Imprecise Beliefs in a Principal Agent Model*

Luca Rigotti†

CentER, Tilburg University

Department of Economics, U.C. Berkeley

August 2001

Abstract

This paper presents a principal-agent model where the agent has multiple, or imprecise, beliefs. We model this situation formally by assuming the agent's preferences are incomplete. One can interpret this multiplicity as limited knowledge of the surrounding environment. In this setting, incentives need to be robust to the agent's beliefs. We study whether robustness implies simplicity. Under mild conditions, we show the unique optimal contract has a two-wage structure. That is, all output levels are divided into two groups, and the optimal incentive scheme pays the same amount for all output levels in each group. When monotonicity occurs, this is a flat payment plus bonus contract. Then, a two-state two-action framework can be thought of as a reduced form of the original model. We solve explicitly the principal's problem in this case, and discuss some implications of our model for firm ownership.

JEL Codes: D82, D81, D23

*I thank Truman Bewley and Ben Polak for invaluable advice; Aldo Rustichini Dirk Bergemann, Subir Bose, Simon Grant, Georg Kirchsteiger, June K. Lee, David Pearce, Matthew Ryan, Chris Shannon, and Dolf Talman gave useful comments at various stages of this work; I also thank seminar audiences at Arizona State, CORE, Universitat Autonoma de Barcelona, ITAM, Stanford, Tilburg, and Yale.

†E-mail: luca@kub.nl

1 Introduction

In some principal-agent settings, the agent may have multiple, or 'imprecise', beliefs. This imprecision arises because the agent is not confident in assessing the possible consequences of his actions. We discuss the characteristics of an incentive scheme in such settings. The main conclusion is that optimal incentive schemes are simple. In particular, we show that the optimal contract takes only two values across all possible output states.

The moral hazard model sometimes generates extremely complex incentive structures. Optimal contracts often involve as many different payments as there are possible levels of output. In addition, small changes in the assumed distribution of outcomes can lead to large changes in the way an optimal scheme depends on output; that is, in its shape. 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. Why is this structure so common across different environments? The standard model's answer is that all of them must share the same stochastic structure of output.

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 than the ones accounted for in the standard model. Once this need for robustness is considered, simple optimal schemes might obtain. We follow this path by introducing a particular robustness requirement. Suppose the agent lacks confidence in judging the stochastic properties of the environment he operates in. This is reflected by non-unique beliefs about possible output levels. An incentive scheme which accounts for this problem is necessarily robust to the agent's different beliefs. In this framework, we show the optimal scheme often has a two-wage structure. Thus, the need for robustness we examine generates simple contracts. Furthermore, these contracts have a shape we commonly observe in the real world.[1]

[1] 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 stated above. 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.

The situation we have in mind is the following. An entrepreneur is considering whether to hire a worker. If she hires him, she cannot observe how much effort he puts into his job. He, however, cannot precisely evaluate the impact of what he does on the production process. Output depends on his effort, but also on many variables beyond his control, like what other workers in the firm do, or even pure luck. In particular, he has no previous experience working at her company. Thus, he may not feel confident evaluating the relation between the effort he devotes to his work and the output produced. She, however, as owner of the production technology can evaluate this relation more precisely.

In this relationship the agent is an outsider who is not familiar with all the details of the production process. The principal is an insider who is familiar with these details. The principal is at an informational disadvantage because she cannot observe the agent's effort choice, but the agent is entering a new environment and cannot evaluate precisely the consequences of his work. The agent's behavior may reflect the uncertainty he faces. The standard principal-agent model, however, neglects this possibility. It assumes both parties can precisely evaluate the stochastic consequences of the agent's action.

We model this situation formally by relaxing the assumption that an agent's preferences are complete. In this case, the agent may be unable to compare alternatives offered to him. Aumann (1962) and Bewley (1986) showed that incomplete preferences can be represented by a Von Neumann-Morgenstern utility function with multiple probability distributions. This multiplicity can be thought of as imprecision of the agent's beliefs over uncertain outcomes. The agent computes an expected utility for each distribution, and they all matter in determining his behavior. One interpretation of this multiplicity follows Knight (1921). Individuals use a single distribution only when they regard events as risky; individuals use a set of distributions when they regard events as uncertain. The term Knightian uncertainty has been associated with the latter situation.

A question that arises with incomplete preference models is what do individuals do when all alternatives are incomparable. Bewley's inertia assumption states that, when faced with incomparable options, an individual sticks with his current behavior, the status quo, unless an alternative is strictly preferred. This 'uncertainty aversion' reflects reluctance to change behavior when the consequences of doing so are difficult to evaluate. In abstract settings, defining the status quo is sometimes hard. This may be one reason why Bewley's model is difficult to apply. In our setting, however, there is a natural candidate for the status quo, the agent's outside option.

We consider the following moral hazard model. A risk-neutral principal has to design an incentive scheme for a risk-neutral agent who has imprecise beliefs over output outcomes. These beliefs are represented by sets of probability distributions, one set for each action. We assume risk-neutrality to focus on the impact of Knightian uncertainty alone. Therefore, multiplicity of beliefs is the only difference between our model and a risk-neutral version of Grossman and Hart (1983). The principal cannot observe the agent's action. Each action has a different disutility to the agent and induces different beliefs over output outcomes. For each action, the principal designs a contract which implements it at the lowest possible expected cost. Then, she selects the action that maximizes the difference between expected output and expected cost.

If agents have imprecise beliefs but do not satisfy the inertia assumption, it is often impossible for the principal to induce the agent to take a privately costly action such as hard work. For an incentive scheme to implement a particular action, the agent must prefer that action to his reservation utility and to all other actions. The agent, however, regards any two actions whose belief sets intersect as not comparable. Thus, an action can only be implemented if the agent's belief set corresponding to it does not intersect any of the belief sets corresponding to the other 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, according to the above result, no action can be implemented.

Implementation is easier when the agent satisfies the inertia assumption. With inertia, an incentive scheme implements an action if this action is preferred to the reservation utility and, for each other action, either the first action is preferred, or the other action is not comparable to the reservation utility. Preferring one action to all others is no longer necessary. With the inertia assumption, implementing an action is sometimes possible even if the agent's belief sets intersect as in the example above.

Optimal contracts under imprecise beliefs are both robust and simple. Regardless of inertia, an optimal contract is robust because it provides incentives for the entire set of probability distributions the agent considers. Under mild conditions on the agent's imprecise beliefs, the unique optimal incentive scheme divides all the possible outputs levels into two groups and pays the same amount in all states belonging to the same group. First, we prove the result when the agent can choose between two actions. Then, we generalize it to the case in which many actions are available to him. Consider the following example. Suppose the number of events a contract is contingent upon is increased by one because the principal decides to make different payments in two output levels that previously corresponded to the same wage. In other words, one event is divided into two separate events. Risk-neutrality implies the agent does not place any premium on receiving different payments. Satisfying the constraints is now more difficult, however, because the new events have, in general, different probabilities for different elements of the agent's belief sets. Thus, dividing events makes it more difficult to provide incentives. Conversely, the formal proof shows that joining events is strictly profitable for the principal. Under mild assumptions, this result holds, regardless of the number of output levels, provided this number is finite. Additional restrictions guarantee it also holds for any finite number of actions available to the agent.

Given this result, the two-action two-state version of the model is a reduced form of the more general formulation. We explicitly solve the principal's problem in this case. We find the optimal incentive scheme with and without inertia. With inertia, we show the agent is unwilling to buy the firm from the principal at a price the principal would accept. Thus, contrary to the standard model, the agency problem cannot be avoided even if the agent is risk-neutral and has unlimited wealth. This result suggests a possible theory of the firm. Firm owners (principals) are the individuals who face less Knightian uncertainty. Workers (agents) are the individuals who face more Knightian uncertainty.

Recently, much attention has been devoted to incomplete contracts. For example, see Moore and Hart (1988), Tirole (1999), and Hart (1995). Tirole argues robustness (in the sense of limited knowledge of the surrounding environment) should be investigated as a source of incomplete contracts. Our main result does not deal with this problem explicitly, but suggests Knightian uncertainty as a useful tool. The main intuition of our result is that asymmetric confidence in beliefs introduces an uncertainty cost in contracts. This cost may depend positively on the number of events the contract is contingent upon. Therefore, it can be reduced by making the contract depend on fewer events. If this is the case, incomplete preference generate contracts that are simple, in the sense that they depend on few contingencies.

Mukerji (1998a) and Ghirardato (1994) present moral hazard models similar in motivation to the one we describe, but unlike us they get rather standard results about incentive schemes. The decision theoretic model they use, however, is different. In both cases, the principal and the agent are Choquet expected utility maximizers (see Schmeidler (1989)). Choquet expected utility is a model in which, loosely speaking, lack of confidence and uncertainty are reflected by the non-additivity of probabilities, not by the imprecision of beliefs. In that framework, there is no simple way to allow for asymmetric confidence, or lack thereof, among across the parties involved and this explain why they do not get simplicity of incentive schemes. On the other hand, Mukerji (1998b) uses that framework to show one can relate uncertainty to contract incompleteness.

The paper is organized as follows. The next section introduces some concepts of individual decision making when preferences are incomplete. Section 3 presents the basic framework and discusses the implementation rules. Section 4 proves that optimal incentive schemes are simple. Section 5 solves the two-state two-action problem explicitly and explores some characteristics of the optimal contracts. Section 6 concludes.

2 Incomplete Preferences and Inertia

We describe briefly individuals' behavior when their preferences are not necessarily complete, and then introduce the inertia assumption. This approach to decision making was pioneered by Bewley (1986), and further developed in Bewley (1987). Incompleteness modifies the expected utility framework to account for multiple probability distributions. The purpose of the inertia assumption, which states that an alternative is chosen only if it is preferred to current behavior, is to explain some choices between incomparable alternatives.

2.1 Preference Representation without Completeness

If a preference ordering satisfies the completeness axiom, the decision maker can compare any two random payoffs and decide which one is better based on their respective expected utilities. Von Neumann and Morgenstern were the first to observe how completeness is not an entirely satisfactory axiom.[2] If a preference ordering does not satisfy completeness, the decision maker is not necessarily able to compute a unique expected utility for each payoff. Incompleteness is reflected by multiplicity of beliefs and alternatives are compared one probability distribution at a time.

Let x and y be random monetary payoffs defined over some state space. Bewley (1986) proves that under standard axioms on preferences, but without completeness, the following holds:

$$x \text{ is preferred to } y \text{ if and only if } E_{\delta}[u(x)] > E_{\delta}[u(y)] \text{ for all } \delta \text{ in } \Delta \tag{1}$$

where $u(\cdot)$ is the Von Neumann-Morgenstern utility derived from the payoffs, and Δ is a closed and convex set whose elements are subjective probability distributions.[3] If the set Δ has only one member, the preference ordering is complete, and the usual representation obtains. When the relevant state-space has only N elements, expression (1) reduces to:

$$x \succ y \qquad \Longleftrightarrow \qquad \sum_{i=1}^{N} \delta_i u(x_i) > \sum_{i=1}^{N} \delta_i u(y_i) \text{ for all } \delta \text{ in } \Delta$$

Bewley argues expression (1) is a possible formulation of Frank Knight's distinction between risk and uncertainty. A payoff is risky when the probabilities of different outcomes are known; if they are unknown, the payoff is uncertain. Hence, payoffs are risky when Δ has only one member and uncertain otherwise. Informally, the size of Δ gauges the amount of uncertainty the individual perceives. It tells us how much imprecision there is in the decision maker's beliefs and can be thought of as measuring confidence in beliefs.

