$ .  $\Delta(\mathcal{N}) := \{\pi \in \mathbf{R}^N : \pi_i \geq 0 \forall i, \sum_i \pi_i = 1\}$  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, Bewley (1986) characterizes incomplete preference relations represented by a unique nonempty, closed, convex set of probability distributions  $\Pi$  and a continuous, strictly increasing, concave function  $u : X \rightarrow \mathbf{R}$ , unique up to positive affine transformations, such that

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

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

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

where  $E_\pi[\cdot]$  denotes the expected value with respect to the probability distribution  $\pi$ , and  $u(x)$  denotes the vector  $(u(x_1), \dots, u(x_N))$ .<sup>6</sup>

The set of probabilities  $\Pi$  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 preferred to another if and only if it is preferred under every probability distribution considered by the agent.<sup>7</sup> Bewley (1986) suggests this approach formalizes the distinction between risk and

---

<sup>3</sup>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.” von Neumann and Morgenstern (1953), Section 3.3.4, pg. 19.

<sup>4</sup>Aumann’s work has been extended and clarified by ? and Shapley and Baucells (1998).

<sup>5</sup>Bewley (1986) has been published recently as Bewley (2002).

<sup>6</sup>This representation has been extended by Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003) in a Savage setting.

<sup>7</sup>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  $\Pi$ .

uncertainty originating in Knight (1921). Informally, the size of  $\Pi$  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.

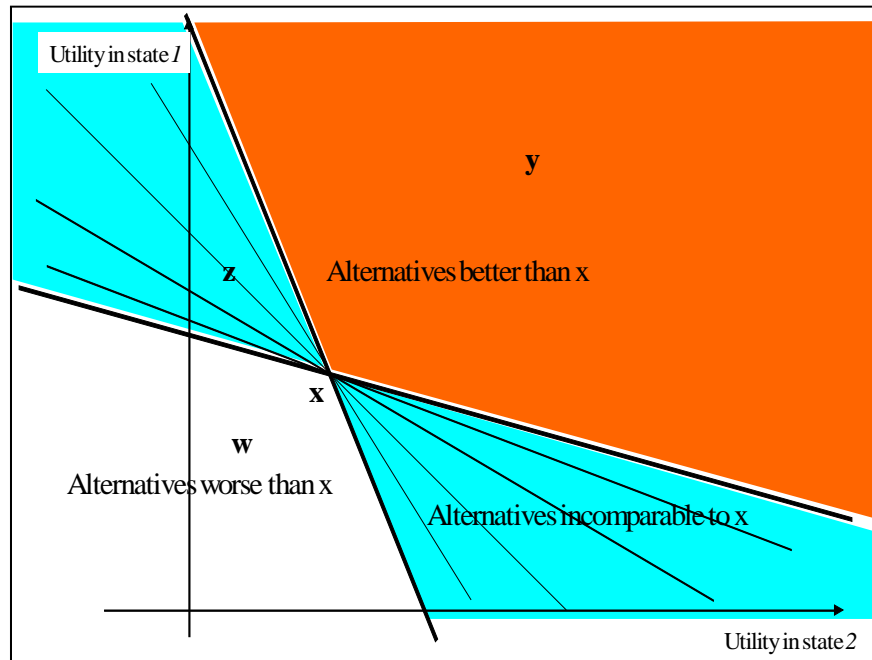


Figure 1: Incomplete Preferences

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  $\Pi$ , and vanishes only when  $\Pi$  is a singleton. As a consequence, the assumption of linear utility generates different behavior with multiple probabilities than in the standard expected utility model. There, linear utility implies a constant marginal rate of substitution and no preference for mixing. Neither of these features follows when there is Knightian uncertainty. In particular, the preference for mixing that is usually associated with risk aversion is

exhibited. 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 which 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 Bewley (1986) 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  $y = (y_1, \dots, y_N)$ , with states labeled so that  $y_N > \dots > y_1$  (throughout, we use subscripts to denote states). A contract is an  $N$ -vector  $w = (w_1, \dots, w_N)$ , where  $w_j$  is the payment from the principal to the agent in state  $j$ . The agent’s reservation utility is  $\bar{w}$ ; we will interpret it as the status quo when we impose the inertia assumption. The agent chooses an action  $a$  from the finite set  $\mathcal{A} := \{1, 2, \dots, M\}$ . The principal’s beliefs are described by a function  $\pi^P : \mathcal{A} \rightarrow \Delta$ , while the agent’s beliefs are described by a correspondence  $\Pi : \mathcal{A} \rightarrow 2^\Delta$ , where  $\Pi(a)$  is nonempty, closed and convex for each  $a$ . In the parametrized distribution formulation of the principal-agent problem pioneered by Holmstrom (1987)  $\Pi$  is a function and is identical to  $\pi^P$ .<sup>8</sup>

Our main assumption is 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  $\pi^P$  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  $a > a'$ ,

---

<sup>8</sup>We discuss in section 5 some undesirable effects of this formulation with Knightian uncertainty.

