Essays in Robustness and Mechanism Design

Nemanja Antić

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Economics
Adviser: Professor Stephen Morris

June 2015

Abstract

This thesis is a theoretical study of the design of optimal mechanisms and relevant robustness considerations in game theory. Chapter 1 examines contracting with moral hazard where an agent has available a known, or baseline, production technology but the principal thinks that the agent may also have access to other technologies, and maximizes her worst-case expected utilities under those possible technologies. That is, the principal aims to design a robust contract, where the level of robustness is the unknown technologies the principal thinks are possible. I show that all Pareto-efficient contracts take the form of participating preferred equity, a mixture of debt and equity. As the principal becomes more concerned with robustness the equity component of efficient contracts increases: the contracts move from debt, via participating preferred equity, to equity.

Chapter 2 (with Matias Iaryczower) studies a common feature in the design of agency relationships: that principals can decide both the direction and the scope or scale of implementation of a policy. There is a natural complementarity between these dimensions: the value of expanding the scale of implementation increases when the policy is close to a player's preferred policy. In the absence of transfers the optimal separating contract involves delegation with strings attached: an agent with an upward policy bias can only choose higher policies by reducing the scale. The solution differs qualitatively from standard quasilinear models and is ex-post inefficient, as the highest policies are too low for both parties and are under-implemented.

Chapter 3 (with Marco Battaglini) considers the robustness of inefficiency results in the literature on dynamic contribution games: a class of stochastic games where a player's action (contribution) is assumed to be monotonic. The literature finds that contributions are gradual and efficient outcomes are not achievable. In this chapter, we show that these results are not robust when some depreciation of contributions is allowed. In particular, we prove that the folk theorem holds in this setting and

thus are able to support efficient levels of the public good. This has important implications for modelling public good games, as small modelling choices deliver very different outcomes.

Acknowledgements

I am greatly indebted to my advisor Stephen Morris who has been singularly generous with his time. I am extremely appreciative of Stephen's energy, insight and continuous(ly differentiable) guidance and am proud to have been his student. Marco Battaglini and Sylvain Chassang have each contributed conspicuously to my growth as a researcher and I am gargantuanly grateful for their suggestions, support and stamina.

I have benefited from discussions (as well as classes, squash and water-cooler talk) with Dilip Abreu, Roland Bènabou, Dirk Bergemann, Marc Fleurbaey, Faruk Gul, Matias Iaryczower, Eric Maskin, Wolfgang Pesendorfer, Thomas Romer, Yuliy Sannikov, Satoru Takahashi and Juan Pablo Xandri.

Thank you to Ellen Graf and Laura Hedden for providing assistance on heterogeneous matters.

I am grateful to Ben Brooks, Lucia del Carpio, Gonzalo Cisternas, Olivier Darmouni, Leandro Gorno, Shaowei Ke, Rohit Lamba, Doron Ravid, Jay Lu, Kai Steverson, Takuo Sugaya, Cheng Sun, Juha Tolvanen and Dan Zeltzer, with whom I shared some of the lightest and brightest moments in the basement of Fisher Hall over the past five years.

I am fortunate to have a loving and understanding[1] family: my cousins Jovana, Griška and Ivan; my aunts and uncles Snežana, Zorica, Zoran and Ljova; Bika moja; my in-laws Dina, Sladja and Ljuba; my kindred spirit Rastko; and my parents Gordana and Miroljub. A benediction to Ana, whose love and support were the key ingredients in this recipe.

[1]Please take this acknowledgment as recompense for some of the neglect you experienced during the writing of this dissertation.

To my (grand)parents,

who tolerated my questions.

Contents

List of Figures

Chapter 1

Contracting with Unknown Technologies

1.1 Introduction

The theory of moral hazard is a staple of information economics. The idea that agents must be given appropriate incentives when their actions cannot be perfectly observed occurs in a wide array of applications, e.g., insurance, franchising, employment contracts, unemployment benefits, CEO compensation, financial contracting, etc. However, it is fair to say that the theory has not been very successful in explaining the types of incentive schemes we observe. This criticism is aptly summarized by Holmström and Milgrom (1987):

> ”Real world incentive schemes appear to take less extreme forms than the finely tuned rules predicted by the basic theory... Agents in the real world typically face a wider range of alternatives and principals a more diffuse picture of circumstances than is assumed in the usual models.”

In this paper, I show that the naturally appealing idea of moral hazard can indeed be used to explain the sorts of contracts we observe if we assume principals have less information about the agents available actions and require contracts robust to this.

To take a concrete example, consider the problem of security design with moral hazard. While the textbook model of this type of financial contracting attempts to argue for the efficiency of debt contracts, this result only holds if an ad-hoc restriction, that contracts are monotonic in cash flows, is added (Innes, 1990). In particular, the efficient contract without this restriction is strikingly unrealistic[1]. This leaves open the question of what are good microfoundations for a complete theory of financial contracting.

Relative to the classical literature, the present paper relaxes the assumption that at the time of contracting the principal (or investor) knows exactly the set of technologies available to the agent (or entrepreneur) to convert effort into profits. Instead, I assume that the principal knows two things: (i) a specific 'baseline' technology which will be available to the agent, and (ii) a 'lower-bound' technology that yields less surplus (total profits) than any other technology. She evaluates other possible technologies with a maxmin criterion. I show that in this 'robust contracting' setting monotonic contracts emerge because the principal is concerned that the agent might have access to a technology that exploits any non-monotonicity. I show that debt is an efficient robust contract when the lower-bound technology is close to the baseline technology. In this case, debt provides the best incentives for the agent to work hard by leaving all profits to him after a certain threshold.

However, I show that equity contracts are efficient when the principal fears that arbitrarily bad technologies could be realized. Intuitively, equity financing guarantees that the agent will not chose a technology that excessively hurts the principal, since the incentives of the two are perfectly aligned. In this case, maxmin considerations

[1]The Pareto optimal contract is a live-or-die contract, where the principal gets paid the entire profit up to some cut-off level (lives) and gets zero above that level (dies).

dominate the value of providing incentives, consistent with the results of the recent literature on robust contracting, e.g., Chassang (2013) and Carroll (2014). In fact, the shape of efficient contracts changes as the lower-bound improves towards the baseline technology: it is equity for arbitrarily bad lower-bounds, participating preferred equity (a mixture of debt and equity) for intermediate cases, and is debt when the lower-bound is close to the baseline. The provision of incentives to the agent plays an increasingly important role as the Knightian uncertainty of the principal diminishes. My framework therefore reconciles two key economic forces that determine contracts in practice: incentives and robustness concerns.

The present paper makes four contributions to the literature. I first show that in a general robust contracting framework debt contracts are efficient, in one extreme case. As an intermediary step, I provide an ambiguity foundation for the monotonicity assumption commonly made in the security design literature.[2] Secondly, I show that in another extreme case equity is efficient, in line with the robust contracting literature. Third, I show that in a general environment efficient contracts take the form of participating preferred equity: a mixture of debt and equity, including both as special cases. The difference between the worst-case and baseline technology is the key simple parameter that determines whether the optimal contract is debt, equity or a mixture of the two. Finally, I prove a technical result: in sufficiently rich maxmin contracting environments, it is without loss of generality to focus on contracts which are lower semicontinuous. There is no need for strong ex-ante restrictions on the set of allowable contracts;[3] and this technical result justifies the use of simple constructive techniques.

[2]This type of monotonicity assumption has been used by an array of authors, including DeMarzo & Duffie (1999), Matthews (2001), Biais & Mariotti (2005), DeMarzo (2005), DeMarzo, Kremer & Skrzypacz (2005), Inderst & Mueller (2006), Axelson (2007), Poblete & Spulber (2012) and Dang, Gorton and Holmstrom (2012).

[3]For example, Carroll (2014) assumes contracts are continuous.

The rest of the paper is organized as follows: Section 2 defines the model and makes some remarks about the MLRP; Section 3 makes initial general observations which are applied throughout the analysis which follows; Section 4 considers the "smallest ambiguity" extreme case and shows the Pareto optimality of debt; section 5 considers the largest possible ambiguity case and shows the efficient contract is simple equity; Section 6 provides general results that encompass the preceding observations and shows that in general participating preferred equity is optimal; Section 7 concludes.

1.2 Model

I develop a moral hazard model where the agent may have technologies which are unknown to the principal at the ex-ante contracting stage. A principal (she) contract with an agent (he), who is to take a costly, private action which will randomly produce a publicly observable profit outcome $\pi \in [0, \overline{\pi}] =: \Pi$.

More formally, an action is a pair $(e, F) \in [0, \overline{e}] \times \Delta(\Pi)$, where $e \in [0, \overline{e}]$ is interpreted as a level of effort, F is a cumulative distribution function (CDF) over profit outcomes and $\Delta(\Pi)$ is the set of Borel measures over Π, which we endow with the topology of weak convergence. The function mapping effort levels to utility cost for the agent, $c\colon [0, \overline{e}] \to \mathbb{R}_+$, is common knowledge, strictly increasing and convex. We normalize c so that $c(0) = 0$. A technology for the agent is a method for converting effort into random profit outcomes, i.e., a technology is a function $F : [0, \overline{e}] \to \Delta(\Pi)$. Instead of writing $(F(e))(\pi)$ we write $F(\pi \mid e)$. Since functions can be represented by their graphs, we can think of technology F as the graph of F:

$$\Gamma(F) = \{(e, F(\cdot \mid e)) \in [0, \overline{e}] \times \Delta(\Pi) : e \in [0, \overline{e}]\},$$

that is, technology F is simply a set of actions (where effort levels are not repeated). We assume that F is continuous in e and satisfies a stochastic concavity property

(Jewitt, 1988; Athey, 2000). These technical assumptions are sufficient to guarantee the existence of solutions to the agent's problem and are common in the classic moral hazard literature. Where it causes little confusion we will abuse notation and denote $\Gamma(F)$ by F.

The textbook models of moral hazard, starting with the classic paper by Holmström (1979), assume that there is a single profit technology, F_0, which is common knowledge. This literature requires further assumptions on the technology to deliver general results; in particular, these papers assume that F_0 satisfies the monotone likelihood ratio property (MLRP). MLRP is a natural regularity condition on the profit technology which formalizes the idea that more effort should lead to better profit distributions: it assumes that higher effort results in better distributions over profit outcomes. Consistent with this literature, we will impose that each technology F satisfies the MLRP. Note however, that we will need a more general version of the MLRP than is frequently used as we want to allow for minimization over a rich set of measures and in particular measures which do not have densities. The general definition of the MLRP, due to Athey (2002), and a discussion is given at the end of this section.

Definition 1 (Technology) *A technology is $F : [0, \overline{e}] \to \Delta(\Pi)$, a continuous map from effort levels to distributions over profit, such that F satisfies the monotone likelihood ratio order in e and F satisfies stochastic concavity, i.e., for all π, $-\int_0^{\pi} F(\pi' \mid e)\, \mathrm{d}\pi'$ is concave in e.*

I consider a robust moral hazard problem in which the assumption that there is a single common knowledge profit technology, F_0, is relaxed. In particular, the principal knows that some baseline technology F_0 is available to the agent, but there could be other, unknown, profit technologies also available. This robust contracting assumption is a version of the assumption made by Carroll (2014).[4]

[4] We will discuss the precise relationship in section 5.

On top of the baseline technology F_0, we assume that the principal knows a lower-bound CDF,[5] G, such that any realized technology (first-order) stochastically dominates G. Let the set of all possible technologies be:

$$\mathcal{D}_G := \left\{ F \in \Delta(\Pi)^{[0,\bar{e}]} : F \text{ satisfies MLRP, } \Gamma(F) \text{ compact, } F(\cdot \mid e) \leq G \text{ for all } e \right\}.$$

Note that if $G = \delta_0$, then the constraint holds trivially for any CDF F. Note that as G approaches F_0, the Knightian uncertainty of the principal is diminishing. We will consider the problem a generic lower-bond CDF, G.[6]

A contract, $B\colon [0, \overline{\pi}] \to \mathbb{R}_+$, specifies the payment made to the principal as a function of the realized profit. We assume B is measurable with respect to the Lebesgue σ-algebra (i.e., the completion of the Borel σ-algebra) and $B(\pi) \in [0, \pi]$ for all π (i.e., the investor's liability is limited to the initial investment and the entrepreneur's liability is limited to his entire profit).