[2] 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.

[3] This is Theorem 1.2 in Bewley (1986); a similar result is found in Aumann (1962).

A graph helps to understand this model. In Figure 1 the axes measure money in each of two possible states. Given a probability distribution over these states, an indifference curve through y represents all the random payoffs that have the same expected utility as y according to this distribution. As the probability distribution changes, we obtain a family of these indifference curves; it represents 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.

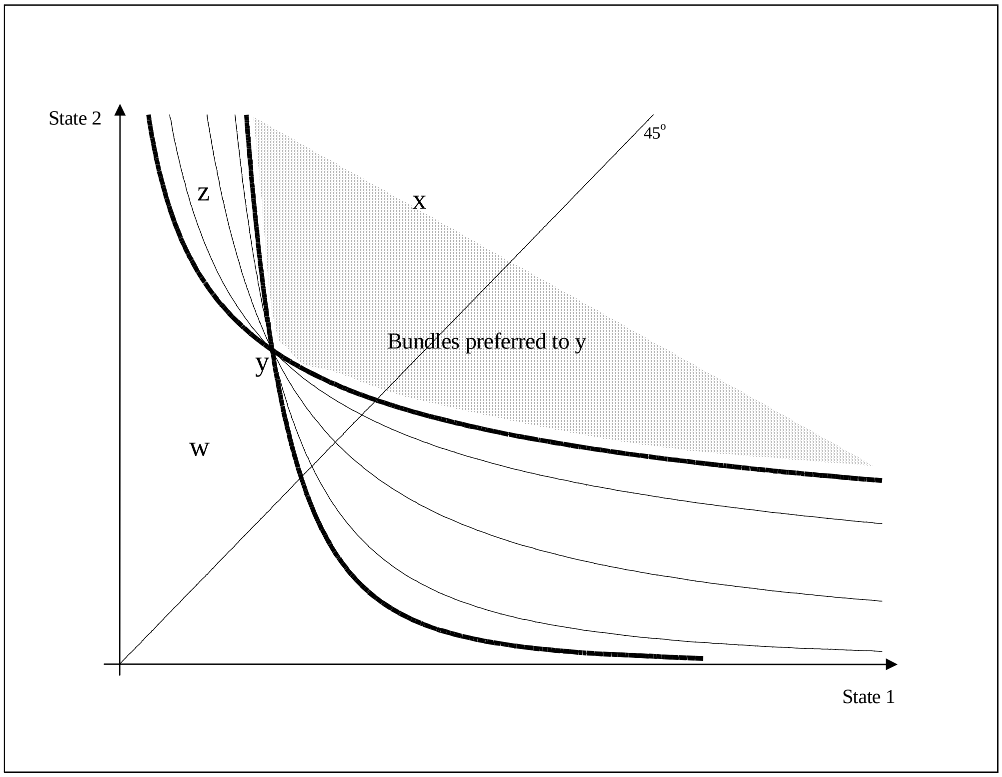

Figure 1: Incomplete Preferences

x is preferred to y since it lies above all of the indifference curves corresponding to some expected utility of y. Also, y is preferred to w since w lies below all of the indifference curves through y. Finally, z is not comparable to y since it lies above some expected values of y and below others. Incompleteness thus 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.

2.2 The Inertia Assumption

When preferences are not complete, we cannot explain choice among incomparable alternatives. If x is chosen when y is available, we cannot say x is revealed preferred to y; we can only say y is not revealed preferred to x. To address this problem, Bewley (1986) introduces the concepts of status quo and inertia. The inertia assumption states that planned behavior, defined as the status quo, is abandoned only for alternatives preferred to it. Therefore, if x is chosen when y is available and y is the status quo, x is revealed preferred to y. If an additional choice z is also available, but z is not preferred to y, the individual still chooses x. However, x is not thereby revealed preferred to z. In Figure 1, the inertia assumption implies that if y is the status quo, alternatives like z will not be chosen since they are incomparable to y.

The inertia assumption is a behavioral statement, but there is evidence consistent with it in economics and psychology. The classic reference is Samuelson and Zeckhauser (1988), who find evidence of status quo biases in both field and experimental data. In an experimental setting, they find significant status quo bias in investment decisions regarding the portfolio composition following a hypothetical inheritance. Ameriks and Zeldes (2000) find similar evidence in data from TIAA-CREF and Surveys of Consumer Finance; almost half of their sample made no change in portfolio composition over the course of their 9 year sample, while the same period saw drastic changes in the returns to bond and equity holdings. Einhorn and Hogarth (1985) find evidence supporting a status quo bias in initial probability assessments in a number of experiments (see also Fox and Tversky (1995), and Heath and Tversky (1991)).

In many economic contexts, there is a natural candidate for the status quo. For example, consider a bargaining game in which each player has an outside option. If we interpret these options as the players' actions before entering the game, defining each player's outside option as his status quo seems natural. In the moral hazard model that follows, the status quo corresponds to the agent's reservation utility. This is the payoff the agent receives if he rejects the contract the principal offers. We interpret it as the reward corresponding to the agent's planned behavior before the possibility of dealing with the principal materialized.

3 The Moral Hazard Setup

We present a moral hazard model where the agent is less confident in beliefs than the principal. Formally, the agent's preferences are not complete and the principal's preferences are complete. Because the inertia assumption is somewhat arbitrary, we establish conditions for a contract to implement some action with and without inertia. We show this has different implications for the model; namely, the latter case imposes undesirable restrictions. Finally, we define an optimal incentive scheme, and derive some of its basic properties.

The principal owns resources that yield output. An agent has to perform some action for production to take place. The principal and the agent observe the realized output level, but the principal cannot observe the action performed by the agent. An incentive scheme is a contract that induces the agent to perform a particular action.

N states of nature are distinguished by the amount of output produced. Output is an N-vector denoted by $y = (y_1, .., y_N)$, with $y_N > .. > y_1$. An incentive scheme is an N-vector denoted by $w = (w_1, .., w_N)$, where the payment from the principal to the agent is w_j in state j. The agent chooses an action a from a discrete set of available actions $\mathcal{A} \equiv \{1, 2, .., M\}$, interpreted as effort levels. The agent's reservation utility, or outside option, is $\overline{w}$. It represents the reward the agent receives from his current behavior, the status quo. Throughout, we use subscripts to denote states and superscripts to denote actions. Each action has two consequences: it imposes a cost on the agent, measured by disutility of effort, and it generates beliefs about the likelihood of different output levels.

Let π^a be the principal's beliefs induced by action a; π^a is a probability distribution over possible output levels. Let Δ^a be the agent's beliefs induced by action a; Δ^a is a closed and convex set whose elements, denoted δ^a, are probability distributions over the possible output levels. More formally, we assume that the principal's beliefs are described by a function $\pi : \mathcal{A} \to \mathcal{P}$, and the agent's beliefs are described by a correspondence $\Delta : \mathcal{A} \to \mathcal{P}$, where $\mathcal{P} \equiv \left\{x \in \Re^N | x \geq 0, \sum_{i=1}^N x_i = 1\right\}$ denotes all the probability distributions over y. This description follows the parametrized distribution formulation of the principal-agent problem pioneered by Holmstrom (1987). There, Δ is restricted to being a function.

We assume π and Δ are common knowledge. We also assume the principal's probability distribution agrees with one of the agent's distributions; formally, for each a in $\mathcal{A}$, π^a is an element of Δ^a. There is asymmetric confidence in beliefs. The principal is more confident in evaluating the stochastic relationship between the agent's effort and output. This is the most relevant feature of the model, and uniqueness of the principal's beliefs is assumed for analytical tractability. Most of the analysis would be the same if π^a was also a correspondence, as long as one maintains asymmetric confidence. Finally, π^a being an element of Δ^a rules out asymmetric information in the standard sense; one of the agent's priors agrees with the principal's prior.

The cost of an action is denoted c^a. We assume actions are ordered such that $c^a > c^{a'}$ and $\sum_{i=1}^N \pi_i^a y_i \geq \sum_{i=1}^N \pi_i^{a'} y_i$ whenever $a > a'$. More expensive actions increase the expected value of output. Both assumptions are standard in the principal-agent literature.

The principal and the agent are risk-neutral. The agent's utility function is given by the expected value of the contract minus the cost of the action he performs. The agent computes many expected values, one for each probability distribution in the belief set induced by the chosen action. His behavior depends on all of them. Formally, for each δ^a in Δ^a, the agent's utility is:

$$E_{\delta^a}[w] - c^a = \sum_{j=1}^N \delta_j^a w_j - c^a.$$

The principal's utility function is given by the expected value of output minus the expected cost of the contract. These expectations are computed according to the probability distribution induced by the effort level chosen by the agent. Formally, the principal's utility is:

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

The set up parallels the standard principal-agent model. There, the principal is assumed risk-neutral while the agent is assumed 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 the same beliefs, and evaluate risk in the same way, but they have different attitudes toward uncertainty.

When the agent's action is observable (and/or verifiable), the principal can make the contract contingent on it. Define the cost of this incentive scheme as C^a_{FB}, where FB stands for first-best. As in the standard model, $C^a_{FB} = \overline{w} + c^a$; the principal gets the agent to pick a by offering a contract that says: in every state, I'll pay you C^a_{FB} if you choose a, $-\infty$ otherwise.

Given this setup, the principal and the agent do not value the firm equally. In particular, the agent is more cautious about this value. Therefore, the usual way of dealing with moral hazard when the parties are risk-neutral does not apply. For a given action, the highest price the agent is willing to pay for the firm equals the lowest expected value of output minus the cost of taking that action. This price is lower than what the firm is worth to the principal.[4]

3.1 Two Implementation Rules

The principal wants the agent to perform a specific action a^*. Because she cannot observe the agent, she must rely on the contract alone to provide the desired incentives. Obviously, different assumptions about the agent's behavior imply different requirements the contract must satisfy. In any case, these requirements must leave no doubts the agent's choice is a^*. A contract is a take-it-or-leave-it offer, and we say a contract implements a^* if the agent accepts the contract and chooses a^* among all possible actions.

3.1.1 Implementing without inertia

Without inertia, a contract implements a^* if it satisfies the conditions one imposes in the standard model, modified to take multiplicity of beliefs into account. The agent participates if a^* is preferred to the reservation utility, and chooses a^* if this action is preferred to all others.

