ESSAYS IN ROBUST MECHANISM DESIGN

by

Maxwell Rosenthal

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by ***Maxwell Rosenthal***, titled *Essays in Robust Mechanism Design* and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Andreas Blume Date: 03/28/2019

Asaf Plan Date: 03/28/2019

Stanley Reynolds Date: 03/28/2019

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

We hereby certify that we have read this dissertation prepared under our direction and recommend that it be accepted as fulfilling the dissertation requirement.

Andreas Blume Date: 03/28/2019
Professor
Department of Economics

Asaf Plan Date: 03/28/2019
Assistant Professor
Department of Economics

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

ABSTRACT

Each of the three chapters of my dissertation Essays in Robust Mechanism Design belong to a developing subfield of mechanism design that seeks to provide foundations for simple and intuitive mechanisms that perform well in a wide variety of settings. In particular, my work seeks to identify contracts and mechanisms that exhibit robustness to large-scale uncertainty about agent preferences.

My job market paper A Revealed Preference Approach to Multidimensional Screening develops a model of data-driven multidimensional screening, with applications to multi-good monopoly pricing and the design of complex products. In this paper, the monopolist observes a population of consumers each purchase products from one or more sets of alternatives, and she uses this choice data as the basis for her beliefs about the distribution of the buyer's preferences. However, there are many distributions of preferences that rationalize the data, and the monopolist evaluates product lineups according to their worst-case payoff against this set of distributions, in the spirit of robustness. I identify circumstances under which the monopolist can do better than to simply re-create one of the product lineups in her data set, and more broadly show how the form of the optimal selling mechanism changes under various natural restrictions to the buyer's preferences. In particular, if the monopolist is uncertain about complementarity or substitutability between product attributes, I show that it is optimal for her to sell only products that are vertically but not horizontally differentiated from the products in her data set.

As part of larger project studying incentives for risk-taking, the second chapter of my dissertation Robust Contracting with Uncertain Risk Preferences studies a moral hazard problem in which the principal does not know the agent's risk preferences. The agent chooses not only how much effort to exert, but also potentially from a variety of safe and risky actions. While the existing moral hazard literature has almost uniformly employed the assumption that the principal exactly knows the agent's risk preferences, I consider a principal who seeks a contract that performs well for all types of agents. I show that all such contracts are capped bonus contracts which do not reward the agent for producing output that is either very small

or very large. As a special case of my model, I identify restrictions to effort costs under which binary contracts that pay the agent a base salary and reward him with a fixed bonus payment for achieving a specified output quota are optimal.

In a related paper, the third chapter of my dissertation Risk Alignment considers a moral hazard problem in which the principal is slightly uncertain about the agent's risk preferences. As before, the agent chooses not only how much effort to exert, but also what sort of risks to take. The principal's payoff depends on both the output produced by the agent and on transfers, and her risk preferences generally differ from those of the agent. In the first part of the paper, I completely characterize the set of risk aligned contracts, which provide the agent with incentives to choose risks as if his goal were to maximize the principal's payoff. Very generally, all such contracts "pay the agent in probability", akin to the lottery mechanisms that are used by researchers in experimental economics to induce risk preferences in laboratory subjects. In the second part of the paper, I show how risk aligned contracts are worst-case optimal for the principal.

CHAPTER 1

A REVEALED PREFERENCE APPROACH TO MULTIDIMENSIONAL SCREENING

1.1 Introduction

A monopolist hires an analyst to design and price a lineup of products to be sold to consumers. The monopolist possesses population-level purchasing data, and she wishes for the analyst to incorporate this data into his designs. In particular, the data is useful for the analyst because it reflects information about the distribution of the consumers' preferences. Although the analyst might himself have prior beliefs about consumer tastes, the monopolist desires an objective analysis that is entirely data-driven. What might the analyst infer about the consumers' preferences from data? How can he use these inferences to design an optimal product lineup for the monopolist? In particular, are there product lineups that perform better than those that merely recreate one of the choice environments in the monopolist's data?

This paper studies a general multidimensional screening problem in which a principal makes allocations to an agent with private information. Allocations consist of a *characteristic* and a price. For example, in the application of our model to multi-good monopoly selling, a characteristic might be a bundle consisting of various quantities of two or more distinct goods. Alternatively, a characteristic might describe the quality of each feature of a multi-attribute good, as in Mussa and Rosen (1978); Wilson (1993); Amstrong (1996); Rochet and Chone (1998). More broadly, we interpret the multidimensional screening exercise as one in which the set of objects to be potentially allocated to the agent is an abstract partially-ordered set, and abstain from tailoring our analysis to the details of any particular application.

In our model, prior to interacting with the agent, the principal observes the joint distribution of choices made by a population of decision makers from one or more sets of exogenous *reference allocations.* Because decision makers maximize their utility and the agent is a typical member of the population, the decision makers' choices are informative about the agent's preferences. However, the set of preferences potentially held by the decision makers is rich. Consequently, there are many distributions of preferences that are consistent with the choice data, and the principal ranks mechanisms according to the expected payoff that they *guarantee* her against this set of distributions.

This paper is primarily concerned with the relationship between the principal's optimal mechanisms and the choice environments constituent to her data. In particular, an immediate question is the matter of whether or not there are mechanisms that outperform those mechanisms that merely recreate one of the choice environments in the data. If there are, what sort of characteristics do they allocate to the agent? What prices do they use?

In the leading specification of our model — which we call the *monotone environment* — the principal knows only that the agent prefers better characteristics to worse ones.[1] Even in this arguably extreme case, mechanisms that re-create one of the choice environments in her data yield predictable distributions of choices by the agent. This feature is attractive for a principal who faces distributional uncertainty about the agent's preferences, as does the principal in our paper. However, when the principal's data includes more than one choice environment, there are sometimes mechanisms that actually outperform every mechanism in her data.[2]

To see why it is sometimes possible for the principal to find a mechanism that outperforms the choice sets in her data, suppose for example that the principal is a two-good monopolist facing unit demand. She observes in her data two choice environments: first, in which good A is sold at price p, and second in which good B is sold at price q. Every member of the population in her data obeys one of two choice patterns: the first group of consumers purchases good A but not good B, and the second group purchases B but not A. Both goods are profitable for the monopolist to sell at these prices. In this situation, it is strictly more profitable for the monopolist to sell *both* goods than it is for her to sell only one. To see

[1] Formally, the principal knows that the agent's utility function is increasing in the partial order on the set of characteristics. For instance, in the multi-good monopoly application, the agent e.g. prefers more of any particular good to less of that same good.

[2] Even if the principal only observes choices from a *single* choice set, there might be mechanisms that outperform those that recreate the choice environment in the data. This is true even in the monotone environment. However, this is *only* true if the data includes allocations that yield a strictly negative payoff for the principal. Our results hold in the more interesting cases in which there are multiple choice environments in the data and every allocation yields a positive payoff for the principal.

why this is true, note that the first group of consumers revealed that they are unwilling to purchase good B at price q, and similarly the second group revealed that they are unwilling to purchase A at price p. Accordingly, the monopolist need not be concerned about unprofitable substitutions between the two goods, and selling both goods is better for her than selling only one.

As our example demonstrates, our problem is interesting to study even when the set of agent preferences is completely unstructured. However, the monotone environment might be viewed as a situation of extreme uncertainty, and in some cases the principal might desire a mechanism that is designed to accommodate some additional restrictions on the agent's preferences. Accordingly, we allow for arbitrary *rich* sets of agent preferences, and show that various forms of richness are sufficient for the optimality of various classes of mechanisms. After developing this general optimality result, we study in detail the following three informational environments. First, we consider the monotone environment described above. Second, we consider the *homogeneous environment* in which the agent's preferences are known to be homogeneous. Third, we analyze the *diminishing returns environment,* in which the agent's utility is known to suffer diminishing returns to quantity or quality.

For reasons that we describe in detail below, we view the first of these three environments as privileged. In accordance with this view, we devote more attention to the monotone environment than either the homogeneous environment or the diminishing returns environment. Nevertheless, one unifying theme of these three special cases of our problem is that the principal possesses some information about the relative valuations that the agent assigns to vertically differentiated allocations, but has no such information about horizontally differentiated allocations. For example, consider an automobile manufacturer who sells a lineup of vehicles to consumers. The manufacturer believes that consumer preferences are homothetic[3], so that e.g. consumers who value safety more than performance do so for both low- and high-performing vehicles. Suppose that in most years, the manufacturer *updates* vehicle models by increasing the quality of each attribute of the vehicle. It is intuitive that updating her

[3]Recall that preferences are said to be homothetic if they admit a homogeneous degree 1 utility representation: i.e., a function is homothetic if it is the monotonic transformation of a homogeneous degree 1 function. Some authors impose the stronger definition that a homothetic utility function is itself homogeneous degree 1. We clarify that in this paper, we adopt the former definition.

product lineup in this way makes the vehicle unambiguously better from the perspective of the consumer, and provided that products are appropriately priced, the manufacturer should not necessarily expect substitutions between vehicle types, or to the outside option.

We formalize updates of the sort described above as *update mechanisms,* via which the firm screens the agent with allocations that are vertically differentiated from the reference allocations in her data. Although we describe additional features of optimal mechanisms for each of the three environments described above, all of these mechanisms belong to the broader family of update mechanisms.

This paper aspires to make two types of contributions. First, from a practical perspective, our model speaks descriptively and perhaps even prescriptively to the wide-spread exercise of "A-B testing" — along with other types of direct market research — in which firms use experiments to gauge consumer reactions to new or modified products. In particular, we hope that our observation that the principal can sometimes do better than simply adopting the best-performing experiment might be valuable for firms. More generally, we imagine that our analysis might be valuable to principals who possess choice data and want to use this data as the basis for conducting a real-world screening exercise.

Second, this paper hopes to make a theoretical contribution by proposing a new way to think about information in screening. In particular, the informational environment in our paper is considerably differentiated from not only the usual Bayesian approach in which the principal evaluates mechanisms against a single prior, but also from other prior-free work in which the principal evaluates mechanisms against a set of distributions that is *specified by the analyst*. In our model, the set of preference distributions considered possible by the principal is exactly the set of distributions that are compatible with her data.

What might be the value of our approach, then, from a theoretical perspective? The lead specification of our model — in which the principal imposes only minimal technical restrictions on the agent's preferences — might be viewed as a novel extension of the standard revealed preference exercise to a certain class of mechanism design problems. To see why, note that in general, there are many preferences that are consistent with any particular revealed preference. In our paper, this multiplicity is resolved without an appeal to additional exogenous information: i.e., in the monotone environment, the principal uses *only* the revealed preference information contained in the choice data. This lies in contrast to Bayesian approaches, in which the choice data is used to update a subjective prior belief about the agent's preferences. Accordingly, our paper arguably provides a "pure" revealed preference approach to screening, to an extent that existing work does not.[4]

This paper joins a growing literature that studies mechanism design problems in which the principal has worst-case objectives (López-Cunat (2000); Chung and Ely (2007); Bergemann and Schlag (2008, 2011); Chassang (2013); Frankel (2014); Garrett (2014); Carroll (2015, 2017); Madarász and Prat (2017); Auster (2018); Antic (2014); Bergemann et al. (2017); Marku and Díaz (2017); Dai and Toikka (2017); Rosenthal (2018, 2019)). One goal of this literature is to identify simple mechanisms that perform well in the face of one or more dimensions of uncertainty about the environment, and provide foundations for these mechanisms by identifying domains in which they are worst-case optimal. We take up this goal, and do so in an environment in which the principal's beliefs are naturally founded in choice data.

Several papers that belong to a partially overlapping body of work identify simple optimal — or approximately optimal — mechanisms for multidimensional screening. For example, Carroll (2017) shows that if the agent's utility is additively separable and the principal knows the marginal distributions of each component of his preferences but not the joint distribution, it is worst-case optimal to screen the agent separately along each dimension of his private information. In particular, this general separation result yields a "no-bundling" result for multi-product monopolists. Elsewhere, there is a literature studying approximately optimal mechanisms for selling multiple indivisible goods: recently published work develops revenue bounds for simple selling mechanisms (Hart and Nisan (2017)), and related work by other authors does so when the buyer has sub-additive utility (Rubinstein and Weinberg (2015)).

Finally, this paper complements existing work that studies data-driven mechanism design (Braverman and Chassang (2015)), and more broadly various forms of learning by principals

[4]There are a variety of alternative prior-free criteria for evaluating mechanisms. However, unlike ad hoc criteria, the worst-case criterion of this paper can be justified either on the basis of robustness — as appears throughout the literature on prior-free mechanism design cited below — or alternatively via behavioral axiomatizations, as in e.g. Gilboa and Schmeidler (1989). Furthermore, while one might reasonably instead consider regret minimization with either additive or multiplicative criteria, there are no finite-regret mechanisms for the monotone environment. Consequently, regret-based criteria are ill-suited for our problem.

(Segal (2003); Caillaud and Robert (2005); Brooks (2013); Cole and Roughgarden (2014); Morgenstern and Roughgarden (2015); Chawla, Hartline, and Nekipelov (2017)). In particular, Braverman and Chassang (2015) study the usefulness of large data sets in reducing adverse selection in capitation schemes. Aside from differences in setting, one important distinction between the authors' work and our own lies with our consideration of population-level data. This is not merely a technical distinction: while inference is difficult in both environments, the cited work emphasizes finite sample issues which do not arise in our asymptotic framework. Less closely related is applied work that studies the identification of parametric models which allow for both complementarity between goods and correlated tastes (Gentzkow (2007)), and also work that considers stochastic revealed preference (Kitamura and Stoye (2018)).

The paper is organized as follows. In section 1.2, we describe the model in full detail. Section 1.3 provides examples. In section 1.4, we develop general tools to analyze our problem. In section 1.5, we study the monotone environment in detail. Section 1.6 treats the homogeneous and diminishing returns environments. Finally, we conclude in section 1.7. Proofs with content of interest are mostly given in the body of the paper, while purely technical material is presented in appendix A.

1.2 Model

Allocations $(x, p) \in X \times \mathbb{R}_+$ consist of *characteristics* x and non-negative *prices* p. The set of characteristics $(X, \leq)$ is a complete lattice.[5] We label as 0 the minimal element of X, and interpret the allocation $(0, 0)$ loosely as the agent's outside option. Finally, we call comparable characteristics *vertically differentiated*, incomparable characteristics *horizontally differentiated,* and say that allocation (x, p) *dominates* (y, q) if $x \geq y$ and $p \leq q$.

[5] Recall that a complete lattice is a partially ordered set $(X, \leq)$ with the property that every subset of X has both a supremum and an infimum. Elements $x, y \in X$ are said to be comparable if $x \leq y$ or $y \leq x$. If neither relation holds, x and y are said to be incomparable. Our model is most interesting to study when X includes incomparable elements. We give X the order-interval topology and the Borel sigma algebra.

1.2.1 The agent's preferences

The agent's payoff from allocation (x, p) is as follows, for $u : X \to \mathbb{R}_+$:

$$u(x) - p.$$

Assumption 1. *(Monotonicity) The agent's utility function u is weakly increasing, upper semicontinuous, and satisfies $u(0) = 0$.*

We call u the agent's *type*. The agent's type belongs to some set of *feasible* types U[6], the identity of which is known to the principal.

1.2.2 Mechanisms

We restrict attention to direct mechanisms, as justified by the revelation principle. Accordingly, *mechanisms* $M = (y, q)$ consist of a measurable assignment $y : U \to X$ of characteristics and a measurable assignment $q : U \to \mathbb{R}_+$ of transfers to agent types u. The agent's outside option yields utility zero, and mechanisms are *incentive compatible* and *individually rational:*

$$\begin{aligned} &u\Big(y(u)\Big) - q(u) \geq u\Big(y(\widehat{u})\Big) - q(\widehat{u}) \quad \forall u, \widehat{u} \in U \\ &u\Big(y(u)\Big) - q(u) \geq 0 \quad \forall u \in U. \end{aligned}$$

In places, given a mechanism $M = (y, q)$, we use the notation $(x, p) \in M$ to indicate that $(x, p) = \Big(y(u), q(u)\Big)$ for some feasible type u.

1.2.3 Information

The principal interacts with only a single agent. However, she observes choice data for a continuum *population* of anonymous decision makers with idiosyncratic preferences. Formally, let $\mathcal{B}_1, ..., \mathcal{B}_K \subset X \times \mathbb{R}_+$ be exogenously-specified sets of allocations, with $1 \leq K < \infty$. Each set $\mathcal{B}_k$ is itself finite, includes the outside option $(0, 0)$, and at least one distinct allocation

[6] We give U the sup-norm topology and the Borel sigma algebra.

(x, p). Define as follows:

$$\mathcal{B} := \mathcal{B}_1 \times ... \times \mathcal{B}_K$$
$$X_0 := \Big\{ x \in X | (x, p) \in \mathcal{B}_1 \cup ... \cup \mathcal{B}_K \Big\}.$$

The principal observes the distribution $\mu_0 \in \Delta(\mathcal{B})$ of choices by the population of decision makers.[7] The decision makers' choices are informative about their preferences, because each decision maker maximizes his utility. Just as the agent's preferences belong to U, so too does each decision maker's preference. We use the notation $b_k(u)$ for the choice by u from choice set $\mathcal{B}_k$, and $b(u)$ for the corresponding list of choices $\Big(b_1(u), ..., b_K(u)\Big)$.[8] For each $b \in \mathcal{B}$, define as follows:

$$U(b) := \Big\{ u \in U | b(u) = b \Big\}.$$

The set $U(b)$ is the set of types that choose b from $\mathcal{B}$, and collectively these sets partition U. We will periodically refer to these sets $U(b)$ as *cells* of agent types. Define as follows:

$$\Delta_0 := \Big\{ \mu \in \Delta(U) | \mu\Big(U(b)\Big) = \mu_0(b) \; \forall b \in \mathcal{B} \Big\}.$$

The agent is representative of the population of decision makers, in the sense that the distribution of his preferences is identical to the population distribution of preferences. Consequently, because Δ_0 is the set of distributions that rationalize the choice data μ_0, it is also the set of potential distributions of the agent's type u.

Assumption 2. *(Consistency) The set Δ_0 is non-empty.*

Consistency should be viewed as a joint restriction to U and μ_0. In particular, because we require monotonicity, consistency rules out distributions μ_0 that assign positive probability

[7] We write $\Delta(\mathcal{B})$ for the set of (discrete) probability distributions over $\mathcal{B}$, and $\Delta(U)$ for the set of Borel probability distributions over U.

[8] We assume that ties are broken uniformly in the principal's data $\mathcal{B}$: i.e., every feasible type u breaks ties according to some hierarchy on the set of reference characteristics X_0 that is independent of prices. For example, if $u(x) - p = u(y) - q$ and $u'(x) - p' = u'(y) - q'$, then ties are resolved in favor of (x, p) by u if and only if ties are resolved in favor of (x, p') by u'.

to dominated alternatives. One natural objection is thus that we are imposing restrictions on data μ_0. However, we interpret μ_0 as a population-level statistic, and if the principal's interpretation of the data requires the assignment of positive probability at the population level to dominated alternatives, then it seems plausible that her model might be misspecified.

1.2.4 The principal's preferences

The principal's costs for characteristic x are given by $c : X \to \mathbb{R}_+$, where c is weakly increasing. The principal's payoff from allocation (x, p) is as follows:

$$\pi(x, p) = p - c(x).$$

Given mechanism M and agent type distribution $\mu \in \Delta(U)$, the principal's payoff is as follows:

$$\pi(M|\mu) := \int_U \pi\Big(y(u), q(u)\Big)\mathrm{d}\mu.$$

The set Δ_0 generically contains multiple distributions, and the principal does not assign primacy to any one of these distributions. Instead, she evaluates mechanisms M by their worst-case performance against the entire set of distributions that are compatible with the choice data μ_0 :

$$\pi(M) := \inf_{\mu \in \Delta_0} \pi(M|\mu).$$

We study the problem of identifying mechanisms that maximize the principal's *guarantee* $\pi(M)$.

1.3 Examples and counterexamples

This section provides three examples, and its purpose is two-fold. First, we illustrate the informational environment studied in our paper in simple settings. Second, we consider and reject a series of conjectures about properties that the optimal mechanism might have. Suppose that the set of feasible types U is as follows:

$$U = \Big\{u : X \to \mathbb{R}_+ | u \text{ is increasing, upper semicontinuous, and } u(0) = 0\Big\}.$$

Given that the principal only knows that the agent's preferences are consistent with the order $\geq$, it seems natural to expect that the principal might not be able to do better than recreating one of the choice environments in her data set. In general, however, this turns out not to be the case.

Example 1. *The principal observes each member of the population's preferred alternative from each of the four following choice sets:*

1. purchase good A for a price of 1, or pay nothing and receive nothing;
2. purchase good A for a price of 2, or pay nothing and receive nothing;
3. purchase good B for a price of 1, or pay nothing and receive nothing;
4. purchase good B for a price of 2, or pay nothing and receive nothing.

The principal observes that every member of the population chooses one of two distinct choice patterns. One half of the population purchases good A at both prices, but does not purchase good B at either price. The other half of the population purchases good B only at the lower price, and does not purchase good A at either price. Production of both goods is free, and thus the principal's objective reduces to revenue maximization.

In example 1, one half of the population is willing to pay at least 2 for good A, but unwilling to pay 1 for good B. The other half of the population is willing to pay between 1 and 2 for good B, but unwilling to pay 1 for good A. Consider the mechanism M that provides the agent with the following choices:

- purchase good A for a price of 2;
- purchase good B for a price of 1;
- pay nothing and receive nothing.

The mechanism M sells the first good for a price of 2 and the second good for a price of 1. See figure 1.1 for illustration.

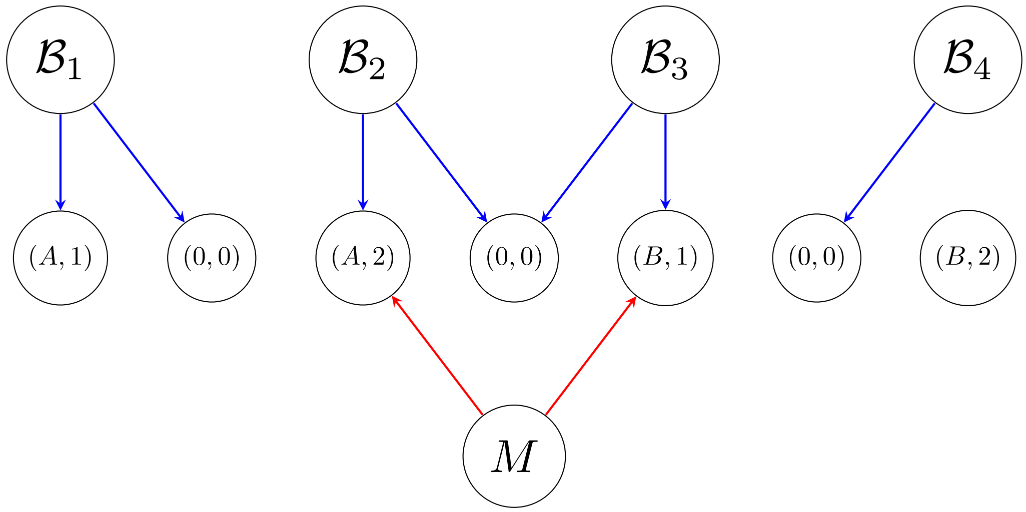


Figure 1.1: The four choice sets in example 1, labeled $\mathcal{B}_1, ..., \mathcal{B}_4$, and mechanism M.

Remark 1. In example 1, the principal's worst-case payoff from selling the first good for a price of 2 and the second good for a price of 1 exceeds her payoff from any mechanism that corresponds directly to one of the four choice sets in her data.

Proof. Given the choice data, it is straightforward to see that M guarantees the principal a payoff at least $3/2$. The largest payoff from a mechanism that corresponds to one of the data sets is given by selling the first good for a price of 2. Because the agent is willing to purchase the first good at a price of 2 with probability $1/2$, this mechanism guarantees the principal a payoff of 1. □

Although we have deliberately chosen the example to be as simple as possible, we nevertheless see that it is not in general optimal for the principal to recreate one of the choice environments in her data. Here, we see that when the principal observes the population make multiple choices, she sometimes gains information that helps her to predict how the agent will respond to choices that are more complicated than those in her data.

We have just seen that the conjecture that optimal mechanisms recreate one of the choice environments in the principal's data is false. But what about the weaker conjecture that they should at least use the same prices as in the data? This too turns out to be false. To see why, consider the following counterexample:

Example 2. *The principal observes each member of the population's preferred alternative from each of the two following choice sets:*

$$\mathcal{B}_1 = \Big\{(A, p), (B, 1), (0, 0)\Big\}$$
$$\mathcal{B}_2 = \Big\{(B, 2), (0, 0)\Big\}.$$

The principal observes that every member of the population purchases good A at price p and good B at price 2. Production of both goods is free, and thus the principal's objective reduces to revenue maximization.

Remark 2. In example 2, if the price p in the principal's data is strictly larger than 1, then it is optimal to sell only good A, for price $p+1$.

Proof. With certainty, the agent's utility function u satisfies the following inequalities:

$$u(A) - p \geq u(B) - 1$$
$$u(B) - 2 \geq 0.$$

Combining these two inequalities yields the following:

$$u(A) \geq p + 1.$$

Accordingly, while it is certainly individually rational to assign the agent the allocation $(A, p+1)$, it is not necessarily individually rational to assign the agent A at any price larger than $p+1$. Similarly, it is certainly individually rational to assign the agent the allocation $(B, 2)$, but not so for any price larger than 2. Because the goods are costless to produce, it is thus better for the monopolist to sell A if she sells only one of the two goods. Moreover, because *every* consumer in the population is willing to purchase good A, the monopolist does not benefit from screening the agent with multiple allocations. □

Finally, we consider the conjecture that optimal mechanisms at least uses the same *number* (or fewer) of allocations as the choice sets in her data. Just as above, this does not hold:

Example 3. *The principal observes each member of the population's preferred alternative from each of the three following choice sets:*

1. purchase good A for a price of 1, good B for a price of 0, or pay nothing and receive nothing;

2. purchase good C for a price of 2, or pay nothing and receive nothing;

3. purchase good B for a price of 1, good D for a price of 0, or pay nothing and receive nothing.

The principal observes that every member of the population chooses one of two distinct choice patterns. One half of the population purchases good A for a price of 1*, good C for a price of* 2*, and good D for a price of* 0*. The other half of the population purchases good B at a price of* 0*, the outside option, and good B for a price of* 1*. Production of both goods is free, and thus the principal's objective reduces to revenue maximization.*

Remark 3. The optimal mechanism sells good A for a price of 2, good C for a price of 2, and good B for a price of 1.

Why more allocations than choice patterns? The presence of $(A, 2)$ serves to ensure that the first group does not choose $(B, 1)$, while the presence of $(C, 2)$ serves to ensure that the first group does not choose the outside option. Neither objective can be accomplished with only one good offered at price 2.

1.4 Analysis

Our formal analysis is organized as follows. First, we propose a class of richness criteria. Second, we show how these criteria are sufficient for the optimality of corresponding classes of mechanisms. Third, we show how various natural restrictions on the agent's preferences lead to certain forms of richness, which in turn yields the optimality of certain types of mechanisms. Although our optimality result is more general, we emphasize the study of environments in which the principal is uncertain about how the agent values horizontally differentiated characteristics, but might have some information about how he values vertically differentiated characteristics.

1.4.1 Richness

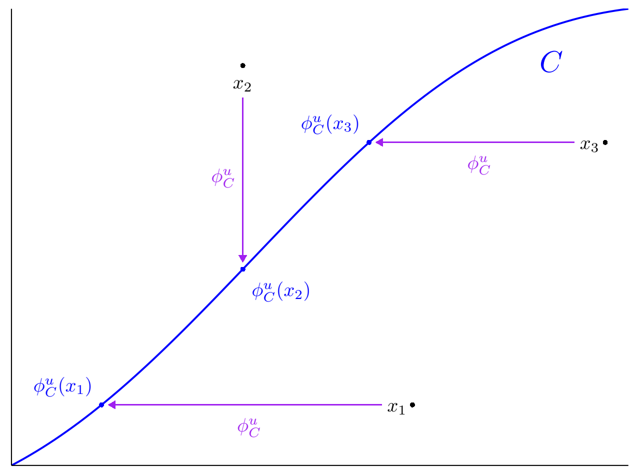


Figure 1.2: The reduction ϕ_C^u onto C with respect to arbitrary utility function u.

Definition 1. Given a closed set of characteristics C that includes every reference allocation and a feasible type u, define the *reduction* $\phi_C^u : X \to C$ as follows*:*

$$\phi_C^u(x) := \arg\max \Big\{ u(y) | y \leq x, y \in C \Big\}.$$

The reduction ϕ_C^u maps characteristics x into the best characteristic $\phi_C^u(x)$ in C that is smaller than x itself, as weighted by the utility function u, with ties broken arbitrarily. See figure 1.2 for illustration. We use the reduction maps to construct families of agent types, which play a central role in our analysis.

Definition 2. Given a closed set of characteristics C that includes every reference allocation and a *parent type* u, define the *dual type* $u_C : X \to \mathbb{R}_+$ as follows:

$$u_C(x) := u\Big(\phi_C^u(x)\Big).$$

The dual type evaluates each characteristic x as if it were the reduction of x onto C. Importantly, the parent type and the dual type assign the same utility to each characteristic in C. These preferences might be viewed as a generalization of Leontief utility:

Example. *Suppose that $X = \mathbb{R}^2_+$. Consider the following utility functions $u, u_C : X \to \mathbb{R}^+$:*

$$u\Big((x,y)\Big) = \frac{1}{2}(x+y)$$
$$u_C\Big((x,y)\Big) = \min\{x,y\}.$$

The Leontief preference u_C is the dual type to the linear preference u, with respect to the set of characteristics $C = \Big\{(x,y) \in X \times X | x = y\Big\}$.

We introduce the following richness criterion, which classifies the set of feasible types U according to whether or not it is sufficiently large to include all of the dual types against a given set of characteristics C.

Definition 3. Given a closed set of characteristics C that includes every reference allocation, the set of feasible types U is $C-$rich if the dual type u_C is feasible for each feasible type u.

Our presentation of richness as a criterion rather than an assumption is deliberate. Rather than assuming richness directly, we show later how various forms of richness are embedded in more natural restrictions on the set of types that the principal considers feasible. Nevertheless, $C-$richness leads to the optimality of the following class of mechanisms:

Definition 4. Mechanism $M = (y, q)$ is a $C-$*mechanism* if $y(u) \in C$ for all feasible types u.

The presence of the dual types in the type space is sufficient for the optimality of C mechanisms, for three reasons. First, the dual type u_C evaluates each characteristic x as if it were instead its reduction $\phi^u_C(x)$ onto C. Because the principal's costs are increasing, it is wasteful to allocate to the dual type any characteristic outside of C. Thus, it is without loss of generality to restrict our search for optimal mechanisms to those that allocate to every dual type some characteristic in C.

Second, the parent type u and the dual type u_C derive the same utility for every characteristic in C. Consequently, if M is a $C-$mechanism, then the allocation $\Big(y(u_C), q(u_C)\Big)$ assigned to the dual type u_C is individually rational and incentive compatible for the parent type u.

Third, that principal can not distinguish between the the parent type u and the dual type u_C. To see why, note simply that if $C \supset X_0$, then $u(x) = u_C(x)$ for each characteristic

$x \in X_0$. Thus, the parent type and the dual type behave identically in the principal's data. We proceed now to the formalization of these arguments.

1.4.2 Optimal mechanisms

Suppose that U is C-rich for some closed set of characteristics C that includes every reference allocation. Given an arbitrary mechanism M, we modify M into a C-mechanism M_C that guarantees the principal a larger payoff. This establishes the optimality of C-mechanisms, up to the existence of an optimal mechanism.

Our construction might be viewed as a two-step exercise. First, given an arbitrary feasible type, we modify the characteristics allocated to its dual type by reducing these characteristics onto the set C. By design, this reduction leaves the dual type agent indifferent, every other type of agent worse off, and the principal better off. Consequently, in addition to providing the principal with a higher payoff, these modified allocations are both individually rational and incentive compatible for the dual type agent.

In the second step, we show that it is individually rational and incentive compatible to assign these modified allocations to the parent type agent. Because the parent type and dual type belong to the same cell of agent types, this is sufficient to establish that the modified mechanism guarantees the principal a better payoff than the original mechanism. Formally, define the *truncation* $M_C = (y_C, q_C)$ of M as follows:

$$y_C(u) := \phi_C^u\Big(y(u_C)\Big)$$
$$q_C(u) := q(u_C).$$

See figure 1.3 for an illustration of this construction.

Lemma 1. *If the set of feasible types U is $C-$rich for some closed set of characteristics C that includes every reference allocation, then the truncation M_C of M is incentive compatible and individually rational.*

Proof. The proof has two parts. First, we show that M_C is incentive compatible and

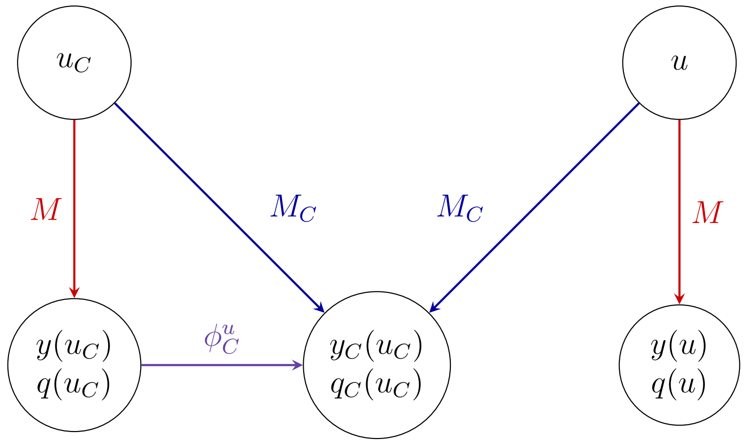


Figure 1.3: Construction of the truncated mechanism M_C.

individually rational on the restriction of U to U_C. That is, we show that the following:

$$u\Big(y(u)\Big) - q(u) \geq u\Big(y(\widehat{u})\Big) - q(\widehat{u}) \quad \forall u, \widehat{u} \in U_C$$
$$u\Big(y(u)\Big) - q(u) \geq 0 \quad \forall u \in U_C.$$

Second, we extend this argument to all of U. Proceeding, let $u \in U$. By construction, we have:

$$u_C\Big(y_C(u_C)\Big) = u_C\Big(y(u_C)\Big). \tag{1.1}$$

Moreover, because $(u_C)_C = u_C$, we have:

$$q_C(u_C) = q(u_C). \tag{1.2}$$

In words, equalities 1.1 and 1.2 establish respectively that the agent with type u_C is indifferent between the characteristics assigned him under the mechanisms M and M_C, and also that his transfer to the principal is the same. It follows immediately that M_C is individually rational for u_C. Furthermore, by construction we have $y_C(\widehat{u}) \leq y\Big(\widehat{u}\Big)$ for each $\widehat{u} \in U_C$. In turn, monotonicity yields:

$$u_C\Big(y_C(\widehat{u})\Big) \leq u_C\Big(y(\widehat{u})\Big). \tag{1.3}$$

Combining 1.1 and 1.3 yields:

$$u_C\Big(y(u_C)\Big) - u_C\Big(y(\widehat{u})\Big) \leq u_C\Big(y_C(u)\Big) - u\Big(y(\widehat{u})\Big). \tag{1.4}$$

Inequalities 1.2 and 1.4 establish that M_C is incentive compatible for u_C. Because we chose u arbitrarily, this establishes that M_C is individually rational and incentive compatible on U_C. To see why this extends to all of U, recall that M_C assigns the same allocation to u as to u_C:

$$y_C(u) = y_C(u_C) \tag{1.5}$$

$$q_C(u) = q(u_C). \tag{1.6}$$

Moreover, because $y_C(u_C) \in C$, 1.5 and our definition of u_C jointly yield the following:

$$u\Big(y_C(u)\Big) = u_C\Big(y_C(u_C)\Big). \tag{1.7}$$

Equalities 1.6 and 1.7 establish that M_C is individually rational and incentive compatible for u. Because we chose u arbitrarily, this establishes that M_C is individually rational and incentive compatible on U. This completes the proof. □

As we have alluded to above, lemma 1 itself comes close to establishing that the truncated mechanism M_C provides a larger guarantee than the original mechanism M. To see why, note that the truncated allocation $y_C(u_C)$ is less in every component than $y(u_C)$. Because costs are increasing and transfers to the principal are identical, the former allocation yields the principal a higher payoff. Moreover, the allocation intended for u_C is also individually rational and incentive compatible for u. Accordingly, in light of the fact that u and u_C belong to the same cell $U(b)$ of agent types, the principal might as well allocate the two agent types the same allocation.