$c(a) > c(a')$  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.

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 risk-averse. Both parties have the same beliefs, and the same attitude toward uncertainty, but they evaluate risk in different ways. In our model, both parties have consistent beliefs and evaluate risk in the same way, but have different attitudes toward uncertainty. Notice that even though the agent's utility function is linear, the agent displays preference for mixing, a feature usually interpreted as a fundamental trait of risk-aversion. Therefore, comparing our results to the standard model in which the agent is risk-averse seems more appropriate than comparing to the risk-neutral case.

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  $C^{FB}(a) = \bar{w} + c(a)$  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.<sup>9</sup>

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

---

<sup>9</sup>Rigotti and Shannon (2005) give a general characterization of Pareto optimal allocations with incomplete preferences.

actions.<sup>10</sup> How this is done depends on whether or not the inertia assumption holds.

### 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:*

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

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

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

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. Definition 1 can be very restrictive, and in Section 5 we show that there are situations in which few actions can be implemented.

### 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*

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

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

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

---

<sup>10</sup>In Lopomo, Rigotti, and Shannon (2009) 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 Lopomo, Rigotti, and Shannon (2009).

or

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

$NC$  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.

### 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  $\hat{w}(a)$  is an *optimal incentive scheme to implement  $a$*  if it is a solution to

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

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

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

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  $E_{\underline{\pi}}[w] - c(a) \geq \bar{w}$  for every  $\underline{\pi} \in \underline{\Pi}(a; w)$ . 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

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

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 1.** *Let  $a$  be the action to be implemented.*

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

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

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

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

(iii) *If  $\hat{w} \in \widehat{W}^I(a)$  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*

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

*or*

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

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

*Similarly, if  $\hat{w} \in \widehat{W}(a)$  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 Are Simple

Our main result shows that a contract can be optimal only if it is two-valued: it divides the states into two groups, and pays the same amount in all states belonging to the each group. Notice that

when the optimal contract is monotone, which follows from standard assumptions, it is not only simple but also very commonly observed: the optional contract consists of a fixed payment plus a bonus if output exceeds a given amount.

Our results suggest Knightian uncertainty is a possible reason for 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 our two-valued contract 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 contrast, our main result shows that a two-tier contract is the *unique* optimal contract with Knightian uncertainty. Therefore, adding Knightian uncertainty to the standard model selects a unique, simple contract. In this sense, Knightian uncertainty can explain why optimal contracts *must* be simple. Moreover, it is not clear that the appropriate standard benchmark for our model is the expected utility model with risk neutrality. As we argued in Section 2, the agent’s better-than sets are not half-spaces but proper cones, since the preferences have a “kink” at the bundle being evaluated. Calling the agent risk neutral in this case is misleading, as he exhibits the preference for mixing that in a standard setting is the definition of risk aversion.

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

The main result depends on two properties of beliefs.

**Assumption A-1:** For each action  $a$  in  $A$ ,  $\Pi(a)$  is the core of a convex capacity  $v^a$  on  $\mathcal{N}$ .<sup>11</sup> That is,

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

Geometrically,  $\Pi(a)$  must be a polyhedron whose boundaries are determined by the linear inequalities in (3).<sup>12</sup> Figure 2 displays some examples with three states:  $\Pi^A$  and  $\Pi^B$  are cores of convex capacities while  $\Pi^C$  is not.

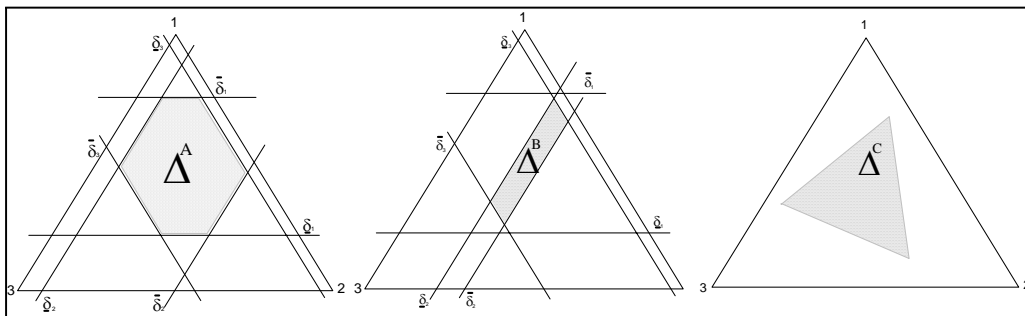


Figure 2: Agent’s Beliefs with 3 Output States

<sup>11</sup>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)  $\forall E, E' \subset \mathcal{N}$ ,  $\nu(E \cup E') \geq \nu(E) + \nu(E') - \nu(E \cap E')$ .