Definition 1 *An incentive scheme w implements a^* when:*

$$\sum_{j=1}^{N} \delta_j^{a^*} w_j - c^{a^*} \geq \overline{w} \qquad \forall \delta^{a^*} \text{ in } \Delta^{a^*} \tag{P}$$

[4] A general proof that 'selling the firm' is not possible in this case, is more complicated than this, since the agent and the principal may not even agree on which action should be chosen to maximize the value of the firm. In Section 5, we show explicitly how selling the firm is not possible when there are only two actions and two states.

and for each a' in $\mathcal{A}$ different than a^*

$$\sum_{j=1}^{N} \delta_j^{a^*} w_j - c^{a^*} \geq \sum_{j=1}^{N} \delta_j^{a'} w_j - c^{a'} \qquad \forall \delta^{a^*} \text{ in } \Delta^{a^*} \text{ and } \forall \delta^{a'} \text{ in } \Delta^{a'} \tag{IC}$$

In words, for any probability distributions induced by a^*, the expected utility the agent receives from 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 *Without inertia, if the belief sets induced by any two actions have an element in common, one of them cannot be implemented.*

Proof: Let the two actions be a and a' and, without loss of generality, assume $a > a'$. Suppose the claim does not hold. Then, there exist an incentive scheme w satisfying Definition 1 and a probability distribution δ that belongs to both Δ^a and $\Delta^{a'}$. Therefore, (IC) must be satisfied when δ appears on both sides of the inequality:

$$\sum_{j=1}^{N} \delta_j w_j - c^a \geq \sum_{j=1}^{N} \delta_j w_j - c^{a'}.$$

This implies $c^a \leq c^{a'}$, and contradicts $a > a'$. ∎

When two belief sets intersect the corresponding actions are not comparable. A possible implication of Proposition 1 is that, when the number of actions is large, only few could be implementable. In particular, no action whose belief set intersects the belief set of a cheaper action can satisfy Definition 1. This problem is mitigated by using the inertia assumption.

3.1.2 Implementing with inertia

The inertia assumption says an alternative is chosen only if it is preferred to the status quo. This can be used to define a sufficient condition for an action not to be chosen: an action not comparable to the reservation utility is not chosen. After an incentive scheme is proposed by the principal, the agent faces two basic decisions: does he accept the offer and if yes, which action does he choose?

By rejecting, the agent opts for the outside option, the status quo. This corresponds to the agent's behavior before the contractual offer was made available to him, that is, the agent's reservation utility.

Inertia is thus useful to define the requirements an incentive scheme must satisfy to implement a^*. Suppose a^* and a' are not comparable, but the first is preferred to the status quo while the second is not. Then, the inertia assumption implies a^* is chosen. Therefore, an incentive scheme does not need to make a^* preferred to all other actions.

Definition 2 *If the inertia assumption holds, an incentive scheme w implements a^* in $\mathcal{A}$ if :*

$$\sum_{j=1}^{N} \delta_j^{a^*} w_j - c^{a^*} \geq \overline{w} \qquad \forall \delta^{a^*} \text{ in } \Delta^{a^*} \tag{P}$$

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

$$\sum_{j=1}^{N} \delta_j^{a^*} w_j - c^{a^*} \geq \sum_{j=1}^{N} \delta_j^{a'} w_j - c^{a'} \qquad \forall \delta^{a^*} \text{ in } \Delta^{a^*} \text{ and } \forall \delta^{a'} \text{ in } \Delta^{a'} \tag{IC}$$

or

$$\sum_{j=1}^{N} \delta_j^{a'} w_j - c^{a'} \leq \overline{w} \quad \text{for some } \delta^{a'} \text{ in } \Delta^{a'} \tag{NC}$$

NC is a non comparability constraint. It says there exists at least one probability distribution induced by a' such that the expected utility the agent derives from the contract is weakly lower than the reservation utility.

Inertia eases the implementation conditions. All incentive schemes that satisfy Definition 1 also satisfy Definition 2. Intuitively, an incentive scheme does not need to make a^* the most preferred action, but only to guarantee no alternative to a^* is attractive enough for the agent.

3.2 The principal's problem

The principal maximizes her expected utility. We can divide her problem into two steps. First, for each action, find the cheapest contract which implements it. Second, decide which action to implement. For a given action a, let $\mathcal{H}^1(a)$ be the set of all incentive schemes which satisfy Definition 1, and $\mathcal{H}^2(a)$ the set of all incentive schemes which satisfy Definition 2. We say $\widehat{w}^a$ is an optimal incentive scheme to implement a when it is a solution to

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

where $i = \{1,2\}$ indicates whether the agent's behavior obeys the inertia assumption. For each a, define $C(a;i) = \sum_{j=1}^{N} \pi_j^a \widehat{w}_j^a$ when $\mathcal{H}^i(a)$ is not empty, and $C(a;i) = -\infty$ otherwise. $C(a;i)$ is the expected cost of the optimal scheme to implement a. The second part of the principal's problem is:

$$\max_{a\in\mathcal{A}} \sum_{j=1}^{N} \pi_j^a y_j - C(a;i).$$

As noted before, $\mathcal{H}^1(a)$ is a (possibly empty) subset of $\mathcal{H}^2(a)$. As a consequence, the principal cannot be worse off if the agent obeys the inertia assumption. Because of the linearity of the problem, existence of a solution is not an issue as long as the $\mathcal{H}^i(a)$ are not empty. Our main objective is to analyze the characteristics of the solutions to (2).

3.3 Some characteristics of the optimal incentive scheme

The agent's behavior depends on all the probability distributions in his beliefs set. Some of them, though, are more relevant for our analysis because they trigger his behavior. For any fixed action a let $\widetilde{\Delta}(w;a) \equiv \left\{\widetilde{\delta}^a \in \Delta^a \mid \widetilde{\delta}^a = \arg\min_{\delta^a\in\Delta^a} \sum_{j=1}^{N} \delta_j^a w_j\right\}$. Then, given the contract w, $\widetilde{\delta}^a$ is a probability distribution yielding its lowest expected value for the agent when he chooses action a. A contract w satisfies the participation constraint for action a if and only if this constraint is satisfied for elements of $\widetilde{\Delta}(w;a)$. A contract w satisfies the non comparability constraint for action a' if and only if this constraint is satisfied for elements of $\widetilde{\Delta}(w;a')$. For any fixed action a let $\widetilde{\widetilde{\Delta}}(w;a) \equiv \left\{\widetilde{\widetilde{\delta}}^a \in \Delta^a \mid \widetilde{\widetilde{\delta}}^a = \arg\max_{\delta^a\in\Delta^a} \sum_{j=1}^{N} \delta_j^a w_j\right\}$. Then, given the contract w, $\widetilde{\widetilde{\delta}}^a$ is a probability yielding its highest expected value 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 this constraint is satisfied for elements of $\widetilde{\Delta}(w;a)$ and $\widetilde{\widetilde{\Delta}}(w;a')$ on the left and right side respectively..

The following proposition specializes to our framework some results about the optimal contract which hold in the standard model (the proof is in the appendix).

Proposition 2 *Let a be the action the principal wants to implement, and let $\widehat{W}^a$ be the set of solutions to (2).*
(i) The participation constraint binds when computed according to $\widetilde{\delta}^a$; formally,

$$\sum_{j=1}^{N} \widetilde{\delta}_j^a \widehat{w}_j^a - c^a = \overline{w}$$

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

(ii) If a is the least costly action, the scheme that pays $\overline{w} + c^1$ in all states is a solution to (2); that is, $a = 1$ implies $\widehat{w}^1 = \overline{w} + c^1 \in \widehat{W}^1$.

(iii) if a is not the least costly action, there exists an action less costly than a such that either (IC) or (NC) bind for this action; formally, if $a > 1$ there exists an $a' < a$ such that

$$\sum_{j=1}^{N} \widetilde{\delta}_j^a \widehat{w}_j^a - c^a = \sum_{j=1}^{N} \widetilde{\widetilde{\delta}}_j^{a'} \widehat{w}_j^a - c^{a'},$$

or

$$\sum_{j=1}^{N} \widetilde{\delta}_j^{a'} \widehat{w}_j^a - c^{a'} = \overline{w}$$

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

4 Optimal Incentive Schemes Are Simple

This section contains the main result. We establish mild conditions for an optimal incentive scheme to be simple. Under these conditions, the solution to (2) divides the N possible states into two groups, and pays the same amount in all states belonging to the each group. The optimal contract distinguishes only between two events: it has a two-wage structure. First, we prove the result when the agent can choose only between two actions. Second, we introduce many actions and show how our result generalizes to this case.

Theorists have conjectured that contracts are simple because they need to be robust. Hart and Holmstrom (1987) argue that, in the real world, a contract must provide incentives across a wider range of circumstances than the ones the standard model considers. The idea of robustness we use represents one way to model this requirement. A different approach is used in Holmstrom and Milgrom (1987). They provide conditions for an optimal scheme to be linear. These are not related our idea of robustness. Furthermore, we adopt a different notion of simplicity. A two-wage scheme is simple because it can be thought of as contingent on only two events.

The main result depends on two characteristics of the beliefs of the parties involved in the contract. Before stating them, we need a bit of additional notation. Let S, with generic element E, be the set of all subsets of $1...N$.

Assumption A-1: For each action a in A, there exists a convex capacity v^a such that:

$$\Delta^a = \{\delta^a \text{ in } \mathcal{P} \mid \delta^a(E) \geq v^a(E) \text{ for each } E \text{ in } S\}, \tag{3}$$

where $\mathcal{P}$ is the set of all probability distributions and $v^a : S \to [0,1]$ is a convex capacity.[5]

This assumption says that Δ^a must be the core of some convex capacity.[6] For each action, A-1 places restrictions on the image of the correspondence Δ. Geometrically, Δ^a must be a polyhedron whose boundaries are determined by the linear inequalities in (3). Figure 2 displays some examples when belief sets are subsets of the two-dimensional simplex. Δ^A and Δ^B satisfy A-1, while Δ^C does not.

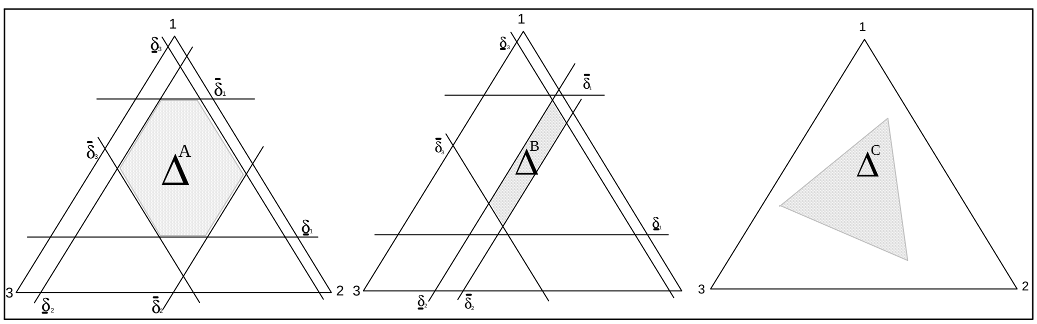

Figure 2: Agent's Beliefs with 3 Output States

Assumption A-2: *For each action a in $\mathcal{A}$, π^a belongs to the interior of Δ^a.*

This makes more stringent the condition that the agent's belief sets contain an element corresponding to the principal's beliefs.

4.1 Simplicity with two actions

In this section, we specialize the framework to the case in which the agent can choose only among two actions, H and L. We interpret them as high and low effort respectively; hence, $c^H > c^L$.

[5]A convex capacity must satisfy the following properties: (i) $v^a(\emptyset) = 0$, (ii) $v^a(S) = 1$, (iii) $\forall E, E' \in S$, $E \subseteq E'$ implies $v^a(E) \leq v^a(E')$, and (iv) $\forall E, E' \in S$, $v(E \cup E') \geq v(E) + v(E') - v(E \cap E')$. v^a is sometimes called a non-additive probability.

[6]In the context of 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.

As seen previously, the principal can implement the low effort action at the first best cost with a contract which promises the same payment in every state. A more interesting problem arises when the principal wants to implement the high effort action. The result is that the optimal contract to do that is two-valued.

The main result is the following.