The agent is a risk-neutral expected utility maximizer: given the set of technologies available to him, $\mathcal{A} = \{F_0, F_1, ..., F_N\} \subset \mathcal{D}_G$, and a contract, B, he solves:

$$\sup_{(e,F)\in\Gamma(\mathcal{A})} \int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F(\pi \mid e) - c(e), \tag{1.2.1}$$

where $\Gamma(\mathcal{A}) = \Gamma(F_0) \cup \Gamma(F_1) \cup ...\Gamma(F_N)$ is a set of actions representing the union of the possible actions under (or graphs of) the various available technologies. We let

[5] We could assume that the principal knows a lower-bound technology. As we will see, the relevant bound for the principal's worst-case analysis is profit distribution the agent can costlessly induce. As such, we can replace this assumption by a lower-bound technology. If the technology is sufficiently unproductive (a lower-bound on how effort gets converted into marginal benefit in terms of profit distributions), the analysis is unchanged.

[6] In what follows we do not need to impose any assumptions on G, except for in the proof of lemma 1.3.1. For that lemma, it is sufficient (but not necessary, in fact much weaker conditions could be given, depending on the contract) that G has a bounded derivative on $(0, \overline{\pi}]$; denote the bound on G' by $K < \infty$. Note that this still allows for non-differentiability at 0, so that G can be δ_0 for example.

$V_A(B \mid \mathcal{A})$ denote the value function of the above. Note that we assume $\mathcal{A}$ is a finite subset of $\mathcal{D}_G$, this is ensures that $\Gamma(\mathcal{A})$ is compact, given our previous assumptions.

Even after the regularity assumptions we have made, note that the supremum in the above problem may not be attained unless we further restrict the set of permissible contracts B. Although this is standard in the literature, e.g., Carroll (2014) assumes B is continuous, one of the technical results in this paper is that it is without loss of generality to assume B is lower semicontinuous, which gives that the supremum in equation 1.2.1 is attained. Thus, it will make sense to talk about the arguments which maximize the agent's utility, $A^*(B \mid \mathcal{A}) \subset \Gamma(\mathcal{A})$.

Principals are extremely ambiguity averse about the potential technologies available to the agent, but are risk-neutral with respect to risks they understand. In particular, the principal's utility:

$$V_P(B \mid F_0) = \inf_{\mathcal{A} \ni F_0} \inf_{(e,F) \in A^*(B \mid \mathcal{A})} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi \mid e),$$

subject to knowing $\mathcal{A} \subset \mathcal{D}_G$. The assumption that the principal is getting the worst possible outcome when the agent is indifferent is largely inconsequential, since the worst-case $\mathcal{A}$ will usually have a single minimizing action.[7] Furthermore, when we show that restricting to lower semicontinuous contracts is without loss of generality, we will have that the infimum above is attained and therefore we may think of it as a minimum.

[7] Brooks (2014) makes the same assumption as above, while Carroll (2014) assumes the agent maximizes the principal's utility when indifferent. The only instance in which the above is consequential is when we have a contract B and a baseline technology F_0, such that at the lowest effort level under F_0 the agent is obtaining the maximum possible profit he can get given B. Carroll (2014) rules these out by requiring contracts to be "eligibile". I make the assumption above predominantly because it avoids special cases and streamlines proofs.

We want to characterize Pareto efficient contracts in this environment. We say that contract B is efficient for technology set $\mathcal{A}$ if $\nexists B'$ such that:

$$\begin{aligned} V_P\left(B' \mid F_0\right) &\geq V_P\left(B \mid F_0\right), \text{ and} \\ V_A\left(B' \mid \mathcal{A}\right) &\geq V_A\left(B \mid \mathcal{A}\right), \end{aligned}$$

with at least one of the above inequalities strict. By varying the outside options of the parties, we hope to get a sense of outcomes under different possible market structures, i.e., a monopolist agent and competitive principals, a monopolist principal and competitive agents, etc. One motivation for looking for Pareto optimal contracts is a central planner who wants to impose efficient outcomes in these markets.

It is not immediate how a Pareto problem should be posed in this case since the agent perfectly knows the technology set $\mathcal{A}$, while the principal faces Knightian uncertainty and is not aware of the realization of $\mathcal{A}$. The idea is to give the agent any "extra" utility that results from the realized $\mathcal{A}$, while satisfying a robust utility constraint for principal. As such, given a specific technology set, $\mathcal{A}$, we want to solve for the Pareto frontier,[8] given by the following problem:

$$\begin{aligned} &\max_B V_A\left(B \mid \mathcal{A}\right) \\ &\text{s.t. } V_P\left(B \mid F_0\right) \geq R, \end{aligned} \tag{1.2.2}$$

where $R \in [0, R_{\max}]$ denotes the location on the frontier and $R_{\max}$ is the maximum payment the principal can be guaranteed (the point at which the agent's participation constraint binds). The reverse problem makes less sense as it assumes that excess utility from the unrealized set $\mathcal{A}$ is going to the principal, who does not even express a preference over this set.

[8]Note that the notion of a Pareto frontier in the textbook setting also makes reference to a specific technology; in that case there is a single technology which is common knowledge.

The key difference between a Pareto problem and a decentralized version of the above is that we are assuming away the possibility of screening or signaling. We focus on the centralized problem in the paper and I will discuss ways of decentralizing the model in section 8. The decentralization involves the agent proposing a set of contracts, from which, if the principal accepts, he can later choose any contract. This has the flavor of reverse convertible bonds/equity, where the issuer has the right to convert the contract given to the investor in some pre-agreed way.

Since our $\mathcal{A}$ is very general, and in particular does not inherit the MLRP from individual technologies,[9] we will typically need to assume some additional structure to be able to solve the above Pareto problem. In problems of this type in the classical literature, starting with Holmström (1979), without the MLRP assumption we cannot hope to provide general results. The same thing is true in the robust contracting problem, unless the robustness of the principal's preferences is simplifying the problem significantly. While this is indeed true in the largest ambiguity case, we want to consider what happens when we place limits on the principal's ambiguity. Therefore, most of the results in the paper will assume that the agent is choosing from an MLRP set of technologies, i.e., the case where $\mathcal{A}$ can be represented by some F which satisfies the MLRP. One sufficient assumption that guarantees this is that there order on technologies i, such that they respect the MLRP in this order, i.e., for $\mathcal{A} = F_0 \cup F_1 \cup F_2 \cup ... \cup F_N$ we could assume that there is a reordering of the set $\{0, 1, ..., N\}$, denoted by r, such that for each i there exists an $e_i \geq e_{i-1}$ such that:

$$
\begin{aligned}
&F_{r(i-1)}(\pi \mid e) \leq F_{r(i)}(\pi \mid e) \text{ for } e < e_i, \\
&F_{r(i-1)}(\pi \mid e_i) \overset{MLRP}{\leq} F_{r(i)}(\pi \mid e_i), \text{ and} \\
&F_{r(i-1)}(\pi \mid e) \geq F_{r(i)}(\pi \mid e) \text{ for } e > e_i,
\end{aligned}
$$

[9]We will describe this in detail in section 6.

where $e_0 = 0$.

In summary: the key features of the above assumptions is that (1) there is common knowledge of a lower-bound CDF and a baseline technology that the agent can choose and (2) we will characterize solutions to the Pareto problem, as stated in program 1.2.2, and mostly focus on the case where the agent is choosing from an MLRP set of technologies $\mathcal{A}$.

1.2.1 Aside on MLRP

In this section, I make some basic remarks regarding a key assumption underlying most classical moral hazard models, including that of Holmström (1979) and Innes (1990)—the monotone likelihood ratio property (MLRP). The simplest version considers a family of CDFs, indexed by e, i.e., $F(\pi \mid e)$, which is twice-differentiable with respect to both π and e (as is the case in Innes (1990) and most existing models). In this case, the monotone likelihood ratio property (MLRP) states that:

$$\frac{\partial}{\partial \pi}\left(\frac{f_e(\pi \mid e)}{f(\pi \mid e)}\right) \geq 0,$$

where f is the density of F.

A slightly more general definition of the MLRP, but still requiring the existence of densities, is that the likelihood ratio:

$$\frac{f(\pi \mid e_H)}{f(\pi \mid e_L)},$$

is non-decreasing for any $e_H \geq e_L$. An equivalent way to state this is to assume that f is log-supermodular, i.e., for all $\pi_H \geq \pi_L$ and $e_H \geq e_L$:

$$\frac{f(\pi_H \mid e_H)}{f(\pi_H \mid e_L)} \geq \frac{f(\pi_L \mid e_H)}{f(\pi_L \mid e_L)}.$$

Recall that a non-negative function defined on a lattice, $h\colon X \to \mathbb{R}$ is log-supermodular if, for all $x, y \in X$, $h(x \wedge y)\, h(x \vee y) \geq h(x)\, h(y)$. Note that in this version of the definition, we can also treat f as the PMF if the measure is discrete.

However, we want to allow for general distributions in the present model—e.g., distributions which involve mixtures of continuous and discrete parts. As such, we work with general probability measures from the outset and require a general MLRP. The idea is to provide a similar definition using Radon-Nikodym derivatives instead of densities, however we need to be careful to ensure the absolute continuity condition in the Radon-Nikodym theorem is satisfied.

This exact problem is addressed by Athey (2002), who gives the right generalization of the MLRP (see definition A1). We now recount this definition, specialized to our setting. For any $e_L < e_H \in \mathbb{R}_+$, define a *carrying measure* as follows:

$$C(\pi \mid e_L, e_H) = \frac{1}{2} F(\pi \mid e_L) + \frac{1}{2} F(\pi \mid e_H).$$

Importantly, note that both $F(\cdot \mid e_L)$ and $F(\cdot \mid e_H)$ are absolutely continuous with respect to $C(\cdot \mid e_L, e_H)$. We say that a family of CDFs, F, satisfies the *monotone likelihood ratio property (MLRP)* if for any $e_L < e_H$, the Radon–Nikodym derivative $h(\pi, e) : (\pi, e) \mapsto \frac{dF(\pi|e)}{dC(\pi|e_L, e_H)}$ is log-supermodular for C-a.e. (π, e), where $e \in \{e_L, e_H\}$.

To give a little intuition for this, consider the special case of differentiable CDFs. We have that:

$$\begin{aligned} \frac{dF(\pi \mid e)}{dC(\pi \mid e_L, e_H)} &= \frac{dF(\pi \mid e)/d\pi}{dC(\pi \mid e_L, e_H)/d\pi} = \frac{f(\pi \mid e)}{\frac{1}{2} f(\pi \mid e_L) + \frac{1}{2} f(\pi \mid e_H)} \\ &= 2\frac{f(\pi \mid e)}{f(\pi \mid e_L) + f(\pi \mid e_H)}. \end{aligned}$$

The MLRP states that the Radon-Nykodym derivative above is log-supermodular, or:

$$\frac{dF(\pi_H \mid e_H)}{dC(\pi_H \mid e_L, e_H)} \frac{dF(\pi_L \mid e_L)}{dC(\pi_L \mid e_L, e_H)} \geq \frac{dF(\pi_H \mid e_L)}{dC(\pi_H \mid e_L, e_H)} \frac{dF(\pi_L \mid e_H)}{dC(\pi_L \mid e_L, e_H)}.$$

We write $F(\cdot \mid e_H) \overset{MLR}{\geq} F(\cdot \mid e_L)$ if the above holds. Note that in the differentiable CDF case reduces to:

$$\begin{aligned} & \frac{f(\pi_H \mid e_H)}{f(\pi_H \mid e_L) + f(\pi_H \mid e_H)} \frac{f(\pi_L \mid e_L)}{f(\pi_L \mid e_L) + f(\pi_L \mid e_H)} \\ \geq & \frac{f(\pi_H \mid e_L)}{f(\pi_H \mid e_L) + f(\pi_H \mid e_H)} \frac{f(\pi_L \mid e_H)}{f(\pi_L \mid e_L) + f(\pi_L \mid e_H)} \end{aligned}$$

or:

$$\begin{aligned} f(\pi_H \mid e_H) f(\pi_L \mid e_L) & \geq f(\pi_H \mid e_L) f(\pi_L \mid e_H) \\ \frac{f(\pi_H \mid e_H)}{f(\pi_H \mid e_L)} & \geq \frac{f(\pi_L \mid e_H)}{f(\pi_L \mid e_L)}, \end{aligned}$$

which is one of the standard definitions given above.

1.3 Preliminary Analysis

This section makes some key observations, which will greatly simplify the proofs of the major results. While some results in this section may be of independent interest, the section may be skipped in its entirety on first reading. The key results of this section are:

- Lemma 1.3.1 which allows us to consider lower semicontinuous contracts without loss of generality, so that maximizers in the agent's problem and minimizers in the principal's problem exist;

- Lemma 1.3.2 which allows for a simpler representation of the principal's problem given by equation 1.3.2; and

- Theorem 1.3.3 which shows that only monotonic contracts are robust.