<sup>12</sup>In the literature on Choquet expected utility (see Schmeidler (1989)), convexity is assumed in almost all applications. For example, both Mukerji (1998a) and Ghirardato (1994) assume the beliefs of both parties involved in a moral hazard model are represented by convex capacities.

**Assumption A-2:** For each action  $a \in \mathcal{A}$ ,  $\pi^P(a)$  is an element of the relative interior of  $\Pi(a)$ .<sup>13</sup>

This requires that the agent perceives sufficiently rich uncertainty so that  $\Pi(a)$  has a non empty relative interior for each action  $a$ .

## 4.1 Simplicity 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 1, 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 2.** *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.*

The inertia assumption plays no role in this result. We provide the details of the proof for the inertia case because the other case can be proved very similarly. Intuitively, this result follows from a simple consequence of assumption A-2. 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 in this model, the optimal contract must “accommodate” this probability and doing so is costly to the principal. Since this is true for all states, 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.

## 4.2 Simplicity 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 2 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 must 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

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

---

<sup>13</sup>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  $\pi \in O \cap \Delta(\mathcal{N}) \subset \Pi(a)$ .

satisfies the monotone likelihood ratio property and concavity of the distribution function. That is, if  $\pi^e : \mathcal{A} \rightarrow \Delta(\mathcal{N})$  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  $j$ .

**Proposition 3.** *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 A Primitive Model

Without the inertia assumption, an incentive scheme implements a particular action if the agent prefers that action to his reservation utility and to all other actions. Using the parametrized distribution approach, this implies a costlier action cannot be implemented if its belief set intersects any of the belief sets corresponding to cheaper actions. In many interesting situations, however, the agent's beliefs intersect. For example, if the agent chooses the lowest effort action, his beliefs may be extremely imprecise but the harder the agent works, the more precisely he evaluates his influence on the production process. In this case, all agent's belief sets intersect and no action can be implemented.

This observation means that it is important to understand the connection between the parameterized distribution model and a primitive model in which actions and states determine output jointly, while the probability distribution over states is given. In the expected utility framework, the primitive model and its parameterized distribution version can be connected easily, and the former is just a shortcut for the latter. Under Knightian uncertainty, this connection is less clear. In particular, establishing that our main assumptions about beliefs have a corresponding meaning in the primitive world is important, as is understanding how action-based beliefs over outcomes are derived from a fixed set of beliefs over states. This is what we do next.

Output is determined by a production function which depends on the agent's effort and on stochastic factors. Let  $\mathcal{S}$  be the state space, with elements  $s = 1, \dots, S$ . For each action  $a \in \mathcal{A}$ ,  $y(a) : \mathcal{S} \rightarrow \mathbf{R}$  denotes the stochastic output that results from the action  $a$ . For each action  $a, a'$ , we assume  $\text{supp } y(a) = \text{supp } y(a') := \{y_1, \dots, y_N\}$  (different output levels across different 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  $c : \mathcal{A} \rightarrow \mathbf{R}$  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:

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

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

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

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

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

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

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

or

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

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  $\Pi$  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  $\Pi$  (the beliefs over the primitive state space).

For each  $E \subset \{1, \dots, N\}$  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  $E \subset \{1, \dots, N\}$ .

**Proposition 4.** Let  $\nu : 2^{\mathcal{S}} \rightarrow [0, 1]$  be a convex capacity on  $\mathcal{S}$  and let  $\Pi$  be the core of  $\nu$ . For each  $a \in \mathcal{A}$  define  $\nu^a : 2^{\mathcal{N}} \rightarrow [0, 1]$  by  $\nu^a(E) = \nu(E(a))$  for each  $E \subset \{1, \dots, N\}$ . Then for each  $a \in \mathcal{A}$ ,  $\nu^a$  is a convex capacity on  $\{1, \dots, N\}$  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  $\Pi$  has non empty relative interior. Here we assume that  $\Pi$  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 5.** *If  $\pi^P \in \text{rel int } \Pi$ , then  $\pi^P(a) \in \text{rel int } \Pi(a)$  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:**  $\Pi$  is the core of a convex capacity  $\nu$  on  $\mathcal{S}$ .