Proposition 3 *Assume A-1 and A-2 hold. Let the set of all contracts that implement H against L be non-empty. Then, the unique optimal incentive scheme to implement H divides the N states in two groups and pays the same amount in all states belonging to each group.*

This Proposition does not depend on which definition of implementation is used. We provide the details of the proof for the case in which the agent obeys the inertia assumption. The other case can be proved very similarly. Loosely, the following example drives the formal argument in the proof. Consider a contract contingent on only two groups of output levels. Suppose the principal divides the group of output levels corresponding to one payment into two, and makes two different payments in each group. As we show in the Appendix, the distribution yielding the agent's extreme expectation of the old contract also yields an extreme expectation of the new one. This extreme distribution, though, determines only the aggregate probability of the two new payments. Given this probability, the lowest expectation of the new contract assigns the smallest weight to the largest of the two new payments. Therefore, the new contract is not as efficient as the old one to provide incentives.

Intuitively, this result follows from a simple consequence of assumption A-2. For any given state, there is one belief in the agent's belief set such that the agent 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 Bewley's model, the optimal contract must "accommodate" this probability. Since everyone is risk-neutral, the principal can always slightly modify a payment scheme to take advantage of the disagreement about any one particular state. In particular, this will be done by making the incentive scheme "flatter", that is, more equal across states. Eventually, the process has to stop because the incentive scheme must have enough 'variation' to distinguish among the two actions.

Note that assumption A-2 is crucial for the argument we developed. It is not, however, necessary to obtain the optimality of simple contracts. The following Proposition provides a simple example where a simplicity result is obtained anyway.

Proposition 4 *Suppose there exists a state j' such that $\underline{\delta}_{j'}^{L} = \pi_{j'}^{L} = \overline{\delta}_{j'}^{L} > 0$ and $\underline{\delta}_{j'}^{H} = \pi_{j'}^{H} = \overline{\delta}_{j'}^{H} > 0$. Then, there exists an optimal incentive scheme that pays the same in all states but j'.*

Proof: Such a scheme makes the contract conditional on two events which have precise, i.e. unique, probabilities. Hence, by writing a contract contingent on j' on one hand and all states other than j' on the other, the problem is reduced to the moral hazard model with a risk neutral agent. The solution to this yields a contract which achieves the first-best allocation. ■

Because there are N possible states and the optimal contract is two-valued, Proposition 3 implies this contract can be found looking among 2^N systems of two equations in two unknowns. In Section 5, we study the features of this optimal contract. An interesting question for future research is how the principal selects among these 2^N possibilities.

4.2 Simplicity with many actions

In this Section we provide conditions under which optimal incentive schemes are simple when the agent chooses among many actions. There are two possible way of proceeding, both are relatively straightforward. First, we indicate how the main result of the previous section generalizes to more actions, without additional restrictions. Then, we look at how the optimal contract may take only two values even when there are many actions.

Proposition 3 says the number of signals an optimal contract considers is the minimal number necessary to find feasible contracts. In the case of two actions, this is exactly two. If there are M possible actions, there will be at most M binding constraints. In fact, after this observation one can formally repeat the argument used to prove Proposition 3, since it shows that reducing the number of states a contract depends on is always profitable for the principal. Clearly, this becomes impossible when all constraints are binding. In the end, this shows that the optimal contract is contingent on at most M events.

A different way to proceed is to find conditions such that the same two-wage structure is optimal regardless of the number of 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. Not surprisingly, the requirements are similar.

Suppose the constraint corresponding to only one alternative action binds at the optimum. Then, the solution to the two-action problem in which this action is the only alternative also solves the more general problem. Therefore, the following proposition provides assumptions such that there is only one action whose constraint binds.

Proposition 5 *Let monotone likelihood ratio (MLR) and concavity of the distribution function (CDFC) hold for all extreme points of the agent's belief sets. Then, the optimal contract to implement an action a^* according to Definition 2 has a two-wage structure.*[7]

Proof:

Let $\widehat{w}$ be the optimal contract, and let $\widetilde{\Delta}(\widehat{w};\cdot)$ and $\widetilde{\delta}^a$ be as defined previously. Without loss of generality, label payments so that $\widehat{w}_N$ corresponds to the highest $\widehat{w}_{N-1}$ to the second highest and so on. We claim there exists only one action a' different from a^* whose constraint binds at $\widetilde{\delta}^{a'}$, and $a' < a^*$. Suppose not. Then, there exist two actions a' and a'' different from a^* such that $\sum_{j=1}^N \widetilde{\delta}_j^{a''} \widehat{w}_j - c^{a''} = \sum_{j=1}^N \widetilde{\delta}_j^{a'} \widehat{w}_j - c^{a'} = \sum_{j=1}^N \widetilde{\delta}_j^{a^*} \widehat{w}_j - c^{a^*} = \overline{w}$. By construction $\widetilde{\delta}^{a^*}$,$\widetilde{\delta}^{a'}$, and $\widetilde{\delta}^{a''}$ are extreme points of the corresponding belief sets. From here on, one can exactly follow the argument in Grossman and Hart (1983) to get the claim. If only the constraint relative to one action binds, $\widehat{w}$ must also be optimal in a problem where all other actions are dropped from the constraints. Therefore, Proposition 3 applies to that problem and the optimal contract 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. In our framework, though, this has nothing to do with output. As

[7] MLR holds if for any two actions a', a, $c^a > c^{a'}$ implies that $\frac{\delta_i^{a'}}{\delta_i^a}$ is decreasing in i. CDFC holds if for any three actions a'', a', a such that $c^a = \lambda c^{a'} + (1-\lambda) c^{a''}$, with $0 \leq \lambda \leq 1$, the following holds: $\sum_{i=1}^j \delta_i^a \leq \lambda \sum_{i=1}^j \delta_i^{a'} + (1-\lambda) \sum_{i=1}^j \delta_i^{a''}$ for all j.

we show in the next section, in our framework the optimal contract is not necessarily monotone even if there are only two actions and two states.

5 Two States and Two Actions

In this section, we specialize the model to the case where only two output levels are possible and the agent can choose between two actions. Because in the general case the optimal contract is contingent on only two events, this section can be thought of as a reduced form of the general problem. We solve explicitly for the optimal contract, and its properties are studied using diagrams. We also prove that the agent is unwilling to buy the firm from the principal at a price the principal would accept. Thus, contrary to the standard model, the agency problem cannot be avoided even if the agent is risk-neutral and has unlimited wealth. This result suggests a possible theory of the firm. Firm owners (principals) are the individuals who face less Knightian uncertainty. Workers (agents) are the individuals who face more Knightian uncertainty.

Notation is as follows. There are only two output levels y_1 and y_2, with $y_2 > y_1$. Thus, we refer to state 2 as the good state. The agent chooses between high and low effort, denoted H and L respectively. High effort is more costly, $c^H > c^L$. Because there are only two states, we can describe the agent's beliefs as probability intervals for the good state. Then, $\delta_2^a \in \left[\underline{\delta}_2^a, \overline{\delta}_2^a\right]$ with a equal to either H or L. To ease notation, from now on we drop the subscript on probabilities, and refer to δ and π as probabilities of the good state.

One can characterize the agent's beliefs according their relative position under the two different actions according to the following definition.

Definition 3 *We say the agent's beliefs about the good state are: optimistic if $\underline{\delta}^L < \underline{\delta}^H$ and $\overline{\delta}^L < \overline{\delta}^H$; extremely optimistic if $\underline{\delta}^L < \overline{\delta}^L < \underline{\delta}^H < \overline{\delta}^H$; less uncertain if $\underline{\delta}^L < \underline{\delta}^H < \overline{\delta}^H < \overline{\delta}^L$; more uncertain if $\underline{\delta}^H < \underline{\delta}^L < \overline{\delta}^L < \overline{\delta}^H$; pessimistic if $\underline{\delta}^L > \underline{\delta}^H$ and $\overline{\delta}^L > \overline{\delta}^H$ holds; extremely pessimistic if $\overline{\delta}^L > \underline{\delta}^L > \overline{\delta}^H > \underline{\delta}^H$.*

Figure 3 shows examples of these cases. Notice that the intervals are disjoint when the agent's beliefs are extremely optimistic or pessimistic. If beliefs are less uncertain, the agent feels he has more control over the production process by choosing high effort. That is, his beliefs are more

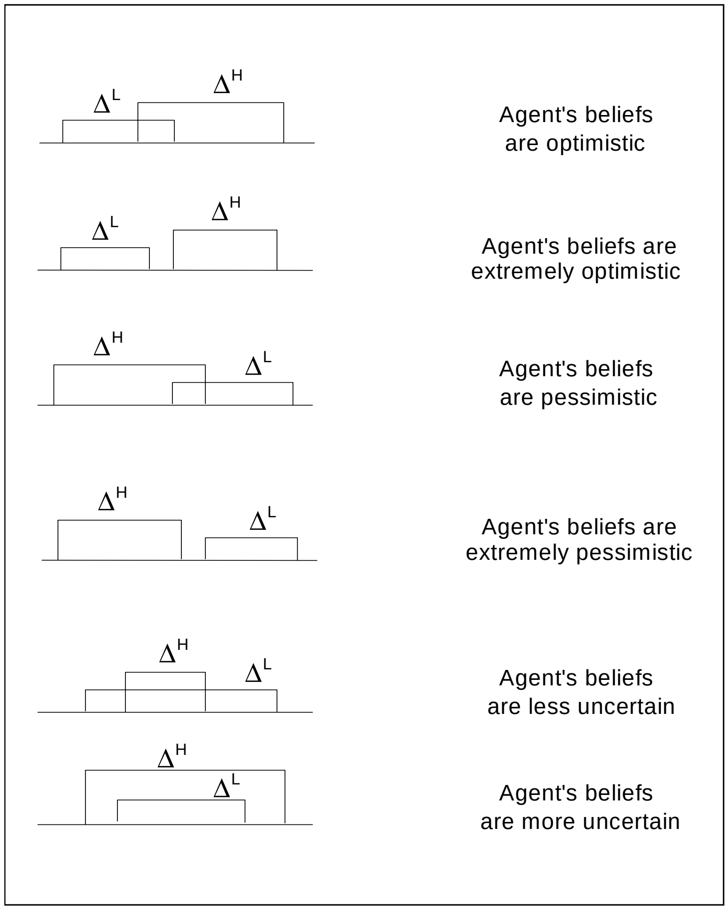

Figure 3: Characterization of the Agent's Beliefs

precise, and his probability interval shrinks. The relative position of the agent's belief influence the solution to the principal's problem. If they are optimistic, the payment the agent receives in the good state is higher. If they are pessimistic, the payment he receives in that state is lower. If they are less uncertain, depending on the position of the principal's beliefs relative to his, the agent receives payments as if he were optimistic or pessimistic. This is formally stated by the following proposition, proven in the appendix.