Theorem 1. *Suppose that the set of feasible types U is $C-$rich for some closed set of characteristics C that includes every reference allocation. Given any mechanism M, the truncation M_C of M yields a better guarantee than M:*

$$\pi(M) \leq \pi(M_C).$$

Proof. It will be technically convenient to restrict our attention to discrete type distributions, as justified by lemma 5, presented in the appendix. Our proof approach is to take an arbitrary discrete type distribution $\mu_C \in \Delta_0$, and construct a (potentially distinct) discrete type distribution $\mu \in \Delta_0$ such that:

$$\pi(M|\mu) \leq \pi(M_C|\mu_C).$$

Because we do so for an arbitrary distribution $\mu_C \in \Delta_0$, in doing so we establish the following inequality, as desired:

$$\pi(M) \leq \pi(M_C)$$

Accordingly, let $\mu_C \in \Delta_0$ have finite support. For each u in the support of μ_C, let μ be the distribution that satisfies the following equality:

$$\mu(u_C) = \mu_C(u).$$

In words, μ assigns the same probability to the dual type u_C as μ_C does to u. In light of our hypothesis that $C \supset X_0$, we have that $u(x) = u_C(x)$ for all $x \in X_0$. In turn, our assumption of uniform tie-breaking implies that there exists $b \in \mathcal{B}$ such that $u, u_C \in U(b)$. Accordingly, we have $\mu \in \Delta_0$. Moreover, for each $u \in U$ we have the following inequality, by construction:

$$\pi\Big(y(u_C), q(u_C)\Big) \leq \pi\Big(y_C(u), q_C(u)\Big).$$

Consequently, we have the following:

$$\pi(M|\mu) \leq \pi(M_C|\mu_C).$$

Because μ_C was chosen arbitrarily, this is sufficient to establish $\pi(M) \leq \pi(M_C)$, as desired. □

Theorem 1 establishes the optimality of $C-$mechanisms, up to the existence of an optimal mechanism. For the sake of completeness, we include an existence proof in the appendix, under the additional hypotheses that the set of characteristics X is finite and the set of feasible types U is compact.

1.5 The monotone environment

In this section, we formalize the intuitive result that if the principal has no information about the agent's preferences outside of monotonicity, then it is optimal to only offer allocations that coincide with the reference allocations in the principal's data set. First, define as follows:

$$U_0 := \Big\{u : X \to \mathbb{R}_+ | u(0) = 0, u \text{ is upper semicontinuous and increasing}\Big\}.$$

We refer to the situation in which the set of feasible types U is equal to U_0 as the *monotone environment.* Knowledge that the agent's type belongs to U_0 reflects only the knowledge that the agent prefers larger characteristics to smaller ones: e.g., that better products are weakly preferred to worse products.

1.5.1 Characteristics

Given the paucity of information suffered by the principal in the monotone environment, the following result follows naturally:

Proposition 1. *Consider the monotone environment. Given any mechanism M, there exists an X_0-mechanism M_0 that yields a better guarantee than M:*

$$\pi(M) \leq \pi(M_0).$$

The proof of proposition 1 is structured as follows. First, we argue that the monotone environment is X_0-rich. Given the permissive nature of the set U_0, this is straightforward. Having established the appropriate form of richness, we apply theorem 1 to obtain the result.

1.5.2 Incentive compatibility

In light of the lack of restrictions that the monotone environment imposes on the agent's preferences, proposition 1 should not be viewed as a theoretically surprising result. However, the proposition is useful for obtaining stronger optimality results, the development of which is the subject of the remainder of this section. Proceeding, it is useful to first introduce some additional terminology:

Definition. Given choice list $b \in \mathcal{B}$, if either (x, p) dominates (y, q) or if there exists constant Δ and index k such that:

$$(x, p + \Delta) = b_k$$
$$(y, q + \Delta) \in \mathcal{B}_k,$$

then we say that (x, p) is *revealed directly preferred to* (y, q) for agent types in $U(b)$.

The allocation (x, p) is revealed directly preferred to (y, q) if the choice data (or monotonicity of the agent's preferences) directly implies that (x, p) is preferred to (y, q). Because the agent's preferences are transitive, it is furthermore possible to string together sequences of direct revealed preferences to make more subtle inferences:

Definition. Given choice list $b \in \mathcal{B}$, if there exists a finite sequence $(x_1, p_1), ..., (x_J, p_J)$ such that:

1. $(x_1, p_1) = (x, p)$;
2. $(x_J, p_J) = (y, q)$;
3. (x_j, p_j) is directly revealed preferred to (x_{j+1}, p_{j+1}) for each $j < J$,

then we say that (x, p) is *revealed preferred* to (y, q) for agent types in $U(b)$ and write $(x, p) \succeq_b (y, q)$.

We say that (x, p) is revealed preferred to (y, q) if it is possible to construct a sequence of allocations directly revealed preferred to one another that begins with (x, p) and ends with (y, q).

Given an X_0-mechanism M, what can we say about which allocations included in M are individually rational and incentive compatible for particular types of agents? Define as follows:

$$C_b(M) := \Big\{(x, p) \in M | (y, q) \in M, (y, q) \succeq_b (x, p) \implies (y, q) = (x, p)\Big\}.$$

The set $C_b(M)$ is the set of allocations $(x, p) \in M$ with the property that no distinct allocation $(y, q) \in M$ is revealed preferred by the agent types in $U(b)$ to (x, p). By construction, we have the following:

Fact. *Suppose that M is an X_0-mechanism. If allocation $(x, p) \in C_b(M)$, then there exists an agent type $u \in U(b)$ to whom the allocation (x, p) is assigned by mechanism M.*

Thus, if the principal's choice data does not *rule out* the allocation of (x, p) to some agent type in the group $U(b)$, then the mechanism M necessarily allocates (x, p) to some type $u \in U(b)$. Furthermore, notwithstanding a caveat pertaining to tie-breaking that is presented in the technical appendix as part of our proof of lemma 2, the converse essentially obtains: *every* allocation that is assigned by mechanism M to some agent type in $u \in U(b)$ belongs to the choice set set $C_b(M)$. Define as follows:

$$\Pi_b(M) := \min_{(x,p)\in C_b(M)} \pi(x, p).$$

The quantity $\Pi_b(M)$ is the worst-case payoff for the principal against the set $C_b(M)$. Thus, we obtain an expression of the principal's payoff:

Lemma 2. *Consider the monotone environment, and suppose that X_0-mechanism M breaks ties in favor of the principal. Then M guarantees the principal the following payoff:*

$$\Pi(M) := \int_{\mathcal{B}} \Pi_b(M) \mathrm{d}\mu_0(b). \tag{1.8}$$

Thus-far, we have shown both that X_0-mechanisms are optimal, and also how to compute the payoff to a given X_0-mechanism. The next step is to study optimal pricing.

1.5.3 Prices

Optimal mechanisms use prices that are tightly related to the bounds on utility differences implied by the principal's data. It will be useful to introduce a notation for these bounds. For each choice list $b \in \mathcal{B}$ and reference characteristic $x \in X_0$, define the set $\mathcal{P}_b(x)$ as follows:

$$\mathcal{P}_b(x) := \{y \in X_0 | \text{there exist prices } p, q \text{ such that } (x, p) \succeq_b (y, q)\}.$$

Given reference characteristic $x \in X_0$, the set $\mathcal{P}_b(x)$ is the set of reference characteristics y such that (x, p) is certainly preferred to (y, q) by agents in group $U(b)$, for appropriate prices p and q. Put differently, these are the characteristics y for which the following difference is bounded above:

$$u(x) - u(y).$$

It will be useful to introduce a notation for this upper bound. For each $x \in X_0$ and $y \in \mathcal{P}_b(x)$, define $V_b(x, y)$ as follows:

$$V_b(x, y) := \max\{p - q | (x, p) \succeq_b (y, q)\Big\}.$$

The significance of this quantity is as follows:

$$(x, p) \succeq_b (y, q) \iff V_b(x, y) \geq p - q.$$

Thus, if a mechanism M includes both allocations and the principal would like to allocate (x, p) to agents who choose choice list $b \in \mathcal{B}$ in her data, it must be the case that prices p, q satisfy the above inequality. Moreover, as we demonstrate, optimal mechanisms use prices for which this bound is tight.

Definition 5. Given an X_0-mechanism M, we call M a *reference mechanism* if for each $(x, p) \in M$ there exists a sequence of allocations $(x_1, p_1), ..., (x_J, p_J) \in M$ and a sequence of choice lists $b_1, ..., b_J$ such that:

1. $(x_1, p_1) = (x, p)$;

2. $(x_J, p_J) = (0, 0)$;

3. $(x_1, p_1) \succeq_{b_1} ... \succeq_{b_J} (x_J, p_J)$;

4. $p = \sum_{j=1}^{J-1} V_{b_j}(x_j, x_{j+1})$.

Reference mechanisms use prices that are derived from "summing up" the inequalities implied by the principal's data.

Lemma 3. *Given any mechanism M_0, there exists a reference mechanism M that yields a better guarantee than M_0:*

$$\pi(M_0) \leq \pi(M).$$

1.5.4 Optimal mechanisms

We offer two remarks. First, our proof of lemma 3, which is presented in the appendix, provides guidance on how to construct optimal prices. Second, because there are only finitely many reference mechanisms, and we have already shown how to compute the payoff to a given reference mechanism, we have thus established both that there exists a reference mechanism that is optimal for the monotone environment, and provided a procedure for identifying it.

Theorem 2. *There exists a reference mechanism that yields the highest-possible guarantee. Equivalently, this mechanism is a maximizer for the function* Π.

1.6 Extrapolation

This section studies the form of optimal mechanisms for environments other than the monotone environment considered in section 1.5. In particular, we consider two cases: first, in which the principal knows that utility is homogeneous degree t for some range of values T, and second in which the principal knows that utility exhibits diminishing returns. These environments are unified by the feature that the principal has some information about how the agent's preferences scale across vertically differentiated allocations, but has no information about how the agent values horizontally differentiated allocations. Because we restrict attention to the case in which the set of characteristics is a vector space — e.g., a characteristic x is a vector of quality or quantity attributes — we interpret this exercise as one in separability-robust multidimensional screening. It will be convenient to introduce the following notation:

$$X^* := \Big\{x \in X | x \text{ is vertically differentiated from some nonzero characteristic } \widehat{x} \in X_0\Big\}.$$

In words, X^* is the set of characteristics that are unambiguously better (or worse) than some reference characteristic $\widehat{x}$.

Definition. Call mechanism $M = (y, q)$ an *update mechanism* if M is an X^*-mechanism.

All of the optimal mechanisms of this section will be update mechanisms.

1.6.1 The homogeneous environment

Thus-far, we have focused our attention on the case in which the set of feasible types U includes every increasing utility function. As we have just seen, this informational environment naturally yields optimal mechanisms that only allocate to the agent characteristics in the principal's data set. However, this specification of the type space might be viewed as an extreme case, and our model yields richer results when the principal has more information.

For a first example, consider the additional hypothesis that the agent's utility is homogeneous.[9] Given set $T \subset \mathbb{R}_+$, define as follows:

$$U^{\updownarrow} := \Big\{ u \in U_0 | u \text{ is homogeneous degree } t \text{ for some } t \in T \Big\}.$$

We refer to the situation in which the set of feasible types U is equal to $U^{\updownarrow}$ as the *homogeneous environment.*

In the homogeneous environment, the principal's data is informative about not only the agent's valuations for characteristics in X_0, but also for many other characteristics. Consequently, in general it is *not* the case that optimal mechanisms only allocate to the agent reference allocations. Define as follows:

$$X^{\updownarrow} := \Big\{ x \in X | x = \lambda \widehat{x} \text{ for some reference allocation } \widehat{x} \in X_0 \text{ and some scalar } \lambda \Big\}.$$

Proposition 2. *Suppose that the set of characteristics X is convex, and consider the homogeneous environment. Given any mechanism M, there exists an $X^{\updownarrow}-$mechanism $M_{\updownarrow}$ that yields a better guarantee:*

$$\pi(M) \leq \pi(M_{\updownarrow}).$$

[9] Throughout this section, we assume that X is a subset of an arbitrary vector space V. We say that $u : X \to \mathbb{R}_+$ is *homogeneous degree* t if u is the restriction to X of some function $v : V \to \mathbb{R}_+$ with the property that $v(\lambda x) = \lambda^t v(x)$ for each $x \in V$ and each $\lambda \geq 0$. We use t for degree rather than the typical k to avoid confusion with the index for our choice sets $\mathcal{B}_k$.

In section A.4 of the appendix, we use proposition 2 to explicitly solve a special case of the homogeneous environment, with some additional restrictions on the principal's data and her cost function c. More broadly, the proposition shows that if the set of characteristics X is convex and the agent is known to have homogeneous utility of some degree t, then there are optimal mechanisms that allocate to the agent characteristics that are proportional to the reference allocations. The interpretation of this result is straightforward: while homogeneity provides the principal with information about how the agent's utility might scale vertically, it says nothing about how the agent values horizontally differentiated characteristics. Because monotonicity also imposes no requirements on how the agent values horizontally differentiated characteristics, the principal does best to avoid assigning such allocations to the agent.

1.6.2 The diminishing returns environment

One natural environment in which a particular class of update mechanisms arise as optimal is when the agent's utility exhibits diminishing returns.[10] Given constant u_0, define the *diminishing returns* environment as follows:

$$U^{\downarrow} := \Big\{ u \in U_0 | u \text{ suffers diminishing returns} \Big\}.$$

Much of our discussion in this section mirrors our discussion of the homogeneous environment. As before, knowledge that the agent is willing to pay price p for allocation x contains non-trivial information about the agent's willingness to pay for every allocation that is smaller than x. Unlike in the homogeneous environment, however, the principal's data is not informative about how the agent values characteristics that are larger than the reference allocations. Define as follows:

$$X^{\downarrow} := \Big\{ x \in X | x = \lambda \widehat{x} \text{ for some reference allocation } \widehat{x} \in X_0 \text{ and some scalar } \lambda \leq 1 \Big\}.$$

Proposition 3. *Suppose that the set of characteristics X is convex, and consider the*

[10]Throughout this subsection of the paper, we again assume that X is a subset of a vector space V. We say that u exhibits *diminishing returns* if $u : X \to \mathbb{R}_+$ is the restriction to X of some function $v : V \to \mathbb{R}_+$ with the property that $v(\lambda x)/\lambda$ is decreasing in λ for each $x \in V$.

diminishing returns environment. Given any mechanism M, there exists an $X^{\downarrow}-$mechanism $M_{\downarrow}$ that yields a better guarantee:

$$\pi(M) \leq \pi(M_{\downarrow}).$$

1.7 Discussion

Stepping back from the particulars of our model, what is the economic content of our finding that update mechanisms are optimal? Simply put, update mechanisms are optimal because they only assign allocations to the agent that are unambiguously better or worse than the reference allocations in the data set. Because the agent's preferences are monotone — which corresponds to free disposal under quantity interpretations — the principal has information about how the agent might value two goods when the first is superior to the second in every component. Consequently, there are circumstances in which it is payoff-improving for her to offer allocations that differ from those in her data set, as is generically the case when she uses an update mechanism.

However, because the agent's utility in our model is not separable, it is not possible for the principal to make useful extrapolations about his willingness-to-pay for alternatives that are horizontally differentiated from those in her data set, even when she observes population-level choices from many data sets. Consequently, the principal does not have any information about how the agent might value two goods when the first is superior to the second in some components but inferior in others, unless she directly observes the agent choose between the two goods in her data set (or between some dominating or dominated alternatives). This unpredictably in the agent's preferences can lead to both failures of individual rationality and also unpredictable substitution patterns.

How should our results be interpreted? One way to evaluate our contribution from a normative perspective is to consider the complexity of our optimal mechanisms. Arguably, update mechanisms are simple for two reasons. First, update mechanisms assign the same allocation to every agent type u with a common dual type u^*. Interpreting the mechanism design exercise as one in which the agent literally announces his type to the principal, update mechanisms thus use a (potentially much) smaller message space than some other types of mechanisms.

A second way in which our optimal update mechanisms might be viewed as simple is at the level of the individual allocation. Our result that update mechanisms are optimal becomes increasingly sharper as the complexity of the set of characteristics X increases. For example, if X is the Cartesian product of N totally ordered sets $(X_n, \leq_n)$, then the ratio of the size of the set of vertically differentiated characteristics X^* to the size of entire design space X declines exponentially as the number of reference allocations $|X_0|$ and N increase proportionally. Of course, this observation is more meaningful for some applications — e.g., for a monopolist who sells complex multi-attribute products — than for others.

As a positive description of behavior, the predictions of our model seems to be at least anecdotally consistent with observation. For example, automobile manufacturers seem to in many (but not all) years offer updated model-year versions of their products that more closely resemble an improved version of last year's model than an entirely new vehicle. Similarly, technology companies that sell a variety of e.g. computers or cell phones might regularly offer improved versions of existing products, and less frequently drop existing products or introduce inexpensively-priced products that are horizontally differentiated from flagship offerings. At the very least, our results highlight virtues of these strategies.

Finally, this paper hopes to be useful by providing an example of how the tools of robust mechanism design can be integrated with choice data to provide solutions to economic design problems. In particular, this approach seems to be promising for problems in which data contains useful but somehow limited information about preferences or other features of the environment.

CHAPTER 2

ROBUST INCENTIVES FOR RISK

2.1 Introduction

Consider the problem faced by a principal in designing incentives for a risk-taking agent with unknown risk preferences. The principal would like for the agent to exert effort[1], which is unobserved and costly for the agent, and so she uses a contract that conditions transfers to the agent on the output that he produces. Because output is stochastic, such contracts expose the agent to risk. In the face of risk, different types of agents take different types of risks: risk-neutral agents prefer actions that yield large average transfers, risk-averse agents seek safety, and risk-seeking agents choose high-risk actions that are likely to produce undesirable outcomes. If the principal desires a contract that performs well in all possible circumstances, what sort of contract should she use?

This paper studies a general moral hazard problem in which the principal is uncertain about the agent's risk preferences and production technology. The principal evaluates contracts according to their worst-case payoff against the set of potential agent preferences and production technologies.

In this environment, contracts that specify transfers that are not constant over any interval of outputs — which we call *fully-contingent* — do not guarantee the principal a payoff that is larger than her payoff if the agent shirks, even if effort is costless for the agent. To see why, suppose that the agent can choose to either shirk and produce no output, or exert high effort and choose from several output distributions. If the agent exerts high effort, he chooses between a risky action that produces high expected output and a safe action that rarely produces low output. If the principal uses a contract with transfers that are strictly increasing in output, then the safe action is less likely to yield a small transfer than the risky action. Accordingly, sufficiently risk-averse agents prefer the safe action, even if it produces

[1] We use the term *effort* generically to refer to the agent's disutility from actions.

little output in expectation. For analogous reasons, fully-contingent contracts provide risk-seeking agents with incentives to take all-or-nothing risks that are again undesirable from the principal's perspective.

There are simple conditions under which a contract does not provide the agent with incentives to take extreme risks, regardless of his preferences. We call a contract *flat at the bottom* if it assigns its minimum transfer to an interval that includes the smallest possible level of output. Contracts that are flat at the bottom do not reward the agent for producing low output, and thus do not provide risk-averse agents with incentives to choose very safe actions. Symmetrically, contracts that assign their maximum transfer to an interval that includes the largest possible level of output belong to a class of contracts that we call *flat at the top*. Just as contracts that are flat at the bottom do not provide risk-averse agents with incentives to choose very safe actions, contracts that are flat at the top do not provide risk-seeking agents with incentives to choose very risky actions. There are many *partially-contingent* contracts that satisfy both of these conditions, and all such contracts guarantee the principal a payoff that is larger than if the agent shirks, provided that transfers are appropriately-sized and effort is not too costly for the agent.

The contributions of this paper are twofold. First, we demonstrate that partially-contingent contracts protect the principal from risk-taking, even when the agent has extreme risk preferences. Such contracts are widely used in practice, especially in executive compensation, where they have been identified as typical (Murphy (1999, 2013)). These contracts are sometimes identified as "80/120 plans", which refers to an especially prevalent contractual form in which the executive's compensation varies only when his performance measure is between 80% and 120% of a specified target. The use of these contracts in practice is not theoretically well-understood[2], and they are regarded as sub-optimal by researchers who propose that they be replaced with piece-rate compensation schemes (Murphy and Jensen (2011)). In particular, these bonus plans do not provide stronger incentives for performance at 75% of the target level than they do for 10%, nor do they reward the executive who doubles his target any more than one who exceeds it by 25%. While this feature is a deficiency

[2]Partially-contingent contracts do appear as solutions to special cases of canonical contracting models — for example, in the setting studied in Hölmstrom (1979) — if e.g. the monotone likelihood ratio property fails to hold but the principal nevertheless demands a contract with transfers that are strictly increasing in output.

when the agent's risk preferences are known, it is also exactly *why* these contracts exhibit robustness to risk-taking by an agent with unknown risk preferences.

The second contribution of the paper lies in complementing the surprisingly limited literature on contracting with diversely risk-averse agents (de Meza and Webb (2001); Jullien, Salanié, and Salanié (2007)). To the best of our knowledge, this paper is the first to study risk-taking by an agent with unknown risk preferences.[3]

There is a growing body of work that considers mechanism design problems with worst-case objectives or other related criteria, often where agents have complex private information (López-Cunat (2000); Chung and Ely (2007); Chassang (2013); Frankel (2014); Garrett (2014); Carroll (2015, 2016); Carroll and Meng (2016a,b); Carroll (2017); Auster (2018); Antic (2014); Bergemann, Brooks, and Morris (2017); Marku and Díaz (2017); Dai and Toikka (2017)). This literature includes at least two papers in which the agent's preferences are unknown to the principal (Frankel (2014), Garrett (2014)), in contexts other than moral hazard. Authors in this literature and elsewhere offer a variety of justifications for worst-case analysis, including both tractability (Hansen and Sargent (2007)) and its tendency to select simple mechanisms. More directly, Carroll (2015) argues that the exhibition of mechanisms that perform well in a wide variety of environments is a useful exercise when studying situations in which real-world principals might themselves work with approximate models. We adopt a worst case criterion in the present paper in part because it provides the simplest language in which to communicate our central finding, which is that partially-contingent contracts exhibit robustness to risk-taking in ways that fully-contingent contracts do not.

The paper is organized as follows. We describe our model in section 2.2, and present our analysis in section 2.3. Lastly, we discuss the related literature in greater detail and conclude in section 2.4. Technical material is given in appendix B, along with extensions.

[3]The author is engaged in an ongoing research project that studies the use of stochastic contracts in aligning the agent's risk-taking behavior with the principal's objectives (Rosenthal (2018)). This project considers an environment in which the principal is subject to only local uncertainty about the agent's risk preferences, but has much less information about his choice technology than in the present paper. Furthermore, the present paper restricts attention to deterministic contracts, which seems to be a sensible restriction for many applications.

2.2 Model

2.2.1 Output, actions, and contracts

The set of potential values $\mathcal{Y} \subset \mathbb{R}^+$ for *output* y is an interval $\mathcal{Y} = [0, \overline{\mathcal{Y}}]$, with $\overline{\mathcal{Y}} > 0$. *Actions* are pairs $(F, e) \in \Delta(\mathcal{Y}) \times \mathbb{R}^+$, where output y is distributed according to F and the agent's non-negative private cost e is referred to generically as *effort*. The agent chooses actions from compact[4] set $\mathcal{A} \subset \Delta(\mathcal{Y}) \times \mathbb{R}^+$, which we refer to as the agent's *technology*. The principal is constrained by limited liability, and upper semicontinuous *contracts* $w : \mathcal{Y} \to \mathbb{R}^+$ reward the agent with a non-negative transfer $w(y)$ when his chosen action produces output y.

2.2.2 The agent's preferences

The agent in our model chooses both how much effort to exert and also what risks to take. His preferences thus describe the extent of his disutility from effort, and also his preferences over financial lotteries. The agent is an expected utility maximizer with additively separable preferences:

Assumption 3. *The agent is an expected utility maximizer. His utility function* $v : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ *over transfer-effort pairs* (t, e) *is additively separable:*

$$v(t, e) = u(t) - k(e).$$

The function u *is strictly increasing, continuous, and satisfies* $u(0) = 0$. *The function* k *is non-decreasing, continuous, and satisfies* $k(0) = 0$.

We call v the agent's *type*, and emphasize that both the *risk preference* u and the cost-of-effort function k are idiosyncratic to the agent's type.

[4]We give the set of Borel probability distributions $\Delta(\mathcal{Y})$ the topology of weak convergence, $\mathbb{R}^+$ the Euclidean topology, and $\Delta(\mathcal{Y}) \times \mathbb{R}^+$ the corresponding product topology. We write δ_x for the distribution that assigns probability 1 to the event x, and δ_x^p for the distribution that assigns probability p to event x and probability $1 - p$ to event 0. Weak convergence is naturally compatible with the continuous and bounded utility functions and upper semicontinuous contracts considered in our model: in particular, our choice of topology endows the agent's expected utility functional with upper semicontinuity. In addition to providing for the existence of a solution to the agent's problem, our choice of topology plays a substantive role in our problem, because our results rely on an assumption that the principal is (at least) locally uncertain about the agent's production technology.

In this paper, the principal does not know the agent's type. Instead, she knows the identity of some set $\mathcal{V}$ to which it belongs. We call $\mathcal{V}$ *the environment,* and focus our analysis on cases in which $\mathcal{V}$ is large. In particular, define as follows:

$$\mathcal{U}_A = \Big\{u : \mathbb{R}^+ \to \mathbb{R} | u \text{ is strictly increasing, continuous, weakly concave, and } u(0) = 0\Big\}$$
$$\mathcal{U} = \Big\{u : \mathbb{R}^+ \to \mathbb{R} | u \text{ is strictly increasing, continuous, and } u(0) = 0\Big\}.$$

Assumption 4. *The environment $\mathcal{V}$ includes at least one type v with risk preference u for each risk preference in either $\mathcal{U}_A$ or $\mathcal{U}$.*

We emphasize that throughout the paper, we consider only environments $\mathcal{V}$ that satisfy assumption 4. If $\mathcal{V}$ includes at least one type v with risk preference u for every risk preference u in $\mathcal{U}_A$, we call $\mathcal{V}$ an environment *with risk aversion.* If $\mathcal{V}$ includes at least one type v with risk preference u for every risk preference u in the larger set $\mathcal{U}$, we call $\mathcal{V}$ an environment *without risk aversion.* For the purposes of assumption 4, we identify risk preferences that are affine transformations of one another.

2.2.3 The agent's production technology

In addition to her uncertainty about the agent's type v, the principal is also uncertain about the agent's production technology$\mathcal{A}$. In particular, at the time of contracting, the principal knows only a lower bound $\mathcal{A}_0$ and an upper bound $\mathcal{A}_1$ for the agent's technology$\mathcal{A}$ in the set inclusion order $\subset$.

Assumption 5. *The agent's technology $\mathcal{A}$ satisfies $\mathcal{A}_0 \subset \mathcal{A}$. The lower bound $\mathcal{A}_0$ is compact and satisfies the following criteria:*

1. *(shirking) $\mathcal{A}_0$ contains $(\delta_0, 0)$;*

2. *(costly production) $\mathcal{A}_0$ contains $(F, 0)$ only if $F = \delta_0$;*

3. *(risky production) if $(F, e) \in \mathcal{A}_0$ and $e > 0$, then F has full support on $\mathcal{Y}$;*

4. *(non-triviality) $\mathcal{A}_0$* contains at least one action (F, e) with $F \neq \delta_0$.

Assumption 6. *The agent's technology $\mathcal{A}$ satisfies $\mathcal{A} \subset \mathcal{A}_1$. The upper bound $\mathcal{A}_1$ is open in $\Delta(\mathcal{Y}) \times \mathbb{R}^+$.*

Label as $\mathscr{A}$ the set of *feasible technologies* $\mathcal{A}$ that satisfy assumptions 5 and 6. At the time of contracting, the principal does not know the identity of the agent's technology $\mathcal{A}$, but does know the identity of the set of feasible technologies $\mathscr{A}$.

This environment is a bounded version of the framework developed in Carroll (2015). In particular, assumption 6 admits the special case[5] $\mathcal{A}_1 = \Delta(\mathcal{Y}) \times \mathbb{R}^+$, in which our model of the agent's technology reduces to that of Carroll (2015), augmented with some additional restrictions. With regards to assumption 5, the costly production assumption is for technical convenience only[6], and the non-triviality assumption ensures that our problem is interesting to study. The full support assumption, feasible shirking assumption, and assumption 6 are more substantive. In brief, the combination of these assumptions provides for two important possibilities:

1. the agent faces a trade-off between risky actions that produce high expected output, and safe actions that produce low expected output;

2. the agent faces a trade-off between risky actions that produce high expected output, and all-or-nothing actions that produce very high output with relatively high probability but no output otherwise.

The interaction of the first type of technology with risk-averse agents, and the second type of technology with risk-seeking agents, lead to our necessary conditions for contracts that exhibit robustness to risk-taking. We discuss this point in detail in section 2.3.

The agent's choice correspondence and payoff are respectively indicated as follows, for

[5]Our results hold under all specifications of $\mathcal{A}_1$, except in proposition 8, wherein we study the optimality of binary contracts in a special case of our model that assumes — along with other restrictions — $\mathcal{A}_1 = \Delta(\mathcal{Y}) \times \mathbb{R}^+$. More generally, were we to restrict our attention to this case throughout the paper, the feasible shirking assumption would be unnecessary.

[6]Because the shirking action does not yield an output distribution with full support, requiring that expected output vanishes as effort approaches 0 streamlines some of our proofs. This assumption can be readily relaxed, at the cost of introducing some mild complications.

technology $\mathcal{A}$, type v, and contract w:

$$c(\mathcal{A}|v,w) := \operatorname*{arg\,max}_{(F,e)\in\mathcal{A}} E_F\Big[v\Big(w(y),e\Big)\Big]$$
$$V(\mathcal{A}|v,w) := \max_{(F,e)\in\mathcal{A}} E_F\Big[v\Big(w(y),e\Big)\Big].$$

In lemma 8, stated in the appendix, we demonstrate that $c(\mathcal{A}|v,w)$ is non-empty.

2.2.4 The principal's preferences

The principal has strictly increasing and concave utility function $u_p : \mathbb{R} \to \mathbb{R}$ with $u(0) \equiv 0$. Her payoff when the agent's technology is $\mathcal{A}$ and his type is v is as follows:

$$\pi(\mathcal{A}|v,w) := \min_{(F,e)\in c(\mathcal{A}|v,w)} E_F\Big[u_p\Big(y-w(y)\Big)\Big].$$

We normalize $u_p(0) \equiv 0$. In order to ensure that our problem is well-defined, we break ties against the principal.[7] The principal's worst-case payoff against the environment $\mathcal{V}$ and the set of feasible technologies $\mathscr{A}$ is as follows:

$$\pi(\mathscr{A}|\mathcal{V},w) := \inf_{(v,\mathcal{A})\in\mathcal{V}\times\mathscr{A}} \pi(\mathcal{A}|v,w).$$

We call $\pi(\mathscr{A}|\mathcal{V},w)$ the *guarantee* provided by contract w, and say that w *provides a guarantee* if $\pi(\mathscr{A}|\mathcal{V},w) > 0$.

2.2.5 Timing and summary

The interaction between the principal and the agent is as follows:

1. the principal — knowing the environment $\mathcal{V}$ and the set of feasible technologies $\mathscr{A}$ — chooses a contract w;

[7] We allow for upper semicontinuous contracts because there are discontinuous contracts that warrant discussion in our problem. Because $y - w(y)$ is thus lower semicontinuous but not necessarily continuous, the principal's payoff achieves its minimum on the compact set $c(\mathcal{A}|v,w)$ but does not necessarily achieve its maximum. We break ties against the principal in order to ensure that our problem is well-defined. Technical details aside, our tie-breaking rule seems conceptually compatible with the worst-case criterion considered in this paper.

2. the agent — knowing his type v and his technology $\mathcal{A}$ — chooses the action (F, e) in the set $c(\mathcal{A}|v, w)$ that minimizes the principal's expected payoff $E_F\Big[u_p\Big(y - w(y)\Big)\Big]$;

3. output y is drawn from distribution F;

4. the agent's payoff is $u\Big(w(y)\Big)$ and the principal's payoff is $u_p\Big(y - w(y)\Big)$.

2.3 Analysis

In our model, the agent's production technology is multi-dimensional, in the sense that he chooses both effort and also the distribution of output. Consequently, in addition to the usual problem of providing incentives for effort, the principal must also provide incentives for the agent to take risks that are desirable from her perspective. Both of these considerations are complicated by the principal's uncertainty about the agent's risk preferences. We proceed by first studying the set of contracts that guarantee the principal a positive payoff when effort is costless for the agent. In doing so, we focus our discussion entirely on the problem of providing robust incentives for risk-taking to agents with unknown risk preferences. After doing so, we return to the usual case in which effort is costly for the agent, and discuss the extent to which our results hold in this more general setting. In addition, we exhibit a special case of our model in which binary contracts with only two transfer levels are optimal.

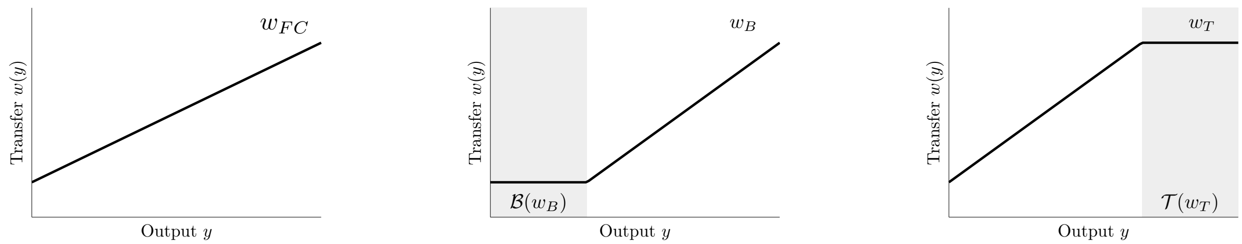
2.3.1 Partially-contingent contracts

First, we describe our contractual forms of interest. Given a contract w, define as follows:

$$\mathcal{B}(w) := \arg\min_{y \in \mathcal{Y}} w(y)$$
$$\mathcal{T}(w) := \arg\max_{y \in \mathcal{Y}} w(y).$$

Simply put, $\mathcal{B}(w)$ is the set of minimizers for w and $\mathcal{T}(w)$ the set of maximizers. We introduce two corresponding criteria, which we show to be necessary conditions for the principal to secure a guarantee.