The first observation is that in finding the principal's worst-case scenario we can, without loss of generality, assume this occurs with zero effort from the agent (at least in the limit, if an argmin does not exist). Since the only guarantee the principal has is that the agent is getting at least the utility guaranteed by F_0, i.e., $V_A(B \mid F_0)$, if the infimum limiting effort level for minimizing technology F_1 was not 0, but $e^* > 0$, we can construct a new technology as follows:

$$F_1^*(\pi \mid e) = \begin{cases} F_1(\pi \mid e + e^*) & \text{if } e \in [0, \overline{e} - e^*] \\ F_1(\pi \mid \overline{e} - e^*) & \text{if } e \in (\overline{e} - e^*, \overline{e}] \end{cases} .$$

Clearly, $V_A(B \mid F_1^*) > V_A(B \mid F_1) \geq V_A(B \mid F_0)$, thus the agent's constraint is not violated. Also note that F_1^* is an MLRP family, since the MLRP is a continuous property which is preserved by limits.

Secondly, when solving for $V_P(B \mid F_0)$ it suffices to consider $\mathcal{A}$ such that $|A^*(B \mid \mathcal{A})| = 1$. We could simply take an alternative $\mathcal{A}$ which removes the technology that leads to multiple optimal choices for the agent, and the principal would be weakly worse-off. In the case where the argmax of the agent's problem does not exist, the same argument to sequences attaining the supremum for the agent. Thus, it is without loss of generality to think of the principal's preferences as:

$$V_P(B \mid F_0) = \inf_{\mathcal{A} \ni F_0} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi \mid e), \text{ subject to } (e, F(\cdot \mid e)) \in A^*(B \mid \mathcal{A}).$$

Lastly, we observe that one can without loss of generality, restrict attention to lower semicontinuous contracts. Since Innes (1990) finds that optimal contracts are not continuous (his live-or-die contract is not continuous, given our definitions), we do not wish to restrict our analysis to purely continuous contracts in the moral hazard problem presented above (for example, Carroll (2014) assumes continuous contracts). However, continuity, as well as some weaker versions of it, ensures that the infimum in the optimization problem of the principal is attained, which simplifies the analysis significantly (and is very useful when constructing worst-case scenarios in subsequent proofs). The next result shows that we can restrict attention to a class of contracts in which representative elements are lower semicontinuous.

Let $\widehat{B}$ denote the lower semicontinuous hull of B, i.e., $\widehat{B}$ is the greatest lower semicontinuous function majorized by B.

Lemma 1.3.1 *We have that* $V_P(B \mid F_0) = V_P\left(\widehat{B} \mid F_0\right)$ *and:*

$$V_P\left(\widehat{B} \mid F_0\right) = \min_{\mathcal{A} \supset A_0} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi)\,,\ \textit{subject to } F(\cdot \mid e) = A^*(B \mid \mathcal{A})\,,$$

where we consider $\mathcal{A} \subset \mathcal{D}_G$*, for some worst-case CDF* G*, where* G *is differentiable on* $(0, \overline{\pi}]$ *with* $G' \leq K < \infty$*.*

We first prove that the minimum problem is well-defined for lower semicontinuous contracts. This follows from a generalization of a classic theorem by Tonelli in the calculus of variations, as stated, for example, in Zeidler (1985) theorem 38.B, also known as the generalized Weierstrass theorem.

The intuition behind the assertion that $V_P(B \mid F_0) = V_P\left(\widehat{B} \mid F_0\right)$ is represented in figure 1.1. The figure plots both CDFs and contracts on the same axis, assuming $\overline{\pi} = 1$. The curve in red is the lower bound, or worst-case, CDF G, and the 45° line is in dashed yellow. A proposed contract, B, is in green and note that B is not lower semicontinuous. The infimum sequence of CDFs, represented in blue, puts mass on π ever closer to 0.4, as figure 1.1 shows. However, we cannot shift the mass all the way to 0.4, since this limiting CDF would result in a higher payoff to the principal. It should be clear that when the limiting CDF is considered with the lower semicontinuous hull

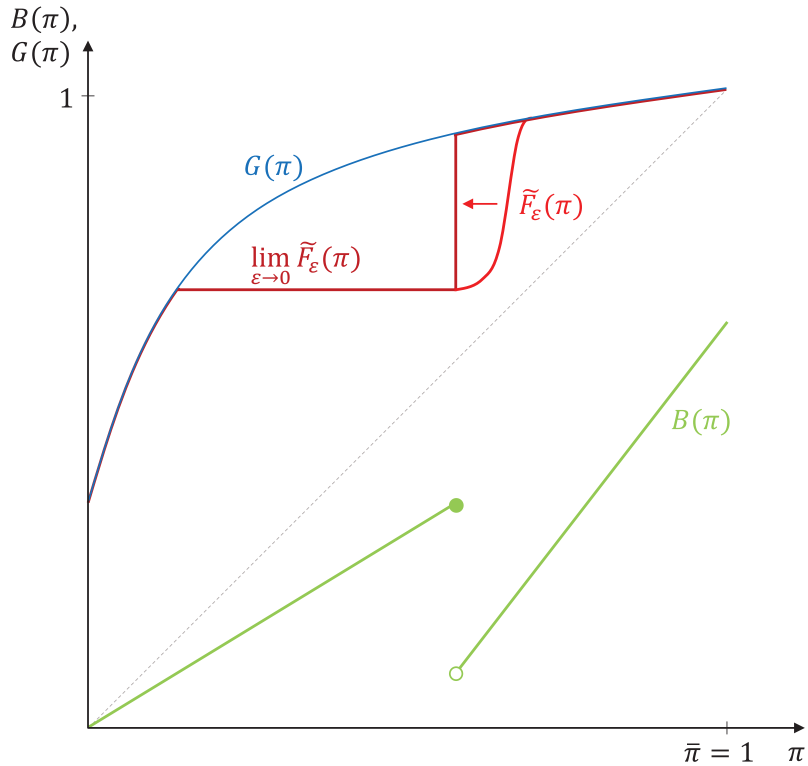


Figure 1.1: Proof idea for lemma 1.3.1.

of B (which in this case just involves moving the point at $\pi = 0.4$ down) we obtain the same payoff as the infimum of CDFs. The significance of this lemma is then also clear—we are able to look at a single minimizing CDF (the limiting CDF) instead of a sequence.

Lemma 1.3.1 therefore shows that replacing a contract by it's lower semicontinuous hull results in the same solution to the principal's problem. This is also always weakly better for the agent, thus there is no loss of generality in focusing on lower semicontinuous contracts. For any contract, B, let $\widehat{B}$ denote the lower semicontinuous hull of B; that is, $\widehat{B}$ is the greatest lower semicontinuous function majorized by B. We say that contracts B and B' are equivalent if $\widehat{B} = \widehat{B'}$, and write $B \sim B'$. We can then define an equivalence class, as follows $[B] = \left\{B' \in \mathbb{R}_+^{[0,\bar{\pi}]} : B \sim B'\right\}$. Thus the lemma implies that it is without loss of generality consider $B \in \mathbb{R}_+^{[0,\bar{\pi}]}/\sim$ and in particular we may take B to be lower semicontinuous.

Hence, we have that the principal's problem can be written as follows:

$$\max_{F\in\mathcal{A},e\in[0,\overline{e}]}\int_0^{\overline{\pi}}(\pi - B(\pi))\ \mathrm{d}F(\pi\mid e) - c(e), \tag{1.3.1}$$

and denote by $V_A(B\mid\mathcal{A})$ and $A^*(B\mid\mathcal{A})$ the value function and argmax of the above, respectively. Note that these are well-defined since $\pi - B(\pi)$ is upper semicontinuous.

The principal who faces unknown technologies can still bound her payoff. In particular she has:

- The knowledge that the agent will not choose something worse than he was getting under F_0, the baseline technology, and
- The knowledge that all technologies must dominate the worst case G.

These lead to a representation of the principal's preferences which makes plain the Gilboa-Schmeidler maxmin preference of the principal, since she is minimizing over a set of measures.

Lemma 1.3.2 *We have that $V_P(B\mid F_0)$ is the solution to:*

$$\min_{F\le G}\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi),\ \textit{subject to}\ \int_0^{\overline{\pi}}(\pi - B(\pi))\ \mathrm{d}F \ge V_A(B\mid F_0). \tag{1.3.2}$$

Furthermore, if $B(\pi)$ and $\pi - B(\pi)$ are monotonic the constraint above holds as an equality.

The lemma is a generalization of similar observations made in theorem 1 in Chassang (2013) and lemma 2.2 in Carroll (2014). In a moral hazard setting both of these papers find that the principal can essentially only bound her utility by the knowledge that the agent will not choose a worse outcome than what he is guaranteed under the known technology. Madarász and Prat (2014) exploit a similar argument in a screening setting. The main differences between my proof and earlier literature arises from complications when there is a non-trivial lower-bound (when $G \neq \delta_0$) and the assumption that the principal fears getting the worst possible outcome when the agent is indifferent.[10]

We can generally think of the constraint in program 1.3.2 as tight. Holmström (1979) and Shavell (1979) point out that the monotonicity of $\pi - B(\pi)$ follows directly from the definition of the MLRP for optimal B, and we will show next that it is without loss of generality to focus on monotone B.

1.3.1 Robustness of Monotone Contracts

Given the above preliminaries, the key assertion of this subsection is that robustness considerations lead to monotonic contracts. The intuition for this is that a principal facing a non-monotonic contract will assume that a productive technology which exploits the non-monotonicity will be available to the agent and therefore disregard any non-monotonic aspects of the contract.

Theorem 1.3.3 *For any G and any non-monotonic contract $B(\pi)$ there exists a monotonic contract $B_m(\pi)$ such that:*

$$\inf_{\mathcal{A} \subset \mathcal{D}_G} \int_0^{\overline{\pi}} B_m(\pi) \, \mathrm{d}F_m^{\mathcal{A}}(\pi) = \inf_{\mathcal{A} \subset \mathcal{D}_G} \int_0^{\overline{\pi}} B(\pi) \, \mathrm{d}F^{\mathcal{A}}(\pi),$$

subject to $F_m^{\mathcal{A}} \in A_A^*(B_m \mid \mathcal{A})$ and $F^{\mathcal{A}} \in A_A^*(B \mid \mathcal{A})$, i.e., the principal is indifferent between the two contracts, and $B(\pi) \geq B_m(\pi)$, i.e., the agent's prefer the monotonic contract.

The intuition for the above result is given in figure 1.2.

The idea is that if a principal is offered a non-monotonic contract, contract B in figure 1.2, she would discount the non-monotonic part, since in the worst-case analysis

[10]The latter assumption allows us to state the lemma without reference to "eligible" contracts.

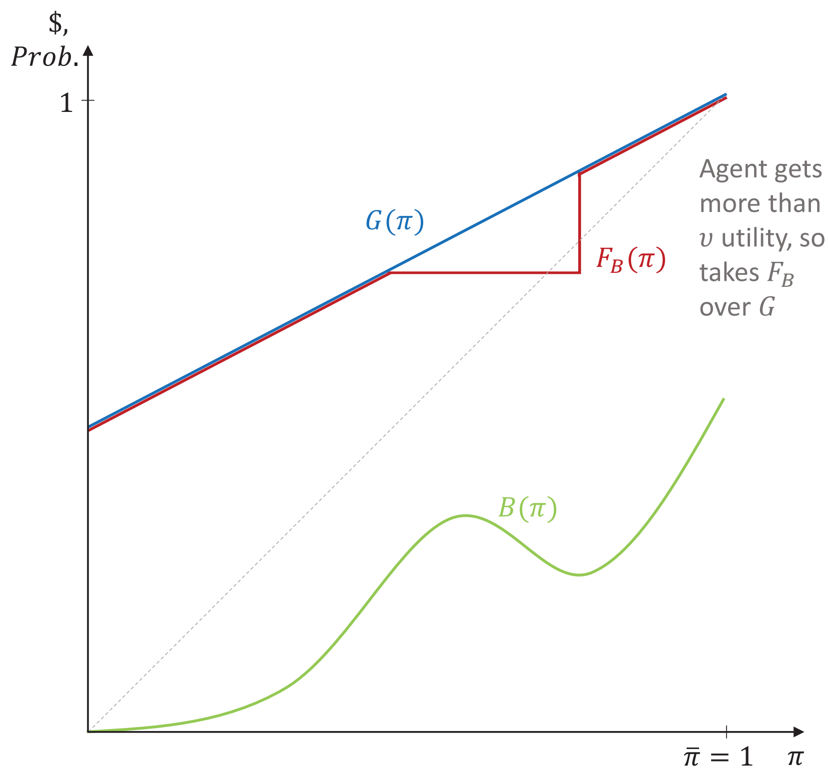


Figure 1.2: Proof idea for theorem 1.3.3.

she thinks that nature will endow the agent with a good technology which puts no mass on the non-monotonic part.