Proposition 6 *Assume each $\mathcal{H}^i$ is not empty and that A-2 holds. Then, the optimal incentive scheme without inertia is:*

$$\widehat{w}^1 = \left(\overline{w} + c^H - \frac{\left(c^H - c^L\right)\underline{\delta}^H}{\underline{\delta}^H - \overline{\delta}_2^L}, \overline{w} + c^H + \frac{\left(c^H - c^L\right)\left(1 - \underline{\delta}^H\right)}{\underline{\delta}^H - \overline{\delta}_2^L} \right),$$

while the optimal incentive scheme with inertia is given by

$$\widehat{w}^2 = \left(\overline{w} + c^H - \frac{(c^H - c^L)\underline{\delta}^H}{\underline{\delta}^H - \underline{\delta}^L}, \overline{w} + c^H + \frac{(c^H - c^L)\left(1 - \underline{\delta}^H\right)}{\underline{\delta}^H - \underline{\delta}^L}\right)$$

if the agent is optimistic, or he is less uncertain and $\pi^H \leq \frac{\underline{\delta}^H \overline{\delta}^L - \overline{\delta}^H \underline{\delta}^L}{\overline{\delta}^L - \overline{\delta}^H + \underline{\delta}^H - \underline{\delta}^L}$*; or*

$$\widehat{w}^2 = \left(\overline{w} + c^H + \frac{\left(c^H - c^L\right)\overline{\delta}^H}{\overline{\delta}^L - \overline{\delta}^H}, \overline{w} + c^H - \frac{\left(c^H - c^L\right)\left(1 - \overline{\delta}^H\right)}{\overline{\delta}^L - \overline{\delta}^H}\right)$$

if the agent is pessimistic, or he is less uncertain and $\pi^H > \frac{\underline{\delta}^H \overline{\delta}^L - \overline{\delta}^H \underline{\delta}^L}{\overline{\delta}^L - \overline{\delta}^H + \underline{\delta}^H - \underline{\delta}^L}$.

Neither optimal contract depends on the principal's beliefs. Therefore, it is robust to some imprecision in them, as long as this does not change the action she wants to implement, or the choice between the two possible contracts when the agent's beliefs are less uncertain. Neither optimal contract depends on the amount of uncertainty the agent faces (the length of the probability intervals), but rather on the difference between the agent's worst expectation when he works hard and one of the extreme expectations when he does not work hard.

With inertia, the optimal incentive scheme is not necessarily monotone in output. Intuitively, the agent is paid more in the bad state when he thinks the bad state is more likely (pessimistic beliefs); or when he thinks the good state is less uncertain and the principal thinks the good state is very likely. In both cases, neither side of the contract likes monotonicity. Without inertia, on the other hand, the optimal incentive scheme is always monotone in output because the agent's beliefs are extremely optimistic for $\mathcal{H}^1$ not to be empty.

Figure 4 represents the principal's problem when the agent obeys the inertia assumption. The agent's preferences are represented by one cone for each action. Contracts inside cone H are preferred to the status quo when the agent works hard. Contracts outside cone L are (at least) not comparable to the status quo when the agent does not work hard. Thus, $\mathcal{H}^2$ is composed of the two regions where these two overlap. Because the principal's beliefs are an element of the agent's belief set, the expected cost of a contract is represented by a family of lines with slope in between the slopes of the sides of cone H. The intersection between the lowest of these lines and $\mathcal{H}^2$ then identifies the

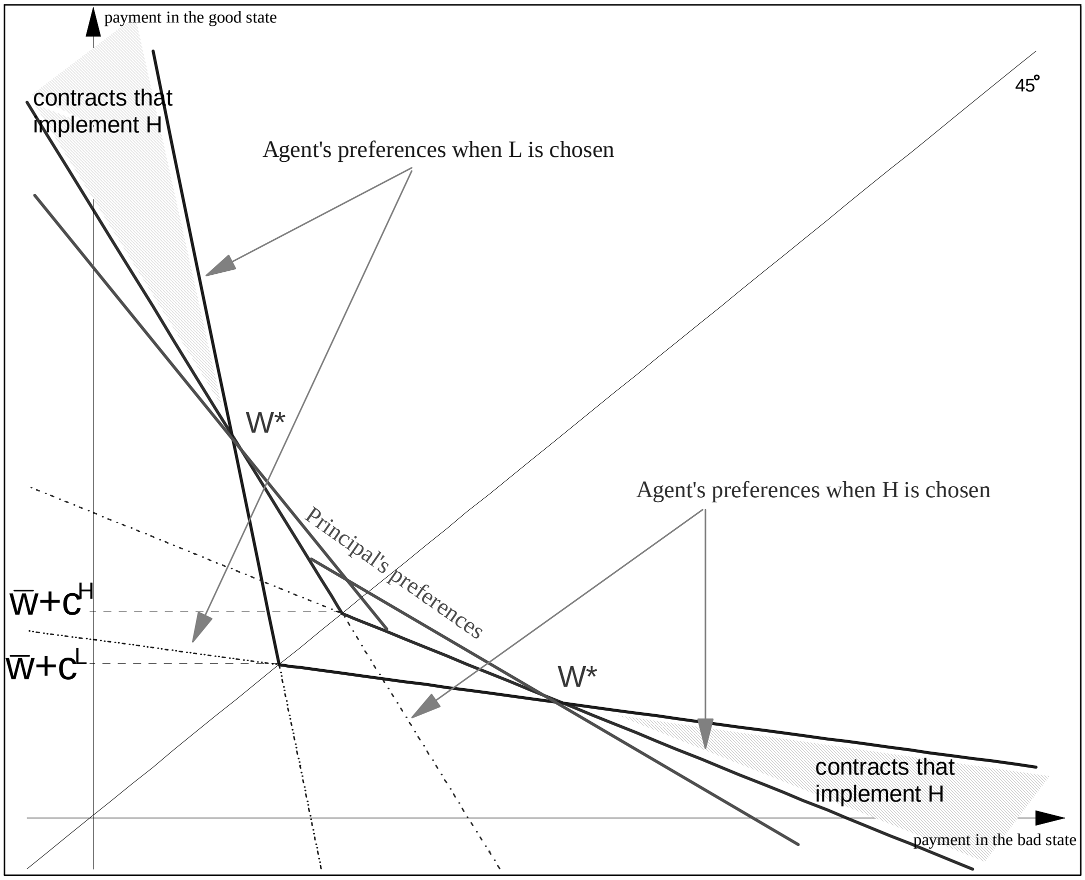

Figure 4: The Optimal Incentive Scheme with Uncertainty Aversion

(unique) optimal scheme $\widehat{w}^2$.[8] Inspection of Figure 4 reveals that the expected cost of $\widehat{w}^2$ depends negatively on the lower probability of the good state induced by L, and positively on the lower probability of the good state induced by H.

Figure 5 represents the principal's problem when the agent does not obey the inertia assumption. The agent's preferences are represented by one cone for each action. Contracts inside cone H are preferred to the status quo when the agent works hard. The status quo is preferred to contracts inside cone L when the agent does not work hard. Thus, $\mathcal{H}^1$ is the region where these two overlap. The principal's preferences are represented by a family of lines which have slope in between the slopes

[8] A similar reasoning with more than two states is what drives the proof of our main result. If the contract is contingent on three states, we can always find one where this argument applies.

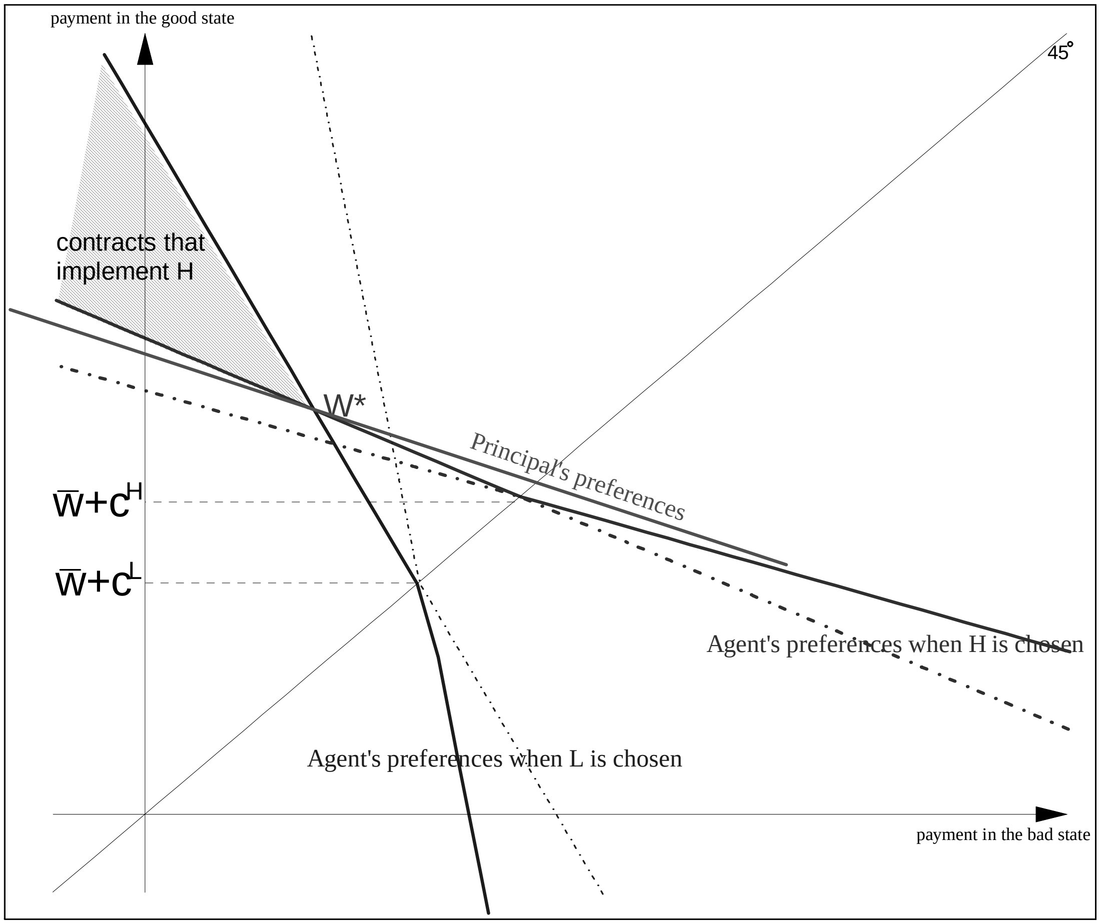

Figure 5: The Optimal Incentive Scheme with Multiple Beliefs

of the sides of the cone H. The intersection between the lowest of these lines and $\mathcal{H}^1$ identifies the (unique) optimal scheme $\widehat{w}^1$.

Inspection of Figure 5 reveals that the expected cost of $\widehat{w}^1$ depends negatively on the upper probability of the good state induced by L and positively on the lower probability of the good state induced by H. Furthermore, if the belief sets corresponding to the two actions are not disjoint, one can see that the hatched region in the diagram is empty, and there are no contracts which implement H.

From the previous diagrams, one can see how relaxing some assumptions would change the optimal contract. If the agent's cost of performing each action and/or the status quo are random, the position of the two cones would differ: their vertices would no longer be on the 45^0 line. If the agent is risk-averse, the cones have non-linear boundaries.

Finally, notice how one can easily use a diagram to show that the principal would rather employ an agent who obeys the inertia assumption than one who does not. This is obvious since, as seen in Figure 6, $\mathcal{H}^1$ is a subset of $\mathcal{H}^2$. Intuitively, an agent who does not obey the inertia assumption threatens to take the wrong action more easily, and thus is more costly for the principal. In other words, the inertia assumption helps the principal because it diminishes the appeal alternative actions have for the agent. Inertia means reluctance to abandon the status quo. Thus, the agent's conservatism prevents him from threatening to choose the low effort action.