**Assumption A-5:**  $\pi^P$  is an element of the relative interior of  $\Pi$ .

**Proposition 6.** *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.*

## Appendix

### Proof of Proposition 1

We prove each statement separately.

*Proof of (i):* Suppose not. Then  $\widehat{w}(a)$  is optimal and  $\sum_{j=1}^N \underline{\pi}_j(a) \widehat{w}_j(a) - c(a) > \bar{w}$ . Reduce the payment in each state by  $\varepsilon > 0$ ; that is, for all  $j$ , let  $\widetilde{w}_j = \widehat{w}_j^a - \varepsilon$ . For  $\varepsilon$  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  $\widehat{w}_j = \bar{w} + c^1$  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  $\bar{w} + c^1 = \sum_{j=1}^N \underline{\pi}_j(1) \widehat{w}_j(1) = \sum_{j=1}^N \pi_j^P(1) \widehat{w}_j(1)$ .

*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

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

for some  $\pi(a') \in \Pi(a')$ . The latter inequality implies  $\sum_{j=1}^N \underline{\pi}_j(a') \widehat{w}_j(a) - c(a') < \bar{w}$  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  $\bar{w} + c(a)$  in all states is optimal. That is,  $\hat{w}_j = \bar{w} + c(a)$  for each  $j$ . Thus,  $\bar{w} + c(a) = \sum_{j=1}^N \pi_j(a) \hat{w}_j(a) = \sum_{j=1}^N \bar{\pi}_j(a') \hat{w}_j(a) = \sum_{j=1}^N \pi_j(a') \hat{w}_j(a)$ . Hence  $c(a') > c(a)$ , contradicting  $a' < a$ .

## Proofs for Section 4

We will use the following results adapted from Chateauneuf and Jaffray (1989).

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

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

*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  $z_1 \leq \dots \leq z_N$  and  $z'_1 \leq \dots \leq z'_N$ , 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 Chateauneuf and Jaffray (1989), 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 2

The main step is to show, by contradiction, that 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  $\hat{w}$ . 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  $\hat{w}_k$ ,  $K - 1$  to the second largest, and so on until event 1 corresponds to the smallest value of  $\hat{w}_k$ . By construction,  $K$  is the number of events the optimal contract is contingent upon, and  $\hat{w}_1 < \hat{w}_2 < \dots < \hat{w}_K$ . We need to show  $K$  equals 2.

By Proposition 1,  $\widehat{w}$  must satisfy

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

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

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

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

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

where  $\varepsilon$  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  $\pi$ , the expected values of  $\widetilde{w}$  and  $\widehat{w}$  are related by the following:

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

$\pi = \underline{\pi}(H)$  implies the right hand side of (6) is equal to 0; thus,  $\widetilde{w}$  satisfies (4). For some  $\varepsilon$  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,  $\underline{\pi}_K(H) < \pi_K^P(H)$  and  $\underline{\pi}_1(H) \geq \pi_1^P(H)$ ; 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  $\bar{w} + c(L) = \bar{w} + c(H)$ , 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:

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

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

These are well defined unless

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

In that case:

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

Using this result in (9):

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

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

$$\begin{aligned}
\widehat{\alpha} &= \frac{\pi_K^P(H) \underline{\pi}_{k'}(L) - \pi_{k'}^P(H) \underline{\pi}_K(L) + \pi_{k'}^P(H) \underline{\pi}_K(H) - \pi_K^P(H) \underline{\pi}_{k'}(H)}{\underline{\pi}_{k'}(L) \underline{\pi}_K(H) - \underline{\pi}_K(L) \underline{\pi}_{k'}(H)} \frac{1}{w} \\
&+ \frac{\pi_K^P(H) \underline{\pi}_{k'}(L) - \pi_{k'}^P(H) \underline{\pi}_K(L)}{\underline{\pi}_{k'}(L) \underline{\pi}_K(H) - \underline{\pi}_K(L) \underline{\pi}_{k'}(H)} c(H) \\
&+ \frac{\pi_{k'}^P(H) \underline{\pi}_K(H) - \pi_K^P(H) \underline{\pi}_{k'}(H)}{\underline{\pi}_{k'}(L) \underline{\pi}_K(H) - \underline{\pi}_K(L) \underline{\pi}_{k'}(H)} c(L)
\end{aligned}$$

and

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

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),

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

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

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

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:

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

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:

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

where

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

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 3

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

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

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 Grossman and Hart (1983) 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 2 applies to that problem and the optimal contract has a two-wage structure.