Since $F_0(\pi \mid e) = G(\pi)$ for all e, the two ways the principal can bound her payoff, the lower-bound and the agent's utility under the reference technology, are one and the same. In particular, if any technology better than G was available, the principal could only improve her payoff since contracts are monotonic. Thus the principal's worst-case in this extreme of the model is simply that only G is available. Note however that ambiguity still has a role: it is critical in proving theorem 1.3.3, which says that robust contracts are monotonic. Aside from this however, the key concern is the provision of incentives to the agent, as in the textbook model.

1.4 Smallest Ambiguity, $F_0(\cdot \mid e) = G$

This section and the next consider the two extreme cases of the model and build intuition for the results. This section considers the smallest ambiguity case, where the lower-bound (G) and reference (F_0) technologies are the same. I will show that debt contracts are optimal in this extreme. We will subsequently analyze the largest ambiguity case and with the intuition of these extremes proceed to the general results.

The main result of this section will be the optimality of debt contracts when the agent is choosing from an MLRP set. The proof idea is the same as in Innes (1990). We show that a non-debt contract induces a lower effort choice than a debt contract which gives the principal the same payoff, and that this is below the first-best level of effort. There are several complications in this version: we need to generalize the argument to allow for non-differentiability of CDFs and we need to be careful since ambiguity considerations are important when we are replacing contracts.[11]

1.4.1 Remarks about the Model

The model we have presented is sufficiently well-behaved. This section makes some general remarks that could be skipped on first reading.

Remark 1 *For any bounded, continuous function ϕ, $\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e)$ is continuous in e.*

Since $F(\pi \mid \cdot)$ is continuous in e, this observation follows directly from the portmanteau theorem. The first-best effort level would be the amount of effort chosen if the agent owned the firm. This is of course not feasible due to limited liability assumption. Note that the first-best problem, if the agent is using technology F, is:

$$\max_{e \geq 0} \left\{ \int_0^{\overline{\pi}} \pi\ \mathrm{d}F(\pi \mid e) - R - c(e) \right\}.$$

[11] In particular, replacements need to be done with respect to the worst-case of the principal, as opposed to the commonly known technology.

Remark 2 *The above has a unique solution with a positive effort level $e^* > 0$.*

This is true since c is strictly convex in e and $\int_0^{\pi'} F(\pi \mid e)\ d\pi$ is convex in e for all π'. By Athey (2002), the latter condition implies that $\int_0^{\overline{\pi}} \pi\ dF(\pi \mid e)$ is concave in e, which is sufficient for the existence and uniqueness of a solution. Under these assumptions, the first-order condition[12] for this problem is:

$$\frac{\partial}{\partial e}\left(\int_0^{\overline{\pi}} \pi\ dF(\pi \mid e)\right) - c_e(e) = 0.$$

This equation is strictly decreasing in e and is positive by assumption for $e = 0$, thus by the mean value theorem there is a unique e^* which solves the above first-order condition.

Through a similar argument we see that contracts, like equity and debt lead to unique solutions for the agent's problem. In particular, the following has a unique solution:

$$\max_{e \geq 0} \int_0^{\overline{\pi}} (\pi - \min(\pi, z))\ dF(\pi \mid e) - c(e),$$

for $z \geq 0$. Denote the solution by $e^*(z)$ and note $e^*(z)$ is continuous in z, by Berge's maximum theorem.

1.4.2 Result

The main result of this section is that Pareto optimal contracts take the form of debt.

Theorem 1.4.1 *For any $\mathcal{A} \subset \mathcal{D}_G$ where the agent is choosing from an MLRP set, a solution to:*

$$\max_B V_A(B \mid \mathcal{A}),$$

$$\textit{subject to} \qquad V_P(B \mid F_0) \geq R,$$

[12]The display expression is assuming differentiability with respect to e, which is a special case of our model. In general the idea is exactly the same and the special case is shown in this case for simplicity.

is $B^D_z(\pi) := \min(\pi, z)$ *for some* $z \in [0, \overline{\pi}]$.

The proof of the theorem goes by showing that when a monotonic contract is replaced by an appropriately chosen debt contract the agent is induced to put in more effort because of the MLRP. This is a key property of the MLRP and is summarized in the following lemma.

Lemma 1.4.2 *Let* $\phi(\pi)$ *be a function such that* $\phi(\pi) \geq 0$ *for* $\pi \leq \pi_B$, $\phi(\pi) \leq 0$ *for* $\pi \geq \pi_B$ *and either:*

1. $\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) = 0$, *or*

2. $\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) \leq 0$ *and* $\phi(\pi)$ *decreasing for* $\pi \geq \pi_B$.

Then, for any $e_H > e_L$ *and any MLRP family* F, *we have that* $\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) \geq \int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_H)$.

This is a generalization of lemma 1 from Innes (1990). The proof technique is similar, but needs to take care of technical difficulties arising from the non-existence of densities. The lemma is key in the proof of the main theorem, since it says that replacing generic monotone contracts by debt contracts implies higher marginal benefits of effort.

Note that the inequality in the above theorem holds strictly if the MLRP is strict. This observation will be useful for the uniqueness result that follows.

Corollary 1.4.3 *The repayment level,* z, *in the optimal contract,* $B^D_z(\pi) = \min(\pi, z)$, *is increasing in* R *and decreasing in* G.[13]

The above corollary follows since the level of debt is chosen so as to guarantee the principal the required utility R under the worst-case scenario where G and only G is available. This implies that the level of repayment z is increasing in R. Furthermore if $G \leq G'$, the level of repayment required under G' would be greater than under G.

[13]If we think of potential G CDFs as being ordered by stochastic dominance.

Corollary 1.4.4 *Debt is the unique solution to the above problem if $R \in (0, R_{\max})$, F satisfies strict MLRP and G has full support.*

1.4.3 Numerical Example

To demonstrate why the efficiency question is interesting, let us consider a simple example. Let $\Pi = [0,1]$, $e = [0,1]$, $c(e) = \frac{1}{10}e^2$ and $G = F_0 = U[0,1]$. Fix a level of principal utility R. As discussed, the worst-case scenario for the principal is that only the (constant) technology F_0 is available to the agent[14]. The principal is thus indifferent between many contracts. In particular, the principal is indifferent between an equity and debt contract defined as follows:

$$\begin{aligned} B^E_\alpha(\pi) &= \alpha\pi, & \text{with } \alpha = 2R \\ B^D_z(\pi) &= \min(\pi, z), & \text{with } z = 1 - \sqrt{1-2R}, \text{ and} \end{aligned},$$

since:

$$\int_0^1 B^D_z(\pi)\ \mathrm{d}G(\pi) = 2R\int_0^1 \pi\ \mathrm{d}\pi = R,$$

and:

$$\begin{aligned} \int_0^1 B^D_z(\pi)\ \mathrm{d}G(\pi) &= \int_0^{1-\sqrt{1-2R}} \pi\ \mathrm{d}\pi + \left(1-\sqrt{1-2R}\right)\left(1 - G\left(1-\sqrt{1-2R}\right)\right) \\ &= \frac{\left(1-\sqrt{1-2R}\right)^2}{2} + \left(1-\sqrt{1-2R}\right)\sqrt{1-2R} \\ &= 1-\sqrt{1-2R} - R \quad + \quad 2R + \sqrt{1-2R} - 1 = R. \end{aligned}$$

Consider now an agent with the following technology set $\mathcal{A}$:

$$F(\pi \mid e) = \pi^{e+1}, \text{ for } e \in [0,1].$$

[14] This is because contracts have to be monotonic and the agent gets his "promised" utility under G.

Note that this is an MLRP technology set and that $F(\pi \mid e) \leq G(\pi)$ for all e. Figure 1.3 plots the utilities of the agent under the two contracts above, given different possible reservation utilities of the principal R.

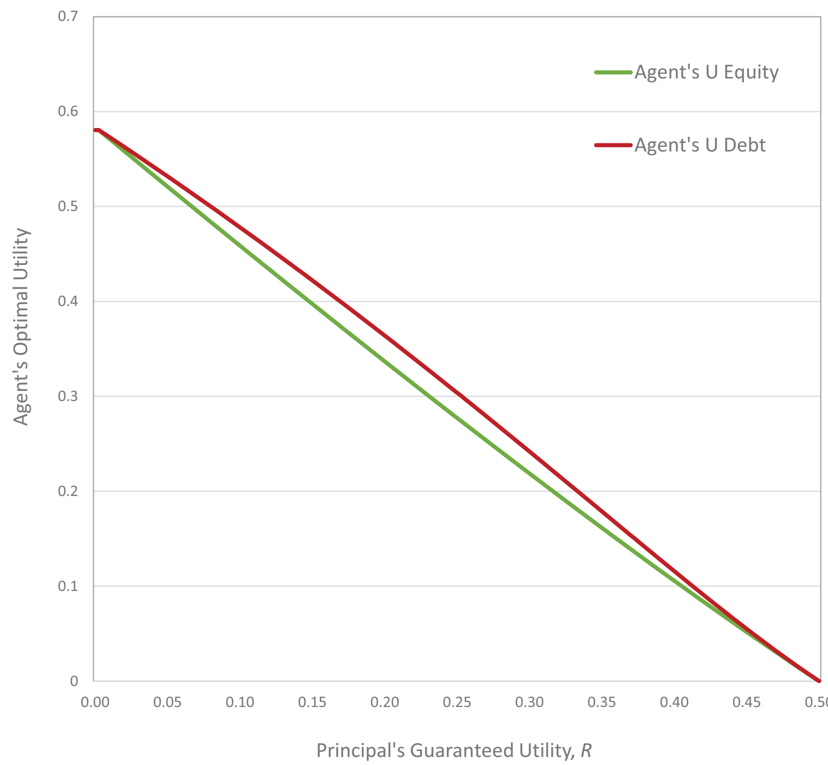
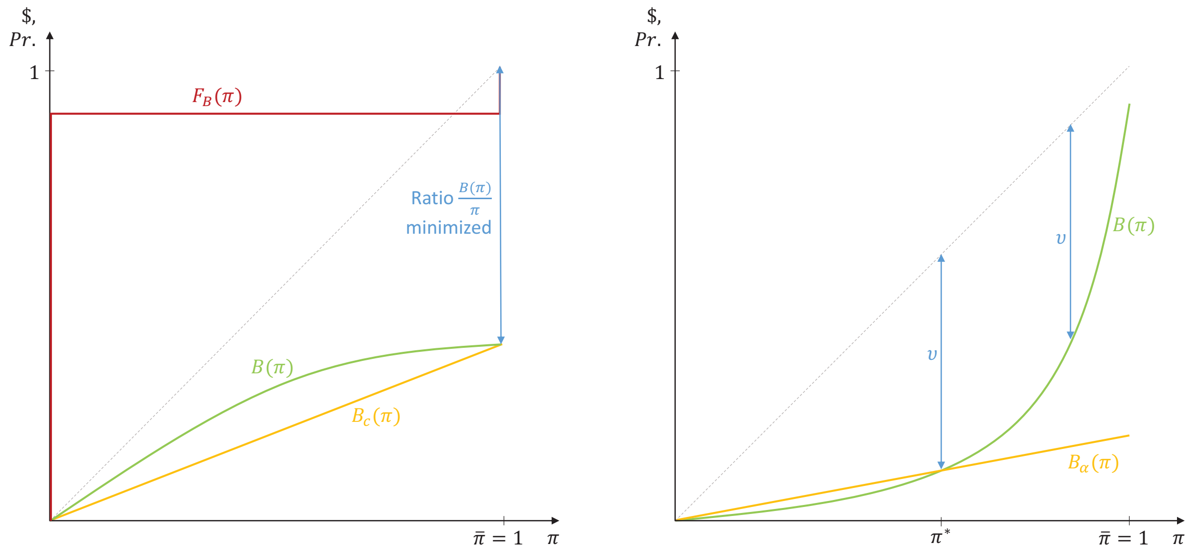


Figure 1.3: Numerical example illustrating Pareto Efficiency

We see in figure 1.3 that although the principal is indifferent between the contracts, the agent clearly prefers the debt contract for all $R \in (0, R_{\max})$. Note that in this case $R_{\max} = 1/2$. When $R = 0$ or $R = R_{\max}$ the debt and equity contracts are the same—they either award all profit to the agent or principal.

1.5 Largest Ambiguity, $G = \delta_0$

We now consider the case where the lower-bound CDF is trivial. In this extreme version of the model the maxmin aspect of the principal's preferences really restricts



Optimal contract is (weakly) convex. Optimal contract is linear.

Figure 1.4: Proof idea for theorem 1.5.1.

what is achievable and we do not need to make further assumptions about $\mathcal{A}$. As such we can consider general sets $\mathcal{A} \subset \mathcal{D}$, and in particular we do not need to assume that the agent is choosing from an MLRP set of technologies.