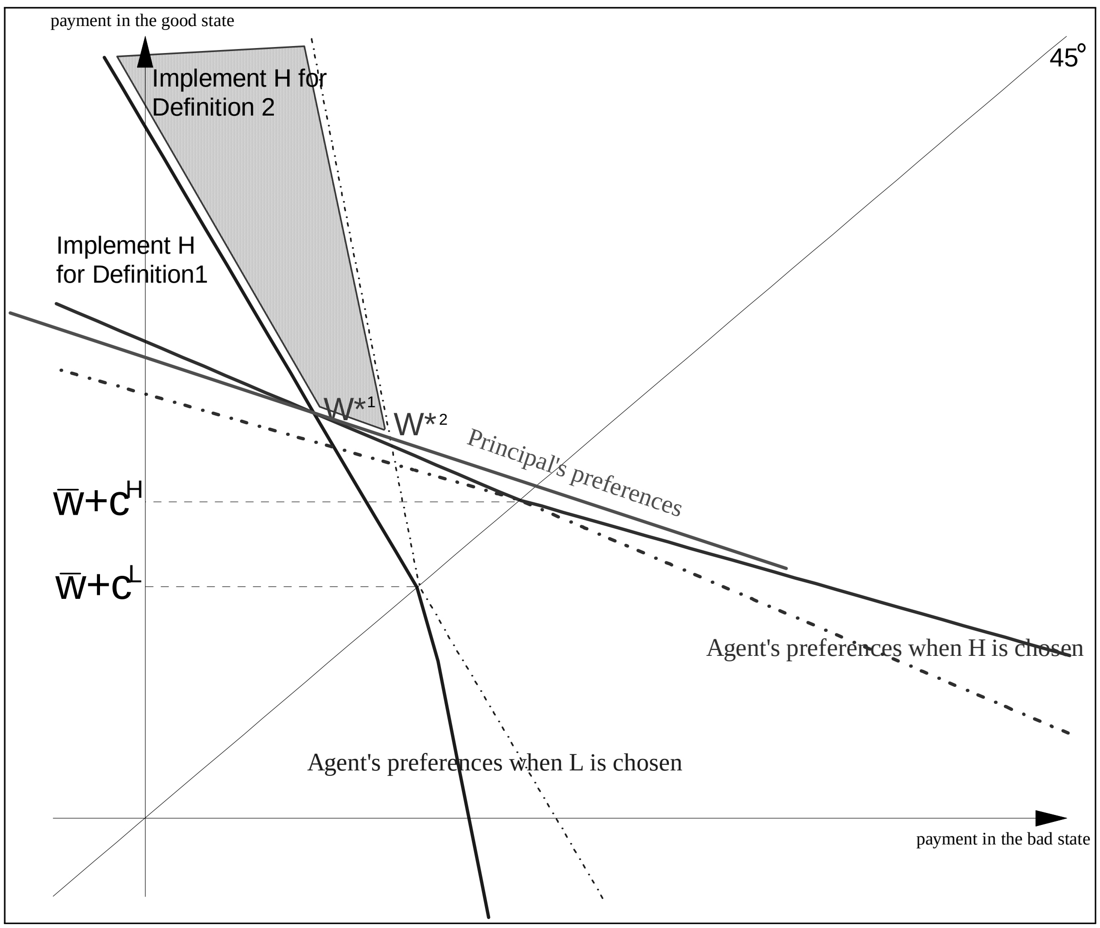

Figure 6: The Difference between Definition 1 and Definition 2

5.1 Ownership of the Production Process

We conclude this section analyzing whether risk aversion and/or wealth constraints are necessary for moral hazard to force a second best allocation. In any principal-agent model, the moral hazard problem can be avoided if the principal and the agent are willing to exchange ownership of the production process. After the transaction, the inability of the principal to observe the agent's action is no longer a factor, and the first best can be achieved. In the standard model, if the agent is risk-neutral and has enough wealth, this transaction is possible.

In our model, even though the agent is risk-neutral and no wealth constraint is imposed, there is no price both parties would accept to exchange the production process if the agent's behavior displays inertia. Knightian uncertainty aversion provides an additional rationale for the existence of agency relationships. The possible gain from trade is measured by the difference in the cost of implementing H for the two parties. This difference is never large enough to compensate for the fact that the principal and the agent do not agree on the value of the firm. Hence, the transaction will never take place. The following proposition states the result.

Proposition 7 *Suppose the agent obeys the inertia assumption, and an optimal incentive scheme exists. Then, the lowest price the principal is willing to accept to sell the firm to the agent is higher than the highest price the agent is willing to pay.*

The intuition for the proof relies on the idea that the two parties do not agree on what the firm is worth. In addition, the agent buys the firm only if doing so is preferred to the status quo. These two observations imply there is no price at which the agent is willing to buy and the principal is willing to sell.

5.2 An Example: Constant Imprecision

As an example of the different way in which imprecision affects the efficiency of the optimal contract, one can look at the case in which imprecision is constant. That is, the amount of imprecision in the agent's beliefs does not depend on the action he chooses. We think of this situation as lack of confidence independent of the agent's actions.[9] Formally, this case is described by $\overline{\delta}^a = \pi^a + \varepsilon$ and

[9] This is called a 'contagion' probability model.

$\underline{\delta}^a = \pi^a - \varepsilon$. By construction, this model represents the case where the agent's beliefs are optimistic. The constraint set for the problem without inertia is non-empty as long as $\varepsilon < \frac{\pi^H - \pi^L}{2}$.

In this case, the optimal contracts without and with inertia are:

$$\left(\overline{w} + c^H + \frac{\varepsilon\left(c^H - c^L\right) - \left(c^H - c^L\right)\pi^H}{\pi^H - \pi^L - 2\varepsilon}, \overline{w} + c^H + \frac{\varepsilon\left(c^H - c^L\right) + \left(c^H - c^L\right)\left(1 - \pi^H\right)}{\pi^H - \pi^L - 2\varepsilon}\right)$$

and

$$\left(\overline{w} + c^H + \frac{\varepsilon\left(c^H - c^L\right) - \left(c^H - c^L\right)\pi^H}{\pi^H - \pi^L}, \overline{w} + c^H + \frac{\varepsilon\left(c^H - c^L\right) - \left(c^H - c^L\right)\pi^H}{\pi^H - \pi^L}\right)$$

Imprecision affects the optimal contract symmetrically. Intuitively, this is because, by assumption, imprecision does not depend on the action the agent chooses. Therefore, relative to the standard case, it does not introduce additional incentive problems.

We now look at the actual form of the contract. Before proceeding, we need to rewrite the optimal contract in a form that makes the comparison easier. When there are only two states, any contract can be described by a fixed payment and a share of realized output. Let f be the fixed amount and s the share of realized output. Then, for any contract $w = w_1, w_2$, we have

$$\begin{aligned} f &= w_1 - y_1 \frac{w_2 - w_1}{y_2 - y_1} \\ s &= \frac{w_2 - w_1}{y_2 - y_1}. \end{aligned}$$

Using the formula above, one can easily determine the optimal fixed payments in the standard model and in our model without and with inertia as:

$$\begin{aligned} \widehat{f} &= \overline{w} + c^H + \frac{-\pi^H\left(c^H - c^L\right) - \frac{y_1\left(c^H - c^L\right)}{y_2 - y_1}}{\pi^H - \pi^L} \\ \widehat{f}^1 &= \overline{w} + c^H + \frac{\varepsilon\left(c^H - c^L\right) - \pi^H\left(c^H - c^L\right) - \frac{y_1\left(c^H - c^L\right)}{y_2 - y_1}}{\underline{\delta}^H - \overline{\delta}^L - 2\varepsilon} \\ \widehat{f}^2 &= \overline{w} + c^H + \frac{\varepsilon\left(c^H - c^L\right) - \pi^H\left(c^H - c^L\right) - \frac{y_1\left(c^H - c^L\right)}{y_2 - y_1}}{\pi^H - \pi^L} \end{aligned}$$

The optimal shares in the standard model and in our model without and with inertia are:

$$\widehat{s} = \frac{\frac{c^H - c^L}{y_2 - y_1}}{\pi^H - \pi^L}$$

$$\widehat{s}^1 = \frac{\frac{c^H - c^L}{y_2 - y_1}}{\pi^H - \pi^L - 2\varepsilon}$$

$$\widehat{s}^2 = \frac{\frac{c^H - c^L}{y_2 - y_1}}{\pi^H - \pi^L}$$

An agent whose beliefs are imprecise receives a higher base payment; furthermore, the fixed payment increases with the amount of imprecision faced by the agent. Without inertia, the share of output is higher than in the standard model, and increases with imprecision. With inertia, the share of realized output does not depend on imprecision, and is equal to the one in the standard model. Therefore, in this case the main effect of imprecision is to require a higher base payment to the agent. Intuitively, because imprecision of beliefs is independent of the agent's choice, the incentive part of the contract is unaffected. This result does not hold without inertia, because the agent evaluates probabilities asymmetrically.

Summarizing, in the special case where uncertainty does not depend on the effort choice, the loss in efficiency of the optimal contract is always positive, and increasing with imprecision of the agent's beliefs. The optimal contract requires a higher base payment, while the payment dependent on realized output may be unaffected.

6 Conclusions

In some agency situations, the agent may not feel as confident as the principal in evaluating uncertainty. For example, the agent may be an outsider who does not control all aspects of the production process. If this is the case, the agent lacks confidence about his influence on the possible output outcomes. This aspect of agency relationships has been generally neglected in economic theory. The framework we adopt is useful to introduce robustness requirements for optimal contracts.

In this setting, an incentive scheme is robust by definition. We showed how the main consequence of this demand for robustness is simplicity of the optimal contract. Furthermore, if the agent's behavior displays inertia, the moral hazard problem cannot be solved by selling the firm to the agent. Contrary to the standard model, this result holds in spite of the agent's being risk-neutral and having unlimited wealth.

Our last result states that an agent who has multiple beliefs will not buy the firm from the principal. This has an interesting implication if interpreted as theory of the firm. A division of tasks where the individual who faces more Knightian uncertainty is the residual claimant is in some sense efficient. Our model suggest a description of the firm like Knight's [1921]. Entrepreneurs are individuals who perceives the business as risky, while workers perceive it as uncertain. Although in a different decision theoretic framework, an first attempt in this direction is made in Rigoti, Ryan, and Vaithianathan (2001).

In this paper, we have analyzed a partial equilibrium framework while disregarding any issues related to risk aversion. Rigoti and Shannon (2001) present a full analysis of the role of uncertainty (and risk aversion) for mutual insurance, and obtain conditions for optimal uncertainty *and* risk sharing. An interesting direction for future research is to study the connection between our partial equilibrium setting and their general equilibrium model.

A Appendix

A.1 Proof of Proposition 2

We prove each statement separately.

Proof of (i). Suppose not. Then, $\widehat{w}^a$ is optimal and $\sum_{j=1}^N \widetilde{\delta}_j^a \widehat{w}_j^a - c^a > \overline{w}$. 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, $\sum_{j=1}^N \pi_j^a \widetilde{w}_j < \sum_{j=1}^N \pi_j^a \widehat{w}_j^a - \varepsilon$, contradicting the optimality of $\widehat{w}^a$.