Definition 6. Call a contract w *flat at the bottom* if $[0, x] \subset \mathcal{B}(w)$ for some $x > 0$.



(a) fully-contingent (b) flat at the bottom (c) flat at the top

Figure 2.1: Contracts which satisfy our flatness criteria.

Definition 7. Call a contract w *flat at the top* if $P_F[\mathcal{T}(w)] > 0$ for some $(F, e) \in \mathcal{A}_0$ with $e > 0$.

Contracts that are flat at the bottom are constant on some neighborhood of 0 output. Moreover, the transfer assigned to this neighborhood is the minimum transfer assigned by the contract. Contracts that are flat at the top assign their maximum transfer to a set of outputs $\mathcal{T}(w)$ that the agent is capale of producing with positive probability. We call contracts that satisfy both of these criteria *partially-contingent,* and contracts which satisfy neither *fully-contingent.* See figure 2.1 for illustration.

There are asymmetries between our two criteria. First, flat at the bottom is a physical requirement, while flat at the top is statistical. Second, the precise analogue to our flat at the bottom criterion is an alternative version of definition 7 in which we require that $[x, \overline{\mathcal{Y}}] \subset \mathcal{T}(w)$ for some $x < \overline{\mathcal{Y}}$. Accordingly, there is a sense in which our flat at the top criterion is strictly more permissive than our flat at the bottom criterion.[8] Why?

When the principal uses a contract that fails to be flat at the bottom, the agent might choose safe actions that produce a small quantity of positive output with high probability, as we demonstrate. However, when she uses a contract that fails to be flat at the top, the agent might choose actions that produce a high level of output with low probability — and no output otherwise — as we also demonstrate. In the former situation, what drives the principal's payoff is the low *quantity* of output. Hence, the physical nature of our flat at the

[8]Definition 7 and the alternative criterion proposed in this paragraph are identical if we restrict our attention to monotone contracts and assume that actions $(F, e) \in \mathcal{A}_0$ other than $(\delta_0, 0)$ have continuous output distributions. Although the latter assumption seems innocuous, the restriction to monotone contracts seems more heavy-handed, and is not pursued here.

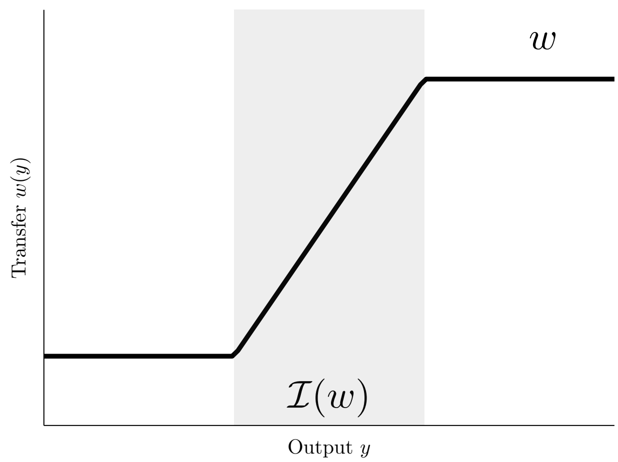


Figure 2.2: The incentive region $\mathcal{I}(w)$, shaded.

bottom criterion. However, in the latter situation, what drives the principal's low payoff is the low *probability* that positive output is produced. Hence, the statistical nature of our flat at the top criterion.

We proceed with a concrete example. It will be useful to introduce some terminology related to definitions 6 and 7. Define as follows:

$$\underline{\mathcal{I}}(w) := \inf \mathcal{Y} \setminus \mathcal{B}(w)$$
$$\overline{\mathcal{I}}(w) := \inf \mathcal{T}(w).$$

In words, $\underline{\mathcal{I}}(w)$ and $\overline{\mathcal{I}}(w)$ are the smallest and largest levels of output that the agent has strict incentives to produce, respectively. We call $\mathcal{I}(w) := [\underline{\mathcal{I}}(w), \overline{\mathcal{I}}(w)]$ the *incentive region.* See figure 2.2 for illustration.

Consider full support output distribution $F \in \Delta(\mathcal{Y})$. We ask: what sort of distributions G might the agent prefer to F under a particular non-constant contract w? If the agent is risk-averse, he might prefer distributions G that assign a smaller probability than F to the bottom of the output range $\mathcal{B}(w)$:

$$P_F[\mathcal{B}(w)] > P_G[\mathcal{B}(w)]. \tag{2.1}$$

Inequality 2.1 yields a lower bound for $E_G[y]$:

$$E_G[y] > \Big(1 - P_F[\mathcal{B}(w)]\Big) \cdot \underline{\mathcal{I}}(w).$$

A necessary condition for this quantity to be positive is $\underline{\mathcal{I}}(w) > 0$, which in turn yields our flat at the bottom criterion.[9] Relatedly, suppose that the agent is risk-seeking. He might prefer distributions G that assign a larger probability than F to the top of the output range $\mathcal{T}(w)$:

$$P_G[\mathcal{T}(w)] > P_F[\mathcal{T}(w)]. \tag{2.2}$$

As above, inequality 2.2 yields a lower bound for $E_G[y]$:

$$E_G[y] > P_F[\mathcal{T}(w)] \cdot \overline{\mathcal{I}}(w).$$

A necessary condition for this quantity to be positive is $P_F[\mathcal{T}(w)] > 0$, which in turn yields our flat at the top criterion.[10] There are two situations in which $\mathcal{T}(w)$ might not be a non-degenerate interval $[x, \overline{\mathcal{Y}}]$ for a contract w that is nevertheless flat at the top. First, if F has a mass point at the top of the output range $\overline{\mathcal{Y}}$, then $\mathcal{T}(w)$ might only include $\overline{\mathcal{Y}}$. Second, if w only achieves its maximum on the interior of $\mathcal{Y}$, then $\mathcal{T}(w)$ might exclude $\overline{\mathcal{Y}}$.

Our discussion thus far suggests that contracts that do not satisfy our flatness criteria might give the agent an incentive to choose either extremely safe or extremely risky actions. This turns out to indeed be the case, and we formalize this observation in section 2.3.2. Conversely, is it true that contracts that *do* satisfy these criteria do not provide perverse incentives for risk-taking, regardless of the agent's preferences? It turns out that the answer is yes, up to some mild additional mild conditions.

Definition 8. Call w a *capped bonus contract* if w is flat at the bottom; flat at the top; increasing; non-constant; and satisfies $w(y) \leq y$ for all y, with equality if and only if $y = 0$.

In addition to satisfying both of our flatness criteria, capped bonus contracts reward the agent with appropriately-sized transfers. We also require monotonicity, for expository

[9] It is also necessary that $1 - P_F[\mathcal{B}(w)] > 0$, but the full support assumption ensures that this is vacuously satisfied by any contract that is not almost-everywhere constant.

[10] It is also necessary that $\overline{\mathcal{I}}(w) > 0$, but this is vacuously satisfied by any contract that is not constant.

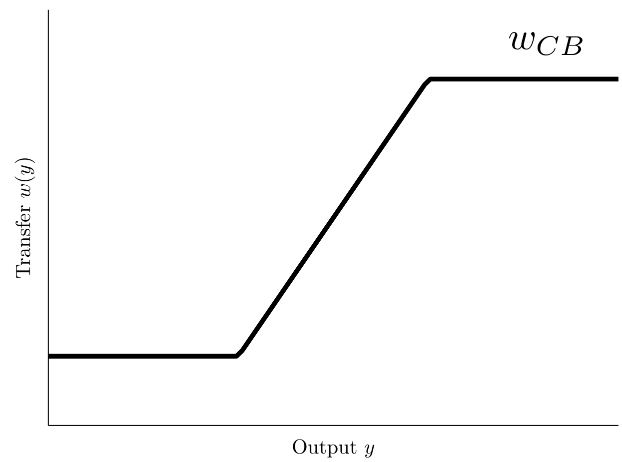


Figure 2.3: A capped bonus contract w_{CB}.

purposes only. See figure 2.3 for illustration.

To see why capped bonus contracts might perform well in our environment, return to our example with full support output distribution F. Suppose that w is a capped bonus contract, and that some alternative output distribution G is preferred by the agent to F under w. First, because w is increasing, it must be the case that the following inequality holds for some quantity of output x in the incentive region $\mathcal{I}(w)$:

$$P_G[y \geq x] \geq P_F[y \geq x]. \tag{2.3}$$

Second, because w is a capped bonus contract — i.e., because transfers are not too large — by definition we have that $y - w(y) \geq 0$ for all output levels y, with equality only if $y = 0$. In turn, we obtain the following bound for the principal's payoff under output distribution G:

$$E_G\Big[u_p\Big(y - w(y)\Big)\Big] \geq P_G[y \geq x] \cdot u_p\Big(x - w(x)\Big). \tag{2.4}$$

Because F has full support and w is flat at the top, $P_F[y \geq x] > 0$. Accordingly, combining 2.3 and 2.4 reveals that the payoff to the principal from G is bounded below by some quantity strictly larger than 0. Thus, we see that capped bonus contracts yield at least partial alignment between the principal's objectives and the agent's risk-taking behavior, regardless of his preferences.

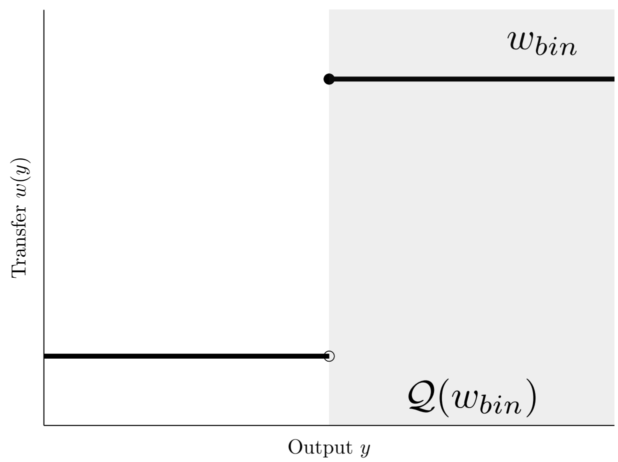


Figure 2.4: A binary contract w_{bin}.

Of course, there are many examples of capped bonus contracts, some of which are quite complex. As a final matter, we introduce a particularly simple class of contracts that that satisfy both of our flatness criteria. Define as follows:

Definition 9. Call w a *binary contract* if w is as follows, for $\mathcal{Q} \subset \mathcal{Y}$ closed and $b > 0$:

$$w(y) = \begin{cases} s & y \notin \mathcal{Q} \\ s + b & y \in \mathcal{Q}. \end{cases}$$

Binary contracts pay the agent a salary s, and a bonus payment b for achieving quota $\mathcal{Q}$. See figure 2.4 for illustration. Provided that the principal chooses a reasonable quota $\mathcal{Q}$, binary contracts do not provide incentives for the agent to take extreme risks, regardless of his preferences. For example, return for a final time to our example with full support output distribution F. Given alternative output distribution G and risk preferences u, u', we have the following:

$$\begin{aligned} E_G\Big[u\Big(w(y)\Big)\Big] \geq E_F\Big[u\Big(w(y)\Big)\Big] &\iff P_G[\mathcal{Q}] \geq P_F[\mathcal{Q}] \\ &\iff E_G\Big[u'\Big(w(y)\Big)\Big] \geq E_F\Big[u'\Big(w(y)\Big)\Big]. \end{aligned}$$

Thus, binary contracts provide the same incentives for risk-taking to each type of agent. As a partial consequence of this fact, there are special cases of our model in which these contracts are optimal for the principal.

2.3.2 Costless effort

This section formalizes the discussion of section 2.3.1. First, we make concrete the notion that effort is costless for the agent. Define as follows:

$$\mathcal{V}_A^0 := \Big\{v | v(t,e) = u(t) \text{ for } u \in \mathcal{U}_A\Big\}$$
$$\mathcal{V}^0 := \Big\{v | v(t,e) = u(t) \text{ for } u \in \mathcal{U}\Big\}.$$

We clarify that $\mathcal{V}_A^0$ includes the agent type $v(t,e) = u(t)$ for *every* $u \in \mathcal{U}_A$, and similarly for $\mathcal{V}^0$ and $u \in \mathcal{U}$. We call the case of our model in which $\mathcal{V} = \mathcal{V}_A^0$ the *costless effort environment with risk aversion,* and similarly the case with $\mathcal{V} = \mathcal{V}^0$ the *costless effort environment without risk aversion.*

We proceed now to the statement of our formal results. First, contracts that provide guarantees are flat at the bottom, in both the costless effort environment with risk aversion and the costless effort environment without risk aversion:

Proposition 4. *Consider either the costless effort environment with risk aversion or the costless effort environment without risk aversion. If contract w guarantees the principal a positive payoff, then w is flat at the bottom.*

Contracts that are not flat at the bottom provide some types of agents with incentives to choose safe actions that produce little output in expectation, provided that such actions are feasible.

To see why, suppose for the purposes of simplicity that $\mathcal{A}_0$ contains only two actions: a high-effort risky action (F, e), and the shirking action $(\delta_0, 0)$. In our model, the agent's technology $\mathcal{A}$ might include additional actions that are similar to those in $\mathcal{A}_0$. For any *fixed* agent type, if the principal's uncertainty about the agent's technology is sufficiently minor — that is, if $\mathcal{A}_1$ is not much larger than $\mathcal{A}_0$ — then the principal is assured that the agent will choose either the high-effort action (F, e), or a different action with a similar output distribution. This is a natural consequence of the continuity of our choice of topology on $\Delta(\mathcal{Y}) \times \mathbb{R}^+$, in which the agent's payoff is upper semicontinuous.

However, regardless of the extent of the principal's uncertainty about the agent's technology, if the principal uses a contract that is not flat at the bottom, there are low-output feasible actions that *some* agent types strictly prefer to (F, e). In particular, sufficiently risk-averse agents prefer actions $(\delta_x, 0)$ for which x is assigned a non-minimal transfer $w(x)$. To see why, note that because output distribution F has full support over $\mathcal{Y}$, the agent receives transfers that are strictly smaller than $w(x)$ with positive probability if he chooses the action (F, e). This is not the case when he chooses the risk-less[11] action $(\delta_x, 0)$. Thus, if the agent is sufficiently risk-averse, he strictly prefers the latter action to the former. Because we assume that the principal is at least somewhat uncertain about the agent's production technology, the safe action $(\delta_x, 0)$ is feasible for all x sufficiently small. Straightforward generalizations of this argument yield a proof of proposition 4.

One interpretation of proposition 4 is that contracts that are not flat at the bottom do not provide the agent with adequate incentives to take on risk, because such contracts are too generous in rewarding the agent for producing low levels of output. However, the principal only likes for the agent to choose risky actions that produce high output in expectation. This is not a feature of all high-risk actions. If the agent is not necessarily risk-averse, fully-contingent contracts provide incentives for some types of agents to engage in excessive risk-taking:

Proposition 5. *Consider the costless effort environment without risk aversion. If contract w guarantees the principal a positive payoff, then w is flat at the top.*

The mechanics of proposition 5 are symmetric to that of proposition 4. If the principal uses a contract that provides the strongest incentives only for output levels that the agent is unlikely to produce, the principal provides some types of agents with incentives to choose high-risk, low-expected-output actions, provided that such actions are feasible.

[11] There is inconsistency in our application of the full support assumption: in particular, we require F to have full support for $(F, e) \in \mathcal{A}_0$, but do not extend this requirement to other actions in $\mathcal{A}_1$. While this is a natural objection, *every* output distribution $F \in \Delta(\mathcal{Y})$ can be continuously transformed into a full support output distribution by mixing F with full support distribution $G \in \Delta(\mathcal{Y})$. Consequently, extending the full support assumption to $\mathcal{A}_1$ is vacuous and does not change our results.

To see why, suppose again that $\mathcal{A}_0$ contains only the high-effort action (F, e) and the shirking action $(\delta_0, 0)$. Just as is in our discussion of proposition 4: for any *fixed* agent type, the principal is protected from extreme risk-taking if she is only slightly uncertain about the agent's production technology. This is true regardless of what sort of contract that the principal uses.

However, if the principal uses a contract that is not flat at the top, there are again low-output feasible actions that *some* agent types prefer to (F, e): in particular, actions $(\delta_x^p, 0)$ with $w(x)$ large and p small. To see why, suppose for the purposes of illustration that w is strictly increasing, and further that F is continuous. The probability with which the agent earns a transfer that is at least $w(x)$ if he chooses the high-output action (F, e) vanishes to 0 as x approaches the maximum feasible output level $\overline{\mathcal{Y}}$. Consequently, if the agent is sufficiently risk-seeking, he might prefer long-shot actions $(\delta_x^p, 0)$ with $p > P_F[y \geq x]$ to (F, e), even if p is small. Because the principal is at least slightly uncertain about the agent's technology, it is again the case that $(\delta_x^p, 0)$ is feasible for all p sufficiently small, regardless of the size of output x. The proof of proposition 5 is then a generalization of this logic.[12]

The above discussions suggest that the presence of even minor uncertainty about the agent's technology has a considerable role to play in our problem. In appendix B.6, we demonstrate by counterexample that propositions 4 and 5 do not hold as stated if there is instead no uncertainty about technology. Accordingly, if the principal uses a fully-contingent contract, then she suffers discontinuous payoff losses from the introduction of uncertainty about the agent's technology. However, capped bonus contracts do not share this feature:

Proposition 6. *Consider the costless effort environment with risk aversion or the costless effort environment without risk aversion. If contract w is a capped bonus contract, then w guarantees the principal a positive payoff.*

In presenting proposition 6 as stated, we acknowledge that we have not provided an exact converse to propositions 4 and 5. In particular, although we require that capped bonus

[12]The full support assumption does not play a prominent role in the proof of proposition 5. Instead, one can imagine a loosely analogous role to be played by a continuity restriction to each $(F, e) \in \mathcal{A}_0$ with $e > 0$. Under this alternative hypothesis, it is possible to strengthen our flat at the top criterion, so that contracts which are flat at the top are required to achieve their maximum on a set of outputs with positive Lebesgue measure. This is not pursued here.

contracts are increasing, this is for expository purposes only. Separately, contracts that guarantee the principal a positive payoff need not be flat at the top if the agent is known to be risk-averse. Finally, recall that definition 8 stipulates restrictions to the size of transfers to the agent. Without such restrictions, it need not be true that capped bonus contracts guarantee the principal a positive payoff, even in the costless effort environment.[13]

Having described the set of contracts that provide the principal with a guarantee in the costless effort environment, what can we say about which contracts provide the *largest-possible* guarantee? In fact, no such contracts exist in this setting, because the principal does best to use a contract with arbitrarily small transfers. However, if the principal's objective depends only on output, rather than net output, then binary contracts guarantee her the largest-possible quantity of expected revenue:

Proposition 7. *Consider the costless effort environment without risk aversion, and suppose that the agent's production technology is unbounded above, so that $\mathcal{A}_1 = \Delta(\mathcal{Y}) \times \mathbb{R}^+$. Suppose further that the principal's objective is to maximize expected revenue, so that her payoff when she uses contract w is as follows:*

$$\pi^*(\mathcal{A}|v, w) := \min_{(F,e) \in c(\mathcal{A}|v,w)} E_F[y]$$
$$\pi^*(\mathscr{A}|\mathcal{V}, w) := \inf_{(v,\mathcal{A}) \in \mathcal{V} \times \mathscr{A}} \pi^*(\mathcal{A}|v, w).$$

There exists a binary contract that provides the largest-possible revenue guarantee for the principal: i.e., a binary contract that maximizes the quantity π^*.

The substance of the proof of proposition 7 lies in establishing that there are risk preferences that respond to any contract *as if* that contract were instead a binary contract. Consequently, the principal does just as well in terms of revenue to indeed use a binary contract.[14]

[13]It is possible to reformulate our flatness criteria and the definition of a capped bonus contract so that proposition 8 is a tight converse to propositions 4 and 5, but this is not pursued here.

[14]In interpreting the proposition, it should be acknowledged that in the context of economic theory, revenue maximization is perhaps an unusual objective. On the other hand, it seems plausible that there might be practical circumstances for which it is relevant.

2.3.3 Costly effort

In propositions 4 and 5, we provide necessary conditions for contracts that guarantee the principal a positive payoff, in the form of our two flatness criteria. For expository purposes, we state these propositions for the special cases of the costless effort environments. However, both propositions hold broadly, as long as assumption 4 is satisfied:

Remark 4. Consider any environment with risk aversion or any environment without risk aversion. If contract w guarantees the principal a positive payoff, then w is flat at the bottom.

Remark 5. Consider any environment without risk aversion. If contract w guarantees the principal a positive payoff, then w is flat at the top.

The proofs of propositions 4 and 5 presented in the appendix are written for the general cases described in remarks 4 and 5, respectively.

Relatedly, in proposition 6 we show that if effort is costless for the agent, then capped bonus contracts guarantee the principal a positive payoff. It is *not* the case that this result can be extended to arbitrary environments, as is the case with the earlier two propositions. However, it is also not necessary that effort be costless in order for there to be contracts that guarantee the principal a positive payoff. The principal might also obtain a guarantee if the agent has more moderate preferences.

In particular, provided that the quota and bonus payment are sufficiently generous and effort is not too burdensome for the agent, binary contracts provide strong enough incentives for effort to guarantee the principal a positive payoff. Moreover, there are special cases in which these contracts provide the largest-possible guarantee:

Proposition 8. *There exists an environment $\mathcal{V}$ without risk aversion and a binary contract w such that:*

1. *w guarantees the principal a positive payoff for environment $\mathcal{V}$; and*
2. *if the agent's production technology is unbounded above — that is, $\mathcal{A}_1 = \Delta(\mathcal{Y}) \times \mathbb{R}^+$ — then w yields the largest-possible payoff guarantee for the principal.*

Proposition 8 should be interpreted with caution: binary contracts are not *in general* optimal in our model. Instead, our result should be viewed as a finding of optimality for *particular* environments $\mathcal{V}$, which we describe in appendix B.3. Specifically, binary contracts are optimal when the agent is not necessarily risk-averse, the agent has quasilinear preferences, the set of feasible potential production technologies is unbounded above[15], and certain types of risk-seeking agent types suffer (weakly) higher disutility from effort than agents with other types of risk preferences.

In detail, the optimality of binary contracts in the environment of proposition 8 arises as a consequence of three considerations. First, sufficiently risk-seeking agents interpret *every* contract approximately as if it were a binary contract. More precisely, we show that for any contract w, there exists a similarly generous binary contract w' under which these risk-seeking agent types choose (approximately) identical risks. Second, because binary contracts provide uniform incentives for risk-taking, the principal thus provides no worse incentives for risk-taking when she uses the binary contract w than she does when she uses the original contract w', regardless of the agent's preferences. Third, if the same agents that choose identical risks under w and w' also exert less effort than other agents, then w' guarantees the principal a payoff that is no worse than that guaranteed by w. While the first two elements of the above argument are true in general, the third element follows from our choice of the environment $\mathcal{V}$. Thus, the relationship between risk preferences and effort costs is an important determinant of the form of the optimal contract.[16]

2.4 Discussion and conclusions

This paper develops a model of moral hazard in which the principal does not know the agent's risk preferences. We establish circumstances under which fully-contingent contracts do not guarantee the principal a larger payoff than her payoff if the agent shirks, regardless of how costly effort is for the agent. Conversely, we see that contracts which are flat at the bottom and flat at the top, as appropriate, do not share this feature, provided that transfers are not too large. Partially-contingent contracts that satisfy these criteria are widely used in practice, and we highlight virtues of these incentive schemes.

Our results depart from classic results in agency theory that fully-contingent contracts are

[15]That is, $\mathcal{A}_1 = \Delta(\mathcal{Y}) \times \mathbb{R}^+$, as in Carroll (2015).

[16]The characterization of optimal contracts for our model in general is difficult and we do not pursue this endeavor here.

generically optimal for strictly risk-averse agents (Hölmstrom (1979)).[17] Here, the agent's risk preferences are known to the principal and his choice technology is one-dimensional, so that risk-taking is not an ingredient of the model. The principal would like to insure the agent because doing so lessens the cost of providing incentives for effort, and the optimal way to provide both insurance and incentives is to use a contract with transfers that vary everywhere with output. On the other hand, providing insurance in the form of strict incentives for even low levels of productivity is dangerous for our principal, as we have seen.

Recent papers have considered environments in which the principal knows the agent's risk preferences, but the agent's production technology is multi-dimensional (Chassang (2013); Carroll (2015); Antic (2014); Barron et al. (2017)). In such environments, it is possible for the principal to tailor a fully-contingent contract so that the agent chooses actions that maximize her own payoff (conditional on effort), and this is sometimes a first-order consideration for the principal. For example, there are natural environments in which linear contracts are optimal when the principal and the agent are both financially risk-neutral (Chassang (2013); Carroll (2015)).

More generally — and of particular relevance for our paper — an early version[18] of Carroll (2015) establishes that *utility-affine* contracts are optimal if the principal and the agent have non-linear utility functions. Just as linear contracts align the risk-taking behavior of a risk-neutral agent with a risk-neutral principal, utility affine contracts do so for fixed, non-linear preferences. However, utility-affine contracts do *not* satisfy either of our flatness criteria, and furthermore our partially-contingent contracts are not utility affine for any specification of preferences. This is easy to see: because the principal strictly prefers more output to less, utility-affine contracts reward the agent with transfers that are strictly increasing in output. Of course, such contracts are neither flat at the bottom nor flat at the top.

Thus, we see that if a principal seeks a contract that provides a guarantee in the environment of this paper, she must give up not only on providing the agent with insurance,

[17]Unlike in our model, in Hölmstrom (1979) the agent does not enjoy limited liability. Elsewhere, researchers have established contracts that satisfy our flat at the bottom criterion as optimal, in the presence of limited liability (Innes (1990); Poblete and Spulber (2012)). However, our results obtain without limited liability, as we demonstrate in section B.5 of the appendix.

[18]Version dated December 21, 2012. Accessed online at http://www.bu.edu/econ/files/2013/03/May-4-Caroll.pdf.

but also on aligning his risk-taking behavior with her own preferences.

Contracting with agents with uncertain risk preferences has received only limited attention in the moral hazard literature, regardless of the other details of the environment. One approach to this problem that is more standard than ours is to consider a single-parameter family of agent types and solve for the menu of contracts that maximizes the principal's expected payoff against an exogenous prior. This is the setting of Jullien, Salanié, and Salanié (2007), and this is the only other paper of which we are aware that considers a general moral hazard problem in which the principal does not know the agent's risk preferences. A related paper by de Meza and Webb (2001) considers an insurance problem, also with only two agent types. While these papers employ technical assumptions that might be regarded as restrictive, we consider a richer model. More substantively, while we identify contracts that provide incentives for effort and also account for risk-taking by the agent, the aforementioned papers consider only effort provision. Less closely related work studies the implications of extreme risk preferences for the size of the rationalizable set in games (Weinstein (2016)).

Given that we emphasize the agent's risk-taking behavior throughout the paper, it seems appropriate to offer some justification for this focus. In particular, our model has little to say about how to design contracts when risk is not an element of the agent's decision problem. Bolton and Dewatripont (2005) offer the following remarks about the role of risk-taking as an ingredient of principal-agent problems:

> At first sight, the overall structure of this package is difficult to relate to the optimal contracts considered in this chapter. This difficulty should not come as a surprise given that the problem of providing adequate incentives to managers is much richer than the stylized principal-agent problems we have considered [...] perhaps most importantly, the managerial incentive problem does not just boil down to eliciting more effort from the manager. It also involves issues of risk taking [...]

We embrace the authors' comments as motivation for our exercise, particularly given that partially-contingent contracts are widely-used for managerial compensation, as we discuss in our introduction.

This paper hopes to serve as a useful starting point for the study of incentive provision to risk-taking agents with unknown preferences. Except in proposition 8, we have avoided making assumptions about the relationship between risk preferences and effort costs, and we have also not endeavored elsewhere to characterize the principal's optimal contract. Just as we provide restrictions to the agent's preferences that yield optimal binary contracts in a special case of our model, perhaps it is possible to extend this approach in different directions and recover other types of interesting optimal contracts.

This paper employs the admittedly severe assumption that the principal does not have any information whatsoever about the agent's risk preferences. Why go this route, rather than the intermediate route of studying bounded uncertainty? A first answer is that our approach is in some ways a natural counterpart to the assumption that the principal knows exactly the agent's risk preferences, in that we consider perhaps the most simply form of uncertainty about risk preferences that one might imagine. Given that risk-taking by agents with unknown risk preferences is a mostly unstudied topic, it seems reasonable to begin by considering a simple environment.

Second, while most of our formal results depend on the possibility that the agent has extreme risk preferences, it is straightforward to establish that any particular contract provides optimal incentives for risk-taking to at most one type of agent. Consequently, it is indeed possible to develop bounds on the performance of various types of contracts, as a function of the extent of the principal's uncertainty about the agent's preferences. By showing that the payoff consequences of risk-taking by the agent might be severe for the principal who does not know the agent's risk preferences, we provide motivation for the more complicated exercise of considering bounded uncertainty, rather than total uncertainty as we do here. From this perspective, this paper might be viewed as providing clear rationale for considering robustness to risk-taking when designing incentives, even where it is reasonable to rule out extreme risk preferences.

CHAPTER 3

RISK ALIGNMENT

3.1 Introduction

The celebrated informativeness principle (Hölmstrom (1979); Shavell (1979)) establishes circumstances in which optimal contracts for moral hazard condition transfers to the agent only on signals that are informative about effort.[1] Contracts that are contingent on uninformative signals expose the agent to additional risk without providing more powerful incentives, and risk-averse agents demand remuneration from the principal in return. Consequently, individuals should only be compensated on the basis of variables other than their own performance if doing so *reduces* risk for the agent, as in common-shock tournaments (Lazear and Rosen (1981)) and relative performance evaluation (Hölmstrom (1982)). However, evidence on the use of relative performance evaluations in executive compensation is mixed (De Angelis and Grinstein (2018)), salespeople are assigned to separate territories, and team members undertake independent tasks.

Why then might we observe contracts that condition transfers on uninformative signals? We provide a unified explanation that does not rely on the details of the underlying contracting environment or on correlated measures of performance: if the agent's risk preferences are unknown, *all* contracts that align the agent's risk-taking behavior with the principal's own objectives have this feature.

This paper studies a general moral hazard problem. In our model, as in the standard formulation of the problem, the agent chooses an unobserved action. Actions consist of a disutility borne by the agent, which we refer to generically as *effort*, and an output distribution. The principal, who strictly prefers more output to less, uses a contract that compensates the agent on the basis of observable outcomes.

In a departure from the vast majority of the existing literature, the principal in this

[1] Formally, informativeness is defined in terms of sufficient statistics. See Proposition 3, Hölmstrom (1979).

paper is at least slightly uncertain about the agent's risk preferences. We allow for the possibility that the agent is known to be risk-averse, but do not require it. Accordingly, the principal might know that the agent is approximately risk-neutral, or be certain that he is very risk-seeking, or she might have no information whatsoever about the agent's risk preferences. The principal also does not know what sort of risks the agent might take: instead, she only knows the relationship between effort and expected output. In the face of uncertainty, the principal seeks robustness: she evaluates contracts in terms of their worst-case performance against the set of potential risk preferences and production technologies.

This is a challenging environment for the principal to contract in. In addition to the usual problem of motivating the agent to exert effort, the principal must also be careful to avoid providing the agent with incentives to take risks that are undesirable from her perspective. If the principal pays the agent only on the basis of the output that he produces, contracts that provide optimal incentives for risk-taking to some types of agents necessarily provide suboptimal incentives to others. This has payoff consequences for the principal, and the severity of these consequences is proportional to the extent of her uncertainty.

The principal might instead use a contract that additionally conditions transfers to the agent on factors other than his own output. For example, consider a contract that specifies a base salary and pays the agent a fixed bonus transfer when certain exogenous conditions are satisfied. Suppose that the principal designs the contract to reward the agent with the bonus payment with a probability that is proportional to the output that he produces. When the principal uses such a contract, the agent chooses risks that maximize expected net output, regardless of his own preferences. Consequently, if the principal is financially risk-neutral, the agent chooses actions as if his objective were to maximize the principal's payoff, conditional on effort. We call contracts with this feature *risk-aligned.*

We devote the first part of the paper to the characterization of risk alignment for a very general class of principal objectives. Regardless of the principal's preferences, we find that risk-aligned contracts must condition transfers on exogenous factors. In the second part of the paper, we incorporate the agent's choice of effort into our analysis and specialize to a financially risk-neutral principal. In this setting, risk-aligned contracts are optimal.

To see the intuition for why contracts that depend only on the agent's output are not risk-aligned, note that such contracts have two features. First, they provide the same incentives for risk-taking to each type of agent. Contracts that have this feature are necessarily binary contracts that reward the agent with a low prize in some circumstances and a high prize in all others. Second, because the principal strictly prefers more output to less, risk-aligned contracts provide strictly more powerful incentives for the production of high output than for moderate output, and similarly for moderate output and low output. While binary contracts that reward the agent only on the basis of his own performance evidently do not have this feature, there are more complicated contracts that do. For example, an agent who competes against other agents in a rank-order tournament has strict incentives to produce as much output as possible — even when prizes are homogeneous — because doing so improves his chances of winning a prize.[2]

This paper connects to several strands of literature. From a technical perspective, the contracts studied in this paper are related to the lottery mechanisms that are used by researchers to induce risk preferences in laboratory subjects. This approach is pioneered by Roth and Malouf (1979), and its theoretical foundations are further developed in Berg, Daley, Dickhaut, and O'Brien (1986). More recently, these "paying in probability" mechanisms have been identified as useful for belief elicitation (Karni (2009); Hossain and Okui (2013)). Our contribution to the understanding of these mechanisms is three-fold. First, we completely characterize the set of mechanisms that induce the desired preferences, which includes lotteries with non-degenerate prizes. Second, we explicitly incorporate the principal's preferences into our risk alignment criterion, which may or may not depend on transfers to the agent. Third, we show how the alignment achieved by these mechanisms has payoff implications for the principal.

There is a small literature in economic theory that considers moral hazard problems in which the agent has unknown risk preferences and a one-dimensional[3] production technology (de Meza and Webb (2001); Jullien et al. (2007)). Separately, there is a larger literature in which the agent has known risk preferences and a multi-dimensional production technology (Diamond (1998); Chassang (2013); Carroll (2015); Antic (2014); Barron, Georgiadis, and Swinkels (2017)). As in the present paper, an earlier paper by the author studies a moral

[2]More precisely, incentives are everywhere strict provided that the relevant order statistic has full support.
[3]We call production technologies in which there is exactly one action for each level of effort *one-dimensional.*

hazard problem in which the agent has unknown preferences and a multi-dimensional technology (Rosenthal (2019)).[4]

Along with several of the aforementioned papers (Chassang (2013); Antic (2014); Carroll (2015); Rosenthal (2019)), this paper belongs to a growing body of work that studies mechanism design problems with worst-case objectives or other related criteria (López-Cunat (2000); Chung and Ely (2007); Carroll (2016); Carroll and Meng (2016a,b); Carroll (2017); Auster (2018); Bergemann, Brooks, and Morris (2017); Marku and Díaz (2017); Dai and Toikka (2017)). This literature includes at least two additional papers that study environments in which the principal does not know the agent's preferences over deterministic outcomes (Frankel (2014); Garrett (2014)). Our risk alignment criterion is related to the analysis of Frankel (2014) and Carroll (2015) in important ways, and we discuss these relationships in detail in the concluding section of the paper. This paper also adds to the literature studying local robustness in mechanism design, including one aforementioned paper (Carroll and Meng (2016b)) and other less closely related papers (Aghion, Fudenberg, Holden, Kunimoto, and Tercieux (2012); Madarász and Prat (2017)). We emphasize that our local uncertainty framework implies that our results are not driven by the possibility that the agent has extreme preferences.