Theorem 1.5.1 *For any $\mathcal{A} \subset \mathcal{D}$, a solution to:*

$$\max_{B} V_A (B \mid \mathcal{A}),$$

$$\textit{subject to} \qquad V_P (B \mid F_0) \geq R,$$

is $B_\alpha(\pi) = \alpha\pi$ for some $\alpha \in [0, 1]$, i.e., a linear/equity contract.

The intuition for this proof is that an extremely uncertain principal places a huge premium on having preferences perfectly aligned with the agent, which is what happens when the contract is linear. Even if there are efficiency gains from providing stronger incentives for the agent at the upper end of profit outcomes, as is the case when $\mathcal{A}$ is an MLRP set, this benefit is over-ridden by the principal's pessimism.

The proof of theorem 1.5.1 is illustrated in figure 1.4. The left-hand panel gives the intuition for why contracts have to be (weakly) convex. In particular, consider a concave contract B (in green). In performing her worst-case analysis, the principal is wants to find the worst way (for her) that the agent can gain exactly the utility guaranteed by F_0, $\upsilon := V_A(B \mid F_0)$. Given that the set of CDFs she can minimize over is unrestricted, she will put mass on just two points: there will be a lot of mass on 0, since this gives her no payoff, and just enough mass on the point which minimizes $\frac{B(\pi)}{\pi}$, i.e., the point which minimizes what the principal gets relative to what the agent gets. In this case "just enough" means to make the agent choose this constructed CDF (at zero effort cost) over whatever was optimal in F_0. This worst-case CDF is illustrated by F_B in the figure. Now, consider replacing B by the lower convex hull, B_c. Note that at the worst-case the principal is indifferent between B and B_c. Furthermore, since B_c is linear, it satisfies a "no-weak-point" constraint, so that the minimizing CDF for the principal is any CDF which delivers the required utility to the agent—including F_B. This replacement therefore makes the principal no worse off, but makes the agent weakly (and generally strictly) better off.

The right-hand panel in figure 1.4 provides intuition for why contracts have to be linear. In particular, consider the principal's worst-case analysis when faced with a convex contract B, where the agent is guaranteed some level of utility υ. Jensen's inequality implies that the worst-case scenario is a dirac distribution δ_{π^*} at the lowest level of profit which gives the agent exactly utility υ. One can replace B by a linear contract B_α that goes through $(\pi^*, B(\pi^*))$ and we again note that the principal is no worse off. It is not immediate that the agent likes this replacement however, since there is an interval, $[0, \pi^*]$, on which $B_\alpha > B$. The agent does like this replacement however—since the agent's average payoff under whatever technology he was choosing from $\mathcal{A}$ is at least υ, it cannot be the case that the agent is putting much mass on $[0, \pi^*]$ relative to the mass this CDF puts on $[\pi^*, \overline{\pi}]$. Another application of Jensen's inequality ensures that this replacement indeed gives the agent higher utility (and strictly higher if the agent's chosen distribution is not δ_{π^*}).

We say that $\mathcal{A}$ has full support, if for all $F_i \in \mathcal{A}$ and $e \in [0, \overline{e}]$, $supp(F_i(\cdot \mid e)) = [0, \overline{\pi}]$.

Corollary 1.5.2 *Equity is the unique solution to the above problem if* $R \in (0, R_{\max})$ *and* $\mathcal{A}$ *has full support.*

The equity contract is the unique efficient contract if the agent's technologies have sufficiently large support.

1.5.1 Relationship with Carroll (2014)

The robust contracting framework of Carroll (2014) maps directly to the largest ambiguity case analyzed above. One difference is that Carroll (2014) focuses on unknown actions, as opposed to technologies, and does not require MLRP. Our choice of focusing on technologies is inspired by the classical literature on contract theory which imposes natural restrictions such as the MLRP. However, as we noted in the model section, if we assume nothing about how these technologies are inter-related there is no bite to the MLRP assumption.

In particular, an action in Carroll's setup can be converted to a technology as follows. Let (F, e) be an action available to the agent in Carroll's model. We can define an MLRP technology, from which (F, e) will be chosen if it dominates the zero action, as follows:

$$F_i(\cdot \mid e') = \begin{cases} \delta_0 & \text{if } e' < e \\ F & \text{if } e' \geq e \end{cases}.$$

Note that F_i, as defined above, satisfies the generalized MLRP.[15]

The key difference is that Carroll (2014) solves the principal-optimal problem. The main result is presented below.

[15] Note however that F_i is not continuous in e (although this could be easily modified through a standard mollifier construction) and that F_i fails stochasic concavity. Both of these assumptions on technologies could be dropped without affecting any results in this section.

Theorem 1.5.3 (Carroll, 2014) *A solution to:*

$$\max_{B} V_P\left(B \mid F_0\right),$$
$$\textit{subject to} \qquad V_A\left(B \mid \mathcal{A}\right) \geq 0,$$

is $B_\alpha\left(\pi\right) = \alpha\pi$ *for some* $\alpha \in (0, 1)$.

The above is a linear contract or, in our security-design-inspired language, the solution to the principal problem is an equity contract. As we illustrated in the discussion in section 5, this does not necessarily imply that the equity contract is efficient. However, Carroll (2014) also shows a uniqueness result: under the same conditions as in corollary 1.5.2, equity is the unique principal-optimal contract. This implies that equity must also be the efficient contract in that case. Theorem 1.5.1 extends these results by observing that equity is an efficient contract even when the uniqueness results fail.

1.6 General Results

The general model can now be analyzed by combining the insights from the study of the extreme cases in the two preceding sections.

We say that a technology set $\mathcal{A}$ is strongly better than the lower-bound technology G if any $F \in \mathcal{A}$ we have that $G \overset{MLR}{\leq} F$. This implies the first-order stochastic dominance assumption we already had, but adds to it somewhat. The fact that we can get general results only when this stronger dominance condition holds should not be surprising—the MLR was identified by the classic literature as the correct notion of unambiguously improving on a random profit technology.

Theorem 1.6.1 *For any $\mathcal{A} \subset \mathcal{D}_G$ strongly better than G, where the agent is choosing from an MLRP set, a solution to:*

$$\max_B V_A (B \mid \mathcal{A}),$$
$$\text{subject to} \qquad V_P (B \mid F_0) \geq R,$$

is $B^P_{\alpha,z}(\pi) = \min(\pi, z + \alpha\pi)$ for some $z \in [0, \overline{\pi}]$, $\alpha \in [0,1]$ with $z + \alpha\overline{\pi} < \overline{\pi}$.

The Pareto optimal contract $B^P_{\alpha,z}$ is participating preferred equity. Participating preferred equity contracts can be thought of as a mixture of debt and equity. An investor issues a debt component and an equity component—the investor is entitled to all profit up to the repayment level of $\frac{z}{1-\alpha}$ and is then entitled to an additional α share of any profit above this level. The class of preferred equity contracts includes the simple debt and equity contracts we proved were efficient in previous sections. In particular, $\alpha = 0$ implies that $B^P_{\alpha,z} = B^D_z$ or simple debt, while $z = 0$ implies $B^P_{\alpha,z} = B^E_\alpha$ or simple equity.

The intuition for this result is rather simple. We have already seen that debt contracts are good for incentive provision. However, when $V_A (B \mid F_0)$ is sufficiently bigger than $V_A (B \mid G)$ we can make an unambiguous improvement if we are starting from the debt contract, as shown in figure 1.5.

Take any contract B, shown in green in figure 1.5, and consider the minimizing CDF subject to some arbitrary G. The worst-case CDF, F_B, for contract B is shown in red in the figure. This CDF has the feature that it puts mass on an interval of small profit realizations and on $\overline{\pi}$, as $\overline{\pi}$ minimizes the ratio of what the principal gets relative to the agent and thus this is the most costly way for the agent to get at least utility $V_A (B \mid F_0)$. The same logic was used in section 6 when we deduced the Pareto optimality of equity contracts.

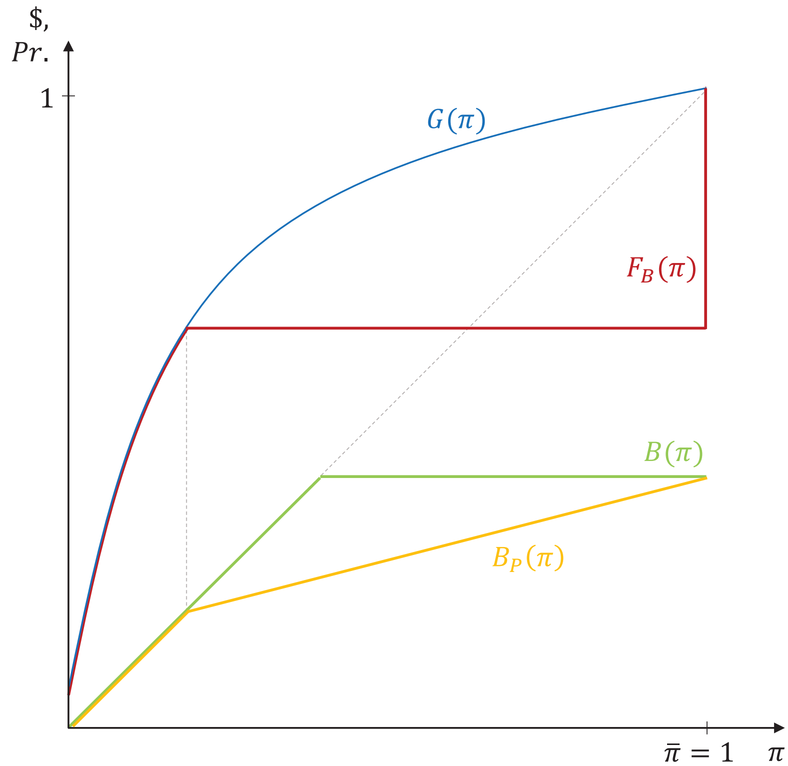


Figure 1.5: Proof idea for theorem 1.6.1.

Given the minimizing CDF, F_B, we see that B can be replaced by B_P, a preferred equity contract that is the lower convex hull of contract B on the region where F_B had no support. Note that a minimizing CDF for contract B_P is still F_B, thus the principal is indifferent to this change. The agent clearly prefers contract B_P since $B_P \leq B$.

Corollary 1.6.2 *The principal's payoff from contract* $B(\pi) = \min\{\pi, z + \alpha\pi\}$ *is:*

$$R = \int_0^{\frac{z}{1-\alpha}} \pi \, \mathrm{d}G + \frac{\alpha V_A(B \mid F_0) + z\left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)}.$$

This corollary gives a relation between α and z, in terms of the known parameters of the model—the lower-bound CDF G and the utility afforded to the agent under technology F_0, $V_A(B \mid F_0)$.

If $G = \delta_0$, then for any $z \geq 0$ we have that:

$$\begin{aligned} R &= \int_0^{\frac{z}{1-\alpha}} \pi \,\mathrm{d}G + \frac{\alpha V_A (B \mid F_0) + z \left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)} \\ &= \frac{\alpha V_A (B \mid F_0)}{(1-\alpha)}, \end{aligned}$$

thus the agent is (at least weakly) better off by setting $z = 0$, since any $z > 0$ is dominated. If $G = F_0$, and since worst-case scenario for Principal is G:

$$\begin{aligned} R &= \int_0^{\frac{z}{1-\alpha}} \pi \,\mathrm{d}G + \frac{\alpha V_A (B \mid F_0) + z \left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)}, \\ &= \int_0^{\frac{z}{1-\alpha}} \pi \,\mathrm{d}G + \int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\alpha\pi + z) \,\mathrm{d}G, \end{aligned}$$

which implies that we must have:

$$\int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\alpha\pi + z) \,\mathrm{d}G = \frac{\alpha V_A (B \mid F_0) + z \left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)}.$$

The above holds when $\alpha = 0$ and, as we argued earlier, such a debt contract provides the best incentives when the agent is choosing from an MLRP set in the smallest ambiguity case.

The characterization in corollary 1.6.2 implies that as G improves towards F_0 the set of z and a pairs which are undominated increases continuously (as the expression is continuous in G). For G sufficiently close to F_0 debt contracts become possible, however they may not be chosen for every realization of the technology set since the repayment level, z, may be too high. When G gets even closer to F_0, the repayment level decreases and debt contracts are certainly efficient when $G = F_0$ for any MLRP realization of $\mathcal{A}$. For sufficiently good realized technology sets, debt becomes Pareto optimal for G where $F_0 < G$.

1.7 Discussion

1.7.1 Decentralization

This paper has considered Pareto problem motivated by a central planner who cares about efficiency. Is there a way to decentralize the problem? The key thing we need to be careful about is avoiding any possibility of signaling.