Proof of (ii). Because π^1 is an element of Δ^1, for any payment scheme w we have $\sum_{j=1}^N \widetilde{\delta}_j^1 w_j \leq \sum_{j=1}^N w_j \pi_j^1$. The scheme $w_j = \overline{w} + c^1$ 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 (2) because $\overline{w} + c^1 = \sum_{j=1}^N \widetilde{\delta}_j^1 \widehat{w}_j^1 = \sum_{j=1}^N \widehat{w}_j^1 \pi_j^1$.

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

$$\sum_{j=1}^N \widetilde{\delta}_j^a \widehat{w}_j^a - c^a > \sum_{j=1}^N \widetilde{\widetilde{\delta}}_j^{a'} \widehat{w}_j^a - c^{a'}$$

or

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

for some $\delta^{a'}$ in $\Delta^{a'}$. The latter inequality implies $\sum_{j=1}^N \widetilde{\delta}_j^{a'} \widehat{w}_j^a - c^{a'} < \overline{w}$. Because none of the respective constraints binds, $\widehat{w}^a$ is a solution for a problem like (2) 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. Thus, $\overline{w} + c^a = \sum_{j=1}^N \widetilde{\delta}_j^a \widehat{w}_j^a = \sum_{j=1}^N \widetilde{\widetilde{\delta}}_j^{a'} \widehat{w}_j^a = \sum_{j=1}^N \widetilde{\delta}_j^{a'} \widehat{w}_j^a$. Hence, the two inequalities above both imply $c^{a'} > c^a$ contradicting $a' < a$. ■

A.2 Proof of the Main Result

In the proof, we need the following result from ?.

Lemma 1 *Let v^a be a convex capacity, and let $\widetilde{\Delta}^a$, with generic element $\widetilde{\delta}^a$, be the set of extreme points of Δ^a. Then, for any N-vector z such that $z_1 \leq z_2 \leq .. \leq z_N$,*

$$\min_{\delta^a \in \Delta^a} \sum_{j=1}^N \delta_j^a z_j$$

is attained for the $\widetilde{\delta}^a$ in $\widetilde{\Delta}^a$ such that $\sum_{s=j}^N \delta_s^a = \sum_{s=j}^N v_s^a$ for each $j = 2, .., N$.

The proof follows from ?, Propositions 10 and 13.

When Assumption A-1 holds, the lemma says provides a characterization of the probability distributions that, among a set, yield the lowest and largest expected values. It can be used to conclude that the lowest expectation the agent can compute for a given z is attained at an extreme point of Δ^a such that $\delta_K^a < \pi_K^a$ and $\delta_1^a \geq \pi_1^a$. A consequence of Lemma 1 is the following corollary.

Corollary 1 *Let z and z' be two an N-vector such that $z_1 \leq .. \leq z_N$ and $z'_1 \leq .. \leq z'_N$. Then, the solution to*

$$\min_{\delta^a \in \Delta^a} \sum_{j=1}^{N} \delta_j^a z_j$$

also solves

$$\min_{\delta^a \in \Delta^a} \sum_{j=1}^{N} \delta_j^a z'_j$$

Proof: We can always write

$$\begin{aligned} \sum_{j=1}^{N} \delta_j^a z_j &= z_1 + \sum_{j=2}^{N} (z_j - z_{j-1}) \sum_{s=j}^{N} \delta_s^a \\ \sum_{j=1}^{N} \delta_j^a z'_j &= z'_1 + \sum_{j=2}^{N} \left(z'_j - z'_{j-1}\right) \sum_{s=j}^{N} \delta_s^a \end{aligned}$$

Lemma 1 says each minimization problem is solved by minimizing $\sum_{s=j}^{N} \delta_s^a$ for each $j = 2, .., N$. Because z and z' are ordered in the same way both minimization problems yield the same solution. ■

When two incentive scheme are 'ordered' in the same way their lowest expectation is attained using the same distribution.

A.2.1 Proof of proposition 3

The main step in the proof is to show, by contradiction, than an optimal contract cannot be contingent on more than two groups 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 $\widehat{w}$ (in this subsection we drop the superscript in denoting the optimal contract). 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 the different k's. 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 $\widehat{w}_k$. By construction, K is the number of events the optimal contract is contingent upon, and $\widehat{w}_1 < \widehat{w}_2 < .. < \widehat{w}_K$. We need to show K equals 2.

Define $\widetilde{\Delta}^H(\widehat{w}) = \left\{ \delta^H \in \Delta^H \mid \delta^H = \arg\min_{\delta^H \in \Delta^H} \sum_{j=1}^{N} \delta_j^H \widehat{w}_j \right\}$, and similarly for $\widetilde{\Delta}^L(\widehat{w})$. Let $\widetilde{\delta}^a$ denote an element of $\widetilde{\Delta}^a(\widehat{w})$ for $a = \{H, L\}$. By Proposition 2, $\widehat{w}$ must satisfy

$$\sum_{k=1}^{K} \widetilde{\delta}_k^H \widehat{w}_k = \overline{w} + c^H \tag{4}$$

We claim $\widehat{w}$ must also satisfy

$$\sum_{k=1}^{K} \widetilde{\delta}_k^L \widehat{w}_k = \overline{w} + c^L \tag{5}$$

Suppose not. Then, $\sum_{k=1}^{K}\delta_k^L \widehat{w}_k < \overline{w}+c^L$. Let $\widetilde{w} = \left(\widehat{w}_1 - \varepsilon\frac{\widetilde{\delta}_K^H}{\widetilde{\delta}_1^H}, \widehat{w}_2, .., \widehat{w}_{K-1}, \widehat{w}_K + \varepsilon\right)$, where ε is positive and small enough so that the ranking of the payments for $\widetilde{w}$ and $\widehat{w}$ is the same. By Corollary 1, $\widetilde{\delta}^H$ and $\widetilde{\delta}^L$ minimize the expected value of $\widetilde{w}$ for the agent. For any distribution δ, the expected values of $\widetilde{w}$ and $\widehat{w}$ are related by the following:

$$\sum_{k=1}^{K}\delta_k\widetilde{w}_k - \sum_{k=1}^{K}\delta_k\widehat{w}_k = \frac{\delta_1\widetilde{\delta}_K^H - \delta_K\widetilde{\delta}_1^H}{\widetilde{\delta}_1^H}\varepsilon \tag{6}$$

$\delta = \widetilde{\delta}^H$ implies the right hand side of (6) is equal to 0; thus, $\widetilde{w}$ satisfies (4). For some ε close enough to zero and $\delta = \widetilde{\delta}^L$ the right hand side of (6) is very small; thus $\widetilde{w}$ satisfies (NC) because $\widehat{w}$ satisfies it strictly. By Lemma 1 and A-1, $\widetilde{\delta}_K^H < \pi_K^H$ and $\widetilde{\delta}_1^H \geq \pi_1^H$; thus, the right hand side of (6) is negative when $\delta = \pi^H$. 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 and $K > 2$. $\widehat{w}$ satisfies Equations (4) and (5). These constitute a system of two equations which can be solved for $\widehat{w}_K$ and some $\widehat{w}_{k'}$, yielding:

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

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

These are well defined, unless

$$0 = \widetilde{\delta}_K^L\widetilde{\delta}_{k'}^H - \widetilde{\delta}_{k'}^L\widetilde{\delta}_K^H \tag{9}$$

for all $k' \neq K$. In that case:

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

Using this result in 9:

$$0 = \widetilde{\delta}_K^H\left(\widetilde{\delta}_{k'}^H - \widetilde{\delta}_{k'}^L\right)$$

Thus, either $0 = \widetilde{\delta}_K^H = \widetilde{\delta}_K^L$, or $\widetilde{\delta}_{k'}^H = \widetilde{\delta}_{k'}^L$ for all $k' \neq K$. If the former happens, $\widehat{w}$ cannot be optimal because it makes the largest payment in a state that does not affect the constraints and, by A-2, has positive probability for the principal. If the latter happens, $\widetilde{\delta}^L = \widetilde{\delta}^H$. Then $\sum_{k=1}^{K} \widetilde{\delta}_k^H \widehat{w}_k = \sum_{k=1}^{K} \widetilde{\delta}_k^L \widehat{w}_k$ 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:

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

where

$$\widehat{\alpha} = \frac{\pi_K^H \widetilde{\delta}_{k'}^L - \pi_{k'}^H \widetilde{\delta}_K^L + \pi_{k'}^H \widetilde{\delta}_K^H - \pi_K^H \widetilde{\delta}_{k'}^H}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} \overline{w} + \frac{\pi_K^H \widetilde{\delta}_{k'}^L - \pi_{k'}^H \widetilde{\delta}_K^L}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} c^H + \frac{\pi_{k'}^H \widetilde{\delta}_K^H - \pi_K^H \widetilde{\delta}_{k'}^H}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} c^L \tag{10}$$

and

$$\widehat{\beta}_k = \pi_k^H - \widetilde{\delta}_k^H \frac{\pi_K^H \widetilde{\delta}_{k'}^L - \pi_{k'}^H \widetilde{\delta}_K^L}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} - \widetilde{\delta}_k^L \frac{\pi_{k'}^H \widetilde{\delta}_K^H - \pi_K^H \widetilde{\delta}_{k'}^H}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} \tag{11}$$

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 (11),

$$\pi_k^H \left(\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H \right) = \widetilde{\delta}_k^H \left(\pi_K^H \widetilde{\delta}_{k'}^L - \pi_{k'}^H \widetilde{\delta}_K^L \right) + \widetilde{\delta}_k^L \left(\pi_{k'}^H \widetilde{\delta}_K^H - \pi_K^H \widetilde{\delta}_{k'}^H \right)$$

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

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

This implies:

$$\frac{\pi_K^H \widetilde{\delta}_{k'}^L - \pi_{k'}^H \widetilde{\delta}_K^L}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} = \frac{\left(\widetilde{\delta}_K^L - \pi_K^H \right)}{\left(\widetilde{\delta}_K^L - \widetilde{\delta}_K^H \right)} \quad \text{and} \quad \frac{\pi_{k'}^H \widetilde{\delta}_K^H - \pi_K^H \widetilde{\delta}_{k'}^H}{\widetilde{\delta}_{k'}^L \widetilde{\delta}_K^H - \widetilde{\delta}_K^L \widetilde{\delta}_{k'}^H} = \frac{\left(\pi_K^H - \widetilde{\delta}_K^H \right)}{\left(\widetilde{\delta}_K^L - \widetilde{\delta}_K^H \right)}$$

We know that $\widetilde{\delta}_K^H < \pi_K^H$. Thus, $\frac{\left(\widetilde{\delta}_K^L - \pi_K^H \right)}{\left(\widetilde{\delta}_K^L - \widetilde{\delta}_K^H \right)} < 1$ and $\frac{\left(\pi_K^H - \widetilde{\delta}_K^H \right)}{\left(\widetilde{\delta}_K^L - \widetilde{\delta}_K^H \right)} > 0$. Hence:

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

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: if $\widehat{\beta}_{k''} > 0$

$$\begin{aligned}
\widetilde{w}_k &= \widehat{w}_k && \text{when} \quad k \neq K, k', k'' \\
\widetilde{w}_K &= \widehat{w}_K + \tfrac{\widetilde{\delta}^H_{k'}\widetilde{\delta}^L_{k''} - \widetilde{\delta}^H_{k''}\widetilde{\delta}^L_{k'}}{\widetilde{\delta}^L_{k'}\widetilde{\delta}^H_{K} - \widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k'}}\varepsilon \\
\widetilde{w}_{k'} &= \widehat{w}_{k'} + \tfrac{\widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k''} - \widetilde{\delta}^H_{K}\widetilde{\delta}^L_{k''}}{\widetilde{\delta}^L_{k'}\widetilde{\delta}^H_{K} - \widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k'}}\varepsilon \\
\widetilde{w}_{k''} &= \widehat{w}_{k''} + \varepsilon
\end{aligned}$$