We aim to make two contributions. First, we provide a theoretical contribution by identifying a simple and robust solution to the complex problem of providing incentives to a risk-taking agent with unknown risk preferences. In particular, when the agent is known to be risk-averse, our undominated optimal contracts are stochastic bonus contracts with only two distinct payment levels. Second, we provide an explanation for the use of stochastic mechanisms for incentive provision. While authors elsewhere have identified a role for stochastic mechanisms in principal-agent problems, recent papers have emphasized the restoration of the optimality of deterministic mechanisms under various conditions (Strausz (2006); Kováč and Mylovanov (2009)).

[4]The research question, model, analysis, and results presented here are neither a special case nor a generalization of the author's earlier work. The present paper emphasizes risk alignment and optimality when the principal is slightly uncertain about the agent's risk preferences, while the earlier paper characterizes the set of contracts that guarantee the principal a positive payoff when the agent's preferences over lotteries are unrestricted. In the earlier paper, the principal has much more information about the agent's production technology than in the present paper, and contracts are restricted to depend only on output.

The body of the paper is organized as follows. We describe our model in section 3.2. In section 3.3, we provide examples and characterize the set of risk-aligned contracts. In section 3.4, we study the optimality of risk-aligned contracts for a financially risk-neutral principal. Lastly, discuss the relationship of this paper to the existing literature and conclude in section 3.5. Technical material is given in appendix C.

3.2 Model

We write $\mathbb{R}^+$ for the positive real numbers and $\mathbb{R}^{++}$ for the strictly positive real numbers. $\mathbb{R}^k$ has the Euclidean topology, and we write $\Delta(\mathcal{K})$ for the set of Borel probability measures over $\mathcal{K} \subset \mathbb{R}^k$, equipped with the topology of weak convergence. Finally, we write δ_k for the discrete distribution that assigns probability 1 to the event $k \in \mathcal{K}$. Product spaces have their natural product topologies, and sets of utility functions have the topology of uniform convergence.

3.2.1 Technology and contracts

The set of potential values $\mathcal{Y} \subset \mathbb{R}^+$ for *output* y is compact. We normalize $\min \mathcal{Y} \equiv 0$, and label $\max \mathcal{Y} := \overline{\mathcal{Y}}$. Our results hold but are vacuous when $\mathcal{Y}$ has only two elements, and so we assume that there are at least three distinct output levels. *Actions* are pairs $(F, e) \in \Delta(\mathcal{Y}) \times \mathbb{R}^+$, where output $y \sim F$ and e determines the agent's disutility from action. We refer to e generically as *effort.* The agent chooses from compact set of actions $\mathcal{A}$, which we call the agent's *technology.*

The principal has some information about the relationship between effort costs and feasible expected output. In particular, for non-empty and compact set of effort levels $\mathcal{E} \subset \mathbb{R}^{++}$ and a continuous map $\mu : \mathcal{E} \to (0, \overline{\mathcal{Y}})$, the agent's technology is known to satisfy the following criterion:

Assumption 7. *For each effort level $e \in \mathcal{E}$, the agent's technology $\mathcal{A}$ includes at least one action (F, e) with $E_F[y] \geq \mu(e)$. Moreover, the agent is always free to exert no effort and produce no output: $(\delta_0, 0) \in \mathcal{A}$.*

The principal knows a set of feasible effort levels $\mathcal{E}$ available to the agent. Second, the principal also knows a lower bound $\mu(e)$ to how much expected output the agent is capable producing when he exerts effort e, for each effort level $e \in \mathcal{E}$. The principal does not know the identity of the agent's technology $\mathcal{A}$, but instead knows the identity of the set $\mathcal{E}$ and the map μ, and that $\mathcal{A}$ satisfies assumption 7. We call such technologies *feasible,* and label the set of feasible technologies $\mathscr{A}$.

The principal and the agent do not contract on actions, which are unobserved. Instead, they contract on output y and *signal* x. The set of potential values $\mathcal{X} \subset \mathbb{R}^N$ for x is compact, convex, and has a nonempty interior. The signal $x \sim G \in \Delta(\mathcal{X})$ is statistically independent from output y: $F \perp\!\!\!\perp G$ for all $F \in \Delta(\mathcal{Y})$. The distribution G is continuous, has full support, and is known to both the agent and the principal at the time of contracting.

The agent enjoys limited liability, and the principal provides incentives to the agent by paying him with an upper semicontinuous *contract* $w : \mathcal{Y} \times \mathcal{X} \to \mathcal{T}$, where $\mathcal{T} \subset \mathbb{R}^+$ is compact[5], convex, and includes $\mathcal{Y}$ as a subset.[6]

3.2.2 Preferences

Assumption 8. *The agent's expected utility[7] preferences over the set of lottery-effort pairs* $\Delta(\mathcal{T}) \times \mathbb{R}^+$ *are represented by continuous Bernoulli utility function* $v : \mathcal{T} \times \mathbb{R}^+ \to \mathbb{R}$*, with* $u : \mathcal{T} \to \mathbb{R}$ *and* $k : \mathbb{R}^+ \to \mathbb{R}$ *both strictly increasing:*

$$v(t, e) = u(t) - k(e).$$

We call u the agent's *risk type*, and allow the effort-cost function k to vary with the agent's *type* v. We normalize $u(0) \equiv 0$ and $k(0) \equiv 0$, label the set of all risk types $\mathscr{U}$, and label the set of all concave risk types $\mathscr{U}_A$.

[5]Our requirement that transfers belong to an exogenously-specified compact set $\mathcal{T}$ facilitates our proof of the existence of a solution to the principal's problem if the agent is not necessarily risk-averse.

[6]One natural alternative to the framework presented here is to directly define contracts as continuous maps $T : \mathcal{Y} \to \Delta(\mathcal{T})$. There are no substantive differences between this formulation of the problem and ours. However, our problem is well-defined under upper semicontinuity of w, which is meaningfully more permissive than continuity of T. In particular, binary quota contracts play a part in our discussion, and these do not correspond to any continuous map T. In lieu of developing ad hoc continuity requirements for T, we opt for the state-space formulation presented here.

[7]Expected utility maximization is not essential, and can be replaced with the weaker requirement that preferences respect first- or second- order stochastic dominance.

In this paper, the principal is uncertain about the agent's risk preferences. The set of *feasible risk types* $\mathcal{U}$ and *feasible types* $\mathcal{V}$ are as follows:

Assumption 9. *The agent's risk type belongs to $\mathcal{U}$, which is open in either $\mathscr{U}$ or $\mathscr{U}_A$. The agent's type belongs to $\mathcal{V}$, which includes at least one type v with risk type u if and only if u belongs to $\mathcal{U}$.*

The principal's preferences are as follows:

Assumption 10. *The principal's expected utility preferences over the set of output-transfer lotteries $\Delta(\mathcal{Y} \times \mathcal{T})$ are represented by continuous Bernoulli utility function $\pi : \mathcal{Y} \times \mathcal{T} \to \mathbb{R}$, with π strictly increasing in y and decreasing in t.*

We call π the principal's *objective,* and normalize $\pi(0,0) \equiv 0$. We say that the principal is *financially risk-neutral* if her objective is as follows:

$$\pi(y,t) = y - t.$$

We suppress the joint distribution of (y,x) in our notation and instead work directly with output distributions. Given a contract w, the map $T^w : \mathcal{Y} \to \Delta(\mathcal{T})$ gives the distribution of transfers $w(y,x)$ to the agent when output is y and signal x is distributed according to G. We identify contracts w and w' if $T^w = T^{w'}$. Define $u^w, \pi^w : \mathcal{Y} \to \mathbb{R}$ as follows:

$$u^w(y) := E_G\Big[u\Big(w(y,x)\Big)\Big] \tag{3.1}$$

$$\pi^w(y) := E_G\Big[\pi\Big(y, w(y,x)\Big)\Big]. \tag{3.2}$$

The agent chooses actions that maximize his expected utility. Label as follows:

$$\begin{aligned} c(\mathcal{A}|v,w) &:= \operatorname*{arg\,max}_{(F,e)\in\mathcal{A}} \Big(E_F[u^w(y)] - k(e)\Big) \\ V(\mathcal{A}|v,w) &:= \max_{(F,e)\in\mathcal{A}} \Big(E_F[u^w(y)] - k(e)\Big) \\ V(\mathscr{A}|v,w) &:= \inf_{\mathcal{A}\in\mathscr{A}} V(\mathcal{A}|v,w). \end{aligned} \tag{3.3}$$

As contracts are upper semicontinuous and technologies are compact, the agent's problem has a solution.[8] In order to ensure that the principal's payoff is well-defined, ties are broken against the principal. She evaluates contracts by their worst-case expected payoff against $\mathscr{A}$ and $\mathcal{V}$. Label as follows:

$$\Pi(\mathcal{A}|v,w) := \min_{(F,e)\in c(\mathcal{A}|v,w)} E_F[\pi^w(y)]$$
$$\Pi(\mathscr{A}|v,w) := \inf_{\mathcal{A}\in\mathscr{A}} \Pi(\mathcal{A}|v,w)$$
$$\Pi(\mathscr{A}|\mathcal{V},w) := \inf_{v\in\mathcal{V}} \Pi(\mathscr{A}|v,w).$$

3.2.3 Timing

The timing of the interaction between the principal and the agent is as follows:

1. the principal — knowing $\mathscr{A}, \mathcal{V}$, and G — chooses a contract w;
2. the agent chooses the action $(F,e) \in c(\mathcal{A}|v,w)$ that minimizes $E_F[\pi^w(y)]$;
3. output y and signal x are drawn independently and simultaneously from F and G, respectively;
4. the principal's payoff is $\pi\Big(y, w(y,x)\Big)$ and the agent's payoff $v\Big(w(y,x), e\Big)$.

3.3 Optimal incentives for risk-taking

In this paper, the agent chooses both effort and an output distribution. Moreover, he has separable tastes: his preferences over lotteries do not vary with effort. We interpret the non-effort component of the agent's decision problem as a choice of *risk.* In this section, we restrict our attention entirely to the agent's risk-taking behavior.

3.3.1 Risk alignment

Our first step is to formalize the notion that a contract provides uniformly optimal incentives for risk-taking to the agent:

[8]See lemma 14.

Criterion 1. *Call contract w* risk-aligned *if it has the following property, for all $u \in \mathcal{U}$ and $F, F' \in \Delta(\mathcal{Y})$:*

$$E_F[u^w(y)] \geq E_{F'}[u^w(y)] \iff E_F[\pi^w(y)] \geq E_{F'}[\pi^w(y)].$$

Contracts are risk-aligned when the agent chooses risks as if his objective were to maximize the principal's payoff, regardless of his preferences. Consider for the moment an alternative environment[9] in which the principal *does* know the agent's risk preferences:

Example 4. The agent is risk-neutral: $\mathcal{U} = \{u_{RN}\}$, with $u_{RN}(t) = t$. The principal is financially risk-neutral.

Claim 1. The set of risk-aligned contracts for example 4 is the set of contracts w with $E_G[w(y, x)] = \alpha y + \beta$, for $\alpha \in (0, 1)$ and $\beta \geq 0$.

Proof of claim 1. First, we show that w is risk-aligned if it has the stated form. Second, we show the converse. Suppose that w is as described. Let $F, F' \in \Delta(\mathcal{Y})$. We have the following:

$$\begin{aligned} E_F[u^w(y)] \geq E_{F'}[u^w(y)] &\iff \alpha E_F[y] + \beta \geq \alpha E_{F'}[y] + \beta \\ &\iff (1-\alpha)E_F[y] - \beta \geq (1-\alpha)E_{F'}[y] - \beta. \\ &\iff E_F[\pi^w(y)] \geq E_{F'}[\pi^w(y)]. \end{aligned}$$

Accordingly, w is risk-aligned. Conversely, suppose that w is not as described. Suppose first that one of the following two inequalities holds for $\lambda \in (0, 1)$:

$$E_G\Big[w\Big(\lambda\overline{\mathcal{Y}}, x\Big)\Big] > \Big(1-\lambda\Big)E_G\Big[w\Big(0, x\Big)\Big] + \lambda E_G\Big[w\Big(\overline{\mathcal{Y}}, x\Big)\Big] \tag{3.4}$$

$$E_G\Big[w\Big(\lambda\overline{\mathcal{Y}}, x\Big)\Big] < \Big(1-\lambda\Big)E_G\Big[w\Big(0, x\Big)\Big] + \lambda E_G\Big[w\Big(\overline{\mathcal{Y}}, x\Big)\Big]. \tag{3.5}$$

[9] Example 4 does not satisfy assumption 9. However, our choice of singleton sets of risk types is purposeful, because it allows us to relate our risk alignment criterion to earlier work on robust contracting. We discuss this relationship in section 3.5.

Set $F = \delta_{\lambda\overline{y}}$ and $F' = (1-\lambda)\delta_0 + \lambda\delta_{\overline{y}}$. If inequality 3.4 obtains, we have the following:

$$E_F[u^w(y)] > E_{F'}[u^w(y)] \tag{3.6}$$

$$E_F[\pi^w(y)] < E_{F'}[\pi^w(y)]. \tag{3.7}$$

If inequality 3.5 obtains, the sense of inequalities 3.6 and 3.7 is reversed. In either situation, w is not risk-aligned. Finally, suppose instead that $E_G[w(y,x)] = \alpha y + \beta$, but $\alpha \notin (0,1)$. Set $F = \delta_0$ and $F' = \delta_{\overline{y}}$ to see that w is not risk-aligned. □

In example 4, risk-aligned contracts are affine in expectation. As a special case, this includes linear contracts $w(y,x) = \alpha y$ with share parameter $\alpha \in (0,1)$. Linear contracts are known to be worst-case optimal for contracting environments with uncertain production technologies and two-sided financial risk-neutrality (Chassang (2013); Carroll (2015)).

In terms of the amount of information that the principal has about the agent's risk preferences, example 4 might be viewed as an extreme case. Consider then the alternative extreme:

Example 5. The principal does not have any information about the agent's risk preferences: $\mathcal{U} = \mathscr{U}$.

Claim 2. Let $w(y,x) = W(y)$ for $W : \mathcal{Y} \to \mathcal{T}$ strictly increasing. Given any pair of distinct output distributions $F_A, F_B \in \Delta(\mathcal{Y})$ that are not ordered by first-order stochastic dominance[10], there exist $u_A, u_B \in \mathcal{U}$ such that:

$$E_{F_A}[u^w_A(y)] > E_{F_B}[u^w_A(y)]$$

$$E_{F_B}[u^w_B(y)] > E_{F_A}[u^w_B(y)].$$

Proof. Our proof is closely related to the utility function characterization of first-order stochastic dominance (Hadar and Russell (1969)). Let $T_A \in \Delta(\mathcal{T})$ be the distribution of $W(y)$ when output $y \sim F_A$ and $T_B \in \Delta(\mathcal{T})$ the distribution of $W(y)$ when $y \sim F_B$. As F_A and F_B are not ordered by first-order stochastic dominance and W is strictly increasing, T_A

[10]Recall that F is said to first-order stochastically dominate F' if $F(x) \leq F'(x)$ for all x.

and T_B are also not ordered by first-order stochastic dominance. Thus, there exist $t_A, t_B \in \mathcal{T}$ such that:

$$T_A(t_A) < T_B(t_A)$$
$$T_B(t_B) < T_A(t_B).$$

For $\theta \in \{t_A, t_B\}$, define $u^\theta = \mathbf{1}(y \geq \theta)$. We have the following:

$$E_{T_A}[u^{t_A}(t)] > E_{T_B}[u^{t_A}(t)]$$
$$E_{T_B}[u^{t_B}(t)] > E_{T_A}[u^{t_B}(t)].$$

As there exist sequences of risk types in $\mathcal{U}$ that converge pointwise to u^{t_A} and u^{t_B}, respectively, Lebesgue's dominated convergence theorem yields the desired risk types u_A, u_B. □

If the principal knows the agent's risk preferences, there are many risk-aligned contracts. Conversely, as example 5 suggests, risk alignment is a more demanding criterion if she is instead uncertain. As in our example, if the principal has no information whatsoever about the agent's risk preferences, there are familiar contractual forms under which the agent might choose either one of a pair of undominated output distributions, even if one produces much more (expected) output than the other. This failure of risk alignment has severe payoff consequences for a principal who seeks a contract that performs well even if the agent has extreme risk preferences (Rosenthal (2019)).

However, regardless of the principal's own preferences and the extent of her uncertainty about $\mathcal{U}$, there are always contracts under which the agent chooses risks as the principal would:

Example 6. The principal — who has any objective π that satisfies 10 — knows that the agent's risk preferences belong to $\mathcal{U}$, for any set $\mathcal{U}$ that satisfies assumption 9.

Claim 3. There exists a risk-aligned contract w.

Proof. Because $\pi(y, t)$ is continuous and strictly increasing in y, there exist transfers $b > 0$

such that $\pi(\overline{\mathcal{Y}}, b) > 0$. Choose such a transfer b. Let w be as follows:

$$\lambda(y) = \frac{\pi(y,0)}{\pi(y,0) + \pi(\overline{\mathcal{Y}}, b) - \pi(y,b)}$$
$$T^w(y) = \Big(1 - \lambda(y)\Big)\delta_0 + \lambda(y)\delta_b.$$

Substituting λ directly into our definitions 3.2 and 3.1 for π^w and u^w, respectively, yields:

$$\pi^w(y) = \lambda(y)\pi^w(\overline{\mathcal{Y}}).$$
$$u^w(y) = \lambda(y)u^w(\overline{\mathcal{Y}}).$$

Thus, we have:

$$\pi^w(y) = \frac{\pi^w(\overline{\mathcal{Y}})}{u^w(\overline{\mathcal{Y}})} u^w(y).$$

Because $\pi^w(\overline{\mathcal{Y}}) = \pi(y, b) > 0$ and $u^w(\overline{\mathcal{Y}}) = u(b) > 0$ per our choice of b, π^w is a positive affine transformation of u^w. Moreover, because we chose u arbitrarily, this holds for each $u \in \mathcal{U}$. By the same argument given in the proof of claim 1, w is risk-aligned. $\square$

In the proof of claim 3, we construct a risk-aligned contract w. The features of this contract do not depend on the identity of the set $\mathcal{U}$ in any way, but do depend on the principal's own preferences. Furthermore — unlike in example 4 — it is not in general true that $\lambda(y)$ is proportional to $\pi(y, 0)$: the principal's problem is more complex than simply inducing her own preferences in the agent, because she needs to account for the cost of providing incentives. Instead, the coincidence of the preference-induction problem and the problem of finding a risk-aligned contract is a special feature of additively separable principal objectives, as with the financially risk-neutral preference discussed above.

3.3.2 Payoff affinity

For the remainder of the paper, we devote our attention to the full development of example 6, in which the principal is locally uncertain about the agent's risk preferences. As in the example, the results of this section do not depend in any way on the identity of the set of feasible risk types.

Criterion 2. *Call contract w* payoff affine *if there exist functions $\alpha : \mathcal{U} \to \mathbb{R}^{++}$ and $\beta : \mathcal{U} \to \mathbb{R}$ such that the following relationship holds for all $u \in \mathcal{U}$:*

$$\pi^w(y) = \alpha(u)u^w(y) + \beta(u).$$

In addition to providing motivation for our exercise, example 4 highlights an important relationship between affinity and risk alignment. When the principal and the agent are both risk-neutral, risk-aligned contracts are affine (in expectation) functions of output. Moreover, these contracts are payoff affine, in the sense of criterion 2. This property characterizes risk alignment:

Proposition 9. *Contract w is risk-aligned if and only if w is payoff affine.*

A basic result of economic theory is that utility representations of preferences are unique up to positive affine transformation. If the principal uses a payoff affine contract, the agent's induced utility function $u^w : \mathcal{Y} \to \mathbb{R}$ is a positive affine transformation of the principal's own induced utility function $\pi^w : \mathcal{Y} \to \mathbb{R}$, regardless of his own preferences. Consequently, risk alignment coincides exactly with payoff affinity.

3.3.3 Stochastic binary contracts

Sections 3.3.1 and 3.3.2 identify properties that contracts might possess. In this section, we identify contracts that possess these properties. It is convenient to first exclude a class of pathological contracts. Call a contract w *aligned at extremes* [11] if it satisfies the following condition for all $u \in \mathcal{U}$:

$$\frac{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}{u^w(\overline{\mathcal{Y}}) - u^w(0)} > 0.$$

Contracts that are not aligned at extremes are not risk-aligned, and furthermore do not yield a positive worst-case payoff for the principal.[12] Proceeding, call a continuous and strictly

[11] This property allows for the case in which the principal and the agent mutually prefer low output to high output. Although we demonstrate in lemma 15, appendix C.2, that such contracts are severely sub-optimal, the permissive nature of assumptions 9 and 10 prevents us from ruling out that such contracts are risk-aligned.

[12] Because we work with an arbitrary open set of risk preferences $\mathcal{U}$ and permissive assumptions about π, it is difficult to provide an informative characterization of the set of contracts that satisfy this criterion. However, provided that the high-output reward $T^w(\overline{\mathcal{Y}})$ is not excessively costly for the principal compared to

increasing function $\lambda : \mathcal{Y} \to [0,1]$ a *weight* if it satisfies $\lambda(0) = 0$ and $\lambda(\overline{\mathcal{Y}}) = 1$. Our contracts of interest are as follows:

Definition 10. Call a contract w which is aligned at extremes a *stochastic binary contract* if its transfer distribution T^w is as follows, for weight λ and $T_L, T_H \in \Delta(\mathcal{T})$:

$$T^w(y) = \Big(1 - \lambda(y)\Big)T_L + \lambda(y)T_H.$$

In this section, we prove two primary results. Given a fixed principal objective π, theorem 3 characterizes the set of risk-aligned contracts as the set of stochastic binary contracts with appropriately chosen weights. Given a fixed stochastic binary contract w, theorem 4 identifies a principal objective for which w is risk-aligned.

Theorem 3. *There exists weight λ^* — depending on the principal's objective and rewards T_L, T_H — such that contract w is risk-aligned if and only if w is a stochastic binary contract with weight λ^*.*

Risk alignment is the conjunction of two criteria. First, risk-aligned contracts unify risk-taking behavior across the set of potential agent types. Second, risk-aligned contracts align the agent's risk-taking behavior with the principal's objective. To see why these criteria can only be simultaneously achieved with a stochastic binary contract, note that contracts that condition transfers only on output only satisfy the first criterion if they specify at most two distinct transfer levels. However, contracts that satisfy the second criterion must provide incentives with strength that is strictly increasing in output, and binary contracts that reward the agent only on the basis of his own output evidently do not have this feature. Accordingly, if the principal seeks risk alignment, she must condition transfers on uninformative signals.

The proof of theorem 3 has two steps. In the first step, we show that if the relative payoff assigned to each level of output does not vary with the agent's risk type, then the distribution of transfers to the agent must be a mixture of some fixed pair of rewards T_L and T_H, for each level of output. At the heart of this argument lies the following lemma:

Lemma 4. *All risk types in $\mathcal{U}$ are indifferent between $T, T' \in \Delta(\mathcal{T}) \iff T = T'$.*

the low-output reward $T^w(0)$, it is sufficient that the former first-order stochastically dominate the latter. For example: the contract w with $T^w(\overline{\mathcal{Y}}) = \delta_b$ and $T^w(0) = \delta_0$ is aligned at extremes for b sufficiently small.

Proof. Suppose that $T \neq T'$. The second-order stochastic dominance[13] relation is antisymmetric on $\Delta(\mathcal{T})$. Consequently, the utility function characterization of second-order stochastic dominance $\Big($Rothschild and Stiglitz (1970)$\Big)$ implies that there exists increasing and concave utility function $u^* : \mathcal{T} \to \mathbb{R}$ for which $E_T[u^*(t)] \neq E_{T'}[u^*(t)]$. Let $u \in \mathcal{U}$ and define u_λ as follows, for $\lambda \in (0,1)$:

$$u_\lambda := \lambda u + (1-\lambda)u^*.$$

First, by construction it is not the case for any λ that both of the following equalities hold:

$$E_T[u(t)] = E_{T'}[u(t)]$$
$$E_T[u_\lambda(t)] = E_{T'}[u_\lambda(t)].$$

Second, u_λ is strictly increasing and concave, and thus belongs to $\mathscr{U}_A$. Because $\mathcal{U}$ is open and u_λ converges uniformly to u as $\lambda \to 1$, $u_\lambda \in \mathcal{U}$ for λ sufficiently large. Accordingly, if $T \neq T'$, then there exists $u \in \mathcal{U}$ such that:

$$E_T[u(t)] \neq E_{T'}[u(t)].$$

Conversely, if $T = T'$, it is clear that $E_T[u(t)] = E_{T'}[u(t)]$ for all $u \in \mathcal{U}$. □

The next step of the proof is to show that for each pair of lotteries T_L and T_H that are compatible with our aligned at extremes requirement, there is a unique weight λ^* that aligns the agent's risk-taking behavior with the principal's objective.[14]

Having characterized the set of contracts which are risk-aligned for a *particular* principal objective, we pose the following question: which contracts are risk-aligned for *some* principal objective?

There exists a principal objective that rationalizes any stochastic binary contract as risk-aligned, provided that the high-output reward is preferred to the low-output reward by

[13]Recall that F is said to second-order stochastically dominate F' if $\int_{-\infty}^{x} F(x) \leq \int_{-\infty}^{x} F'(x)$ for all x.

[14]The risk-aligned weight λ^* depends both on the principal's objective π and potentially the rewards T_L, T_H. Although we do not emphasize the identity of this function in discussion of our results, we provide an explicit expression for λ^* in C.1, appendix C.1.

the agent:

Theorem 4. *For each weight λ and pair of rewards T_L, T_H with T_L first-order stochastically dominated by T_H, there exists a principal objective for which the stochastic binary contract with weight λ and rewards T_L, T_H is risk-aligned.*

Whenever the principal uses a stochastic binary contract and the agent chooses between equal-effort actions, the agent's choice of risk reduces to the problem of either maximizing or minimizing the expected value of λ. Consequently, stochastic binary contracts tie down the agent's choice of actions, conditional on effort. Thus, all such contracts align the agent's risk-taking behavior to some principal objective.[15]

3.4 Optimal incentives for effort and risk-taking

Theorem 3 characterizes the set of risk-aligned contracts for a broad class of principal objectives. With this characterization in hand, we study a moral hazard problem in which risk-aligned contracts are optimal. We maintain the following assumption for the remainder of our analysis:

Assumption 11. *The principal is financially risk-neutral.*

In this section, we identify contracts w that provide the largest-possible *guarantee* $\Pi(\mathscr{A}|\mathcal{V}, w)$ for the principal. That is, we analyze the following maximization problem:

$$\max_{w} \Pi(\mathscr{A}|\mathcal{V}, w). \tag{3.8}$$

The principal's problem 3.8 has a solution, and there are *gains from trade*[16]:

Assumption 12. *For some fixed share $s < \frac{1}{\mathcal{Y}}$ and fixed lottery $T \in \Delta(\mathcal{T})$ with $E_T[t] < \overline{\mathcal{Y}}$, for each $v \in \mathcal{V}$ the following inequality is satisfied for some $e(v) \in \mathcal{E}$:*

$$\frac{k\Big(e(v)\Big)}{\mu\Big(e(v)\Big)} \leq sE_T[u(t)].$$

[15] In constructing the desired principal objective, we hold the set of feasible risk types $\mathcal{U}$ fixed.

[16] See lemma 20 for an existence proof. Although our statement is admittedly more complicated than we would like, see lemma 18 for proof that there are gains from trade if and only if assumption 12 is satisfied.

We emphasize that while the effort level $e(v)$ in the statement of assumption 12 is allowed to vary with the agent's type v, the share s and reference lottery T are fixed. Our results are strongest when the following restriction to $\mathcal{U}$ is satisfied:

Definition 11. Call the agent's preferences *regular* if:

1. the set of feasible risk types $\mathcal{U} \subset \mathscr{U}_A$; and

2. the set of feasible risk types $\mathcal{U}$ includes a risk-neutral type; and

3. if u_n converges to u in $\mathcal{U}$, then v_n converges to v in $\mathcal{V}$.[17]

We say that the agent has regular preferences when he is known by the principal to be risk-averse, the principal does not rule out the possibility that the agent is risk-neutral, and the agent's effort disutility varies continuously with his risk preferences. We establish uniqueness results in the form of *typewise dominance* when the agent has regular preferences:

Definition 12. Contract w' *typewise dominates* contract w if $\Pi(\mathscr{A}|v,w) \leq \Pi(\mathscr{A}|v,w')$ for all $v \in \mathcal{V}$, with strict inequality for some v'.

The first two assumptions presented in definition 11 are substantive, and we discuss their role in our dominance results below. The third assumption is technical.[18]

3.4.1 Output affine contracts

In section 3.3, we define contracts to be payoff affine if they induce an affine relationship between the principal's payoff and the agent's payoff, for each type of agent. In proving the results of this section, we make use of a related criterion:

Criterion 3. *Say that w is* output affine *if there exist functions $a, b : \mathcal{U} \to \mathbb{R}$ and constants $a(\pi)$, $b(\pi)$ such that the following equalities hold, where $a(u)$ and $a(\pi)$ are all strictly positive*

[17] Here, each v_n has risk type v_n and type v has risk type u.

[18] The continuity assumption constituent to definition 11 serves to rule out uninteresting corner cases in which contracts that do not guarantee the principal a positive payoff might nevertheless remain undominated.

or all strictly negative:

$$u^w(y) = a(u)y + b(u)$$
$$\pi^w(y) = a(\pi)y + b(\pi).$$

Output affine contracts are those for which the principal's payoff and the agent's payoff are affine in output, for each type of agent. For general principal objectives, output affine contracts need not exist. However, if the principal is financially risk-neutral, they do exist, and output affinity is equivalent to payoff affinity:

Proposition 10. *Contract w is payoff affine if and only if w is output affine.*

3.4.2 Stochastic bonus contracts

As in section 3.3, the contracts of interest in this section are stochastic binary contracts. There are many such contracts, and all are risk-aligned when their weight λ coincides with the risk-aligned weight λ^*. Substitution of the principal's objective $\pi(y,t) = y - t$ into equation C.1 yields the following:

$$\lambda_{RN} := \lambda^*(y) = \frac{y}{\overline{\mathcal{Y}}}. \tag{3.9}$$

Although risk-neutrality on the part of the principal yields an especially simple form for the risk-aligned weight λ_{RN}, the rewards T_L, T_H associated with risk-aligned contracts may be quite complicated: our only requirement is that they are compatible with our aligned at extremes criterion, which is permissive. However, the set of stochastic binary contracts admits a natural refinement:

Definition 13. Call contract w a *stochastic bonus contract* if its transfer distribution T^w is as follows, for weight λ and $b \in \mathcal{T}$ with $b > 0$:

$$T^w(y) = \Big(1 - \lambda(y)\Big)\delta_0 + \lambda(y)\delta_b.$$

In addition to their simplicity, stochastic bonus contracts are naturally compatible with risk aversion on the part of the agent. We state the main result of this section of the paper:

Theorem 5. *There exists a risk-aligned stochastic binary contract that maximizes the principal's payoff. If the agent has regular preferences, then there exists a risk-aligned stochastic bonus contract that maximizes the principal's payoff, and all typewise undominated contracts are risk-aligned stochastic bonus contracts.*

Theorem 5 summarizes two results. First, we show that stochastic binary contracts are optimal. Second, we show that if the agent has regular preferences, stochastic bonus contracts typewise dominate other types of contracts. The first step of the proof is to improve an arbitrary contract to a stochastic binary contract:

Proposition 11. *If w is not a risk-aligned stochastic binary contract, then there exists a risk-aligned stochastic binary contract w' such that for all $v \in \mathcal{V}$:*

$$\Pi(\mathscr{A}|v, w) \leq \Pi(\mathscr{A}|v, w').$$

The proof of proposition 11 has two steps. First, we compute an upper bound for the guarantee provided by an arbitrary contract w, as a function of the transfer distributions $T^w(0)$ and $T^w(\overline{\mathcal{Y}})$. Second, we show that because stochastic binary contracts with appropriately chosen weights are risk-aligned, this bound is tight for such contracts.

Although proposition 11 yields a strong form of optimality, there are a multiplicity of risk-aligned stochastic binary contracts. However, all of these contracts that fail definition 13 impose more risk on the agent than equally generous stochastic bonus contracts. Consequently, if the agent has regular preferences, typewise undominated contracts are stochastic bonus contracts:

Proposition 12. *If w is not a risk-aligned stochastic bonus contract and the agent has regular preferences, then there exists a risk-aligned stochastic bonus contract w' that typewise-dominates w.*

The proof of proposition 12 has two parts. In the first part, we strengthen proposition 11 by showing that if the agent has regular preferences, contracts that are not risk-aligned stochastic binary contracts are typewise-dominated. To see the role of the risk-neutral type in the proof, let u_{RN} be the risk type with $u_{RN}(t) = t$ for all $t \in \mathcal{T}$. The following relationship holds for *all* contracts w:

$$\pi^w(y) = y - u^w_{RN}(y).$$

Regardless of the form of the contract w, the principal faces a tradeoff between her own payoff π^w and the agent's payoff u^w_{RN}. This need not be the case when the agent is not risk-neutral.

In the second part of the argument, we refine the set of undominated contracts to those with degenerate rewards. This has two steps. First, we show that optimal contracts do not reward the agent for being unproductive, because they satisfy $T^w(0) = \delta_0$. The interpretation of this result is straightforward: contracts that pay the agent when he produces no output provide weaker incentives for effort and are more costly for the principal. Second, we show that contracts with $T^w(\overline{\mathcal{Y}}) = \delta_b$ provide stronger incentives for effort to risk-averse agents than their equally-generous counterparts, at no additional cost to the principal. To see why, suppose that w is a stochastic binary contract with $T_L = \delta_0$ and $T_H = T$, for non-degenerate $T \in \Delta(\mathcal{T})$. Let w' be the stochastic bonus contract with $b = E_{T_H}[t]$, and suppose risk type u is strictly risk-averse. We have the following, for all $F \in \Delta(\mathcal{Y})$ with F distinct from δ_0:

$$E_F[u^{w'}(y)] - E_{\delta_0}[u^{w'}(y)] > E_F[u^w(y)] - E_{\delta_0}[u^w(y)]$$
$$E_F[\pi^{w'}(y)] = E_F[\pi^w(y)].$$

In addition to providing stronger incentives for effort than their more complicated counterparts, stochastic bonus contracts provide the principal with an equally large payoff for any fixed action.

3.4.3 Screening

In our paper, the agent is privately informed about his preferences and production technology. This suggests that the principal might benefit from screening the agent. Suppose that the agent observes both his type v and his technology $\mathcal{A}$ before contracting with the principal, and the principal offers the agent a nonempty menu of contracts $\mathcal{W}$.