### Proofs for Section 5

#### Proof of Proposition 4

Fix  $a \in \mathcal{A}$ . Let  $E, F \subset \{1, \dots, N\}$ . 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  $\nu^a$  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  $\nu^a$ . Since  $\Pi(a)$  is closed and convex, it suffices to show that  $\Pi(a)$  contains all of the ‘‘marginal contribution’’ vectors for  $\nu^a$ , that is, any vector  $\pi$  of the form  $\pi_j = \nu^a(\{\sigma(1), \dots, \sigma(j)\}) - \nu^a(\{\sigma(1), \dots, \sigma(j-1)\})$  where  $\sigma$  is a permutation on  $\{1, \dots, N\}$ . Without loss of generality, consider the identity permutation and corresponding vector  $\pi$  in which  $\pi_j = \nu^a(\{1, \dots, j\}) - \nu^a(\{1, \dots, j-1\})$  for each  $j$ . Thus for each  $j$ ,  $\pi_j = \nu(\{y_1, \dots, y_j\}(a)) - \nu(\{y_1, \dots, y_{j-1}\}(a))$ . For each  $j$ , set  $\{s_1^j, \dots, s_{k_j}^j\} := \{y_j\}(a) = \{y_1, \dots, y_j\}(a) \setminus \{y_1, \dots, y_{j-1}\}(a)$ . Define  $\bar{\pi} \in \Delta(\mathcal{S})$  as follows. Set

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

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

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

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

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

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

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

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

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

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

### Proof of Proposition 5

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

- AUMANN, R. J. (1962): "Utility Theory without the Completeness Axiom," *Econometrica*, 30, 445–462.
- BEWLEY, T. F. (1986): "Knightian Decision Theory: Part I," Discussion paper, Cowles Foundation.
- (1987): "Knightian Decision Theory: Part II," Discussion paper, Cowles Foundation.
- (1989): "Market Innovation and Entrepreneurship: A Knightian View," Discussion paper, Cowles Foundation.
- (2002): "Knightian Decision Theory: Part I," *Decisions in Economics and Finance*, 2, 79–110.
- CHATEAUNEUF, A., AND J.-Y. JAFFRAY (1989): "Some Characterizations of Lower Probabilities and other Monotone Capacities through the Use of Möbius Inversion," *Mathematical Social Sciences*, 17, 263–283.

- GHIRARDATO, P. (1994): “Agency Theory with Uncertainty Aversion,” Discussion paper, Caltech.
- GHIRARDATO, P., F. MACCHERONI, M. MARINACCI, AND M. SINISCALCHI (2003): “A Subjective Spin on Roulette Wheels,” *Econometrica*, 71, 1897–1908.
- GROSSMAN, S. J., AND O. D. HART (1983): “An Analysis of the Principal-Agent Problem,” *Econometrica*, 51, 7–45.
- HART, O. D., AND B. HOLMSTROM (1987): “The Theory of Contracts,” in *Advances in Economic Theory*, ed. by T. F. Bewley.
- HOLMSTROM, B. (1987): “Moral Hazard and Observability,” *Bell Journal of Economics*, 10, 74–91.
- HOLMSTROM, B., AND P. MILGROM (1987): “Aggregation and Linearity in the Provision of Intertemporal Incentives,” *Econometrica*, 55, 303–328.
- KNIGHT, F. H. (1921): *Uncertainty and Profit*. Boston: Houghton Mifflin.
- LOPOMO, G., L. RIGOTTI, AND C. SHANNON (2009): “Uncertainty in Mechanism Design,” *mimeo*.
- MUKERJI, S. (1998a): “Ambiguity and Contractual Forms,” Discussion paper, Yale University.
- (1998b): “Ambiguity Aversion and Incompleteness of Contractual Form,” *American Economic Review*, 88, 1207–1231.
- RIGOTTI, L., AND C. SHANNON (2005): “Uncertainty and Risk in Financial Markets,” *Econometrica*, 73, 203–243.
- SCHMEIDLER, D. (1989): “Subjective Probability and Expected Utility without Additivity,” *Econometrica*, 57(3), 571–587.
- SHAPLEY, L. S., AND M. BAUCELLS (1998): “Multiperson Utility,” Discussion paper, University of California, Los Angeles.
- VON NEUMANN, J., AND O. MORGENSTERN (1953): *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.