Consider the following timing:

1. The agent, knowing his realized technology set $\mathcal{A}$, as well as the information available to the principal (the baseline technology F_0, and lower-bound G) prosposes a set of contracts $\mathcal{B}$;

2. The principal accepts or rejects the set of contracts $\mathcal{B}$, based on the understanding that the agent will be able to select any $B \in \mathcal{B}$. The principal has an ex-post utility constraint, so that he will accept the set of contracts if for any $B \in \mathcal{B}$, $V(B \mid F_0) \geq R$;

3. The agent chooses some $B^* \in \mathcal{B}$ and some $(e, F) \in \mathcal{A}$;

4. Nature realizes profits and they are shared: the principal gets $B^*(\pi)$ and the agent gets $\pi - B^*(\pi)$.

In this game it is a weakly dominant strategy for the agent to propose the largest set of contracts that will get accepted, which is any contract B for which $V(B \mid F_0) \geq R$. This is related to Myerson's (1983) *principle of inscrutability* in informed principal models: the informed party (the principal in Myerson's model, the agent here) should not want to reveal their private information if they can help it.

The maxmin preference of the principal actually helps us here. If the principal was Bayesian and had an ex-ante participation constraint, signaling could be helpful as it could indicate to the principal that "certain technologies are unlikely" and can therefore relax the participation constraint in favor of the agent. With a maxmin principal this signaling benefit is not relevant, since for any contract and any technology realization the principal has to get the required return R; there is no incentive compatible sense in which this could be relaxed.

In the financial contracting interpretation, this decentralization looks like a reverse convertible bond. With such a contract the issuer has the right to convert between a pre-agreed set of contracts, e.g., this maps to the choice that the agent gets from the set $\mathcal{B}$.

1.7.2 Concluding Remarks

This paper introduced a general model of robust contracting when the principal does not know ex-ante all of the profit technologies available to the agent. The relaxation of this assumption of the textbook financial contracting model gives us a lot of traction. Firstly, it provides a complete theory of debt contracts which was the goal of this classic literature. Secondly, it shows that other, readily observable, contracts such as equity are Pareto efficient. More generally, these are examples of contracts in the class of participating preferred equity which we find to be Pareto optimal in a general environment. While debt and equity are clearly common contracts, empirical work on venture capital also suggests participating preferred equity is not uncommon in practice.[16] The key empirical implication of the results is that we should see firms in 'new' industries (such as social networking or biotech startups), where investors have little prior experience and face a lot of ambiguity about how the firm is going to generate profits, funded by equity contracts. Conversely, firms in 'established' industries (such as restaurants or accounting offices), where investors have a lot of experience and face less Knightian uncertainty, should be financed by debt.

[16] Kaplan and Strömberg (2003) find that 40 percent of venture capital funding rounds in their data set involve participating preferred equity.

In a very different set of models, focusing on costly information acquisition instead of moral hazard, a similar type of empirical prediction results. Dang, Gorton and Holmstrom (2012) and Yang (2013) find that in cases where information acquisition by the investor is not socially optimal (e.g., if the project is in a well-established industry), debt contracts should be observed as they provide the worst incentives for costly information acquisition. Yang and Zeng (2014) generalize this model and find that if there are enough benefits from information acquisition by the investor (e.g., if the project is in a new industry), the class of participating (convertible) preferred equity contracts might be optimal.

The analysis in this paper makes headway using two key simplifying assumptions. Firstly, in obtaining general results, we are restricting the analysis to cases where the agent is choosing from MLRP technology sets. This is to be expected, as the classic contracting literature also requires MLRP technologies to provide general conclusions, but it does leave open the question of what we can say without any restriction on the technology sets. Secondly, in the principal's minimization problem we are allowing for a rich set of distributions for the principal to minimize over.[17] While we reduce the ambiguity of the principal by decreasing the size of the minizing set (by increasing G), the techniques employed require this set to be rich (i.e., all CDFs that dominate G). It is natural to consider what happens when the richness of the minimizing sets is somehow restricted. These are two possible directions for future research.

[17]We also do not impose an upper-bound technology, but this is of far less improtance.

1.A Omitted Proofs

1.A.1 Proof of Lemma 1.3.1

Lemma 1.A.1 *We have that* $V_P(B \mid F_0) = V_P\left(\widehat{B} \mid F_0\right)$ *and:*

$$V_P\left(\widehat{B} \mid F_0\right) = \min_{\mathcal{A} \supset A_0} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi)\,,\ \textit{subject to}\ F(\cdot \mid e) = A^*(B \mid \mathcal{A})\,,$$

where we consider $\mathcal{A} \subset \overline{\mathcal{D}}_G$*, for some worst-case CDF* G*, where* $G = \delta_0$ *or* G *is differentiable (has a density), with bound* $K < \infty$*.*

Proof. We shall treat the set of relevant CDFs as a subset of $L^2([0, \overline{\pi}])$. Clearly L^2 is a reflexive Banach space, since it is a Hilbert space. Furthermore, the set $\{F : F \leq G\}$ is bounded and closed in L^2. Note that in analyzing the principal's worst-case scenario, we have that the worst case is achieved at 0 effort (observation 1) in the limit and since $\widehat{B}$ is lower semicontinuous, the MLRP restriction plays no role in constraining the set of CDFs as:

$$\lim_{e \to 0} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi \mid e) \geq \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi \mid 0)\,.$$

We also note that the agent's problem with $\widehat{B}$ attains the solution and that by the generalized theorem of the maximum $A^*\left(\widehat{B} \mid \mathcal{A}\right)$ is closed.

We are left to show that the functional above is weak sequentially lower semicontinuous, i.e., for any sequence $\|F^n - F\|_2 \to 0$, we have that:

$$\liminf_{n \to \infty} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F^n(\pi) \geq \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi)\,.$$

To see this, note that by Hölder's inequality $\|F^n - F\|_1 \leq \sqrt{\lambda([0, \overline{\pi}])}\,\|F^n - F\|_2 \to 0$, and L^1 convergence of CDFs on metric spaces of bounded diameter implies weak convergence of measures. Therefore, the above inequality follows directly by the portmanteau theorem, since $\widehat{B}$ is lower semicontinuous and bounded from below by 0[18]. This proves the second claim.

For the first claim, for any $\mathcal{A}$, $V_A(B \mid \mathcal{A}) \leq V_A\left(\widehat{B} \mid \mathcal{A}\right)$, since $B \geq \widehat{B}$, and similarly:

$$\inf_{F \leq G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi) \geq \min_{F \leq G} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi).$$

To complete the proof, we will show that for any $\varepsilon > 0$:

$$\inf_{F \leq G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi) \leq \min_{F \leq G} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi) + \varepsilon.$$

Let $\widehat{F} \in \left\{\arg\min_{F \leq G} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi)\right\} \neq \emptyset$. Note that by definition[19] of the lower semicontinuous hull $\widehat{B}(\pi) = \lim_{\varepsilon \to 0^+} \inf_{\{\pi' :\ |\pi - \pi'| < \varepsilon\}} B(\pi')$. Therefore, for any $\varepsilon > 0$, there exists a $e > 0$ such that for all $\varepsilon' < \min(e, \varepsilon)$, $0 < \inf_{\{\pi' :\ |\pi - \pi'| < \varepsilon'\}} B(\pi') - \widehat{B}(\pi) < \frac{1}{2}\varepsilon$. By the definition of the infimum for any $\varepsilon > 0$, there is a π^* such that $0 < B(\pi^*) - \inf_{\{\pi' :\ |\pi - \pi'| < \varepsilon'\}} B(\pi') < \frac{1}{2}\varepsilon$ and $|\pi - \pi^*| < \varepsilon'$. Therefore:

$$\begin{aligned} B(\pi^*) - \widehat{B}(\pi) &\leq \left(B(\pi^*) - \inf_{\{\pi' :\ |\pi - \pi'| < \varepsilon'\}} B(\pi')\right) + \left(\inf_{\{\pi' :\ |\pi - \pi'| < \varepsilon'\}} B(\pi') - \widehat{B}(\pi)\right) \\ &< \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon. \end{aligned}$$

Summarizing, for any π and any $\varepsilon > 0$, there exists a π^* such that $|\pi - \pi^*| < \varepsilon$ and $B(\pi^*) - \widehat{B}(\pi) < \varepsilon$.

Fix any $\varepsilon > 0$ and consider a partition of Π, $\mathcal{P}_N = \{[\pi_0, \pi_1]\} \cup \{(\pi_{i-1}, \pi_i]\}_{i=2}^N$ so that $N < \infty$ and for all i, $\pi_{i+1} - \pi_i < \frac{1}{2K}\varepsilon$ and $\widehat{F}(\pi_i) = \widehat{F}(\pi_i)_-$,[20] i.e., π_i are not mass points of $\widehat{F}$. Let $\pi_i' = \arg\min_{\pi \in [\pi_{i-1}, \pi_i]} \widehat{B}(\pi)$ and note that by the extreme value theorem this is well defined as $\widehat{B}$ is lower semicontinuous. By the above summary, find π_i^* such that $|\pi_i' - \pi_i^*| < \frac{1}{2K}\varepsilon$ and $B(\pi_i^*) - \widehat{B}(\pi_i') < \frac{1}{2K}\varepsilon$. Clearly for any

[18] It is easy to construct counter-examples for functions which are not lower semi-continuous.

[19] See Rockafeller (1970) or Penot (2013) proposition 1.21. Note that $\widehat{B}$ is well defined, since $\inf_{\{\pi' :\ |\pi - \pi'| < \varepsilon\}} B(\pi')$ is a decreasing in ε and bounded below by 0.

[20] This works as there can only be countably many mass points.

$\pi \in [\pi_{i-1}, \pi_i]$, we have that $|\pi - \pi_i^*| < \frac{1}{2K}\varepsilon$. Now, let $f_\varepsilon(\pi_i^*) = \widehat{F}(\pi_i) - \widehat{F}(\pi_{i-1})$. Define $F_\varepsilon(\pi) = \int_0^\pi f_\varepsilon(\pi')\ \mathrm{d}\pi'$ and:

$$F_\varepsilon^G(\pi) = \begin{cases} \sum_{j\le i} f_\varepsilon\left(\pi_j^*\right) & \text{if } \pi \in [\pi_{i-1}, \pi_i] \text{ and } \sum_{j\le i} f_\varepsilon\left(\pi_j^*\right) \le G(\pi) \\ G(\pi_{i-1}) & \text{if } \pi \in [\pi_{i-1}, \pi_i] \text{ and } \sum_{j\le i} f_\varepsilon\left(\pi_j^*\right) > G(\pi) \end{cases},$$

and note that by construction $F_\varepsilon^G \le G$. Let $P_i = \mathbf{1}_{\sum_{j\le i} f_\varepsilon\left(\pi_j^*\right) \le G(\pi_{i-1})}$ and note that:

$$\begin{aligned}
& \inf_{F\le G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi) - \min_{F\le G} \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}F(\pi) \\
\le\ & \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F_\varepsilon^G(\pi) - \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}\widehat{F}(\pi) \\
=\ & \sum_{i=1}^N f_\varepsilon(\pi_i^*) B(\pi_i^*) P_i + \sum_{i=1}^N f_\varepsilon(\pi_i^*) B(\pi_i^*) (1 - P_i) - \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}\widehat{F}(\pi) \\
\le\ & \sum_{i=1}^N \left(\widehat{F}(\pi_i) - \widehat{F}(\pi_{i-1})\right) B(\pi_i^*) - \int_0^{\overline{\pi}} \widehat{B}(\pi)\ \mathrm{d}\widehat{F}(\pi) \\
\le\ & \sum_{i=1}^N \left(\widehat{F}(\pi_i) - \widehat{F}(\pi_{i-1})\right) \left(B(\pi_i^*) - \widehat{B}(\pi_i')\right) \le \frac{1}{2}\varepsilon.
\end{aligned}$$

Note that there is still $\frac{1}{2}\varepsilon$ to squeeze in the constraint $F_\varepsilon \le G$. Clearly if $G = \delta_0$, the inequality is satisfied. Otherwise, it works by a Taylor approximation given that the derivative of G is bounded by $K < \infty$. ■

1.A.2 Proof of Lemma 1.3.2

Lemma 1.A.2 *We have that $V_P(B \mid F_0)$ is the solution to:*

$$\min_{F\le G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi)\,, \text{ subject to } \int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F \ge V_A(B \mid F_0)\,. \tag{1.A.1}$$

Furthermore, if $B(\pi)$ and $\pi - B(\pi)$ are monotonic the constraint above holds as an equality.

Proof. For any $\mathcal{A}$, consider any $(e, F_1) \in A^*(B \mid \mathcal{A})$. It must be the case that:

$$\int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F_1(\pi \mid e) \geq \int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F_1(\pi \mid e) - c(e) \geq V_A(B \mid F_0),$$

thus:

$$V_P(B \mid F_0) \geq \min_{F \leq G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi), \text{ s.t. } \int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F \geq V_A(B \mid F_0).$$

To show the reverse, let F attain the minimum on the RHS of the above inequality. Consider $\mathcal{A} = F_0 \cup F_c$, where $F_c(\pi \mid e) = F(\pi)$ for all e. Since c is strictly increasing the agent chooses action $(0, F_c)$ as this gives him (at least) the utility $V_A(B \mid F_0)$ and the principal is worse off in the case of indifference.