Under what circumstances might the principal benefit at all from screening the agent? Only when there are no types of agents that are uniformly less willing to exert effort than other types of agents. In particular, suppose that the agent has *quasilinear preferences:* that is, each $v \in \mathcal{V}$ has the following form:

$$v(t, e) = u(t) - \gamma e,$$

where the constant $\gamma > 0$ is chosen so that $\mathcal{V}$ satisfies our gains from trade assumption 12. Given a set of quasilinear agent types $\mathcal{V}$, call $v(t, e) = u(t) - \gamma e$ a *low-effort type* if the following condition is satisfied for all $v' \in \mathcal{V}$ with $v'(t, e) = u'(t) - \gamma' e$:

$$\frac{u(t)}{\gamma} \leq \frac{u'(t)}{\gamma'} \forall t \in \mathcal{T}.$$

If $\mathcal{V}$ includes a low-effort type, then the principal does not benefit from screening the agent:

Proposition 13. *Suppose that $\mathcal{V}$ includes a low-effort type v. Then the principal's payoff when she screens the agent is no larger than when she uses an optimal contract.*

More generally, if these conditions are not satisfied, the principal does at least as well to screen the agent with a menu consisting exclusively of risk-aligned stochastic binary contracts.

Proposition 14. *For each menu of contracts, there exists a menu of risk-aligned stochastic binary contracts that provides a larger guarantee.*

Consequently, our model retains explanatory power for stochastic binary contracts under screening.

3.5 Conclusions

This paper presents a model of moral hazard in which the principal is uncertain about the agent's risk preferences and production technology. We characterize the set of risk-aligned contracts that provide the agent with incentives to choose risks as if he shared the principal's objectives, regardless of his own preferences. Very generally, such contracts condition transfers to the agent on variables that do not directly affect the principal's payoff. When the agent is known to be either risk-neutral or risk-averse, undominated optimal contracts are stochastic bonus contracts with only two transfer levels.

Why might the principal be concerned with risk alignment? As suggested by example 5, when the principal knows little about the agent's risk preferences, certain types of contracts provide him with incentives to choose risks that are *very* undesirable from her perspective. In particular, if she only knows that the agent is risk-averse but does not know bounds to his risk aversion, the agent might choose actions that are — up to approximation — no better for the principal than when he shirks, regardless of how much effort he is willing to exert. This topic is explored in depth in Rosenthal (2019). We take this result as motivation for the exercise undertaken in the present paper, wherein we show how to align the agent's behavior with the principals objective, but place relatively little emphasis on developing the *extent* of the consequences for the principal when she uses an unaligned contract.

Our emphasis on alignment mechanisms is shared with Frankel (2014), who considers a transfer-free mechanism design problem in which a principal delegates multiple decisions to an agent with unknown preferences. Here, the author calls a mechanism an *aligned delegation* mechanism if all types of agents choose actions as the principal would, were she given his information about the state of the world. As in the author's aligned delegation mechanisms, our risk-aligned contracts exploit commonly-held preference monotonicities to provide uniformly-interpreted incentives to agents with heterogeneous tastes. While ranking mechanisms achieve aligned delegation in Frankel's environment because all agents prefer higher actions in higher states, stochastic binary contracts achieve risk alignment in our environment because all agent preferences respect first- or second- order stochastic dominance.

Risk alignment is also related in important ways to Carroll (2015), who studies a model of moral hazard in which the principal is uncertain about the agent's production technology[19] and there is two-sided risk-neutrality. Here, the author presents an intuition for the worst-case optimality of linear contracts that is conceptually very similar to our risk alignment criterion. Furthermore, in an unpublished version of the paper, Carroll includes an analysis of the optimality of *utility-affine* contracts if the principal and the agent are not necessarily

[19]Although assumption 7 is arguably the simplest assumption under which the principal's problem is interesting to study, it is admittedly the case that the principal in Carroll (2015) has strictly more information about the agent's production technology than in our model. In Carroll's model, the principal knows a non-empty lower bound $\mathcal{A}_0$ to the agent's technology $\mathcal{A}$, in the set inclusion order. The results of section 3.3 of our paper obtain in Carroll's environment, as do the results of section 3.4 under restrictions to the relationship between risk preferences and effort costs. These arguments are not presented here.

financially risk-neutral.[20] Linear contracts — and utility-affine contracts more broadly — are risk-aligned for the appropriate preferences if we generalize assumption 9 to include the case in which $\mathcal{U}$ contains a single risk type, as in example 4. Moreover, as the author emphasizes, the optimality of linear contracts in his environment is driven by the alignment that they achieve between the principal's interests and the agent's payoff, which is precisely the content of our own risk alignment criterion.

[20] Version December 21, 2012. Accessed online at http://www.bu.edu/econ/files/2013/03/May-4-Caroll.pdf.

APPENDIX A

Appendix to Chapter One

This appendix is organized as follows. First, section A.1 provides technical results not stated in the body of the paper. Second, section A.2 treats the monotone environment. Third, section A.3 provides omitted proofs for section 1.6. Finally, section A.4 provides an explicit solution for the homogeneous environment.

A.1 Omitted technical material, section 1.4

The following technical lemma supports our proof of theorem 1:

Lemma 5. *Given any mechanism M and type distribution $\mu \in \Delta_0$, there exists discrete type distribution $\mu^D \in \Delta_0$ such that:*

$$\pi(M|\mu^D) \leq \pi(M|\mu).$$

Proof. Let M and μ be as described. For each $b \in \mathcal{B}$, there exists $u_b \in U(b)$ such that:

$$\pi\Big(y(u_b), q(u_b)\Big)\mu\Big(U(b)\Big) \leq \int_{U(b)} \pi\Big(y(u), q(u)\Big)\mathrm{d}\mu.$$

Define μ^D as follows:

$$\mu^D(u) = \begin{cases} \mu\Big(U(b)\Big) & u = u_b \\ 0 & else. \end{cases}$$

We have $\pi(M|\mu^D) \leq \pi(M|\mu)$, as desired. □

Next, we provide a proof of the existence of an optimal mechanism, under some technical hypotheses not imposed in the body of the paper:

Lemma 6. *Suppose that the set of characteristics X is finite and the set of feasible types U is compact. If the set of feasible types is $C-$rich, then there exists an optimal $C-$mechanism.*

Proof of lemma 6. First, we show that the principal's payoff $\pi(M)$ has a finite supremum. Because the set of feasible types U is uniformly bounded, there exists a finite price P such that all individually rational mechanisms include only allocations (x, p) with $p < P$. Accordingly, $\sup \pi(M) < \infty$.

Second, let M_n be a sequence of mechanisms with guarantees that approach this supremum. Without loss of generality, suppose that each M_n breaks ties in favor of the principal, for every type of agent. Note that this is a well-defined criterion, because each M_n includes only finitely many distinct allocations. It will be convenient to interpret mechanisms as sets of allocations rather than as direct mechanisms.

Proceeding, because X is finite, there are only finitely many subsets of X. Consequently, there exists a fixed set $A \subset X$ and subsequence n_k of n such that the set of distinct characteristic vectors allocated to some type of agent is exactly A for each M_{n_k}. Thus, each M_{n_k} can be expressed as follows:

$$M_{n_k} = \Big\{ \Big(x, p_{n_k}(x) \Big) | x \in A \Big\},$$

where $p_{n_k} : A \to \mathbb{R}_+$ gives the price assigned to characteristic $x \in A$ by mechanism M_{n_k}. Furthermore, because the sequence p_{n_k} is uniformly bounded and has a finite domain, p_{n_k} converges uniformly to some $p^\infty : A \to \mathbb{R}_+$. Define the mechanism M^∞ as follows, with ties again broken in favor of the principal:

$$M^\infty = \Big\{ \Big(x, p^\infty(x) \Big) | x \in A \Big\}.$$

We claim that $\pi(M^\infty) \geq \lim \pi(M_{n_k})$. To see why, let $b \in \mathcal{B}$, and consider the cell of agent types $U(b)$. Because M^∞ includes only finitely many distinct allocations, there are consequently only finitely many allocations assigned by M^∞ to types in $U(b)$. Let $u \in U(b)$. Denote by $\Big(x, p^\infty(x) \Big)$ his allocation under M^∞, and similarly $\Big(x_{n_k}, p_{n_k}(x_{n_k}) \Big)$ his allocation under M_{n_k}. We claim the following:

$$\liminf_k \pi\Big(x_{n_k}, p_{n_k}(x_{n_k}) \Big) \leq \pi\Big(x, p^\infty(x) \Big). \tag{A.1}$$

To see why, note that because A is finite, there exists subsequence n_{k_j} of n_k such that $x_{n_{k_j}} = x'$ for some $x' \in A$ for all j. Moreover, we have the following:

$$u(x') - p_{n_{k_j}}(x') \geq u(x) - p_{n_{k_j}}(x)\, \forall k \implies u(x') - p^{\infty}(x') \geq u(x) - p^{\infty}(x).$$

Consequently, $\big(x', p^{\infty}(x')\big)$ is individually rational and incentive compatible for u, under mechanism M^{∞}. Finally, we have the following limit:

$$\pi\Big(x', p_{n_{k_j}}(x')\Big) \to \pi\Big(x', p^{\infty}(x')\Big).$$

Thus, because M^{∞} breaks ties in favor of the principal, inequality A.1 obtains. Assembling the pieces, we have established that for any allocation assigned to some type $u \in U(b)$ by M^{∞}, we can find a sequence of allocations assigned to u by $M_{n_{k_j}}$ that do no better (in the limit) for the principal than the M^{∞} allocation. Because there are only finitely many distinct allocations assigned to each cell of agents, and only finitely many cells of agents, this is sufficient to establish that $\pi(M^{\infty}) \geq \lim \pi(M_{n_k}) = \sup \pi(M)$. Having established the existence of an optimal mechanism, the existence of an optimal update mechanism follows directly from theorem 1. $\square$

A.2 Omitted technical material, section 1.5

A.2.1 Subsection 1.5.1

Proof of proposition 1. In light of theorem 1, it is sufficient to show that U_0 is X_0-rich. Proceeding, let $u \in U_0$: i.e., suppose that $u(0) = 0$, that u is upper semicontinuous, and that u is increasing. Let u_{X_0} be the corresponding dual type with respect to the set of reference allocations X_0. We claim that $u_{X_0} \in U_0$.

First, because 0 is the minimal element of X, we have that $u_{X_0}(0) = 0$, as required. Second, we claim that u_{X_0} is increasing. Suppose that $x \geq y$, and let $z \in X_0$. If $z \leq y$, then the transitivity of $\geq$ implies directly that $z \leq x$. Thus, by definition, u_{X_0} is increasing. Finally, we claim that u_{X_0} is upper semicontinuous. This follows from the upper semicontinuity of u and that X_0 is closed. Accordingly, $u_{X_0} \in U_0$. Thus, U_0 is X_0-rich. $\square$

A.2.2 Subsection 1.5.3

It will be useful to first make an intermediate refinement to the set of potentially optimal mechanisms for the monotone environment. Accordingly, define as follows:

Definition 14. Given X_0-mechanism M, we say that M has *empirical prices* if for each $(x,p) \in M$ there exists a sequence of allocations $(x_1,p_1),...,(x_J,p_J) \in M$ and a sequence of choice lists $b_1,...,b_J$ such that:

1. $(x_1,p_1) = (x,p)$;
2. $(x_J,p_J) = (0,0)$;
3. $(x_1,p_1) \succeq_{b_1} ... \succeq_{b_J} (x_J,p_J)$.

We say that a mechanism M uses empirical prices if each allocation is priced in order to be consistent with the inferences that the principal is able to make from her choice data. There is no loss of optimality to restrict attention to mechanisms that use empirical prices:

Lemma 7. *Suppose that M is an X_0-mechanism that assigns only positive-payoff allocations to the agent. Then there exists an X_0-mechanism M' with empirical prices that guarantees the principal a payoff that is at least as large as M:*

$$\pi(M) \leq \pi(M').$$

Proof of lemma 7. Let M be as described, and fix choice list $b \in \mathcal{B}$. Define M^+ as the set of allocations $(x,p) \in M$ that satisfy the criteria of definition 14, and as M^- the set of allocations in M which do not. How does excluding the allocations in M^- from M change the choice set $C_b(M)$? We claim that one of the two following conditions must hold:

1. $C_b(M^+) \subset C_b(M)$;
2. $(0,0) \in C_b(M)$.

To see why, suppose that the choice set $C_b(M^+)$ contains some allocation (x,p) that is not contained in $C_b(M)$. By definition of M^+, there exists a sequence $(x,p)_j$ in M satisfying the criteria of definition 14, with $(x,p)_1 = (x,p)$. Furthermore, if $(x,p) \notin C_b(M)$, then there exists $(y,q) \in M^-$ such that:

$$(y,q) \succeq_b (x,p).$$

Combining these two facts yields the existence of a sequence of allocations $(y,q)_j$ in M satisfying the criteria of definition 14, with $(y,q)_1 = (y,q)$. This is contrary to our hypothesis that $(y,q) \in M^-$. Thus, we conclude that one of the two conditions above holds. Given our hypothesis that M assigns only profitable allocations, this is sufficient to establish our claim. □

It is possible to refine lemma 7 considerably. To see why, note that while mechanisms that use empirical prices use prices that lie within some constraints that are implied by the data, this criterion says nothing about whether or not these constraints are tight. This is the content of lemma 3, which we state in the body of the paper.

Proof of lemma 3. In light of proposition 1 and lemma 7, we restrict attention to the case in which M_0 is an X_0 mechanism with empirical prices. Our approach will be to consider a constrained maximization problem that is related to the problem of maximizing the principal's guarantee, and show that the solution to this maximization problem is a reference mechanism M. We then verify that this mechanism M indeed guarantees the principal a larger payoff than the original mechanism M. Define as follows:

$$X^{M_0} := \Big\{x | (x,p_0) \in M_0 \text{ for some } p_0\Big\}.$$

The set $X^{M_0} \subset X_0$ is the set of characteristics that appear in the mechanism M_0. For the entirety of the proof, we restrict attention to mechanisms M with the property that $(x,p) \in M$ for some price p if and only if $x \in X^{M_0}$. Thus, we restrict attention to mechanisms M that allocate exactly the same characteristics to the agent as does the mechanism M_0, but at potentially different prices than the latter.

It will be convenient in places to interpret mechanisms as a price lists $p : X^{M_0} \to \mathbb{R}_+$, rather than as sets of allocations $M \subset X^{M_0} \times \mathbb{R}_+$.We reserve the notation p_0 for the price vector that corresponds to mechanism M_0, so that $\big(x, p_0(x)\big) \in M_0$ for each $x \in X^{M_0}$.

Proceeding, for each pair of characteristics $x, y \in X^{M_0}$ and each price list $p : X^{M_0} \to \mathbb{R}_+$, there are potentially between 0 and $|\mathcal{B}|$ choice lists $b \in \mathcal{B}$ for which the following binary relation is satisfied:

$$\big(x, p(x)\big) \succeq_b \big(y, p(y)\big). \tag{A.2}$$

Equivalently, $p(x)$ and $p(y)$ satisfy the following inequality:

$$V_b(x, y) \geq p(x) - p(y). \tag{A.3}$$

Given a pair of allocations $x, y \in X^{M_0}$ and a choice list $b \in \mathcal{B}$, we label inequality A.3 as $[x \to y|b]$. Finally, we label the set of all such inequalities that are satisfied by the given price list p_0 as $\mathcal{C}$. We interpret these inequalities as constraints, and study the set $\mathcal{P}$ of price lists p that satisfy each constraint in $\mathcal{C}$. In particular, we claim that $\mathcal{P}$ has a maximum element in the product order:

$$\exists p^* \in \mathcal{P} : p^*(x) \geq p(x)\, \forall x \in X^{M_0}\, \forall p \in \mathcal{P}.$$

Establishing the existence of this maximum price list is straightforward. First, note that while increasing the price $p(x)$ assigned to characteristic x tightens each constraint $[z \to y|b] \in \mathcal{C}$ with $z = x$, it either relaxes or does not affect every constraint $[z \to y|b] \in \mathcal{C}$ with $z \neq x$. Consequently, if price lists p, p' belong to the set $\mathcal{P}$, then so does their pointwise-maximum. Second, by construction the set $\mathcal{P}$ is compact. These two facts establish the existence of a maximal element p^*, as desired.

There are two remaining details to our argument. First, we need to establish that the mechanism M^* that corresponds to the price list p^* provides a larger guarantee than M, as promised. To see why this is true, note that by construction M^* satisfies every constraint in $\mathcal{C}$. This is sufficient to establish the following:

$$C_b(M^*) \subset C_b(M_0)\, \forall b \in \mathcal{B}.$$

Because $p^*(x) \geq p_0(x)$, this in turn establishes the following:

$$\Pi_b(M^*) \geq \Pi_b(M_0) \forall b \in \mathcal{B}.$$

This establishes that M^* provides a larger guarantee than M_0, as desired. Finally, we are left to show that M^* is indeed a reference mechanism. To see why this holds, note that because p^* is the maximum of $\mathcal{P}$, for every characteristic $x \in X^{M_0}$ there exists a characteristic $y \in X^{M_0}$ and a choice list $b \in \mathcal{B}$ such that the constraint $[x \to y|b]$ both belongs to $\mathcal{C}$ and is binding. This is sufficient to establish that M^* is a reference mechanism. □

A.2.3 Subsection 1.5.4

Proof of theorem 2. Suppose that M is an X_0-mechanism, and let M break ties in favor of the principal. The substance of the proof lies in establishing that the map Π indeed gives the correct payoff for an X_0-mechanism M. That is, we first demonstrate that the following equality holds for M:

$$\pi(M) = \Pi(M).$$

Fix choice list $b \in \mathcal{B}$. Because $U = U_0$ includes *every* utility function $u : X \to \mathbb{R}_+$ that is consistent with monotonicity, there exists a type $u \in U(b)$ that corresponds to any choice pattern that is consistent with both the choice list b and the monotonicity requirement.[1] Because the set $C_b(M)$ excludes exactly the allocations that are inconsistent with the data and monotonicity, every allocation $(x,p) \in C_b(M)$ is assigned by M to some type of agent $u \in U(b)$.

As a final detail, there are knife-edge circumstances in which it might be incentive compatible to assign allocations $(y,q) \notin C_b(M)$ to agent types $u \in C_b(M)$, in cases of indifference. Our hypothesis that M breaks ties in favor of the principal ensures that the principal's payoff is unaffected by this complication. To see why, note that if $\pi(y,q) \geq \Pi_b(M)$, then doing so does not affect the principal's *worst-case* payoff. If alternatively it is true that $\pi(y,q) < \Pi_b(M)$, then M breaks ties against the allocation (y,q). Thus, $\pi(M) = \Pi(M)$, as desired.

Finally, lemma 3 justifies restricting our search for an optimal mechanism to reference mechanisms. Because there are only finitely many reference mechanisms, the existence of a reference mechanism that maximizes the principal's payoff is immediate. This completes the

[1] Because we work with X_0-mechanisms — i.e., mechanisms with only finitely many distinct allocations — our requirement that u be upper semicontinuous does not complicate our argument.

proof. □

A.3 Omitted technical material, section 1.6

Proof of proposition 2. In light of theorem 1, it is sufficient to show that $U^{\updownarrow}$ is $X^{\updownarrow}-$rich. Proceeding, let $u \in U^{\updownarrow}$: i.e., suppose that $u(0) = 0$, that u is increasing, that u is homogeneous degree t for some $t \in T$, and that u is upper semicontinuous. We write $u_{\updownarrow}$ as shorthand for the dual type $u_{X^{\updownarrow}}$ with respect to the set of allocations $X^{\updownarrow}$. We claim that $u_{\updownarrow}$ belongs to $U^{\updownarrow}$.

First, we claim that $u_{\updownarrow} \in U_0$. That $u(0) = 0$ and that $u_{\updownarrow}$ is increasing is inherited directly from u; see the proof of proposition 1 for details. Furthermore, because u is upper semicontinuous and $X^{\updownarrow}$ is closed, $u_{\updownarrow}$ is upper semicontinuous. Thus, $u_{\updownarrow} \in U_0$.

Second, we claim that $u_{\updownarrow}$ is homogeneous of the same degree t as the parent type u. Thus, we show that for each $x \in X$ and each $\lambda \geq 0$ with $\lambda x \in X$, we have the following:

$$u_{\updownarrow}(\lambda x) = \lambda^t u_{\updownarrow}(x). \tag{A.4}$$

Proceeding, choose $x \in X$ and $\lambda \geq 0$ with $\lambda x \in X$. There are two cases to consider. Suppose first that $x \in X^{\updownarrow}$. By construction, $\lambda x \in X^{\updownarrow}$. Because $u = u_{\updownarrow}$ on $X^{\updownarrow}$, equality A.4 follows immediately. Suppose alternatively that $x \notin X^{\updownarrow}$. The following equality obtains straightforwardly from e.g. argument by contradiction:

$$\phi^u_{X^{\updownarrow}}(\lambda x) = \lambda \phi^u_{X^{\updownarrow}}(x).$$

This is sufficient to establish that equality A.4 again holds, as desired. Accordingly, $u_{\updownarrow}$ is homogeneous degree t. Because we have already argued that $u_{\updownarrow} \in U_0$, this is sufficient to establish that $u_{\updownarrow} \in U^{\updownarrow}$. Thus, $U^{\updownarrow}$ is $X^{\updownarrow}-$rich. □

Proof of proposition 3. In light of theorem 1, it is sufficient to show that $U^{\downarrow}$ is $X^{\downarrow}-$rich. Proceeding, let $u \in U^{\downarrow}$: i.e., suppose that $u \in U_0$ and that u exhibits diminishing returns, . We write $u_{\downarrow}$ as shorthand for the dual type $u_{X^{\downarrow}}$ with respect to the set of allocations $X^{\downarrow}$.

We claim that $u_{\downarrow}$ belongs to $U^{\downarrow}$. As in the proof of proposition 2, it is straightforward to establish that $u_{\downarrow} \in U_0$. More substantively, we claim that $u^{\downarrow}$ suffers diminishing returns. Let

$x \in X$ and let $\lambda > 1$ with $\lambda x \in X$. We claim the following:

$$u_{\downarrow}(\lambda x) \leq \lambda u_{\downarrow}(x).$$

To see why, note that by definition of $X^{\downarrow}$, there exist characteristic $x_0 \in X_0$ and scalar $\lambda_0 \leq 1$ such that:

$$u_{\downarrow}(\lambda x) = u(\lambda_0 x_0).$$

Moreover, ${}^{\lambda_0}/_{\lambda} x_0 \in X^{\downarrow}$. Consequently, by definition we have the following:

$$\begin{aligned} u_{\downarrow}(x) &\geq u_{\downarrow}\Big(\frac{\lambda_0}{\lambda}x_0\Big) \\ &= u\Big(\frac{\lambda_0}{\lambda}x_0\Big). \end{aligned}$$

Because $\lambda \geq 1$ and u exhibits diminishing returns, this is sufficient to establish that $u_{\downarrow}(\lambda x) \leq \lambda u^{\downarrow}(x)$. Thus, $u_{\downarrow}$ exhibits diminishing returns. Accordingly, $u_{\downarrow} \in U^{\downarrow}$. We conclude that $U^{\downarrow}$ is $X^{\downarrow}-$rich, as desired. □

A.4 Explicit solution for special case of the homogeneous environment

Suppose that the principal is a monopolist who sells complex products to a consumer with unit demand. Each product has N quality dimensions, for N finite, and the set of products $X = \mathbb{R}^N_+$. The monopolist's marginal costs for each quality dimension are constant, and the agent's utility is homogeneous degree t for $t < 1$ known to the monopolist:

$$U = \Big\{u : X \to \mathbb{R}_+ | u \text{ is increasing,} u(0) = 0, \text{ and } u \text{ is homogeneous degree } t\Big\}.$$

The monopolist observes choices from only one set of products. Moreover, these products are strongly horizontally differentiated, prices are strictly positive, and more expensive products have a larger markup than their cheaper counterparts.[2]

[2]By strongly horizontally differentiated, we mean that for each allocation $(x, p), (x', p') \in \mathcal{B}$, there exist quality dimensions i, i' such that $x(i), x'(i') > 0$ and $x(i'), x'(i) = 0$, where we write $x(i)$ for the i^{th} quality

Proposition 15. *There exists an optimal mechanism with only as many distinct allocations as there are contained in the data set $\mathcal{B}$. These allocations are as follows, for each $(x, p) \in \mathcal{B}$:*

$$\lambda = \left(\frac{p}{tc(x)}\right)^{\frac{1}{1-t}}$$
$$(y, q) = \Big(\lambda x, \lambda^t p\Big).$$

Proof. Proposition 2 justifies restricting our search for an optimal mechanism to a search for a mechanism that proportionally scales each quality component. We proceed by first maximizing the principal's worst-case payoff in a relaxed program without incentive compatibility constraints. Second, we show that the resulting mechanism yields the same payoff in the full program as it does in the relaxed program. In doing so, we establish that this mechanism is optimal for the principal.

Consider the following maximization problem:

$$\max_{y,q}\Big\{\pi(y, q)|(y, q) \text{ is individually rational for each } u \in U(b)\Big\}. \tag{A.5}$$

For each $b = (x, p) \in \mathcal{B}$, define as follows:

$$u_b(z) := \Big(\max\{r|z \geq rx\}\Big)^t p.$$

Because $x, \widehat{x}$ are strongly horizontally differentiated for each pair $x, \widehat{x}$, type u_x satisfies the following equality for each $z \in X_0$:

$$u_b(z) = \begin{cases} p & z = x \\ 0 & z \neq x. \end{cases}$$

Consequently, $u_b \in U(b)$. Moreover, individual rationality demands that $u(x) \geq u_b(x)$ for each $u \in U(b)$. Thus, individual rationality of (y, q) for u_b is necessary and sufficient for individual rationality for *all* $u \in U(b)$. Accordingly, the problem presented in A.5 is equivalent to the

dimension of x. By markup, we mean the quantity $p/c(x)$.

following relaxed problem:

$$\max_{y,q}\Big\{\pi(y,q)|(y,q) \text{ is individually rational for } u_b\Big\}.$$

This problem is equivalent to the following:

$$\max_{\{\lambda|\lambda x\in X\}} \lambda^t p - \lambda c(x)$$

Necessary and sufficient conditions for a solution λ are as follows:

$$\lambda = \lambda_b := \left(\frac{p}{tc(x)}\right)^{\frac{1}{1-t}}.$$

The corresponding allocation (y,q):

$$y = y_b := \lambda_b x$$
$$q = q_b := \lambda_b^t p.$$

By hypothesis, λ_b is strictly increasing in p. A direct verification reveals that for $u \in U(b)$ and $b' = (x', p') \in \mathcal{B}$ with $p > p'$, we have:

$$u(y_b) - q_b \geq u(y_{b'}) - q_{b'}. \tag{A.6}$$

Moreover, if $p > p'$, a direct calculation reveals that:

$$\pi(y_b, q_b) > \pi(y_{b'}, q_{b'}). \tag{A.7}$$

Assembling the pieces, let M be the mechanism that assigns allocations (x_b, q_b) to each $u \in U(b)$, where it is incentive compatible to do so. Inequalities A.6 and A.7 jointly imply that if $(y_{\widehat{b}}, q_{\widehat{b}})$ is incentive compatible for $u \in U(b)$, then the following inequality must hold:

$$\pi(y_{\widehat{b}}, q_{\widehat{b}}) \geq \pi(y_b, q_b).$$

Thus, the mechanism that we have just constructed has the same payoff in the relaxed problem without incentive compatibility constraints as it does in the full problem. Because this mechanism is optimal for the relaxed problem, it is thus optimal for the full problem. □

We emphasize two features of this solution. First, our assumption about markups has the implication that more profitable allocations $(x, p) \in \mathcal{B}$ receive larger scaling factors λ. Consequently, if the allocation intended for type $u \in U(b)$ is incentive compatible for type $u' \in U(b')$, then it must be the case that the allocation intended for those in $U(b)$ yields a strictly higher payoff for the principal than the allocation intended for those in $U(b')$. Accordingly, any failures of incentive compatibility are profit improving for the principal, and this feature drives why it is optimal to include only one allocation intended for each cell of agents.

Second — and more importantly — we have deliberately chosen our hypotheses to deliver a closed-form solution to the principal's optimization problem. Inspection of the expression for λ above reveals that it is generically optimal to choose λ so that $\lambda y \neq x$ for any $x \in X_0$. Accordingly, this example again provides a demonstration that it is not optimal in general for the principal in our model to use a mechanism M that corresponds exactly to one of the observed choice sets $\mathcal{B}_k$.

APPENDIX B

Appendix to Chapter Two

This appendix is organized as follows. First, we present technical lemmas not stated in the body of the text in section B.1. Second, we present proofs to propositions 4, 5, and 6 in section B.2, and treat proposition 8 in section B.3. Finally, section B.4 addresses screening by the principal, and section B.5 addresses the role of limited liability in our problem. Before proceeding, we establish some notation and facts about our problem that will play a role in our arguments throughout this appendix:

- Our proofs are complicated somewhat by the presence of the $(\delta_0, 0)$ action in $\mathcal{A}_0$, as δ_0 evidently does not have full support. Accordingly, we use the following notation:

$$\mathcal{A}(e) := \{(F, e') \in \mathcal{A}_0 | e' \geq e\}.$$

  Because $\mathcal{A}(e)$ is compact, the full support assumption implies the following for positive effort levels $e > 0$, for all open sets $\mathcal{O}$:

$$\min_{(F,e) \in \mathcal{A}(e)} P_F[\mathcal{O}] > 0.$$

- Given that contracts $w : \mathcal{Y} \to \mathbb{R}^+$ are upper semicontinuous and the agent enjoys limited liability, the range of w is compact for each contract w. We will make repeated use of this fact throughout our arguments.

- For each $\tau \in \mathbb{R}^+$, define u_A^τ as follows:

$$u_A^\tau(t) := \begin{cases} t & t < \tau \\ \tau & t \geq \tau. \end{cases}$$

- For each $\theta \in \mathbb{R}^+$, define u^θ as follows:

$$u^\theta(t) := \begin{cases} 0 & t < \theta \\ 1 & t \geq \theta. \end{cases}$$

Although the risk preferences u_A^τ and u^θ are both excluded from our model because they fail to be strictly increasing (and u^θ furthermore fails to be continuous), these types are nearby to types in $\mathcal{U}_A$ and $\mathcal{U}$, respectively, and we make use of this fact in several of our proofs.

B.1 Lemmas omitted from text

In this section, we prove three technical lemmas. First, we establish the existence of a solution to the agent's maximization problem, using standard compactness and continuity arguments. Second, we prove continuity results for both risk preference domains considered in this paper, which support the proofs of the propositions presented in the body of the paper. We emphasize that lemma 9 provides for the existence of a risk-averse risk preference with the desired property, while lemma 10 potentially yields a risk-seeking risk preference. Consequently, lemma 9 is not supplanted by its counterpart.

Lemma 8. *For each type v, compact technology $\mathcal{A}$, and contracts w, we have $c(\mathcal{A}|v, w) \neq \emptyset$.*

Proof. Fix type v, technology $\mathcal{A}$, and contract w. Define as follows:

$$M := \sup_{(F,e)\in\mathcal{A}} E_F\Big[v\Big(w(y), e\Big)\Big].$$

Because w is upper semicontinuous, it is bounded above. Thus, because v is continuous and decreasing in e, we have that $M < \infty$. Let (F_n, e_n) be such that:

$$E_{F_n}\Big[v\Big(w(y), e_n\Big)\Big] \to M.$$

As $\mathcal{A}$ is compact, there exists subsequence n_k of n and $(F^*, e^*) \in \mathcal{A}$ such that $(F_{n_k}, e_{n_k}) \to (F^*, e^*)$. Because w is upper semicontinuous, we have that $E_{F^*}\Big[v\Big(w(y), e^*\Big)\Big] \geq M$. This completes the proof. □

Lemma 9. *Let* $\mathcal{K} \subset \mathbb{R}^+$ *and* $\mathcal{F} \subset \Delta(\mathcal{K})$ *each be compact. If* $\max_{\mathcal{F}} P_F[t \geq \tau] < 1$*, then there exists* $u \in \mathcal{U}_A$ *such that* $\max_{\mathcal{F}} E_F[u(t)] < E_{\delta_\tau}[u(t)]$.

Proof. Let $\mathcal{F}$ and τ be as described. Define F^* as follows:

$$F^* := \inf_{F \in \mathcal{F}} F(t).$$

Note that $P_{F^*}[t \geq \tau] = \max_{F \in \mathcal{F}} P_F[t \geq \tau]$. Because $\max_{F \in \mathcal{F}} P_F[t \geq \tau] < 1$ by hypothesis, we have the following:

$$E_{\delta_\tau}[u_A^\tau(t)] > E_{F^*}[u_A^\tau(t)].$$

Let $u_n \in \mathcal{U}_A$ converge pointwise to u_A^τ on $\mathcal{K}$. Lebesgue's dominated convergence theorem implies that for N sufficiently large, we have the following:

$$E_{\delta_\tau}[u_N(t)] > E_{F^*}[u_N(t)].$$

By construction, F^* first-order stochastically dominates each $F \in \mathcal{F}$. Thus, we have the following:

$$E_{\delta_\tau}[u_N(t)] > \max_{F \in \mathcal{F}} E_F[u_N(t)].$$

□

Lemma 10. *Let* $\mathcal{K} \subset \mathbb{R}^+$ *and* $\mathcal{F} \subset \Delta(\mathcal{K})$ *each be compact. If* $\max_{\mathcal{F}} P_F[t \geq \theta] < p$*, then there exists* $u \in \mathcal{U}$ *such that* $\max_{\mathcal{F}} E_F[u(t)] < E_{\delta_\theta^p}[u(t)]$.

Proof. Let $\mathcal{F}, \theta$, and p be as described. Define F^* as follows:

$$F^* := \inf_{F \in \mathcal{F}} F(t).$$

Note that $P_{F^*}[t \geq \theta] = \max_{F \in \mathcal{F}} P_F[t \geq \theta]$. Because $p > \max_{F \in \mathcal{F}} P_F[t \geq \theta]$ by hypothesis, we have the following:

$$E_{\delta_\theta^p}[u^\theta(t)] > E_{F^*}[u^\theta(t)].$$

Let $u_n \in \mathcal{U}$ converge pointwise to u^θ on $\mathcal{K}$. Lebesgue's dominated convergence theorem implies that for N sufficiently large, we have the following:

$$E_{\delta_\theta^p}[u_N(t)] > E_{F^*}[u_N(t)].$$

By construction, F^* first-order stochastically dominates each $F \in \mathcal{F}$. Thus, we have the following:

$$E_{\delta_\theta^p}[u_N(t)] > \max_{F \in \mathcal{F}} E_F[u_N(t)].$$

□

B.2 Proofs omitted from text

B.2.1 Proposition 4

We prove proposition 4 for the general case in which $\mathcal{V}$ is *any* environment with risk aversion. We offer two points of clarification. First, because every environment without risk aversion includes as a subset an environment with risk aversion, the proof is valid for environments without risk aversion, as well. Second, the proof admits the costless effort environments (with or without risk aversion) as special cases.

Proof of proposition 4. Let $\mathcal{V}$ be an environment with risk aversion. Fix $\epsilon > 0$ and let x be a maximizer for w on the set $[0, \epsilon]$. Let $\tau = w(x)$. We exhibit a type $v \in \mathcal{V}$ and a technology $\mathcal{A} \in \mathscr{A}$ such that $c(\mathcal{A}|v, w) \subset \{(\delta_0, 0), (\delta_x, 0)\}$. After doing so, taking the limit $x \to 0$ yields the desired result.