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. Corollary 1 applies, and $\widetilde{\delta}^H$ and $\widetilde{\delta}^L$ yield the lowest expected values of $\widetilde{w}$ for H and L. Hence, we know that, if one defines $\widetilde{\alpha}$ and $\widetilde{\beta}$ using equations (10) and (11), $\widetilde{\alpha} = \widehat{\alpha}$ and $\widetilde{\beta} = \widehat{\beta}_k$. Moreover,

$$\begin{aligned}
0 &= \widetilde{\delta}^L_K \frac{\widetilde{\delta}^H_{k'}\widetilde{\delta}^L_{k''} - \widetilde{\delta}^H_{k''}\widetilde{\delta}^L_{k'}}{\widetilde{\delta}^L_{k'}\widetilde{\delta}^H_{K} - \widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k'}} + \widetilde{\delta}^L_{k'} \frac{\widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k''} - \widetilde{\delta}^H_{K}\widetilde{\delta}^L_{k''}}{\widetilde{\delta}^L_{k'}\widetilde{\delta}^H_{K} - \widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k'}} + \widetilde{\delta}^L_{k''} \\
0 &= \widetilde{\delta}^L_K \frac{\widetilde{\delta}^H_{k'}\widetilde{\delta}^L_{k''} - \widetilde{\delta}^H_{k''}\widetilde{\delta}^L_{k'}}{\widetilde{\delta}^L_{k'}\widetilde{\delta}^H_{K} - \widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k'}} + \widetilde{\delta}^L_{k'} \frac{\widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k''} - \widetilde{\delta}^H_{K}\widetilde{\delta}^L_{k''}}{\widetilde{\delta}^L_{k'}\widetilde{\delta}^H_{K} - \widetilde{\delta}^L_{K}\widetilde{\delta}^H_{k'}} + \widetilde{\delta}^L_{k''}
\end{aligned}$$

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

$$\begin{aligned}
\sum_{k=1}^{K} \pi^H_k \widetilde{w}_k &= \widetilde{\alpha} + \sum_{k\neq K,k'} \widetilde{\beta}_k \widetilde{w}_k = \widehat{\alpha} + \sum_{k\neq K,k'} \widehat{\beta}_k \widetilde{w}_k \\
&= \widehat{\alpha} + \sum_{k\neq K,k'} \widehat{\beta}_k \widehat{w}_k + \widehat{\beta}_{k''}\varepsilon
\end{aligned}$$

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 the latter. 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 $K = 2$. ■

A.3 Proof of Proposition 6

Because there are only two actions, Proposition 2 implies participation and incentive compatibility are binding when computed according to the extreme expectations. Since there are only two states, the smallest (largest) expected value is reached by putting as much weight as possible on the smallest (largest) payment. These constitute a system of two equations in two unknowns, and $\widehat{w}^1$ is the solution to this system. A similar argument applies for $\widehat{w}^2$, taking into account that during the proof of Proposition 3 we showed that the non-comparability constraint must bind at the optimum. There may be two contracts that solve the system constituted by participation and non comparability constraints, according to the relative positions of the action's respective belief intervals. In this case, the principal chooses the cheapest of these two feasible schemes depending on the position of her beliefs relative to the agent's. ■

A.4 Proof of Proposition 7

The proof is tedious because we need to allow the agent to consider any possible action after he acquires the firm. We then show that there is no mutually acceptable trading price. The agent buys the firm only if there exists an action a such that, for all $\delta^a \in \Delta^a$

$$E_{\delta^a}[y] - c^a - P \geq \overline{w}$$

where P is the price he pays, and a is either H or L. Because $y_2 > y_1$, this condition is satisfied if and only if it holds for $\delta^a = (1 - \underline{\delta}^a, \underline{\delta}^a)$. Let P^a be the highest price the agent would pay. This is:

$$P^a = (1 - \underline{\delta}^a)\, y_1 + \underline{\delta}^a y_2 - c^a - \overline{w}$$

The principal sells if and only if she receives at least her valuation of the firm. This is the difference between the expected value of output and the expected cost of implementing action H. That is:

$$E_{\pi^H}[y] - C^2(H)$$

Where the cost of the incentive scheme is

$$C^2(H) = \begin{cases} \overline{w} + c^H + \frac{\pi^H - \underline{\delta}^H}{\underline{\delta}^H - \underline{\delta}^L}\left(c^H - c^L\right) \\ \text{or} \\ \overline{w} + c^H + \frac{\overline{\delta}^H - \pi^H}{\overline{\delta}^L - \overline{\delta}^H}\left(c^H - c^L\right) \end{cases} \tag{12}$$

depending on which of the two possible schemes is optimal.

We need to show that $P^a < E_{\pi^H}[y] - C^2(H)$. Suppose not. Then,

$$P^a \geq E_{\pi^H}[y] - C^2(H) \tag{13}$$

Depending on the action the agent takes and the beliefs of the parties, we have several cases.

Suppose the agent chooses the low effort level. Then, (13) becomes:

$$\left(1 - \underline{\delta}^L\right) y_1 + \underline{\delta}^L y_2 - c^L - \overline{w} \geq \left(1 - \pi^H\right) y_1 + \pi^H y_2 - C^2(H) \tag{14}$$

CASE L-1. The agent's beliefs are optimistic; or they are less uncertain, and the principal's beliefs are relatively pessimistic. Substituting for $C^2(H)$ from equation (12) and rearranging, (14) yields:

$$\frac{c^H - c^L}{\underline{\delta}^H - \underline{\delta}^L} \geq y_2 - y_1 \tag{15}$$

If the principal prefers to implement H, we have:

$$y_2 - y_1 > \frac{(\pi^H - \pi^L)(y_2 - y_1)}{\pi^H - \underline{\delta}^L} \geq \frac{c^H - c^L}{\underline{\delta}^H - \underline{\delta}^L}$$

which proves (15) does not hold.

CASE L-2. The agent's beliefs are pessimistic; or they are less uncertain, and the principal's beliefs are relatively optimistic. Substituting $C^2(H)$ from equation (12) and rearranging, (14) yields:

$$\frac{\left(c^H - c^L\right)\left(\overline{\delta}^L - \pi^H\right)}{\overline{\delta}^L - \overline{\delta}^H} \geq \left(\pi^H - \underline{\delta}^L\right)(y_2 - y_1) \tag{16}$$

If the principal prefers to implement H, and $\underline{\delta}^L \leq \pi^L$, we have:

$$(\pi^H - \underline{\delta}^L)(y_2 - y_1) > (\pi^H - \pi^L)(y_2 - y_1) \geq \frac{(c^H - c^L)(\overline{\delta}^L - \pi^H)}{\overline{\delta}^L - \overline{\delta}^H}$$

which proves (16) does not hold.

Suppose the agent chooses the high effort level. Then, (13) is:

$$\left(1 - \underline{\delta}^H\right) y_1 + \underline{\delta}^H y_2 - c^H - \overline{w} \geq \left(1 - \pi^H\right) y_1 + \pi^H y_2 - C^2(H) \tag{17}$$

CASE H-1. The agent's beliefs are optimistic; or they are less uncertain, and the principal's beliefs are relatively pessimistic. Substituting $C^2(H)$ from equation (12) and rearranging, (17) yields:

$$\frac{c^H - c^L}{\underline{\delta}^H - \underline{\delta}^L} \geq y_2 - y_1$$

which cannot hold because of CASE L-1.

CASE H-2. The agent's beliefs are pessimistic; or they are less uncertain, and the principal's beliefs are relatively optimistic. Substituting $C^2(H)$ from equation (12) and rearranging, equation (17) yields:

$$\frac{(c^H - c^L)(\overline{\delta}^H - \pi^H)}{\overline{\delta}^L - \overline{\delta}^H} \geq (\pi^H - \underline{\delta}^H)(y_2 - y_2) \tag{18}$$

We distinguish two subcases.

H-2a. The agent's beliefs are pessimistic. Because the principal prefers to implement H, and $\underline{\delta}^H < \underline{\delta}^L < \pi^L$:

$$\left(\pi^H - \underline{\delta}^H\right)(y_2 - y_1) > \left(\pi^H - \pi^L\right)(y_2 - y_1) \geq \frac{\left(c^H - c^L\right)\left(\overline{\delta}^L - \pi^H\right)}{\overline{\delta}^L - \overline{\delta}^H}$$

which implies equation (18) does not hold.

H-2b. The agent's beliefs are less uncertain and the principal's are relatively optimistic. Because the principal prefers to implement H:

$$\left(\pi^H - \underline{\delta}^H\right)(y_2 - y_1) \geq \frac{\left(\pi^H - \underline{\delta}^H\right)\left(\overline{\delta}^L - \pi^H\right)\left(c^H - c^L\right)}{(\pi^H - \pi^L)\left(\overline{\delta}^L - \overline{\delta}^H\right)}$$

If the principal is relatively optimistic, $\left(\overline{\delta}^H - \pi^H\right)\left(\pi^H - \pi^L\right) < \left(\pi^H - \underline{\delta}^H\right)\left(\overline{\delta}^L - \pi^H\right)$ implying that equation (18) cannot hold.

We ruled out all possibilities; thus, the agent does not prefer buying the firm to the status quo, and the proof is complete. ■

References

Ameriks, J., and S. P. Zeldes (2000): "How Do Household Portfolio Shares Vary with Age?," Discussion paper, Columbia University.

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.

Einhorn, H. J., and R. M. Hogarth (1985): "Ambiguity and Uncertainty in Probabilistic Inference," *Psychological Review*, 92, 433–461.

Fox, C. R., and A. Tversky (1995): "Ambiguity Aversion and Comparative Ignorance," *Quarterly Journal of Economics*, 110, 586–603.

Ghirardato, P. (1994): "Agency Theory with Uncertainty Aversion," Discussion paper, Caltech.

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

Hart, O. D. (1995): *Firms, Contracts, and Financial Structure*. Oxford: Oxford University Press.

Hart, O. D., and B. Holmstrom (1987): "The Theory of Contracts," in *Advances in Economic Theory*, ed. by T. F. Bewley.

Heath, C., and A. Tversky (1991): "Preference and Belief: Ambiguity and Competence in Choice Under Uncertainty," *Journal of Risk and Uncertainty*, 4, 5–28.

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.

Moore, J., and O. D. Hart (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755–785.

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.

Rigoti, L., M. Ryan, and R. Vaithianathan (2001): "Entrepreneurial Innovation," Discussion paper, U.C. Berkeley.

Rigoti, L., and C. M. Shannon (2001): "Uncertainty and Risk in Financial Markets," Discussion paper, U.C. Berkeley.

Samuelson, W., and R. Zeckhauser (1988): "Status Quo Bias in Decision Making," *Journal of Risk and Uncertainty*, 1, 7–59.

Schmeidler, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57, 571–587.

Tirole, J. (1999): "Incomplete Contracts: Where Do We Stand?," *Econometrica*, 67, 741–781.

von Neumann, J., and O. Morgenstern (1953): *Theory of Games and Economic Behavior.* Princeton: Princeton University Press.