To see the second claim, assume that the minimizing $F(\pi) < G(\pi)$ for some π (since CDFs are monotonic and upper-semicontinuous the strict inequality holds on some interval). Then if:

$$\int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F > V_A(B \mid F_0),$$

we can find an $\widetilde{F}$, $F \leq \widetilde{F} \leq G$ such that $F(\pi) < \widetilde{F}(\pi) < G(\pi)$ when $F < G$ and

$$\int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}\widetilde{F} \geq V_A(B \mid F_0).$$

The last part is possible by the monotonicity of $\pi - B(\pi)$. Furthermore, by the monotonicity of $B(\pi)$:

$$\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi) < \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}\widetilde{F}(\pi),$$

but then F could not have been a solution to the minimization problem, which is a contradiction. If the minimizing $F = G$, then $F_0 = G$ by the monotonicity of

$\pi - B(\pi)$. Thus:

$$\int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F = \int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}G = V_A(B \mid F_0),$$

and the constraint holds as an equality. ■

1.A.3 Proof of Theorem 1.3.3

Theorem 1.A.3 *For any G and any non-monotonic contract $B(\pi)$ there exists a monotonic contract $B_m(\pi)$ such that:*

$$\inf_{\mathcal{A} \subset \mathcal{D}_G} \int_0^{\overline{\pi}} B_m(\pi)\ \mathrm{d}F_m^{\mathcal{A}}(\pi) = \inf_{\mathcal{A} \subset \mathcal{D}_G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F^{\mathcal{A}}(\pi),$$

subject to $F_m^{\mathcal{A}} \in A_A^*(B_m \mid \mathcal{A})$ and $F^{\mathcal{A}} \in A_A^*(B \mid \mathcal{A})$, i.e., the principal is indifferent between the two contracts, and $B(\pi) \geq B_m(\pi)$, i.e., the agent's prefer the monotonic contract.

Before giving a proof of the theorem, we present some lemmas. Note that in the current version these proofs do not rely on the lower-semicontinuity theorem observed in the preliminaries.

Lemma 1.A.4 *Fix any B and assume there exists $F^B = \arg\min_{F \leq G} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi)$. For any interval (π_1, π_2) such that there exists $\pi^* > \pi_2$ for which $B(\pi^*) < B(\pi')$ for all $\pi' \in (\pi_1, \pi_2)$, F^B is constant on $[\pi_1, \pi_2)$.*

Proof. Assume by way of contradiction $\lim_{\pi' \uparrow \pi_2} F^B(\pi') - F^B(\pi_1) = \gamma > 0$. Consider:

$$F^*(\pi) = \begin{cases} F^B(\pi) & \text{if } \pi < \pi_1 \\ F^B(\pi_1) & \text{if } \pi \in [\pi_1, \pi_2) \\ F^B(\pi) - \gamma & \text{if } \pi \in [\pi_2, \pi^*) \\ F^B(\pi) & \text{if } \pi \geq \pi^* \end{cases},$$

and note that $F^*(\pi) \leq F^B(\pi) \leq G(\pi)$ for all π and thus satisfies the constraint. Furthermore, by construction we have that:

$$\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F^B(\pi) > \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F^*(\pi),$$

and thus F^B could not have solved the minimization problem, which is a contradiction. ■

A final lemma which says that the lower-semicontinuous hulls of contracts preserve order.

Lemma 1.A.5 *Let $B \geq B'$. If $\widehat{B}(\pi) > \widehat{B'}(\pi)$, then for every $\varepsilon > 0$, there exists a π', $|\pi - \pi'| < \varepsilon$, such that $B(\pi') > B'(\pi')$.*

Proof. We shall prove the contrapositive. Thus, assume that for some $\varepsilon > 0$, for all π', $|\pi - \pi'| < \varepsilon$ we have that $B(\pi') = B'(\pi')$. Now, by definition we have that:

$$\sup\left\{\widehat{b}(\pi) : \widehat{b} \text{ lsc}, \widehat{b} \leq B\right\} = \widehat{B}(\pi) > \widehat{B'}(\pi) = \sup\left\{\widehat{b'}(\pi) : \widehat{b'} \text{ lsc}, \widehat{b'} \leq B'\right\},$$

and if B and B' are equal in a neighborhood of π, any lower semicontinuous function which is below one, must be below the other. ■

We are now ready to prove the theorem.

Proof of Theorem 1.3.3. Consider the following contract:

$$B_m(\pi) = \inf_{\pi' \in [\pi, \overline{\pi}]} B(\pi').$$

Clearly $B_m(\pi) \leq B(\pi)$ for all π and $B_m(\pi)$ is monotone. We are left to show that:

$$\inf_{\mathcal{A} \supset \mathcal{A}_0} \int_0^{\overline{\pi}} B_m(\pi)\ \mathrm{d}F_m^{\mathcal{A}}(\pi) = \inf_{\mathcal{A} \supset \mathcal{A}_0} \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F^{\mathcal{A}}(\pi).$$

We claim that it is enough to show that:

$$\inf_{F\leq G_m}\int_0^{\overline{\pi}} B_m(\pi)\ \mathrm{d}F(\pi)=\inf_{F\leq G}\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F(\pi),$$

where $(G,c')\in\arg\max_{(F,c)\in\mathcal{A}_0}\int_0^{\overline{\pi}}\pi-B(\pi)-c\,\mathrm{d}F(\pi)$, i.e., G is the distribution over profit outcomes which would be chosen by the agent from $\mathcal{A}_0$ under contract B, and similarly G_m is the action that would have been chosen under B_m.

By lemma 1.3.1, we are left to show that:

$$\min_{F\leq G}\int_0^{\overline{\pi}}\widehat{B}(\pi)\ \mathrm{d}F(\pi)=\min_{F\leq G_m}\int_0^{\overline{\pi}}\widehat{B_m}(\pi)\ \mathrm{d}F(\pi).$$

Let $F^B=\arg\min_{F\leq G}\int_0^{\overline{\pi}}\widehat{B}(\pi)\ \mathrm{d}F(\pi)$, by lemma 1.3.1 this exists. We first claim that:

$$\int_0^{\overline{\pi}}\widehat{B}(\pi)\ \mathrm{d}F^B(\pi)=\int_0^{\overline{\pi}}\widehat{B_m}(\pi)\ \mathrm{d}F^B(\pi).$$

If $\widehat{B}(\pi)>\widehat{B_m}(\pi)$, by lemma 1.A.5 there is some π' such that $B(\pi')>B_m(\pi')$. By definition, there exists some π^* such that $B(\pi'')>B_m(\pi'')$ for all $\pi''\in(\pi',\pi^*)$.

Next, we argue that for any other $F\leq G_m$, $\int_0^{\overline{\pi}}\widehat{B_m}(\pi)\ \mathrm{d}F(\pi)\geq\int_0^{\overline{\pi}}\widehat{B_m}(\pi)\ \mathrm{d}F^B(\pi)$. To see this, assume by way of contradiction the opposite. It's clear that by lemma 1.A.4, F must assign zero measure to the set where B and B_m do not agree, but then F^B could not be a minimizer for $\widehat{B}$.

That yields the final result. ■

1.A.4 Proof of Lemma 1.4.2

Lemma 1.A.6 *Let $\phi(\pi)$ be a function such that $\phi(\pi)\geq 0$ for $\pi\leq\pi_B$, $\phi(\pi)\leq 0$ for $\pi\geq\pi_B$ and either:*

1. $\int_0^{\overline{\pi}}\phi(\pi)\ \mathrm{d}F(\pi\mid e_L)=0$, *or*

2. $\int_0^{\overline{\pi}}\phi(\pi)\ \mathrm{d}F(\pi\mid e_L)\leq 0$ *and $\phi(\pi)$ decreasing for $\pi\geq\pi_B$.*

Then, for any $e_H > e_L$ and any MLRP family F, we have that $\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) \geq \int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_H)$.

Proof. Note that under case 1 we have:

$$\int_0^{\pi_B} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) = -\int_{\pi_B}^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) > 0.$$

Take any $e_H > e_L$ and consider:

$$\begin{aligned}
&\left(\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) - \int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_H)\right) \int_0^{\pi_B} \phi(\pi_L)\ \mathrm{d}F(\pi_L \mid e_L) \\
= \;& \int_0^{\pi_B} \phi(\pi_L)\ \mathrm{d}F(\pi_L \mid e_L) \int_0^{\overline{\pi}} \phi(\pi) \left(\frac{dF(\pi \mid e_L)}{dC(\pi)} - \frac{dF(\pi \mid e_H)}{dC(\pi)}\right) \mathrm{d}C(\pi) \\
= \;& \left(-\int_{\pi_B}^{\overline{\pi}} \phi(\pi_H) \frac{dF(\pi_H \mid e_L)}{dC(\pi_H)}\ \mathrm{d}C(\pi_H)\right) \int_0^{\pi_B} \phi(\pi_L) \left(\frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} - \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)}\right) \mathrm{d}C(\pi_L) \\
&+ \left(\int_0^{\pi_B} \phi(\pi_L) \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)}\ \mathrm{d}C(\pi_L)\right) \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H) \left(\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} - \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}\right) \mathrm{d}C(\pi_H),
\end{aligned}$$

where we write $C(\pi)$ for $C(\pi \mid e_L, e_H)$. By Fubini's theorem (applies since the above are integrable and C is a probability measure and therefore σ-finite) the above equals:

$$\begin{aligned}
&-\int_0^{\pi_B} \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\phi(\pi_L) \frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} \left(\frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} - \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)}\right) \mathrm{d}C(\pi_H)\ \mathrm{d}C(\pi_L) \\
&+\int_0^{\pi_B} \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\phi(\pi_L) \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} \left(\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} - \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}\right) \mathrm{d}C(\pi_H)\ \mathrm{d}C(\pi_L) \\
= \;& \int_0^{\pi_B} \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\phi(\pi_L) \left[\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)} - \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}\right] \mathrm{d}C(\pi_H)\ \mathrm{d}C(\pi_L) \\
\geq \;& 0,
\end{aligned}$$

where the last inequality follows since $\phi(\pi_L) \geq 0$, $\phi(\pi_H) \leq 0$ and by the generalized MLRP:

$$\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)} \leq \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}.$$

Thus:

$$\begin{pmatrix} \int_0^{\overline{\pi}} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_L\right) \\ -\int_0^{\overline{\pi}} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_H\right) \end{pmatrix} \int_0^{\pi_B} \phi\left(\pi_L\right)\ \mathrm{d}F\left(\pi_L \mid e_L\right) \geq 0,$$

and since $\int_0^{\pi_B} \phi\left(\pi_L\right)\ \mathrm{d}F\left(\pi_L \mid e_L\right) > 0$, we have that:

$$\int_0^{\overline{\pi}} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_L\right) \geq \int_0^{\overline{\pi}} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_H\right).$$

Case 2 follows similarly (using first-order stochastic dominance of $F\left(\cdot \mid e_H\right)$ over $F\left(\cdot \mid e_L\right)$ and the fact that ϕ is decreasing for $\pi \geq \pi_B$). May need to "split up" mass at π^*.

In particular, under case 2, there exists some π^* and $\alpha \in (0,1]$ such that:

$$\begin{aligned} & \int_0^{\pi_B} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_L\right) \\ = & -\lim_{\pi' \to \pi^*_-} \int_{\pi_B}^{\pi'} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_L\right) - \alpha\left[F\left(\pi^* \mid e_L\right) - F\left(\pi^*_- \mid e_L\right)\right]\phi\left(\pi^*\right). \end{aligned}$$

We can then repeat the above, replacing $\overline{\pi}$ by π^*, with the alpha-mass adjustment. We have then shown that:

$$\begin{pmatrix} \int_0^{\pi^*} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_L\right) - \int_0^{\pi^*} \phi\left(\pi\right)\ \mathrm{d}F\left(\pi \mid e_H\right) \\ -\alpha\phi\left(\pi^*\right)\left[f\left(\pi^* \mid e_L\right) - f\left(\pi^* \mid e_H\right)\right] \end{pmatrix} \int_0^{\pi_B} \phi\left(\pi_L\right)\ \mathrm{d}F\left(\pi_L \mid e_L\right) \geq 0, \tag{1.A.2}$$

where

$$f\left(\pi^* \mid e_L\right) = F\left(\pi^* \mid e_L\right) - F\left(\pi^*_- \mid e_L\right).$$

Because $F\left(\pi \mid e_H\right)$ dominates $F\left(\pi \mid e_L\right)$ with respect to the monotone likelihood ratio order, it also conditionally first-order stochastically dominates it (conditioning on any set). Conditioning on (π^*, ∞) and π^* implies that:

$$\begin{aligned} & \int_{\pi^*_+}^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_L) + \alpha f(\pi^* \mid e_L)\,\phi(\pi^*) \\ \geq\ & \int_{\pi^*_+}^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e_H) + \alpha f(\pi^* \mid e_H)\,\phi(\pi^*), \end{aligned} \tag{1.A.3}$$

since $\phi(\pi)$ decreasing for $\pi \geq \pi_B$. Combining 1.A.2 and 1.A.3 we have the desired result. ■

1.A.5 Proof of Theorem 1.4.1

Theorem 1.A.7 *For any $\mathcal{A} = G \cup F(\cdot \mid e) \subset \overline{\mathcal{D}}_G$, a solution to:*

$$\begin{aligned} & \max_B V_A(B \mid \mathcal{A}), \\ \textit{subject to} \quad & V_P(B \mid F_0) \geq R, \end{aligned}$$

is $B^D(\pi, z) := \min(\pi, z)$ for some z.