As w is upper semicontinuous and not flat at the bottom, there exists open set $\mathcal{O}$ and constant $m < \tau$ such that $w(y) < m$ for $y \in \mathcal{O}$. The following inequality holds for any $e > 0$, per the full support assumption:

$$1 > \max_{(F,e) \in \mathcal{A}(e)} P_F[w(y) \geq \tau].$$

Lemma 9 implies that the following inequality holds for some risk preference $u \in \mathcal{U}_A$:

$$E_{\delta_x}\Big[u\Big((w(y)\Big)\Big] > \max_{(F,e) \in \mathcal{A}(e)} E_F\Big[u\Big(w(y)\Big)\Big].$$

Let $v \in \mathcal{V}$ have risk preference u. As v is decreasing in e, we have the following:

$$E_{\delta_x}\Big[v\Big(w(y), 0\Big) > \max_{(F,e)\in\mathcal{A}(e)} E_F\Big[v\Big(w(y), e\Big)\Big]. \tag{B.1}$$

Define $\mathcal{A}_x = \mathcal{A} \cup \{(\delta_x, 0)\}$. Note that because $(\delta_x, 0) \to (\delta_0, 0)$, we have that $\mathcal{A}_x \subset \mathcal{A}_1$ for all x sufficiently small.

Let $E_n \searrow 0$. Inequality B.1 establishes that for each n, there exists agent type $v_n \in \mathcal{V}$ such that $c(\mathcal{A}_x|v_n, w) \cap \mathcal{A}(E_n) = \emptyset$. Let $(F_n, e_n) \in c(\mathcal{A}_x|v_n, w)$. As $\mathcal{A}_x$ is compact, there exists a subsequence n_k of n and $(F^*, e^*) \in \mathcal{A}_x$ such that $(F_{n_k}, e_{n_k}) \to (F^*, e^*)$. Moreover, $(F^*, e^*) \in \{(\delta_0, 0), (\delta_x, 0)\}$. Accordingly, we have the following:

$$\pi(\mathscr{A}|\mathcal{V}, w) \leq \max\Big[u_p\Big(0 - w(0)\Big), \liminf_{x\to 0} u_p\Big(x - w(x)\Big)\Big].$$

As u_p is continuous and $w(y) \geq 0$ for all y, we conclude that $\pi(\mathscr{A}|\mathcal{V}, w) \leq 0$. □

B.2.2 Proposition 5

We prove proposition 5 for the general case in which $\mathcal{V}$ is any environment without risk aversion. Of course, this includes the costless effort environment without risk aversion as a special case.

Proof of proposition 5. Let $\mathcal{V}$ be an environment without risk aversion. As $\mathcal{Y}$ is compact and w is upper semicontinuous, w attains its maximum M on $\mathcal{Y}$. Let x be such that $w(x) = M$ and let $\theta = M$. As w is not flat at the top, the following inequality holds for any $e > 0$ and $p > 0$:

$$p > \max_{(F,e)\in\mathcal{A}(e)} P_F[y \geq \theta].$$

Lemma 10 implies that the following inequality holds for some risk preference $u \in \mathcal{U}$:

$$E_{\delta_x^p}\Big[u\Big((w(y)\Big)\Big] > \max_{(F,e)\in\mathcal{A}(e)} E_F\Big[u\Big(w(y)\Big)\Big].$$

Let $v \in \mathcal{V}$ have risk preference u. As v is decreasing in e, we have the following:

$$E_{\delta_x^p}\Big[v\Big(w(y), 0\Big] > \max_{(F,e)\in\mathcal{A}(e)} E_F\Big[v\Big(w(y), e\Big)\Big]. \tag{B.2}$$

Define $\mathcal{A}_p = \mathcal{A} \cup \{(\delta_x^p, 0)\}$. Note that because $(\delta_x^p, 0) \to (\delta_0, 0)$, we have that $\mathcal{A}_p \subset \mathcal{A}_1$ for all p sufficiently small.

Let $E_n \searrow 0$. Inequality B.2 establishes that for each n, there exists agent type $v_n \in \mathcal{V}$ such that $c(\mathcal{A}_p|v_n, w) \cap \mathcal{A}(E_n) = \emptyset$. Let $(F_n, e_n) \in c(\mathcal{A}_p|v_n, w)$. As $\mathcal{A}_p$ is compact, there exists a subsequence n_k of n and $(F^*, e^*) \in \mathcal{A}_p$ such that $(F_{n_k}, e_{n_k}) \to (F^*, e^*)$. Moreover, $(F^*, e^*) \in \{(\delta_0, 0), (\delta_x^p, 0)\}$. Accordingly, we have the following:

$$\pi(\mathscr{A}|\mathcal{V}, w) \leq \max\left[u_p\Big(0 - w(0)\Big), \liminf_{p\to 0}\Big((1-p)u_p\Big(0 - w(0)\Big) + (p)u_p\Big(x - w(x)\Big)\Big)\right].$$

As u_p is continuous and $w(y) \geq 0$ for all y, we conclude that $\pi(\mathscr{A}|\mathcal{V}, w) \leq 0$. □

B.2.3 Proposition 6

We prove proposition 6 for the case in which $\mathcal{V}$ is the costless effort environment without risk aversion. Because the costless effort environment with risk aversion is a subset of the costless effort environment without risk aversion, our proof is valid for the former case, as well.

Proof of proposition 6. Let $\mathcal{V}$ be the costless effort environment without risk aversion, and suppose that w be as described. Define as follows:

$$y_L := \max \mathcal{B}(w)$$
$$y_H := \min \mathcal{T}(w).$$

Next, let $(F, e) \in \mathcal{A}_0$ such that $P_F[y \geq y_H] > 0$, noting that such an action must exist per our hypothesis that w is flat at the top. Suppose that the following inequality holds for some risk preference u and some $F' \in \Delta(\mathcal{Y})$:

$$E_{F'}\Big[u\Big(w(y)\Big)]\Big] \geq E_F\Big[u\Big(w(y)\Big)\Big].$$

Because w is increasing, it must be the case that $P_{F'}[y \geq x] \geq P_F[y \geq x]$ for some $x \in [y_L, y_H]$. As $w(y) \leq y$ for all $y \in \mathcal{Y}$, we obtain the following:

$$E_{F'}[u_p\Big(y - w(y)\Big)] \geq \inf_{x \in [y_L, y_H]} \Big(P_F[y \geq x] u_p\Big(x - w(x)\Big)\Big).$$

As w is upper semicontinuous, $x - w(x)$ is lower semicontinuous. Consequently, $x - w(x)$ achieves its minimum on $[y_L, y_H]$. Moreover, we have $w(y) < y$ for all $y > 0$. Thus, we have:

$$\min_{x \in [y_L, y_H]} u_p\Big(x - w(x)\Big) > 0.$$

Finally, as $P_F[y \geq x] \geq P_F[y \geq y_H] > 0$ for all $x \in [y_L, y_H]$, we have that there exists $\pi > 0$ such that $E_{F'}[u_p\Big(y - w(y)\Big)] \geq \pi$, as desired. □

B.2.4 Proposition 7

Proof of proposition 7. Let $\mathcal{V}$ and $\mathscr{A}$ be as described. Consider an arbitrary contract w with at least two distinct transfer levels. We show how to improve w into a binary contract w' that provides a larger revenue guarantee than w. Proceeding, choose $y \in \mathcal{Y}$ so that the following inequality holds:

$$w(y) > \inf_{z \in \mathcal{Y}} w(z).$$

Let $\theta = w(x)$, and label as $T \subset \mathcal{Y}$ the set of output levels that are assigned a transfer at least θ:

$$T = \{y | w(y) \geq \theta\}.$$

Finally, let $x = \min T$. For $\epsilon > 0$ small, consider the following technology:

$$\mathcal{A} = \mathcal{A}_0 \cup \Big\{(\delta_x^{p+\epsilon}, 0)\Big\}.$$

Lemma 10 establishes the existence of some type $v \in \mathcal{V}$ such that:

$$c(\mathcal{A} | v, w) = \Big\{(\delta_x^{p+\epsilon}, 0)\Big\}.$$

Thus, we obtain the following upper bound for the principal's revenue guarantee when she uses contract w:

$$\pi^*(\mathscr{A}|\mathcal{V}, w) \leq p \cdot x.$$

Consider alternatively the following binary contract w', with quota $\mathcal{Q} = T$ and bonus payment $b > 0$:

$$w'(y) = \begin{cases} 0 & y \notin \mathcal{Q} \\ b & y \in \mathcal{Q}. \end{cases}$$

It is straightforward to verify that $\pi^*(\mathscr{A}|\mathcal{V}, w) = p \cdot x$. Thus, we have the following inequality, as desired:

$$\pi^*(\mathscr{A}|\mathcal{V}, w) \leq \pi^*(\mathscr{A}|\mathcal{V}, w').$$

Thus, we have established that there is no loss to the principal to restrict her search for a contract that maximizes the revenue guarantee π^* to binary contracts. Furthermore, it is simple to show that it is optimal to restrict attention further to increasing binary contracts. Existence of an optimal binary contract then follows from the upper semicontinuity of the quantity $\max_{(F,e) \in \mathcal{A}_0} P_F[y \geq x] \cdot x$ in x. □

B.3 Costly effort

This section is devoted to the proof of proposition 8. We assume that $\mathcal{A}_1 = \Delta(\mathcal{Y}) \times \mathbb{R}^+$, as is consistent with the statement of the second part of proposition 8. In order to see that the first part of the proposition holds without this restriction, observe that if $\mathscr{A}^* \subset \mathscr{A}$, then $\pi(\mathscr{A}^*|\mathcal{V}^*, w) \geq \pi(\mathscr{A}|\mathcal{V}^*, w)$ for all contracts w.

We proceed to the construction of the environment $\mathcal{V}$ without risk aversion referenced in the statement of proposition. As usual, we call type v *quasilinear* if it is of the following form:

$$v(t, e) = u(t) - e.$$

Let $(F^*, e^*) \in \mathcal{A}_0$ with $e^* > 0$. For output y^* and fixed transfer $b^* < y^*$, let $\mathcal{V}^*$ be exactly the set of quasilinear types v that satisfy the following inequality for some risk preference $u \in \mathcal{U}$:

$$\frac{e^*}{P_{F^*}[y \geq y^*]} \leq u(b^*). \tag{B.3}$$

Recall that for the purposes of assumption 4, we identify risk preferences that are affine transformations of one another, and so $\mathcal{V}^*$ is indeed a valid environment. One interpretation of $\mathcal{V}^*$ is as the set of agent types that consists of exactly the set of quasilinear agent types who prefer action (F^*, e^*) to shirking under the following binary contract w^*:

$$w^*(y) = \mathbf{1}(y \geq y^*)b^*.$$

We emphasize, however, that w^* is itself sub-optimal. The first step of the proof of proposition 8 is to improve an arbitrary contract into a binary contract:

Lemma 11. *For any contract w, there exists a binary contract w' that provides a larger guarantee against $\mathcal{V}^*$.*

Proof. Let w be a contract. If w is not flat at the bottom and flat at the top, $\pi(\mathscr{A}|\mathcal{V}^*, w) \leq 0$, and the improvement is vacuous; see propositions 4 and 5, respectively. Similarly, if w is constant, we have $(\delta_0, 0) \in c(\mathcal{A}_0|v, w)$ for all v, and hence $\pi(\mathscr{A}|\mathcal{V}, w) \leq 0$. Suppose instead that w is flat at the bottom, flat at the top, and not constant. Define as follows:

$$\mathcal{Q}^* = \Big\{y|w(y) \geq b^*\Big\}.$$

We construct our improved contract. Define $w' : \mathcal{Y} \to \mathbb{R}^+$ as follows:

$$w'(y) = \mathbf{1}\Big(\mathcal{Q}^*\Big)b^*.$$

Because w is a contract and thus upper semicontinuous, the set $\mathcal{Q}^*$ is closed. Thus, w' is itself upper semicontinuous and consequently a valid contract. Proceeding, choose $\mathcal{A} \in \mathscr{A}$ and $v \in \mathcal{V}^*$ arbitrarily. Let u be the risk preference belonging to type v i.e, $v(t, e) = u(t) - e$.

Let $(F^*, e^*) \in c(\mathcal{A}|v, w')$. We obtain the following inequality:

$$P_{F^*}[\mathcal{Q}^*]u(b^*) - e^* \geq \max_{(F,e)\in\mathcal{A}} P_F[\mathcal{Q}^*]u(b^*) - e. \tag{B.4}$$

For $p > P_{F^*}[\mathcal{Q}^*]$, define as follows:

$$\begin{aligned} F_p &= (1-p)\delta_0 + p \cdot \delta_{\min \mathcal{Q}^*} \\ \mathcal{A}_p &= \mathcal{A} \cup \Big\{(F_p, e^*)\Big\}. \end{aligned}$$

By choice of F_p, we have:

$$\liminf_{p \to P_F[\mathcal{Q}^*]} E_{F_p}\Big[u_p\Big(y - w(y)\Big)\Big] \leq E_{F^*}\Big[u_p\Big(y - w'(y)\Big)\Big].$$

Consequently, it is sufficient for the purposes of our proof to establish the existence of an agent type $v_p \in \mathcal{V}^*$ such that $(F_p, e^*) \in c(\mathcal{A}_p|v_p, w)$. We proceed accordingly. First, define u^* as follows:

$$\begin{aligned} u^*(t) &:= \mathbf{1}(t \geq b^*)u(b^*) \\ v^*(t, e) &:= u^*(t) - e. \end{aligned}$$

Note the following:

$$E_{F_p}\Big[v^*\Big(w(y), e^*\Big)] > \max_{(F,e)\in\mathcal{A}} E_F\Big[v^*\Big(w(y), e\Big)].$$

Because u^* is not strictly increasing, v^* does not belong to $\mathcal{V}^*$. However, there are types nearby to v^* that do belong to $\mathcal{V}^*$, and this turns out to be sufficient for our purposes. The remainder of the proof is devoted to establishing this approximation. Let $u_n \to u^*$ in $\mathcal{U}$, with $u_n \geq u^*$. Let $v_n(t, e) = u_n(t) - e$, noting that $v_n \in \mathcal{V}^*$, as desired. Our argument has two steps. First, Lebesgue's dominated converge theorem yields the following limit:

$$E_{F_p}\Big[v_n\Big(w(y), e^*\Big)\Big] \to E_{F_p}\Big[v^*\Big(w(y), e^*\Big)\Big]. \tag{B.5}$$

Second, we claim the following:

$$V(\mathcal{A}|v_n, w) \to \max_{(F,e)\in\mathcal{A}} E_F\Big[v^*\Big(w(y), e\Big)\Big]. \tag{B.6}$$

To see why, choose $(F_n, e_n) \in c(\mathcal{A}|v_n, w)$. Because $\mathcal{A}$ is compact, there exists subsequence n_k of n and $(\widehat{F}, \widehat{e}) \in \mathcal{A}$ such that $(F_{n_k}, e_{n_k}) \to (\widehat{F}, \widehat{e})$. To ease notation, we substitute n for n_k. We claim:

$$\liminf_n E_{F_n}[u_n\Big(w(y)\Big)] \leq E_{\widehat{F}}[u^*\Big(w(y)\Big)]. \tag{B.7}$$

Fix $\epsilon > 0$, and let $\tau_k \to b^*$. First, for each k, $u_n \rightrightarrows u^*$ outside of $[\tau_k, b^*)$. Consequently, there exists subsequence n_k of n such that:

$$n \geq n_k \implies \Big(u_n(t) - u^*(t) < \epsilon\Big) \forall t \notin [\tau_k, b^*). \tag{B.8}$$

Second, because $F_{n_k} \to \widehat{F}$ and $\tau_k \to b^*$, we have the following:

$$P_{\widehat{F}}[t \geq b^*] \geq \liminf_k P_{F_{n_k}}[t \geq \tau_k]. \tag{B.9}$$

Inequalities B.8 and B.9 jointly yield inequality B.7. Moreover, because $u_n \geq u$, the limit holds with equality. As $e_n \to \widehat{e}$, this is sufficient to establish B.6. In turn, B.5 and B.6 jointly yield the existence of v_n such that $(F_p, e^*) \in c(\mathcal{A}_p|v_n, w)$. This completes the proof. □

Having constructed an improvement of an arbitrary contract w to a binary contract w', the second step is to compute the guarantee provided by a binary contract, and show that there exists an optimal such guarantee.

Proof of proposition 8. Lemma 11 establishes that if an optimal contract exists, there exists an optimal binary contract of the form:

$$w(y) = \mathbf{1}(y \geq x)b^*.$$

We compute the guarantee associated to such contracts, as a function of x. First, if $x \leq b^*$, then w does not provide a guarantee. To see why, consider technology $\mathcal{A} = \mathcal{A}_0 \cup \{(\delta_x, 0)\}$.

For all $v \in \mathcal{V}^*$, we have:

$$\{(\delta_x, 0)\} \in c(\mathcal{A}|v, w).$$

This is sufficient to establish that w does not provide a guarantee if $x \leq b^*$. More substantively, suppose that $x > b^*$. Let $v \in \mathcal{V}^*$ and suppose that (F', e') is preferred by v to each $(F, e) \in \mathcal{A}_0$:

$$E_{F'}\Big[u\Big(w(y)\Big)\Big] - e' \geq \max_{(F,e)\in\mathcal{A}_0}\Big(E_F\Big[u\Big(w(y)\Big)\Big] - e\Big).$$

In turn, we obtain the following bound for $P_{F'}[y \geq x]$:

$$P_{F'}[y \geq x] \geq \max_{(F,e)\in\mathcal{A}_0}\left(P_F[y \geq x] + \frac{e' - e}{u(b^*)}\right). \tag{B.10}$$

It will be convenient to define $p(x)$ as follows:

$$p(x) := \left(\max_{(F,e)\in\mathcal{A}_0}\left(P_F[y \geq x] - \frac{e}{e^*}P_{F^*}[y \geq y^*]\right)\right).$$

The bound on the right-hand side of B.10 is minimized when $e' = 0$ and when inequality B.3 holds with equality. Substitution yields:

$$P_{F'}[y \geq x] \geq p(x).$$

Proceeding, we have the following lower bound for the principal's guarantee:

$$\pi(\mathscr{A}|\mathcal{V}^*, w) \geq p(x)u_p(x - b^*).$$

To see that this bound is tight, let $v \in \mathcal{V}^*$ be any type for whom inequality B.3 holds with equality. Consider action $(F, e) = \Big(\delta_x^{p(x)}, 0\Big)$. We have the following:

$$(F, e) \in c\Big(\mathcal{A} \cup \Big\{(F, e)\Big\}|v, w\Big).$$

Thus, we obtain:

$$\pi(\mathscr{A}|\mathcal{V}^*, w) = p(x)u_p(x - b^*)$$

Per our construction of $\mathcal{V}^*$, we have that $p(x) > 0$ for each $x < y^*$. Consequently, if $x \in (b^*, y^*)$, we have $\pi(\mathscr{A}|\mathcal{V}^*, w) > 0$. This establishes the first part of our claim: that there exist binary contracts that provide the principal with a guarantee against the environment$\mathcal{V}^*$.

To see that the second part of our claim holds — that there exist binary contracts which provide the principal with the largest possible payoff guarantee — note that because the quantity $P_F[y \geq x]$ is upper semicontinuous in x for each $F \in \Delta(\mathcal{Y})$, $p(x)$ is itself upper semicontinuous per standard arguments. This is sufficient to establish that there exists a binary contract w^* that maximizes $\pi(\mathscr{A}|\mathcal{V}^*, w)$. □

B.4 Screening

In the body of the text, we restrict our attention to analyzing the performance of individual contracts. However, the agent in our problem is privately informed about both his type v and his technology $\mathcal{A}$. Provided that the agent observes his type and technology before choosing a contract, this suggests that the principal might benefit from screening the agent with a menu contracts $\mathcal{W}$.

In this section, we allow for screening by the principal. We require that menus are incentive compatible: that is, the contract w assigned to type v and technology $\mathcal{A}$ satisfies the following set of inequalities:

$$v(\mathcal{A}|v, w) \geq v(\mathcal{A}|v, w') \, \forall w' \in \mathcal{W}.$$

We label as $w^v_{\mathcal{A}}$ the contract assigned to the pair $(v, \mathcal{A})$. The principal's payoff for menu $\mathcal{W}$ is as follows:

$$\pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) := \inf_{\mathcal{V} \times \mathscr{A}} \pi(\mathcal{A}|v, w^v_{\mathcal{A}}).$$

We require menus to be compact in the sup-norm topology, which ensures that there is an incentive compatible contract for each type-technology pair $(v, \mathcal{A})$.

Our necessary conditions for menus are analogous to propositions 4 and 5. First, menus that provide guarantees include contracts that are flat at the bottom:

Proposition 16. *Consider any environment with risk aversion or any environment without* *risk aversion. If menu $\mathcal{W}$ guarantees the principal a positive payoff, then $\mathcal{W}$ includes contracts that are flat at the bottom.*

We prove proposition 16 by considering environments with risk aversion. Because every environment without risk aversion includes an environment with risk aversion as a subset, our proof is valid for environments without risk aversion, as well.

Proof of proposition 16. Let $\mathcal{V}$ be any environment with risk aversion. Define as follows:

$$w_0 = \max_{w \in \mathcal{W}} \Big(\inf_{y \in \mathcal{Y}} w(y) \Big).$$

If $\mathcal{W}$ contains a contract w such that $w(y) = w_0$ for all y in some neighborhood $\mathcal{B}(w)$ of output $y = 0$, then there is nothing to demonstrate. Suppose instead that $\mathcal{W}$ includes no such contract. We claim $\mathcal{W}$ does not provide a guarantee.

By hypothesis, for all $\epsilon > 0$ there exists $x < \epsilon$ and $w^x \in \mathcal{W}$ such that $w^x(x) > w_0$. Set $\tau = w^x(x)$. We claim that the following inequality holds for all $E > 0$:

$$E_{\delta_x}\Big[u_A^\tau\Big(w^x(y)\Big)\Big] > \sup_{w \in \mathcal{W}} \Big(\max_{(F,e) \in \mathcal{A}(E)} E_F\Big[u_A^\tau\Big(w(y)\Big)\Big]\Big). \tag{B.11}$$

To see why, let w_n be an arbitrary sequence in $\mathcal{W}$. As $\mathcal{W}$ is compact, there exists subsequence n_k of n and $w^* \in \mathcal{W}$ such that w_{n_k} converges uniformly to w^*. By our definition of w_0 and the upper semicontinuity of w^*, there exists some open set $\mathcal{O}$ and some $\tau' < \tau$ such that $w^*(y) < \tau'$ for all $y \in \mathcal{O}$. Moreover, as w_{n_k} converges uniformly to w^*, there exists $\tau'' < \tau$ such that $w_{n_k} < \tau''$ for all $y \in \mathcal{O}$ for all k sufficiently large. As each $(F, e) \in \mathcal{A}(E)$ has full support, we conclude that inequality B.11 holds.

Let $u_n \in \mathcal{U}_A$ converge uniformly to u_A^τ. Inequality B.11 extends to u_N for N sufficiently large. The remainder of the argument is identical to the final steps of the proof of proposition 4. □

Similarly, if the agent is not necessarily risk-averse, menus that provide guarantees include contacts that are flat at the top:

Proposition 17. *Consider any environment without risk aversion. If menu$\mathcal{W}$ guarantees the principal a positive payoff, then $\mathcal{W}$ includes contracts that are flat at the top.*

Proof. Let $\mathcal{V}$ be any environment without risk aversion. Define as follows:

$$w_1 = \max_{w \in \mathcal{W}} \Big(\max_{y \in \mathcal{Y}} w(y) \Big).$$

If $\mathcal{W}$ contains a contract w such that $P_F[w(y) = w_1] > 0$ for some $(F, e) \in \mathcal{A}_0$, then there is nothing to demonstrate. Suppose instead that $\mathcal{W}$ includes no such contract. We claim that $\mathcal{W}$ does not provide a guarantee.

Set $E > 0$. Let $\mathscr{T}(E) = \{w(F) | w \in \mathcal{W}, (F, e) \in \mathcal{A}(E)\}$, where we write $w(F)$ for the distribution of transfers $w(y)$ when output $y \sim F$. Next, let $T(E)$ be the pointwise infimum of $\mathscr{T}(E)$.

First, because the pointwise infimum of a family of upper semicontinuous functions is itself upper semicontinuous, we have that $T(E) \in \Delta(\mathbb{R}^+)$. Second, $T(E)$ first-order stochastically dominates T for each $T \in \mathscr{T}(E)$. Third, because menu $\mathcal{W}$ does not include any contracts that are flat at the top, we have that $P_{T(E)}[w_1] = 0$.

Let $(x, w^x) \in \mathcal{Y} \times \mathcal{W}$ be such that $w^x(x) = w_1$. For any $\epsilon > 0$, lemma 10 implies that the following inequality holds for some risk preference $u \in \mathcal{U}$:

$$E_{\delta_x^\epsilon}\Big[u\Big(w^x(y)\Big)\Big] > E_{T(E)}[u(t)].$$

As $T(E)$ first-order stochastically dominates T for each $T \in \mathscr{T}(E)$, we conclude the following:

$$E_{\delta_x^\epsilon}\Big[u\Big(w^x(y)\Big)\Big] > \sup_{w \in \mathcal{W}} \bigg(\max_{(F,e) \in \mathcal{A}(E)} E_F\Big[u\Big(w(y)\Big)\Big] \bigg).$$

The remainder of the argument is identical to the final steps of the proof of proposition 5. □

Our model thus continues to predict the presence of contracts that are flat at the bottom and flat at the top if the principal screens the agent with a menu of contracts.

B.5 Limited liability

Our analysis demonstrates that it is challenging for the principal to provide the agent with incentives for effort without potentially providing him with incentives to take undesirable risks. One prospective solution to these challenges is to sell the firm to the agent: that is, for the principal to use a contract $w(y) = y - \beta$ that designates the agent the sole claimant of output, in return for a fee $\beta > 0$. Under such an arrangement, the principal's payoff is independent of output, and so she need not be concerned with risk-taking by the agent. Although such contracts are excluded by the limited liability constraint, this is inconsequential because sufficiently risk-averse agents refuse to purchase the firm from the principal in the absence of limited liability.

Formally, relax the requirement that $w(y) \geq 0$, and replace it with the requirement that $w(y) \geq -l$ for $l > 0$. Suppose that the agent has outside option $(\delta_{w_0}, 0)$, with $w_0 \geq 0$. This yields the following individual rationality constraint:

$$V(\mathcal{A}|v, w) \geq v(w_0, 0). \tag{B.12}$$

We say that a contract w *provides limited liability* if $w(y) \geq w_0$ for all $y \in \mathcal{Y}$. In our model, contracts that provide a guarantee provide limited liability:

Proposition 18. *Consider any environment with risk aversion or any environment without risk aversion. If contract w guarantees the principal a positive payoff, then w provides limited liability.*

We prove proposition 18 for environments with risk aversion. Because every environment with risk aversion includes an environment without risk aversion as a subset, our proof is valid for environments without risk aversion, as well.

Proof of proposition 18. Let $\mathcal{V}$ be any environment with risk aversion. Suppose that $w(y) < w_0$ for some $y \in \mathcal{Y}$. Because w is upper semicontinuous, there exists some open set $\mathcal{O}$ and some constant $\tau < w_0$ such that $w(y) < \tau$ on $\mathcal{O}$. The following inequality holds for any $e > 0$, per the full support assumption:

$$\max_{(F,e)\in\mathcal{A}(e)} P_F[w(y) \geq \tau] < 1.$$

Lemma 9 implies that the following holds for risk preference $u \in \mathcal{U}_A$:

$$E_{\delta_\tau}[u(t)] > \max_{(F,e)\in\mathcal{A}(e)} E_F\Big[u\Big(w(y)\Big)\Big]. \tag{B.13}$$

Let $v \in \mathcal{V}$ have risk preference u. As v is decreasing in e, we have the following:

$$v(\tau, 0) > \max_{(F,e)\in\mathcal{A}(e)} E_F\Big[v\Big(w(y), e\Big)\Big]. \tag{B.14}$$

Let $E_n \searrow 0$. For each n, let $v_n \in \mathcal{V}$ satisfy inequality B.14 for technology $\mathcal{A}(E_n)$. Let $(F_n, e_n) \in c(\mathcal{A}_0|v_n, w)$. As $\mathcal{A}_0$ is compact, there exists subsequence n_k of n and $(F^*, e^*) \in \mathcal{A}_0$ such that $(F_{n_k}, e_{n_k}) \to (F^*, e^*)$. There are two cases to consider.

Case 1. Suppose first that $e^* > 0$. For all k sufficiently large, we have that $e_{n_k} > E_{n_k}$. Consequently, for all such k, we have $(F_{n_k}, e_{n_k}) \in \mathcal{A}(E_{n_k})$. By construction, inequality B.14 implies the following:

$$v_{n_k}(\tau, 0) > V(\mathcal{A}_0|v_{n_k}, w).$$

As $v_{n_k}(\tau, 0) < v_{n_k}(w_0, 0)$, this is a violation of the individual rationality constraint B.12. Consequently, w does not provide a guarantee.

Case 2. Suppose second that $(F^*, e^*) = (\delta_0, 0)$. There are two subcases.

Case i. First, suppose $\limsup E_{F_{n_k}}[w(y)] < w_0$. Jensen's inequality implies the following, for all k sufficiently large:

$$u_{n_k}(w_0) > E_{F_{n_k}}\Big[u_{n_k}\Big(w(y)\Big)\Big].$$

As v_{n_k} is decreasing in e, we have the following:

$$v_{n_k}(w_0, 0) > V(\mathcal{A}_0 | v_{n_k}, w).$$

As this is a violation of the individual rationality constraint B.12, w does not provide a guarantee.

Case ii. Second, suppose $\limsup E_{F_{n_k}}[w(y)] \geq w_0$. We have the following:

$$\liminf E_{F_{n_k}}[y - w(y)] \leq -w_0.$$

Because u_p is concave, Jensen's inequality then implies the following:

$$\liminf E_{F_{n_k}}[u_p\Big(y - w(y)\Big)] \leq u_p(-w_0).$$

As $w_0 \geq 0$, w does not provide a guarantee.

□

The proof of proposition 18 demonstrates that if the agent is sufficiently risk-averse, he is unwilling to accept contracts that specify transfers smaller than w_0 unless he can obtain a transfer that is at least w_0 by shirking. Consequently, if a contract w does not provide limited liability, sufficiently risk-averse agents either accept the contract and shirk, or reject the contract outright. In either case, the contract does not provide a guarantee.[1]

We emphasize that while proposition 18 provides an explanation for contracts that are flat at the bottom with minimum transfer $w(0) \geq w_0$, our limited liability constraint demands only that $w(y) \geq -l$. Because we allow for the case that $w_0 > -l$, our limited liability constraint does not bind in general. Contrast this result to existing work that finds debt contracts to be optimal under binding limited liability constraints $\Big($Innes (1990); Poblete and Spulber (2012)$\Big)$.

[1]Although we do not pursue this line of analysis, proposition 18 suggests that the principal might screen risk-averse agents by offering contracts that do not provide limited liability.

B.6 Fixed technology

Consider an alternative environment in which $\mathscr{A} = \{\mathcal{A}_0\}$: that is, the principal is not uncertain about the agent's production technology. Propositions 4 and 5 do not continue to hold in general. We provide a counterexample.

Example 7. Suppose that $\mathcal{A}_0$ is ordered by first-order stochastic dominance $\succeq_1$:

$$F \succeq_1 F' \iff e \geq e' \, \forall (F, e), (F', e') \in \mathcal{A}_0.$$

Claim 4. Consider the costless effort environment with risk aversion or the costless effort environment without risk aversion. Suppose contract w is strictly increasing and $w(y) \leq y$ for all output y, with equality only if $y = 0$. Then w guarantees the principal a positive payoff.

We prove claim 4 for the costless effort environment without risk aversion. Because this environment includes the costless effort environment with risk aversion as a subset, the proof is valid for the latter environment, as well.

Proof of claim 4. As w is strictly increasing and $\mathcal{A}_0$ satisfies the first-order stochastic dominance condition stated above, we have the following for every risk preference u:

$$E_{F'}\Big[u\Big(w(y)\Big)\Big] \geq E_F\Big[u\Big(w(y)\Big)\Big] \iff e' \geq e.$$

As $\mathcal{A}_0$ is compact, it has a maximal-effort action (F^*, e^*). Under our stochastic dominance condition, this action is unique. Moreover, because F^* has full support and $w(y) < y$ for all $y > 0$, we have the following:

$$E_{F^*}\Big[u_p\Big(y - w(y)\Big)\Big] > 0.$$

Let $\mathcal{V}$ be the costless effort environment without risk aversion. We have $\pi(\mathscr{A}|\mathcal{V}, w) = E_{F^*}\Big[u_p\Big(y - w(y)\Big)\Big]$. $\square$

Thus, we see that the contract w fails to be flat at the bottom or flat at the top, yet provides a guarantee in some environments.

APPENDIX C

Appendix to Chapter Three

This appendix is divided into two sections. First, section C.1 provides technical material relating to section 3.3. Second, section C.2 does so for section 3.4.

C.1 Proofs, section 3.3

This section proves the results of section 3.3. First, we state and prove two supporting lemmas. After doing so, we prove proposition 9. Next comes a technical lemma, which establishes the existence of our contracts of interest, which is followed by the proof of theorem 3, which characterizes the set of risk-aligned contracts as stochastic binary contracts with weight λ^*. Lastly, we prove theorem 4, which is straightforward. For convenience, we present the risk-aligned weight λ^* here:

$$\lambda^* = \frac{E_{T_L}[\pi(y,t)] - E_{T_L}[\pi(0,t)]}{E_{T_L}[\pi(y,t)] - E_{T_L}[\pi(0,t)] + E_{T_H}[\pi(\overline{\mathcal{Y}},t)] - E_{T_H}[\pi(y,t)]}. \tag{C.1}$$

Note that λ^* depends on the principal's objective π and the rewards T_L, T_H.

C.1.1 Lemmas

Lemma 12. *If w is payoff affine, then $u^w, \pi^w : \mathcal{Y} \to \mathbb{R}$ are injective for all $u \in \mathcal{U}$.*

Proof. Suppose that w is payoff affine. We have the following:

$$\begin{aligned} \pi^w(x) = \pi^w(z) &\iff u^w(x) = u^w(z) \forall u \in \mathcal{U} && \text{(C.2)} \\ &\iff T^w(x) = T^w(z) && \text{(C.3)} \\ &\iff x = z. && \text{(C.4)} \end{aligned}$$

Line C.2 follows from criterion 2. Lemma 4 then yields line C.3. As π is strictly increasing in y and $T^w(x) = T^w(z)$, line C.4 follows. ∎

C.1.2 Proof of proposition 9

Proof of proposition 9 . First, we show that if w is risk-aligned, then w is payoff affine. Second, we show the converse. We proceed by contrapositive. Suppose that w is not payoff affine. There are three cases to consider.

Case 1. Suppose that for some type $u \in \mathcal{U}$, there exists $x \in \mathcal{Y}$ such that:

$$u^w(x) = (1-\lambda)u^w(0) + \lambda u^w(\overline{\mathcal{Y}})$$
$$\pi^w(x) \neq (1-\lambda)\pi^w(0) + \lambda \pi^w(\overline{\mathcal{Y}}).$$

Set $F = (1-\lambda)\delta_0 + \lambda\delta_{\overline{\mathcal{Y}}}$ and $F' = \delta_x$ to see that w is not risk-aligned.

Case 2. Suppose that there is an affine relationship between π^w and u^w for all $u \in \mathcal{U}$, but $\alpha(u) = 0$ for some risk type u. Criterion 2 implies directly that π^w is constant on $\mathcal{Y}$.

We argue that it can not also be the case that u^w is constant for all $u \in \mathcal{U}$. To see why, suppose that u^w were constant for all u. Lemma 4 implies that there exists $T \in \Delta(\mathcal{T})$ such that $T^w(y) = T$ for all y. However, because π is strictly increasing in y, π^w is not constant. We conclude that there exists $u \in \mathcal{U}$ such that u^w is non-constant.

Accordingly, there exist $x, z \in \mathcal{Y}$ such that $u^w(x) \neq u^w(z)$. Set $F = \delta_x$ and $F' = \delta_z$ to see that w is not risk-aligned.

Case 3. Suppose that there is an affine relationship between π^w and u^w for all u, but $\alpha(u) < 0$ for some risk type u. Set $F = \delta_0$ and $F' = \delta_{\overline{\mathcal{Y}}}$ to see that w is not risk-aligned.

Conversely, suppose that w is payoff affine. As $\alpha(u) > 0$ for all $u \in \mathcal{U}$, we have:

$$E_F[\pi^w(y)] \geq E_{F'}[\pi^w(y)] \iff E_F[u^w(y)] \geq E_{F'}[u^w(y)].$$

Accordingly, w is risk-aligned. □

C.1.3 Proof of theorem 3

First, we prove that our stochastic binary contracts are indeed contracts:

Lemma 13. *If* $T(y) = \big(1 - \lambda(y)\big)T_L + \lambda(y)T_H$ *for weight* λ *and* $T_L, T_H \in \Delta(\mathcal{T})$*, then there exists a contract* w *with* $T^w = T$*.*

Proof. We divide our proof into three parts. First, we show how our model embeds a universal probability space. Second, we construct a function $w : \mathcal{Y} \times \mathcal{X} \to \mathcal{T}$ such that $w(y, x) \sim T(y)$ for each y. Third, we verify that w is a contract.

Part 1. Recall that $x = (x_1, ..., x_N) \in \mathcal{X} \subset \mathbb{R}^N$. For $1 \leq n \leq N$, let G_n be the marginal distribution of x_n, and note that G_n inherits continuity from G. Define $f : \mathcal{X} \to [0, 1]$ as $f(x) = G_n(x_n)$. Because G_n is continuous, $f(x)$ is uniformly distributed on $[0, 1]$ when $x \sim G$. Moreover, f is continuous.

Part 2. Let $T_0 \in \Delta(\mathcal{T})$. We construct an upper semicontinuous function $\tilde{w} : \mathcal{X} \to \mathcal{T}$ such that $\tilde{w}(x) \sim T_0$ when $x \sim G$. Let $T^+ : [0, 1) \to \mathcal{T}$ be the upper semicontinuous generalized inverse of T_0:

$$T^+(p) = \inf\{t \in \mathcal{T} | T_0(t) > p\}.$$

Let T^* be the extension of T^+ to domain $[0, 1]$ with $T^*(1) = \max \mathcal{T}$. As $f(x)$ is uniformly distributed over $[0, 1]$, we have that $T^*(f(x)) \sim T_0$ when $x \sim G$. Set $\tilde{w}(x) = T^*(f(x))$ to obtain the desired function $\tilde{w}$.

Finally, for each $y \in \mathcal{Y}$, let $\tilde{w}^y$ be such that $\tilde{w}^y(x) \sim T(y)$, per the construction $\tilde{w}$ just given. Set $w(y, x) = \tilde{w}^y(x)$.

Part 3. We verify that w is upper semicontinuous, so that it is indeed a contract. We show:

$$(y_k, x_k) \to (y, x) \implies \limsup_k w(y_k, x_k) \leq w(y, x).$$

It is notationally convenient to introduce the following shorthand:

$$T_y := T(y)$$
$$T_n := T(y_n).$$

If $f(x) = 1$, then $w(y, x) = \max \mathcal{T} \geq w(y', x')$ for all (y', x'). More substantively, suppose that $f(x) < 1$. We proceed by contradiction. Let $(y_k, x_k) \to (y, x)$. Suppose that:

$$\limsup_k w(y_k, x_k) > w(y, x).$$

First, by our definition of w, there exists $t' \in \mathcal{T}$ and subsequence k_j of k such that:

$$\inf\{t | T_{k_j}(t) > f(x_{k_j})\} > t' > \inf\{t | T_y(t) > f(x)\}.$$

Thus, $T_{k_j}(t') \leq f(x_{k_j})$ for all j. Consequently, for all j we have:

$$P_{T_{k_j}}[t \geq t'] \geq 1 - f(x_{k_j}).$$

Second, there exists $t'' < t'$ and $p > f(x)$ such that $T_y(t'') = p$. Accordingly, as $t' > t''$, we have:

$$P_{T_y}[t \geq t'] \leq P_{T_y}[t > t''] = 1 - p.$$

As f is continuous, $f(x_{k_j}) \to f(x)$. Thus, $P_{T_{k_j}}[t \geq t'] > P_{T_y}[t \geq t']$ for j large. This contradicts that $T_{k_j} \to T_y$. Thus, w is upper semicontinuous, as desired.

□

Second, we state the proof of the result:

Proof of theorem 3. We prove the theorem in three steps. First, we show that if w is payoff affine, then w is a stochastic binary contract with a particular weight λ^*. Second, we verify such contracts are indeed payoff affine. Third, we address remaining technical details.

Part 1. Suppose that w is payoff affine. Label as follows:

$$T_L := T^w(0)$$
$$T_H := T^w(\overline{\mathcal{Y}}).$$

Subtracting $\pi^w(0)$ from $\pi^w(\overline{\mathcal{Y}})$ and rearranging criterion 2 yields:

$$\alpha(u) = \frac{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}{u^w(\overline{\mathcal{Y}}) - u^w(0)}. \tag{C.5}$$

Lemma 12 establishes that u^w, π^w are injective, which ensures that equation C.5 is well-defined. Next, subtracting $\pi^w(0)$ from $\pi^w(y)$ and substituting the right-hand side of equation C.5 into criterion 2 yields:

$$u^w(y) = \left(1 - \frac{\pi^w(y) - \pi^w(0)}{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}\right)u^w(0) + \left(\frac{\pi^w(y) - \pi^w(0)}{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}\right)u^w(\overline{\mathcal{Y}}). \tag{C.6}$$

Define $\Lambda : \mathcal{Y} \to \mathbb{R}$ as follows:

$$\Lambda(y) = \frac{\pi^w(y) - \pi^w(0)}{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}. \tag{C.7}$$

Substituting equation C.7 into equation C.6 yields:

$$\int u(t)\mathrm{d}T^w(y) = \int u(t)\mathrm{d}\Big[\Big(1 - \Lambda(y)\Big)T_L + \Lambda(y)T_H\Big] \; \forall u \in \mathcal{U}. \tag{C.8}$$

Lemma 4 implies that the following equality is equivalent to equation C.8[1]:

$$T^w(y) = \Big(1 - \Lambda(y)\Big)T_L + \Lambda(y)T_H. \tag{C.9}$$

We are left to exhibit an expression Λ that solves the implicit equation C.7, and verify that the corresponding function is a weight. First, direct substitution of C.9

[1] We have not yet argued that $\Lambda \in [0, 1]$, and so it is not assured that the finite signed measure on the right-hand side of equation C.5 is a probability distribution. However, because the set of finite signed measures forms a vector space, we are free to use vector space operations as necessary to derive an equality for which lemma 4 applies.

into definitions 3.1 and 3.2 —wherein we define u^w and π^w — respectively yields the following:

$$\pi^w(y) = \Lambda(y)\Big(E_{T_H}[\pi(y,t)] - E_{T_L}[\pi(y,t)]\Big) + E_{T_L}[\pi(y,t)] \tag{C.10}$$

$$u^w(y) = \Lambda(y)\Big(E_{T_H}[u(t)] - E_{T_L}[u(t)]\Big) + E_{T_L}[u(t)]. \tag{C.11}$$

Second, substitution of C.10 and C.11 into C.5 yields:

$$\alpha(u) = \frac{\Big(E_{T_H}[\pi(\overline{\mathcal{Y}},t)] - E_{T_L}[\pi(\overline{\mathcal{Y}},t)]\Big) - \Big(E_{T_H}[\pi(0,t)] - E_{T_L}[\pi(0,t)]\Big)}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}. \tag{C.12}$$

Next, rearranging C.5 gives the following:

$$\pi^w(y) - \pi^w(0) = \alpha(u)\Big(u^w(y) - u^w(0)\Big). \tag{C.13}$$

Substitution of C.11 into C.13 gives:

$$\pi^w(y) - \pi^w(0) = \alpha(u)\Lambda(y)\Big(E_{T_H}[u(t)] - E_{T_L}[u(t)]\Big). \tag{C.14}$$

In turn, solving for $\Lambda(y)$ yields:

$$\Lambda(y) = \frac{\pi^w(y) - \pi^w(0)}{\alpha(u)\Big(E_{T_H}[u(t)] - E_{T_L}[u(t)]\Big)}. \tag{C.15}$$

Finally, substituting definition 3.2 — in which we define π^w — and C.12 into C.15 yields the following expression for $\Lambda(y)$:

$$\Lambda(y) = \frac{E_{T_L}[\pi(y,t)] - E_{T_L}[\pi(0,t)]}{E_{T_L}[\pi(y,t)] - E_{T_L}[\pi(0,t)] + E_{T_H}[\pi(\overline{\mathcal{Y}},t)] - E_{T_H}[\pi(y,t)]}.$$

Set $\lambda^* := \Lambda(y)$, and observe that λ^* coincides with C.1, presented at the beginning of the appendix. Note that $\lambda^*(0) = 0$, $\lambda^*(\overline{\mathcal{Y}}) = 1$, and λ^* is strictly increasing and continuous. Consequently, λ^* is a weight. Moreover, because $\alpha(u) > 0$, w is aligned at extremes. Thus, w is a stochastic binary contract.

Part 2. Suppose that w is a stochastic binary contract with weight λ^*. By construction, there is an affine relationship between π^w and u^w:

$$\pi^w(y) = \left(\frac{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}{u^w(\overline{\mathcal{Y}}) - u^w(0)}\right)\Big(u^w(y) - u^w(0)\Big) + \pi^w(0).$$

As w is a stochastic binary contract, w is aligned at extremes. By definition, we have the following:

$$\frac{\pi^w(\overline{\mathcal{Y}}) - \pi^w(0)}{u^w(\overline{\mathcal{Y}}) - u^w(0)} > 0.$$

Consequently, w is payoff affine.

Part 3. We have established that a contract is payoff affine if and only if it is an stochastic binary contract with weight λ^*. Proposition 9 establishes that a contract is risk-aligned if and only if it is payoff affine. Consequently, contracts are risk-aligned if and only if they are stochastic binary contracts with weight λ^*. The final step of the argument is to ensure that each stochastic binary contract with weight λ^* is indeed a contract: that is, that there exists a contract $w : \mathcal{Y} \times \mathcal{X} \to \mathcal{T}$ that induces the desired distribution of transfers $T^w : \mathcal{Y} \to \Delta(\mathcal{T})$. This is the content of lemma 13.

□

C.1.4 Proof of theorem 4

Proof of theorem 4. Suppose that T_H first-order stochastically dominates T_L. Consequently, $u^w(\overline{\mathcal{Y}}) > u^w(0)$ for all $u \in \mathcal{U}$. For increasing and continuous cost function $C : \mathcal{T} \to \mathbb{R}$ with $C(0) = 0$ and $E_{T_H}[C(t)] < 1$, define:

$$\pi(y, t) = \lambda(y) - C(t).$$

Note that w is aligned at extremes. Thus, w satisfies the criteria of definition 10. For all $F \in \Delta(\mathcal{Y})$, the principal's payoff is as follows:

$$E_F[\pi^w(y)] = \Big(1 - E_F[\lambda(y)]\Big)\Big(- E_{T_L}[C(t)]\Big) + E_F[\lambda(y)]\Big(1 - E_{T_H}[C(t)]\Big).$$

Accordingly, we have the following, for all $u \in \mathcal{U}$:

$$\begin{aligned} E_F[u^w(y)] \geq E_{F'}[u^w(y)] &\iff E_F[\lambda(y)] \geq E_{F'}[\lambda(y)] \\ &\iff E_F[\pi^w(y)] \geq E_{F'}[\pi^w(y)]. \end{aligned}$$

We conclude that w is risk-aligned. $\square$

C.2 Proofs, section 3.4

The primary objective of this section is to prove theorem 5. The theorem is mostly an implication of propositions 11 and 12, which we prove first. These propositions require a series of lemmas, which we state and prove first, in the order in which they are invoked. After doing so, we present a proof of theorem 14, which is mostly self-contained. First, we prove that the agent's problem has a solution:

Lemma 14. *The agent's problem 3.3 has a solution.*

Proof. Let $v = u(t) - k(e)$ be the agent's type and let $\mathcal{A} \in \mathscr{A}$. Define as follows:

$$M := \sup_{(F,e)\in\mathcal{A}} E_F[u^w(y)] - k(e).$$

Let $(F_n, e_n) \in \mathcal{A}$ be such that $E_{F_n}[u^w(y)] - k(e_n) \to M$. As $\mathcal{A}$ is compact, there exists subsequence n_k of n and $(F^*, e^*) \in \mathcal{A}$ such that $(F_{n_k}, e_{n_k}) \to (F^*, e^*)$. As u^w is upper semicontinuous and k is continuous, $E_{F^*}[u^w(y)] - k(e^*) \geq M$. We conclude that $c(\mathcal{A}|v, w) \neq \emptyset$. $\square$

C.2.1 Notation and preliminaries

We assume throughout this section that the principal is financially risk-neutral, as in assumption 11. Accordingly, risk-aligned contracts are stochastic binary contracts with linear weights. Given agent type v and contract w, define $\psi_w^v, \Pi_w^v : \mathcal{E} \to \mathbb{R}$ as follows, wherein we use the shorthand $T_L := T^w(0)$ and $T_H := T^w(\overline{\mathcal{Y}})$:

$$\psi^v_w(e) := \frac{\mu(e)}{\overline{\mathcal{Y}}}\left(E_{T_H}[u(t)] - E_{T_L}[u(t)]\right) + E_{T_L}[u(t)] - k(e) \tag{C.16}$$

$$\Pi^v_w(e) := \left(\frac{\psi^v_w(e) - E_{T_L}[u(t)]}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}\right)\left(\overline{\mathcal{Y}} - (E_{T_H}[t] - E_{T_L}[t])\right) - E_{T_L}[t]. \tag{C.17}$$

We introduce the following notation:

- $\mathscr{A}^*$ is the set of feasible technologies $\mathcal{A}$ such that F has support $\{0, \overline{\mathcal{Y}}\}$ for all $(F, e) \in \mathcal{A}$;
- $\mathcal{A}^* \in \mathscr{A}^*$ is the feasible technology such that: $(F, e) \in \mathcal{A}^*$ if and only if $e \in \mathcal{E}$ and $E_F[y] = \mu(e)$.

C.2.2 Proof of proposition 10

Proof of proposition 10. First, we show that payoff affine contracts are output affine. This is the substance of the proposition. Second, we prove the converse.

$\Longrightarrow$ Suppose that w is payoff affine.

Part 1. We claim that there exist functions $\alpha, \beta : \mathscr{U} \to \mathbb{R}$ such that the following relationship holds for all $u \in \mathscr{U}$:

$$\pi^w(y) = \alpha(u)u^w(y) + \beta(u). \tag{C.18}$$

We proceed by contrapositive. Suppose that no such functions α, β exist. Accordingly, there exists $\lambda \in (0, 1)$, $u \in \mathscr{U}$, and $a, b, c \in \mathcal{Y}$ such that:

$$u^w(b) = (1 - \lambda)u^w(a) + \lambda u^w(c)$$
$$\pi^w(b) \neq (1 - \lambda)\pi^w(a) + \lambda\pi^w(c).$$

Let $u_0 \in \mathcal{U}$ and define u_α as follows, for each $\alpha \in (0, 1)$:

$$u_\alpha := (1 - \alpha)u_0 + \alpha u.$$

Let $F_\lambda = (1-\lambda)\delta_a + \lambda\delta_c$. For all $F \in \Delta(\mathcal{Y})$, we have:

$$E_F[u^w_\alpha(y)] = (1-\alpha)E_F[u^w_0(y)] + \alpha E_F[u^w(y)].$$

By construction, it can not be the case that both of the following inequalities hold simultaneously:

$$E_{F_\lambda}[u^w_\alpha(y)] = E_{\delta_b}[u^w_\alpha(y)]$$
$$E_{F_\lambda}[u^w_0(y)] = E_{\delta_b}[u^w_0(y)].$$

As $u_\alpha \in \mathcal{U}$ for $\alpha > 0$ sufficiently near 0, w is not risk-aligned. Proposition 9 then implies that w is not payoff affine. We conclude that the desired functions α, β exist.

Part 2. We claim that there exists constants $c \neq 0$, d such that $\pi^w(y) = cy + d$. Let u be the risk type $u(t) = t$. We have the following, for all contracts w:

$$\pi^w(y) = y - u^w(y). \tag{C.19}$$

Equating the right-hand side of equation C.19 with the right-hand side of equation C.18 yields the following:

$$y - b = (1+a)u^w(y). \tag{C.20}$$

We claim that $a \notin \{-1, 0\}$. To see why, note that $a = -1 \implies b = y$. As b is a constant, this can not be the case. We conclude that $a \neq -1$. If $a = 0$, equation C.19 implies that $\pi^w = b$. However, lemma 12 establishes that π^w is injective when w is payoff affine. We conclude that $a \neq 0$. Rearranging equation C.20 and substituting for $u^w(y)$ in equation C.19 yields the following:

$$\pi^w(y) = \left(\frac{a}{a+1}\right)y + \frac{b}{a+1}. \tag{C.21}$$

Equation C.21 gives the desired constants c, d. Substitution into criterion 2 yields the functions $a, b : \mathcal{U} \to \mathbb{R}$ that satisfy criterion 3. Consequently, w is output affine.

$\Longleftarrow$ Suppose that w is output affine. Direct substitution of criterion 3 into criterion 2 reveals that there is an affine relationship between π^w and u^w, for each $u \in \mathcal{U}$. The sign restrictions that criterion 3 imposes on $a(u)$ and $a(\pi)$ ensure that $\alpha(u) > 0$ for each $u \in \mathcal{U}$. Accordingly, w is payoff affine.

$\square$

C.2.3 Proof of proposition 11

For each $F \in \Delta(\mathcal{Y})$, define F^T as follows:

$$F^T := \left(1 - \frac{E_F[y]}{\overline{\mathcal{Y}}}\right)\delta_0 + \frac{E_F[y]}{\overline{\mathcal{Y}}}\delta_{\overline{\mathcal{Y}}}. \tag{C.22}$$

Relatedly, for each $\mathcal{A} \in \mathscr{A}$, define $\mathcal{A}^T$ as follows:

$$\mathcal{A}^T = \Big\{(F^T, e) | (F, e) \in \mathcal{A}\Big\}. \tag{C.23}$$

Note that because $E_{F^T}[y] = E_F[y]$, $\mathcal{A}^T \in \mathscr{A}$. Moreover, by construction, $\mathcal{A}^T \in \mathscr{A}^*$.

Proof of proposition 11. Theorem 3 establishes that w is risk-aligned if and only if w is a stochastic binary contract with weight λ_{RN}, as defined in 3.9. Suppose that w is not a stochastic binary contract with weight λ_{RN}. Define w' as follows:

$$T^{w'}(y) = \Big(1 - \lambda_{RN}(y)\Big)T^w(0) + \lambda_{RN}(y)T^w(\overline{\mathcal{Y}}).$$

First, because $\mathscr{A}^* \subset \mathscr{A}$, we have for all types v and all contracts w_0:

$$V(\mathscr{A}|v, w_0) \leq V(\mathscr{A}^*|v, w_0) \tag{C.24}$$

$$\Pi(\mathscr{A}|v, w_0) \leq \Pi(\mathscr{A}^*|v, w_0). \tag{C.25}$$

Second, by construction we have:

$$V(\mathscr{A}^*|v,w) = V(\mathscr{A}^*|v,w') \tag{C.26}$$

$$\Pi(\mathscr{A}^*|v,w) = \Pi(\mathscr{A}^*|v,w'). \tag{C.27}$$

Third, we show:

$$V(\mathscr{A}^*|v,w') = V(\mathscr{A}|v,w') \tag{C.28}$$

$$\Pi(\mathscr{A}^*|v,w') = \Pi(\mathscr{A}|v,w'). \tag{C.29}$$

Choose $\mathcal{A} \in \mathscr{A}$, and let $\mathcal{A}^T$ be as defined in C.23. Because $E_F[u^{w'}(y)]$ depends on F only through $E_F[y]$, we have:

$$V(\mathcal{A}|v,w') = V(\mathcal{A}^T|v,w'). \tag{C.30}$$

Moreover, because w' is risk-aligned per theorem 3, equation C.30 in turn yields:

$$\Pi(\mathcal{A}|v,w') = \Pi(\mathcal{A}^T|v,w'). \tag{C.31}$$

Because we chose $\mathcal{A} \in \mathscr{A}$ arbitrarily, C.30 is sufficient for C.28, and C.31 is sufficient for C.29. Inequalities C.24, C.26, and C.28 jointly yield the following, as desired:

$$V(\mathscr{A}|v,w) \leq V(\mathscr{A}|v,w').$$

Similarly, C.25, C.27, and C.29 jointly yield:

$$\Pi(\mathscr{A}|v,w) \leq \Pi(\mathscr{A}|v,w').$$

This completes the proof. □

C.2.4 Lemmas in support of proposition 12

Proposition 12 requires a series of technical lemmas. Before stating the lemmas, we describe the dependencies between them. First, lemma 15 and lemma 16 are self-contained. Lemma 17 depends on both of the aforementioned lemmas, and lemma 18 on lemma 17. Finally, lemma 19 depends on lemma 15 and lemma 17. First, lemma 15 provides necessary conditions for contracts that provide a guarantee:

Lemma 15. *If* $\Pi(\mathscr{A}|\mathcal{V}, w) > 0$, *then:*

1. $\pi^w(\overline{\mathcal{Y}}) > \pi^w(0)$; *and*

2. $u^w(\overline{\mathcal{Y}}) > V(\mathscr{A}|v, w) > u^w(0)$ *for all* $v \in \mathcal{V}$.

Proof. First, we claim that if w provides a guarantee, then $\pi^w(\overline{\mathcal{Y}}) > \pi^w(0)$. To see why, suppose that $\pi^w(\overline{\mathcal{Y}}) \leq \pi^w(0)$. As $\pi^w(0) \leq 0$ for all contracts w, we have $\Pi(\mathscr{A}|v, w) \leq 0$.

Second, we claim that if w provides a guarantee, then $u^w(\overline{\mathcal{Y}}) > V(\mathscr{A}|v, w)$. To see why, note that there are feasible technologies $\mathcal{A}$ that satisfy $F = \delta_{\overline{\mathcal{Y}}}$ for each $(F, e) \in \mathcal{A}$. Because $k(e) > 0$ for each $e \in \mathcal{E}$, $V(\mathcal{A}|v, w) < u^w(\overline{\mathcal{Y}})$ for such technologies $\mathcal{A}$. Consequently, $V(\mathscr{A}|v, w) < u^w(\overline{\mathcal{Y}})$ for all v.

Finally, we claim that if w provides a guarantee, then $V(\mathscr{A}|v, w) > u^w(0)$. To see why, suppose that $V(\mathscr{A}|v, w) \leq u^w(0)$. Then there exist feasible technologies $\mathcal{A} \in \mathscr{A}$ such that $(\delta_0, 0) \in c(\mathcal{A}|v, w)$. Because that immediately implies that $\Pi(\mathscr{A}|v, w) \leq 0$, this contradicts our hypothesis that w provides a guarantee. □

Second, lemma 16 shows that except in uninteresting cases, worst-case output distributions have zero effort cost, and provide minimal utility to the agent:

Lemma 16. *If* $\Pi(\mathscr{A}|v, w) > 0$, *then:*

$$E_F[u^w(y)] > V(\mathscr{A}|v, w) \implies E_F[\pi^w(y)] > \Pi(\mathscr{A}|v, w).$$

Proof. Let type v have risk type u. Let $F \in \Delta(\mathcal{Y})$ satisfy $E_F[u^w(y)] > V(\mathscr{A}|v, w)$. Define F' as follows:

$$F' = \left(1 - \frac{V(\mathscr{A}|v, w)}{E_F[u^w(y)]}\right)\delta_0 + \frac{V(\mathscr{A}|v, w)}{E_F[u^w(y)]}F.$$

Note that:

$$E_{F'}[u^w(y)] - k(0) = \left(1 - \frac{V(\mathscr{A}|v, w)}{E_F[u^w(y)]}\right)u^w(0) + V(\mathscr{A}|v, w).$$

Accordingly, $E_{F'}[u^w(y)] - k(0) \geq V(\mathscr{A}|v, w)$. Moreover, we have the following:

$$E_{F'}[\pi^w(y)] = \left(1 - \frac{V(\mathscr{A}|v, w)}{E_F[u^w(y)]}\right)\pi^w(0) + \frac{V(\mathscr{A}|v, w)}{E_F[u^w(y)]}E_F[\pi^w(y)].$$

First, because $\Pi(\mathscr{A}|v, w) > 0$, we have $E_F[\pi^w(y)] > 0$. Moreover, as the agent enjoys limited liability, we have $\pi^w(0) \leq 0$. Thus, $E_{F'}[\pi^w(y)] < E_F[\pi^w(y)]$.

Second, because $E_{F'}[u^w(y)] \geq V(\mathscr{A}|v, w)$, there exists $\mathcal{A}' \in \mathscr{A}$ such that $(F', 0) \in c(\mathcal{A}'|v, w)$. As ties are broken against the principal, this establishes that $\Pi(\mathscr{A}|v, w) \leq E_{F'}[\pi^w(y)]$. We conclude that $\Pi(\mathscr{A}|v, w) < E_F[\pi^w(y)]$. □

Third, lemma 17 computes the guarantee belonging to a stochastic binary contract:

Lemma 17. *If* $\Pi(\mathscr{A}|v, w) > 0$ *for risk-aligned stochastic binary contract* w*, then:*

$$V(\mathscr{A}|v, w) = \max\left(\psi_w^v\right)$$
$$\Pi(\mathscr{A}|v, w) = \max\left(\Pi_w^v\right).$$

Proof. The functions ψ_w^v and Π_w^v are as defined in section C.2.1. Theorem 3 implies that w is risk-aligned if and only if w is a stochastic binary contract with weight λ_{RN}, as defined in 3.9. Let w be such a contract, and suppose $\Pi(\mathscr{A}|v, w) > 0$.

First, we claim that $V(\mathscr{A}|v, w) = \max\left(\psi_w^v\right)$. For all $F \in \Delta(\mathcal{Y})$, we have:

$$\begin{aligned} E_F[u^w(y)] &= \Big(1 - E_F[\lambda_{RN}(y)]\Big)E_{T_L}[u(t)] + E_F[\lambda_{RN}(y)]E_{T_H}[u(t)] \\ &= \left(1 - \frac{E_F[y]}{\overline{\mathcal{Y}}}\right)E_{T_L}[u(t)] + \frac{E_F[y]}{\overline{\mathcal{Y}}}E_{T_H}[u(t)]. \end{aligned}$$

As $\Pi(\mathscr{A}|v, w) > 0$, lemma 15 implies that $E_{T_H}[u(t)] > E_{T_L}[u(t)]$. Accordingly, $V(\mathcal{A}^*|v, w) = V(\mathscr{A}|v, w)$, where $\mathcal{A}^*$ is as defined in section C.2.1. We have:

$$V(\mathscr{A}|v, w) = \max_{e \in \mathcal{E}}\left(\frac{\mu(e)}{\overline{\mathcal{Y}}}\Big(E_{T_H}[u(t)] - E_{T_L}[u(t)]\Big) + E_{T_L}[u(t)] - k(e)\right).$$

Thus, $V(\mathscr{A}|v,w) = \max\left(\psi_w^v\right)$. We claim that $\Pi(\mathscr{A}|v,w) = \max\left(\Pi_w^v\right)$. First, lemma 16 yields:

$$E_F[\pi^w(y)] = \Pi(\mathscr{A}|v,w) \implies E_F[u^w(y)] \leq V(\mathscr{A}|v,w).$$

Second, there exists $\mathcal{A} \in \mathscr{A}$ such that $(F,e) \in c(\mathcal{A}|v,w)$ if and only if:

$$E_F[u^w(y)] - k(e) \geq V(\mathscr{A}|v,w).$$

Thus, we restrict our attention to actions $(F,0)$ with $E_F[u^w(y)] = \max\left(\psi_w^v\right)$. As w is risk-aligned, we are free to consider a single such action. Set F as follows:

$$F = \left(1 - \frac{\max\left(\psi_w^v\right) - E_{T_L}[u(t)]}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}\right)\delta_0 + \left(\frac{\max\left(\psi_w^v\right) - E_{T_L}[u(t)]}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}\right)\delta_{\overline{\mathcal{Y}}}.$$

As desired, $E_F[u^w(y)] = \max\left(\psi_w^v\right)$. We have the following:

$$E_F[\pi^w(y)] = \left(\frac{\max\left(\psi_w^v\right) - E_{T_L}[u(t)]}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}\right)\left(\overline{\mathcal{Y}} - \left(E_{T_H}[t] - E_{T_L}[t]\right)\right) - E_{T_L}[t]. \tag{C.32}$$

The quantity on the right-hand side of 16 is Π_w^v, evaluated at the value of $e \in \mathcal{E}$ that maximizes ψ_w^v. By construction, if $\Pi(\mathscr{A}|v,w) > 0$, then Π_w^v is a positive affine transformation of ψ_w^v. Thus, $E_F[\pi^w(y)] = \max\left(\Pi_w^v\right)$. □

Fourth, lemma 18 establishes the existence of contracts providing a guarantee:

Lemma 18. *There exists a contract w with $\Pi(\mathscr{A}|\mathcal{V},w) > 0$ if and only if assumption 12 is satisfied.*

Proof. We first prove that there exists a contract that provides a guarantee if the gains from trade assumption is satisfied. After doing so, we prove the converse statement.

$\Longleftarrow$ Suppose that the gains from trade assumption is satisfied. Let w be the stochastic binary contract with $T^w(\overline{\mathcal{Y}}) = T$ and $T^w(0) = 0$, where T is as described in the statement of the assumption. Let w have risk-aligned weight λ_{RN}, as defined in

3.9. We have as follows, per lemma 17:

$$\begin{aligned}\Pi(\mathscr{A}|v,w) &= \left(\max_{e\in\mathcal{E}} \frac{\mu(e)E_T[u(t)]}{\overline{\mathcal{Y}}} - k(e)\right)\left(\frac{\overline{\mathcal{Y}} - E_T[t]}{E_T[u(t)]}\right) \\ &\geq \mu\Big(\min \mathcal{E}\Big)\left(\frac{1}{\overline{\mathcal{Y}}} - s\right)\Big(\overline{\mathcal{Y}} - E_T[t]\Big).\end{aligned}$$

Accordingly, $\Pi(\mathscr{A}|\mathcal{V},w) > 0$.

$\Longrightarrow$ Suppose that the gains from trade assumption is not satisfied. Accordingly, if $E_T[t] < \overline{\mathcal{Y}}$, for each $\epsilon > 0$ there exists a type $v_\epsilon \in \mathcal{V}$ with risk type u_ϵ such that the following holds for all $e \in \mathcal{E}$:

$$k(e) \geq (1-\epsilon)\frac{\mu(e)E_T[u_\epsilon(t)]}{\overline{\mathcal{Y}}}.$$

Let $T^w(\overline{\mathcal{Y}}) = T$. Suppose first $E_T[t] < \overline{\mathcal{Y}}$. Let $\mathcal{A}^*$ be as defined in section C.2.1. For all $\epsilon > 0$ there exists a type $v_\epsilon \in \mathcal{V}$ such that $V(\mathcal{A}^*|v_\epsilon, w) < E_T[u_\epsilon(t)]\epsilon$. Define as follows:

$$\begin{aligned}F_\epsilon &:= (1-\epsilon)\delta_0 + \epsilon\delta_{\overline{\mathcal{Y}}} \\ \mathcal{A}_\epsilon &:= \mathcal{A}^* \cup \{(F_\epsilon, 0)\}.\end{aligned}$$

We have $c(\mathcal{A}_\epsilon|v_\epsilon, w) = \{(F_\epsilon, 0)\}$. Thus, $\Pi(\mathscr{A}|\mathcal{V},w) \leq 0$. Suppose instead that $E_T[t] \geq \overline{\mathcal{Y}}$. We have immediately that $\Pi(\mathcal{A}^*|v,w) \leq 0$ for all v. Accordingly, $\Pi(\mathscr{A}|\mathcal{V},w) \leq 0$.

$\square$

Finally, lemma 19 shows that within the class of stochastic binary contracts, it is optimal to pay the agent nothing when he produces zero output:

Lemma 19. *If $\Pi(\mathscr{A}|\mathcal{V},w) > 0$ and $T^w(0)$ differs from δ_0 for risk-aligned stochastic binary contract w, then there exists a risk-aligned stochastic binary contract w' with $T^{w'}(0) = \delta_0$ and $\Pi(\mathscr{A}|\mathcal{V},w') > \Pi(\mathscr{A}|\mathcal{V},w)$.*

Proof. Let w be as described. Define $T_L := T^w(0)$ and $T_H := T^w(\overline{\mathcal{Y}})$. Let w' be as follows, with risk-aligned weight λ_{RN} as defined in 3.9:

$$T^{w'}(y) = \Big(1 - \lambda_{RN}(y)\Big)\delta_0 + \lambda_{R_N}(y)T_H.$$

Theorem 3 implies that w' is risk-aligned. Let $v \in \mathcal{V}$. Let $e_0 \in \arg\max \psi^v_w$ and $e_1 \in \arg\max \psi^v_{w'}$. We have:

$$\frac{\mu(e_1)}{\overline{\mathcal{Y}}} E_{T_H}[u(t)] - k(e_1) \geq \frac{\mu(e_0)}{\overline{\mathcal{Y}}} E_{T_H}[u(t)] - k(e_0). \tag{C.33}$$

Lemma 15 implies $E_{T_H}[u(t)] > 0$. Dividing inequality C.33 by $E_{T_H}[u(t)]$ yields:

$$\frac{\mu(e_1)}{\overline{\mathcal{Y}}} - \frac{k(e_1)}{E_{T_H}[u(t)]} > \frac{\mu(e_0)}{\overline{\mathcal{Y}}} - \frac{k(e_0)}{E_{T_H}[u(t)]}.$$

In turn, because $E_{T_L}[u(t)] > 0$, we have the following:

$$\frac{\mu(e_1)}{\overline{\mathcal{Y}}} - \frac{k(e_1)}{E_{T_H}[u(t)]} > \frac{\mu(e_0)}{\overline{\mathcal{Y}}} - \frac{k(e_0)}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}.$$

Direct substitution into our definition of Π^v_w, presented in C.17, yields:

$$\Pi^v_{w'}(e_1) - \Pi^v_w(e_0) > \left(1 - \frac{\mu(e_0)}{\overline{\mathcal{Y}}} + \frac{k(e_0)}{E_{T_H}[u(t)] - E_{T_L}[u(t)]}\right) E_{T_L}[t]. \tag{C.34}$$

As $\mu(\max \mathcal{E}) < \overline{\mathcal{Y}}$, there exists constant $\Delta > 0$ such that the right-hand side of inequality C.34 is at least Δ for all $u \in \mathcal{U}$. Lemma 17 then establishes that $\Pi(\mathscr{A}|\mathcal{V}, w') > \Pi(\mathscr{A}|\mathcal{V}, w)$. □

C.2.5 Proof of proposition 12

Proof of proposition 12. Suppose that the agent has regular preferences, and suppose that w is not a stochastic binary contract with weight λ_{RN}, as defined in 3.9.

We prove the proposition in two parts. First, we establish the existence of a risk-aligned stochastic binary contract that typewise dominates w. Second, we improve this contract into a stochastic bonus contract.

Part 1. Because w is not risk-aligned, proposition 9 and proposition 10 imply that w is not output affine. Thus, there exists type $v \in \mathcal{V}$ with risk type u and $y^* \in \mathcal{Y}$ for which one of the following inequalities holds:

$$u^w(y^*) < \left(1 - \frac{y^*}{\overline{\mathcal{Y}}}\right)u^w(0) + \frac{y^*}{\overline{\mathcal{Y}}}u^w(\overline{\mathcal{Y}}) \tag{C.35}$$

$$u^w(y^*) > \left(1 - \frac{y^*}{\overline{\mathcal{Y}}}\right)u^w(0) + \frac{y^*}{\overline{\mathcal{Y}}}u^w(\overline{\mathcal{Y}}). \tag{C.36}$$

Define w' as follows:

$$T^{w'}(y) = \Big(1 - \lambda_{RN}(y)\Big)T^w(0) + \lambda_{RN}(y)T^w(\overline{\mathcal{Y}}).$$

Proposition 11 demonstrates that the following inequalities hold for all $v \in \mathcal{V}$:

$$\begin{aligned} V(\mathscr{A}|v, w) &\leq V(\mathscr{A}|v, w') \\ \Pi(\mathscr{A}|v, w) &\leq \Pi(\mathscr{A}|v, w'). \end{aligned} \tag{C.37}$$

We proceed in two cases. We first show that if w provides a guarantee, then C.37 holds strictly for some type $v \in \mathcal{V}$. In this case, we make use of the following facts about w'. First, because proposition 11 implies that $\Pi\mathscr{A}|v, w') \geq \Pi(\mathscr{A}|v, w)$, lemma 15 establishes that:

$$u^{w'}(\overline{\mathcal{Y}}) > V(\mathscr{A}|v, w') > u^{w'}(0). \tag{C.38}$$

Second, because $E_F[u^{w'}(y)]$ only depends on F through $E_F[y]$, we obtain:

$$V(\mathcal{A}^*|v, w') = V(\mathscr{A}|v, w').$$

Third, because w provides a guarantee, we have $\pi^w(\overline{\mathcal{Y}}) > \pi^w(0)$. This inequalities extend to w' by construction, and thus w' is aligned at extremes. Accordingly, w' is risk-aligned per theorem 3.

After treating the case in which the original contract w provides a guarantee, we treat the less interesting cases in which w does not provide a guarantee. In the latter case, the improved contract does not necessarily coincide with w'.

Case 1. Suppose that w provides a guarantee. Suppose further that inequality C.35 holds. First, we claim that:

$$V(\mathscr{A}|v,w) < V(\mathscr{A}|v,w'). \tag{C.39}$$

To see why, let $\mathcal{A} \in \mathscr{A}^*$ satisfy $V(\mathcal{A}|v,w) = V(\mathscr{A}^*|v,w)$.[2] For each $(F,e) \in \mathcal{A}$, choose $\tilde{F}$ with support $\{0, y^*, \overline{\mathcal{Y}}\}$ and $E_{\tilde{F}}[y] = E_F[y]$. Define $\tilde{\mathcal{A}}$ as follows:

$$\tilde{\mathcal{A}} = \Big\{(\tilde{F},e)|(F,e) \in \mathcal{A}\Big\}.$$

By construction, $\tilde{\mathcal{A}} \in \mathscr{A}$. Moreover, $V(\tilde{\mathcal{A}}|v,w') = V(\mathscr{A}|v,w')$. However, inequality C.35 implies that $V(\tilde{\mathcal{A}}|v,w) < V(\tilde{\mathcal{A}}|v,w')$. Thus, we have C.39, as desired.

Second, let F^* have support $\{0, \overline{\mathcal{Y}}\}$ and satisfy:

$$E_{F^*}[u^{w'}(y)] = V(\mathscr{A}|v,w'). \tag{C.40}$$

Note that C.38 ensures the existence of such a distribution. Because w' is risk-aligned per theorem 3, we have:

$$E_{F^*}[\pi^{w'}(y)] = \Pi(\mathscr{A}|v,w'). \tag{C.41}$$

[2] For convenience, exclude $(\delta_0, 0)$ from $\mathcal{A}$. This exclusion is inconsequential, because w provides a guarantee and hence $(\delta_0, 0) \notin c(\mathcal{A}|v,w)$.

By our choice of support for F^*, we have:

$$E_{F^*}[u^{w'}(y)] = E_{F^*}[u^w(y)] \tag{C.42}$$

$$E_{F^*}[\pi^{w'}(y)] = E_{F^*}[\pi^w(y)]. \tag{C.43}$$

Combining C.39, C.40, and C.42 yields $V(\mathscr{A}|v,w) < E_{F^*}[u^w(y)]$. Lemma 16 then implies that:

$$\Pi(\mathscr{A}|v,w) < E_{F^*}[\pi^w(y)]. \tag{C.44}$$

Finally, C.41, C.43, and C.44 yield:

$$\Pi(\mathscr{A}|v,w) < \Pi(\mathscr{A}|v,w').$$

Case 2. Suppose that w provides a guarantee. Suppose further that inequality C.36 and the following inequality both hold:

$$\pi^w(y) \leq \left(1 - \frac{y^*}{\overline{\mathcal{Y}}}\right)\pi^w(0) + \frac{y^*}{\overline{\mathcal{Y}}}\pi^w(\overline{\mathcal{Y}}).$$

Let F have support $\{0, y^*, \overline{\mathcal{Y}}\}$ and F^* have support $\{0, \overline{\mathcal{Y}}\}$. Let F, F^* satisfy the following equations, where $\mathcal{A}^*$ is as defined in section C.2.1:

$$E_F[u^w(y)] = V(\mathcal{A}^*|v,w)$$

$$E_{F^*}[u^w(y)] = V(\mathcal{A}^*|v,w).$$

As before, C.38 ensures the existence of such distributions. Next, by choice of w', we have:

$$E_{F^*}[u^w(y)] = E_{F^*}[u^{w'}(y)]$$

$$E_{F^*}[\pi^w(y)] = E_{F^*}[\pi^{w'}(y)].$$

As discussed at the beginning of the proof, w' is risk-aligned and satisfies

$V(\mathcal{A}^*|v,w') = V(\mathscr{A}|v,w')$. Thus, we have:

$$E_{F^*}[u^{w'}(y)] = V(\mathscr{A}|v,w')$$
$$E_{F^*}[\pi^{w'}(y)] = \Pi(\mathscr{A}|v,w').$$

However, per our choice of F, we have:

$$E_F[\pi^w(y)] < E_{F^*}[\pi^w(y)].$$

Set $\mathcal{A} = \mathcal{A}^* \cup \Big\{(F,0)\Big\}$. As $(F,0) \in c(\mathcal{A}|v,w)$, we have:

$$\Pi(\mathscr{A}|v,w) \leq \Pi(\mathcal{A}|v,w) < \Pi(\mathscr{A}|v,w').$$

Case 3. Suppose that w provides a guarantee, and that inequality C.36 and the following inequality both hold:

$$\pi^w(y^*) > \left(1 - \frac{y^*}{\overline{\mathcal{Y}}}\right)\pi^w(0) + \frac{y^*}{\overline{\mathcal{Y}}}\pi^w(\overline{\mathcal{Y}}). \tag{C.45}$$

As the agent has regular preferences, $\mathcal{U}$ contains a risk-neutral type $\tilde{u}$. Inequality C.45 then implies the following:

$$\tilde{u}^w(y) < \left(1 - \frac{y^*}{\overline{\mathcal{Y}}}\right)\tilde{u}^w(0) + \frac{y^*}{\overline{\mathcal{Y}}}\tilde{u}^w(\overline{\mathcal{Y}}).$$

Thus, this case reduces to case 1.

Case 4. Suppose that w does not provide a guarantee. There are two subcases to consider. First, if $u^w(0) \leq V(\mathscr{A}|v,w)$ for *all* $v \in \mathcal{V}$, then it is sufficient to consider any risk-aligned stochastic binary contract that does provide a guarantee. The existence of such contracts is established in lemma 18.

More substantively, suppose that $\Pi(\mathscr{A}|v_0,w) > 0$ for some $v_0 \in \mathcal{V}$ with risk type u_0. We claim that the continuity assumption constituent to definition 11 ensures the existence of some open set $\mathcal{U}_0 \subset \mathcal{U}$ containing u_0 and some constant $\kappa > 0$ such that:

$$\tilde{u} \in \mathcal{U}_0 \implies \Pi(\mathscr{A}|\tilde{v}, w) > \kappa,$$

where $\tilde{v}$ is the type corresponding to risk type $\tilde{u}$.

We proceed by contradiction. Suppose that no such pair $(\mathcal{U}_0, \kappa)$ exists. Let $v_n \rightrightarrows v_0$ such that $\Pi(\mathscr{A}|v_n, w) < \epsilon$ for ϵ arbitrarily small. There exists a sequence of distributions $F_n \in \Delta(\mathcal{Y})$ such that:

$$E_{F_n}[\pi^w(y)] < \epsilon$$
$$E_{F_n}[u_n^w(y)] = V(\mathscr{A}|v_n, w).$$

It is straightforward to establish that $V(\mathscr{A}|v_n, w) \to V(\mathscr{A}|v_0, w)$. Thus, $E_{F_n}[u_0^w(y)] \to V(\mathscr{A}|v_0, w)$. Because $\Delta(\mathcal{Y})$ is sequentially compact, there exists a sequence n_k of n and a distribution $F_0 \in \Delta(\mathcal{Y})$ such that $F_{n_k} \to F_0$. Moreover, because $u_0^w : \mathcal{Y} \to \mathbb{R}$ is upper semicontinuous and $\pi^w : \mathcal{Y} \to \mathbb{R}$ is lower semicontinuous, we have:

$$E_{F_0}[u_0^w(y)] \geq V(\mathscr{A}|v_0, w).$$
$$E_{F_0}[\pi^w(y)] \leq \epsilon.$$

Thus, there are technologies $\mathcal{A} \in \mathscr{A}$ such that $(F_0, 0) \in c(\mathcal{A}|v_0, w)$. Because $\pi(\mathcal{A}|v_0, w) \leq \epsilon$, this contradicts our hypothesis that $\Pi(\mathscr{A}|v_0, w) > 0$ and thus establishes the existence of the desired set $\mathcal{U}_0$ and constant $\kappa > 0$. Let $\mathcal{V}_0$ be the corresponding set of agent types. The problem of establishing that w' typewise dominates w against $\mathcal{V}_0$ reduces to case 1, case 2, or case 3 above. Because $\mathcal{V} \supset \mathcal{V}_0$, this is sufficient for typewise dominance against $\mathcal{V}$. This completes part 1.

Part 2. Thus far, we have established that if the agent has regular preferences, typewise undominated contracts are stochastic binary contract with weight λ_{RN}. Moreover, if w is stochastic binary contract with weight λ_{RN} that does not satisfy $T^w(0) = \delta_0$, then lemma 19 establishes that w is not an optimal contract. Suppose that w is as follows, for non-degenerate lottery $T \in \Delta(\mathcal{T})$ and risk-aligned weight λ_{RN}, as defined in 3.9:

$$T^w(y) = \Big(1 - \lambda_{RN}(y)\Big)\delta_0 + \lambda_{RN}(y)T.$$

Set $b = E_T[t]$ and let w' be as follows:

$$T^{w'} = \Big(1 - \lambda_{RN}(y)\Big)\delta_0 + \lambda_{RN}(y)\delta_b.$$

First, let $u_{RN} \in \mathcal{U}$ be risk-neutral. We have $E_T[u_{RN}(t)] = u_{RN}(b)$. Thus, for corresponding type v, we have $\Pi(\mathscr{A}|v, w) = \Pi(\mathscr{A}|v, w')$. Next, let $u \in \mathscr{U}$ be strictly risk-averse. For $\gamma \in (0, 1)$, define as follows:

$$u^* := (1 - \gamma)u_{RN} + \gamma u.$$

For γ sufficiently small, $u^* \in \mathcal{U}$. Moreover, because u^* is strictly concave, Jensen's inequality implies that $E_T[u^*(t)] < u^*(b)$. Let v^* be the corresponding type $v \in \mathcal{V}$. Let $e_0 \in \arg\max \psi^v_w$ and $e_1 \in \arg\max \psi^v_{w'}$. A direct calculation reveals the following:

$$\frac{\psi^{v^*}_{w'}(e_1)}{u^*(b)} > \frac{\psi^{v^*}_{w}(e_0)}{E_T[u^*(t)]}.$$

In turn, we obtain that $\Pi^{v^*}_{w'}(e_1) > \Pi^{v^*}_{w}(e_0)$ via direct substitution into C.17, wherein we define Π^v_w. Lemma 17 completes the proof.

□

C.2.6 Proof of theorem 5

First, we establish the existence of a contract maximizing the principal's payoff:

Lemma 20. *The principal's problem 3.8 has a solution.*

Proof. First, proposition 11 and lemma 19 jointly imply that it is without loss of generality to restrict our search for an optimal contract to stochastic binary contracts w with $T^w(0) = \delta_0$ and risk-aligned weight λ_{RN}, as defined in 3.9.

Second, we claim that there exists $m > 0$ such that $\Pi(\mathscr{A}|\mathcal{V}, w) \leq 0$ if $E_{T^w(\overline{\mathcal{Y}})}[t] < m$. Let $v \in \mathcal{V}$, and recall that $v = u(t) - k(e)$, with k strictly increasing. Accordingly, $k\Big(\min \mathcal{E}\Big) > 0$. The continuity of u implies the existence of $m > 0$ such that:

$$E_T[t] < m \implies E_T[u(t)] < k(\min \mathcal{E}\Big).$$

Set $\mathcal{A}_0 = \mathcal{A}^* \cup \{(\delta_0, 0)\}$, where $\mathcal{A}^*$ is as defined in section C.2.1. If $E_T[u(t)] < m$, and $T^w(\overline{\mathcal{Y}}) = T$, we have $c(\mathcal{A}_0|v, w) = \{(\delta_0, 0)\}$. This establishes the existence of the desired constant m.

Consequently, we are free to further restrict our search to risk-aligned stochastic binary contracts w with $T^w(0) = \delta_0$ and $T^w(\overline{\mathcal{Y}}) \in \mathcal{T}^*$, defined as follows:

$$\mathcal{T}^* := \Big\{T \in \Delta(\mathcal{T}) | E_T[t] \in [m, \overline{\mathcal{Y}}]\Big\}.$$

For each $v \in \mathcal{V}$ with risk type u, define $\psi^v : \mathcal{E} \times \mathcal{T}^* \to \mathbb{R}$ and $\phi^v, \Pi^v, \Pi^{\mathcal{V}} : \mathcal{T}^* \to \mathbb{R}$ as follows:

$$\psi^v(e, T) := \frac{\mu(e)}{\overline{\mathcal{Y}}} E_T[u(t)] - k(e) \qquad \Pi^v(T) := \frac{\phi^v(T)}{E_T[u(t)]}\Big(\overline{\mathcal{Y}} - E_T[t]\Big)$$

$$\phi^v(T) := \max_{e \in \mathcal{E}} \psi^v(e, T) \qquad \Pi^{\mathcal{V}}(T) := \inf_{v \in \mathcal{V}} \Pi^v(T).$$

Note that ψ^v is continuous in both of its arguments. Consequently, ϕ^v is continuous, per the Maximum Theorem. In turn, Π^v is continuous. As $\Pi^{\mathcal{V}}$ is the infimum of a family of continuous functions, it is upper semicontinuous on $\mathcal{T}^*$. Moreover, as $\mathcal{T}^*$ is compact, $\Pi^{\mathcal{V}}$ achieves its maximum on $\mathcal{T}^*$.

Let T^* be the maximizer for $\Pi^{\mathcal{V}}$ on $\mathcal{T}^*$, and let w^* be the risk-aligned stochastic binary contract with $T^{w^*}(0) = \delta_0$ and $T^{w^*}(\overline{\mathcal{Y}}) = T^*$. Similarly, let $T \in \mathcal{T}^*$, and let w be the risk-aligned stochastic binary contract with $T^w(0) = \delta_0$ and $T^w(\overline{\mathcal{Y}}) = T$. First, lemma 17 establishes for each type v that:

$$\begin{aligned}\Pi(\mathscr{A}|v, w^*) &= \Pi^v(T^*)\\ \Pi(\mathscr{A}|v, w) &= \Pi^v(T).\end{aligned}$$

Thus, by definition we have:

$$\begin{aligned}\Pi(\mathscr{A}|\mathcal{V}, w^*) &= \Pi^{\mathcal{V}}(T^*)\\ \Pi(\mathscr{A}|\mathcal{V}, w) &= \Pi^{\mathcal{V}}(T).\end{aligned}$$

As T^* is a maximizer for $\Pi^{\mathcal{V}}$, this completes the proof. □

Finally, we assemble the pieces of the argument:

Proof of theorem 5. First, lemma 20 and proposition 11 jointly provide for the existence of an optimal risk-aligned stochastic binary contract. Lemma 18, which establishes the existence of contracts that guarantee the principal a positive payoff, ensures that this optimality is not vacuous

Second, if the agent has regular preferences, proposition 12 establishes that undominated contracts are stochastic bonus contracts. □

C.2.7 Proof of propositions 13 and 14

The contract $\mathcal{W}(\mathcal{A}, v)$ chosen by the agent with technology $\mathcal{A}$ and type v satisfies the following incentive compatibility condition:

$$V\Big(\mathcal{A}|v, \mathcal{W}(\mathcal{A}, v)\Big) \geq V(\mathcal{A}|v, w)\, \forall w \in \mathcal{W}.$$

We employ the following notation:

$$\begin{aligned}\Pi(\mathcal{A}|v, \mathcal{W}) &:= \Pi\Big(\mathcal{A}|v, \mathcal{W}(\mathcal{A}, v)\Big)\\ \Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) &:= \inf_{\mathscr{A}\times\mathcal{V}} \Pi(\mathcal{A}|v, \mathcal{W}).\end{aligned}$$

Proof of proposition 13. Because we have already established via lemma 18 that there exist contracts w with $\pi(\mathscr{A}|\mathcal{V}, w) > 0$, suppose that $\pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) > 0$ as well. Let $\mathcal{V}$ be as described, so that the problem admits a low-effort type v_0. Let $u_0 \in \mathcal{U}$ be risk type associated to type v_0, and let w_0 be the contract in $\mathcal{W}$ that is incentive compatible for type v_0 and technology $\mathcal{A}^*$, as defined in section C.2.1. Define the contract w_1 as follows:

$$T^{w_1}(y) = \Big(1 - \lambda_{RN}(y)\Big)T^{w_0}(0) + \lambda_{RN}(y)T^{w_0}(\overline{\mathcal{Y}}).$$

By definition, we have $\Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) \leq \Pi(\mathscr{A}|v_0, \mathcal{W})$. Moreover, because w_0 is incentive compatible for type v_0 and technology $\mathcal{A}^*$, we have:

$$\Pi(\mathscr{A}|v_0, \mathcal{W}) \leq \Pi(\mathcal{A}^*|v_0, w_0).$$

Second, by construction we have:

$$V(\mathscr{A}^*|v_0, w_1) = V(\mathscr{A}^*|v_0, w_0)$$
$$\Pi(\mathscr{A}^*|v_0, w_1) = \Pi(\mathscr{A}^*|v_0, w_0)$$

Note that because $\mathcal{W}$ provides a guarantee, it must true that $u_0^w(\overline{\mathcal{Y}}) > V(\mathscr{A}|v_0, w_0) > u_0^w(0)$: see lemma 15. Furthermore, per our choice of weight λ_{RN}, $E_F[u_0^{w_1}(y)]$ depends only on $E_F[y]$ for all $F \in \Delta(\mathcal{Y})$. These two facts jointly yield the following:

$$V(\mathscr{A}|v_0, w_1) = V(\mathcal{A}^*|v_0, w_1).$$

In turn, because w_1 is risk-aligned per theorem 3, we obtain the following:

$$\Pi(\mathscr{A}|v_0, w_1) = \Pi(\mathcal{A}^*|v_0, w_1).$$

Thus far, we have established:

$$\Pi(\mathscr{A}|v_0, \mathcal{W}) \leq \Pi(\mathscr{A}|v_0, w_1).$$

Finally, if $T^{w_1}(0) \neq \delta_0$, define w_2 as follows:

$$T^{w_2}(y) = \Big(1 - \lambda_{RN}(y)\Big)\delta_0 + \lambda_{RN}(y) T^{w_1}(\overline{\mathcal{Y}}).$$

The calculations presented in lemma 19 are sufficient to establish that $\Pi(\mathscr{A}|v_0, w_1) \leq \Pi(\mathscr{A}|v_0, w_2)$. Moreover, because v_0 is a low-effort type, lemma 17 establishes the following for all $v \in \mathcal{V}$:

$$\Pi(\mathscr{A}|v_0, w_2) \leq \Pi(\mathscr{A}|v, w_2).$$

Assembling the inequalities established thusfar yields $\Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) \leq \Pi(\mathscr{A}|\mathcal{V}, w_2)$. This completes the proof. □

Proof of proposition 14. Let $\mathcal{W}$ be any menu of contracts. Let $\mathcal{W}^* \subset \mathcal{W}$ be the set of contracts that are incentive compatible for some type $v \in \mathcal{V}$ and some technology $\mathcal{A} \in \mathscr{A}^*$, with $\mathscr{A}^*$ as defined in section C.2.1.

For each $w \in \mathcal{W}^*$, let w' be the stochastic binary contract with $T^{w'}(0) = T^w(0)$, $T^{w'}(\overline{\mathcal{Y}}) = T^w(\overline{\mathcal{Y}})$, and risk-aligned weight λ_{RN}, as defined in 3.9. Let $\mathcal{W}'$ be the set of such contracts w'. We claim that $\Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) \leq \Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}')$. First, for all $\mathcal{A} \in \mathscr{A}^*$ and $v \in \mathcal{V}$, we have:

$$V(\mathcal{A}|v, \mathcal{W}) = V(\mathcal{A}|v, \mathcal{W}')$$
$$\Pi(\mathcal{A}|v, \mathcal{W}) = \Pi(\mathcal{A}|v, \mathcal{W}').$$

Accordingly, we have:

$$\Pi(\mathscr{A}^*|\mathcal{V}, \mathcal{W}) = \Pi(\mathscr{A}^*|\mathcal{V}, \mathcal{W}'). \tag{C.46}$$

Second, because $\mathscr{A}^* \subset \mathscr{A}$, for *any* menu of contracts $\mathcal{W}$ we have:

$$\Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) \leq \Pi(\mathscr{A}^*|\mathcal{V}, \mathcal{W}). \tag{C.47}$$

We claim that inequality C.47 holds with equality for $\mathcal{W}'$. Let $\mathcal{A} \in \mathscr{A}$, $w' \in \mathcal{W}'$, and $v \in \mathcal{V}$.[3]

[3] We choose w' arbitrarily: it is not necessary for the purposes of our argument that w' be incentive compatible for type v and technology $\mathcal{A}$.

For $\mathcal{A}^T$ as defined inC.23, we have by construction:

$$V(\mathcal{A}^T|v, w') = V(\mathcal{A}|v, w').$$

Furthermore, because w' is risk-aligned per theorem 3, we have:

$$\Pi(\mathcal{A}^T|v, w') = \Pi(\mathcal{A}|v, w').$$

As $\mathcal{A} \in \mathscr{A}$, $v \in \mathcal{V}$, and $w' \in \mathcal{W}'$ were each chosen arbitrarily, we conclude that:

$$\Pi(\mathscr{A}^*|\mathcal{V}, \mathcal{W}') = \Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}'). \tag{C.48}$$

Inequalities C.46,C.47, and C.48 jointly imply, as desired:

$$\Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}) \leq \Pi(\mathscr{A}|\mathcal{V}, \mathcal{W}').$$

$\square$

REFERENCES

Philippe Aghion, Drew Fudenberg, Richard Holden, Takashi Kunimoto, and Olivier Tercieux. Subgame-perfect implementation under information perturbations. *Quarterly Journal of Economics*, 127(4):1843–1881, 2012. ISSN 00335533. doi: 10.1093/qje/qjs026.

Mark Amstrong. Multiproduct nonlinear pricing. *Econometrica*, 1996. ISSN 00129682. doi: 10.2307/2171924.

Nemanja Antic. Contracting with Unknown Technologies. Unpublished paper, 2014. URL `http://faculty.chicagobooth.edu/workshops/micro/pdf/AnticJMP.pdf`.

Sarah Auster. Robust contracting under common value uncertainty. *Theoretical Economics*, 13(1):175–204, 2018.

Daniel Barron, George Georgiadis, and Jeroen Swinkels. Optimal Contracts with a Risk-Taking Agent. Unpublished paper, 2017. URL `http://www.kellogg.northwestern.edu/faculty/georgiadis/RiskTakingContracts.pdf`.

J.E. Berg, L.a. Daley, J.W. Dickhaut, and J.R. O'Brien. Controlling preferences for lotteries on units of experimental exchange. *The Quarterly Journal of Economics*, 101(2):281, 1986. ISSN 0033-5533. doi: 10.2307/1891116. URL `http://qje.oxfordjournals.org/content/101/2/281.short`.

Dirk Bergemann and Karl Schlag. Robust monopoly pricing. *Journal of Economic Theory*, 2011. ISSN 00220531. doi: 10.1016/j.jet.2011.10.018.

Dirk Bergemann and Karl H. Schlag. Pricing without priors. *Journal of the European Economic Association*, 2008. ISSN 15424766. doi: 10.1162/JEEA.2008.6.2-3.560.

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. Informationally Robust Optimal Auction Design. Unpublished paper, 2017.

Patrick Bolton and Mathias Dewatripont. *Contract Theory*. 2005. ISBN 0262025760. doi: 10.1093/acprof:oso/9780198765615.001.0001.

Mark Braverman and Sylvain Chassang. Data-Driven Incentive Alignment in Capitation Schemes. Unpublished paper, 2015.

Benjamin Brooks. Surveying and selling: Belief and surplus extraction in auctions. Unpublished paper, 2013.

Bernard Caillaud and Jacques Robert. Implementation of the revenue-maximizing auction by an ignorant seller. *Review of Economic Design*, 2005. ISSN 14344742. doi: 10.1007/s10058-005-0125-y.

Gabriel Carroll. Robustness and linear contracts. *American Economic Review*, 105(2):536–563, 2015.

Gabriel Carroll. Informationally robust trade and limits to contagion. *Journal of Economic Theory*, 166:334–361, 2016. ISSN 10957235. doi: 10.1016/j.jet.2016.09.003.

Gabriel Carroll. Robustness and Separation in Multidimensional Screening. *Econometrica*, 85(2):453–488, 2017. ISSN 0012-9682. doi: 10.3982/ECTA14165.

Gabriel Carroll and Delong Meng. Robust contracting with additive noise. *Journal of Economic Theory*, 166:586–604, 2016a. ISSN 10957235. doi: 10.1016/j.jet.2016.10.002.

Gabriel Carroll and Delong Meng. Locally robust contracts for moral hazard. *Journal of Mathematical Economics*, 62:36–51, 2016b. ISSN 18731538. doi: 10.1016/j.jmateco.2015.11.001.

Sylvain Chassang. Calibrated Incentive Contracts. *Econometrica*, 81(5):1935–1971, 2013.

Shuchi Chawla, Jason Hartline, and Denis Nekipelov. Mechanism Redesign. 2017.

Kim Sau Chung and J. C. Ely. Foundations of dominant-strategy mechanisms. *Review of Economic Studies*, 74(2):447–476, 2007. ISSN 00346527. doi: 10.1111/j.1467-937X.2007.00427.x.

Richard Cole and Tim Roughgarden. The Sample Complexity of Revenue Maximization. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 243–252, New York, 2014. ACM.

Tianjiao Dai and Juuso Toikka. Robust Incentives for Teams. Unpublished paper, 2017.

David De Angelis and Yaniv Grinstein. Relative performance evaluation in CEO compensation: A non-agency explanation. *Working Paper, Rice University, Cornell University*, 2018. ISSN 1556-5068. doi: 10.2139/ssrn.2432473. URL `http://ssrn.com/abstract=2432473`.

David de Meza and David C. Webb. Advantageous Selection in Insurance Markets. *The RAND Journal of Economics*, 32(2):249–262, 2001.

Peter Diamond. Managerial incentives: on the Near Linearity of Optimal Compensation. *Journal of Political Economy*, 106(5):931–957, 1998. ISSN 00223808, 1537534X. doi: 10.1086/250036. URL `http://www.jstor.org/stable/10.1086/250036`.

Alexander Frankel. Aligned delegation. *American Economic Review*, 104(1):66–83, 2014.

Daniel F. Garrett. Robustness of simple menus of contracts in cost-based procurement. *Games and Economic Behavior*, 87:631–641, 2014. ISSN 10902473. doi: 10.1016/j.geb.2013.06.004.

Matthew Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, 2007. ISSN 00028282. doi: 10.1257/aer.97.3.713.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 1989.

Josef Hadar and William R Russell. Rules for Ordering Uncertain Prospects. *American Economic Review*, 59(1):25, 1969. ISSN 0036-8075. doi: 10.1126/science.151.3712.867-a. URL `http://search.ebscohost.com/login.aspx?direct=true{&}db=bth{&}AN=4500168{&}site=ehost-live`.

Lars Peter Hansen and Thomas J Sargent. *Robustness*. Princeton University Press, 2007.

Sergiu Hart and Noam Nisan. Approximate revenue maximization with multiple items. *Journal of Economic Theory*, 2017. ISSN 10957235. doi: 10.1016/j.jet.2017.09.001.

Bengt Hölmstrom. Moral hazard and observability. *Bell Journal of Economics*, 10, repr.(1): 74–91, 1979. ISSN 0361-915X. doi: 10.2307/3003320.

Bengt Hölmstrom. Moral Hazard in Teams. *The Bell Journal of Economics*, 13(2):324, 1982. ISSN 0361915X. doi: 10.2307/3003457. URL `http://www.jstor.org/stable/3003457?origin=crossref`.

Tanjim Hossain and Ryo Okui. The Binarized Scoring Rule. *The Review of Economic Studies*, 80(3):984–1001, 2013.

Robert D. Innes. Limited liability and incentive contracting with ex-ante action choices. *Journal of Economic Theory*, 52(1):45–67, 1990. ISSN 10957235. doi: 10.1016/0022-0531(90)90066-S.

Bruno Jullien, Bernard Salanié, and François Salanié. Screening risk-averse agents under moral hazard: Single-crossing and the CARA case. *Economic Theory*, 30(1):151–169, 2007.

Edi Karni. A Mechanism for Eliciting Probabilities. *Econometrica*, 77(2):603–606, 2009.

Yuichi Kitamura and Jörg Stoye. Nonparametric Analysis of Random Utility Models. *Econometrica*, 86(6):1883–1909, 2018.

Eugen Kováč and Tymofiy Mylovanov. Stochastic mechanisms in settings without monetary transfers: The regular case. *Journal of Economic Theory*, 2009.

Edward P. Lazear and Sherwin Rosen. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, 89(5):841–864, 1981. ISSN 0022-3808. doi: 10.1086/261010. URL `http://www.journals.uchicago.edu/doi/10.1086/261010`.

Javier M López-Cunat. Adverse selection under ignorance. *Economic Theory*, 16(2):379–399, sep 2000. ISSN 1432-0479. doi: 10.1007/PL00004089. URL `https://doi.org/10.1007/PL00004089`.

Kristóf Madarász and Andrea Prat. Sellers with Misspecified Models. *Review of Economic Studies*, 84:790–815, 2017. ISSN 0034-6527. doi: 10.1093/restud/rdw030.

Keler Marku and Sergio Ocampo Díaz. Robust Contracts in Common Agency. Unpublished paper, 2017.

Jamie Morgenstern and Tim Roughgarden. The Pseudo-Dimension of Near-Optimal Auctions. In *Advances in Neural Information Processing Systems*, pages 136–144, 2015.

Kevin J. Murphy. Executive compensation. In *Handbook of Labor Economics*, volume 3, pages 2485–2563. 1999.

Kevin J. Murphy. Executive Compensation: Where We Are, and How We Got There. *Handbook of the Economics of Finance*, (PA):211–356, 2013. ISSN 15740102. doi: 10.1016/B978-0-44-453594-8.00004-5.

Kevin J. Murphy and Michael C Jensen. CEO Bonus Plans: And How To Fix Them. *Harvard Business School NOM Unit Working Paper No. 12-022*, 2011. ISSN 1556-5068. doi: 10.2139/ssrn.1935654.

Michael Mussa and Sherwin Rosen. Monopoly and product quality. *Journal of Economic Theory*, 1978. ISSN 10957235. doi: 10.1016/0022-0531(78)90085-6.

Joaquín Poblete and Daniel Spulber. The form of incentive contracts: Agency with moral hazard, risk neutrality, and limited liability. *RAND Journal of Economics*, 43(2):215–234, 2012. ISSN 07416261. doi: 10.1111/j.1756-2171.2012.00163.x.

Jean-Charles Rochet and Philippe Chone. Ironing, Sweeping, and Multidimensional Screening. *Econometrica*, 1998. ISSN 00129682. doi: 10.2307/2999574.

Maxwell Rosenthal. Risk Alignment. Unpublished paper, 2018.

Maxwell Rosenthal. Robust Incentives for Risk. Unpublished paper, 2019.

Alvin E. Roth and Michael W. Malouf. Game-theoretic models and the role of information in bargaining. *Psychological Review*, 86(6):574–594, 1979. ISSN 0033295X. doi: 10.1037/0033-295X.86.6.574.

Michael Rothschild and Joseph E Stiglitz. Increasing Risk: I. A Definition. *Journal of Economic Theory*, 2(3):225–243, 1970. ISSN 00220531. doi: 10.1016/0022-0531(70)90038-4. URL `https://ideas.repec.org/a/eee/jetheo/v2y1970i3p225-243.html`.

Aviad Rubinstein and S Matthew Weinberg. Simple Mechanisms for a Subadditive Buyer and Applications to Revenue Monotonicity. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, pages 377–394, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3410-5. doi: 10.1145/2764468.2764510. URL http://doi.acm.org/10.1145/2764468.2764510.

Ilya Segal. Optimal pricing mechanisms with unknown demand. *American Economic Review*, 2003. ISSN 00028282. doi: 10.1257/000282803322156963.

Steven Shavell. Risk Sharing and Incentives in the Principal and Agent Relationship. *The Bell Journal of Economics*, 10(1):55–73, 1979. ISSN 0361915X. doi: 10.2307/3003319.

Roland Strausz. Deterministic versus stochastic mechanisms in principal-agent models. *Journal of Economic Theory*, 128(1):306–314, 2006.

Jonathan Weinstein. The Effect of Changes in Risk Attitude on Strategic Behavior. *Econometrica*, 84(5):1881–1902, 2016. ISSN 0012-9682. doi: 10.3982/ECTA13948.

Robert B. Wilson. *Nonlinear Pricing*. Oxford University Press, 1993.