Proof. Let $B(\pi)$ be a monotonic non-debt contract, i.e., $\left\{\pi : B^D(\pi, z) \neq B(\pi)\right\}$ is not G-null for every z. The principal's worst case in this instance is if only technology $F_0 = G$ was available. Let z_0 solve:

$$\int B^D(\pi, z_0)\ \mathrm{d}G(\pi) = \int B(\pi)\ \mathrm{d}G(\pi).$$

Note that such a z_0 exists by Berge's maximum theorem since B^D is continuous in z_0, as remarked in the main text.

Define:

$$\phi(\pi) = B^D(\pi, z_0) - B(\pi).$$

Fix any $F_1(\cdot \mid e) \subset \overline{\mathcal{D}}_G$ such that $\mathcal{A} = F_0 \cup F_1$ is an MLRP set and can be represented by some F satisfying MLRP. Note that $F \leq G$ by assumption and that $F \overset{MLRP}{\geq} F_0 =$

G, so that:

$$\int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F(\pi \mid e) + c(e) - \int_0^{\overline{\pi}} \left(\pi - B^D(\pi, z_0)\right)\ \mathrm{d}F(\pi \mid e) - c(e)$$
$$= \int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e).$$

Now, by definition:

$$\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}G(\pi) = 0,$$

and by lemma 1.4.2, for any $e \geq 0$:

$$\int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}F(\pi \mid e) \leq \int_0^{\overline{\pi}} \phi(\pi)\ \mathrm{d}G(\pi) = 0,$$

so that:

$$\int_0^{\overline{\pi}} (\pi - B(\pi))\ \mathrm{d}F(\pi \mid e) - c(e) \leq \int_0^{\overline{\pi}} \left(\pi - B^D(\pi, z_0)\right)\ \mathrm{d}F(\pi \mid e) - c(e),$$

but then the agent gets weakly higher utility under B^D than under B. Note that the above holds strictly when the MLRP is strict, as this implies a strict version of lemma 1.4.2. Since by definition of B^D we have that the principal's robust constraint is satisfied:

$$\int B^D(\pi, z_0)\ \mathrm{d}G(\pi) = \int B(\pi)\ \mathrm{d}G(\pi) \geq R,$$

we have that B^D is optimal. ■

1.A.6 Proof of Theorem 1.5.1

The statement of the theorem is repeated below for convenience.

Theorem 1.A.8 *For any $\mathcal{A} \subset \mathcal{D}$, a solution to:*

$$\max_{B} V_A(B \mid \mathcal{A}), \tag{1.A.4}$$
$$\textit{subject to} \qquad V_P(B \mid F_0) \geq R,$$

is $B_\alpha(\pi) = \alpha\pi$ for some α, i.e., a linear/equity contract.

The proof proceeds by first showing that a solution to the above must be a (weakly) convex contract B, since the principal will not put any value on concave portions of a contract and thus the lower convex hull of B is evaluated in the same way as the original contract by the principal. We then show that the appropriate linear contract is optimal within the set of convex contracts, since it is no worse for the principal and better for the agent.

Lemma 1.A.9 *In problem 1.A.4 for any non-convex B, there exists a convex B_c such that $V_P(B \mid F_0) \leq V_P(B_c \mid F_0)$ and $B \geq B_c$.*

Proof. Note that it is without loss of generality to consider B which are lower semicontinuous by lemma 1.3.1. Let B_c be the lower convex hull of B, i.e., the largest weakly convex function majorized than B. Clearly $B \geq B_c$ and thus $V_A(B \mid F_0) \leq V_A(B_c \mid F_0)$. It suffices to consider the case where $V_A(B_c \mid F_0) = V_A(B \mid F_0) =: \upsilon$, since by lemma **??** a larger $V_A(B_c \mid F_0)$ decreases the constraint set and thus weakly increases $V_P(B_c \mid F_0)$.

It suffices to show that:

$$
\begin{aligned}
&\left\{\min_{F\in\Delta(\Pi)}\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F \text{ s.t. } \int_0^{\overline{\pi}} \pi - B(\pi)\ \mathrm{d}F \geq \upsilon\right\}\\
= &\left\{\min_{\substack{F\in\Delta(\Pi)\\ supp(F)=D}}\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F \text{ s.t. } \int_0^{\overline{\pi}} \pi - B(\pi)\ \mathrm{d}F \geq \upsilon\right\}, \text{ and}\\
&\left\{\min_{F\in\Delta(\Pi)}\int_0^{\overline{\pi}} B_c(\pi)\ \mathrm{d}F \text{ s.t. } \int_0^{\overline{\pi}} \pi - B_c(\pi)\ \mathrm{d}F \geq \upsilon\right\}\\
= &\left\{\min_{\substack{F\in\Delta(\Pi)\\ supp(F)=D}}\int_0^{\overline{\pi}} B_c(\pi)\ \mathrm{d}F \text{ s.t. } \int_0^{\overline{\pi}} \pi - B_c(\pi)\ \mathrm{d}F \geq \upsilon\right\},
\end{aligned}
$$

where $D = \{x : B(\pi) = B_c(\pi)\}$. Let F_B be the CDF which minimizes the LHS and F_c be the CDF which minimizes the RHS. Clearly, it is without loss of generality to assume that $supp(F_c) \subset D$.[21] We will show that $supp(F_B) \subset D$.

Assume by way of contradiction that there is some $\pi \in supp(F_B)$ and $\varepsilon > 0$ such that $B(\pi') > B_c(\pi')$ for all $\pi' \in \mathcal{N}_\varepsilon(\pi)$. Note that by construction there exist $\pi_L < \pi_H$ such that for all $\pi' \in \mathcal{N}_\varepsilon(\pi)$ there exists an $\beta(\pi') \in (0,1)$ such that $\pi' = \beta(\pi')\pi_L + (1-\beta(\pi'))\pi_H$ and:

$$
\begin{aligned}
B(\pi') > B_c(\pi') &= \beta(\pi') B_c(\pi_L) + (1-\beta(\pi')) B_c(\pi_H)\\
&= \beta(\pi') B(\pi_L) + (1-\beta(\pi')) B(\pi_H).
\end{aligned}
$$

Let $m = F_B(\pi+\varepsilon)_- - F_B(\pi-\varepsilon) = \int_{\mathcal{N}_\varepsilon(\pi)} \mathrm{d}F_B(\pi')$ and note that $m > 0$ since $\pi \in supp(F_B)$. Let:

$$
\beta^* = \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} \beta(\pi')\ \mathrm{d}F_B(\pi'),
$$

[21] To see this, note that any $\pi \notin D$ is a convex combination of two elements in D and we can therefore, instead of putting mass on π, put the appropriate mass on the elements which constitute the convex combination.

so that:

$$\begin{aligned}
1-\beta^* &= 1-\frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} \beta(\pi')\ \mathrm{d}F_B(\pi') \\
&= \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} \mathrm{d}F_B(\pi') - \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} \beta(\pi')\ \mathrm{d}F_B(\pi') \\
&= \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} (1-\beta(\pi'))\ \mathrm{d}F_B(\pi').
\end{aligned}$$

Thus:

$$\begin{aligned}
&\beta^* B_c(\pi_L) + (1-\beta^*) B_c(\pi_H) \\
&\quad= \beta^* B(\pi_L) + (1-\beta^*) B(\pi_H) \\
&\quad= \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} \beta(\pi') B(\pi_L)\ \mathrm{d}F_B(\pi') + \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} (1-\beta(\pi')) B(\pi_H)\ \mathrm{d}F_B(\pi') \\
&\quad= \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} \beta(\pi') B(\pi_L) + (1-\beta(\pi')) B(\pi_H)\ \mathrm{d}F_B(\pi') \\
&\quad< \frac{1}{m}\int_{\mathcal{N}_\varepsilon(\pi)} B(\pi')\ \mathrm{d}F_B(\pi'),
\end{aligned}$$

hence shifting mass to points π_L and π_H leads to a lower expected payoff for the principal and thus F_B could not have been a minimizer. This concludes the proof of equality **??**. ■

We are now ready to prove the theorem. The proof goes by invoking Jensen's inequality and using a revealed preference argument to rule out technologies which the suggested replacement of contract makes worse.

Proof of Theorem 1.5.1. Next we show that for the principal's problem, a minimizing F for convex B puts mass on a single point. Note that for any F, where $\pi_F = \mathbb{E}_F[\pi]$, such that $\int_0^{\overline{\pi}} \pi - B(\pi)\ \mathrm{d}F \geq \upsilon$, Jensen's inequality implies that:

$$\upsilon \leq \int_0^{\overline{\pi}} \pi - B(\pi)\ \mathrm{d}F \leq \pi_F - B(\pi_F) = \int_0^{\overline{\pi}} \pi - B(\pi)\ \mathrm{d}\delta_{\pi_F},$$

since $\pi - B(\pi)$ is concave. Furthermore, since B is convex:

$$\int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}\delta_{\pi_F} = B(\pi_F) \leq \int_0^{\overline{\pi}} B(\pi)\ \mathrm{d}F.$$

Note that in finding the minimizing CDF for the principal, the agent's utility constraint will hold with equality. Now consider replacing $B(\pi)$ by a linear contract $B_\alpha(\pi) = \alpha\pi$, where $\alpha = B(\pi^*)/\pi^*$ and:

$$\pi^* = \min \pi \text{ s.t. } \pi - B(\pi) = \upsilon.$$

Let $(e^*, F_0) = A_A^*(B \mid F_0)$, write $\widehat{F} = F_0(\pi \mid e^*)$ and consider (Careful with e^* here, still needs editing):

$$\begin{aligned} V_A\left(B_\alpha \mid \widehat{F}\right) &= \int_0^{\overline{\pi}} \pi - B_\alpha(\pi)\ \mathrm{d}\widehat{F} = \int_0^{\overline{\pi}} \frac{\pi^* - B(\pi^*)}{\pi^*}\pi\ \mathrm{d}\widehat{F} \\ &= \upsilon \int_0^{\overline{\pi}} \frac{\pi}{\pi^*}\ \mathrm{d}\widehat{F}. \end{aligned}$$

Assume by way of contradiction that $\pi^* > \int_0^{\overline{\pi}} \pi\ \mathrm{d}\widehat{F} = \pi_{\widehat{F}}$, but then by Jensen's inequality and since $\pi - B(\pi)$ is increasing for $\pi < \pi^*$ (by definition of π^*), we have that:

$$\begin{aligned} \upsilon &= \int_0^{\overline{\pi}} \pi - B(\pi)\ \mathrm{d}\widehat{F} \\ &\leq \pi_{\widehat{F}} - B\left(\pi_{\widehat{F}}\right) \\ &< \pi^* - B(\pi^*) \\ &= \upsilon, \end{aligned}$$

which is a contradiction. Thus:

$$V_A\left(B_\alpha \mid F_0\right) = \upsilon \int_0^{\overline{\pi}} \frac{\pi}{\pi^*} \,\mathrm{d}\widehat{F} \geq \upsilon = V_A\left(B \mid F_0\right).$$

Similarly for any F such that $\int_0^{\overline{\pi}} \pi - B\left(\pi\right) \,\mathrm{d}F > \upsilon$, we have that $V_A\left(B_\alpha \mid F\right) \geq V_A\left(B \mid F\right)$. ■

1.A.7 Proof of Theorem 1.6.1

We will need the following generalization of lemma 1.4.2.

Lemma 1.A.10 *Let $C \subset [0, \overline{\pi}]$ be measurable*[22] *and $\phi\left(\pi\right)$ be a function such that $\phi\left(\pi\right) \geq 0$ for $\pi \in [0, \pi_B] \cap C$, $\phi\left(\pi\right) \leq 0$ for $\pi \in [\pi_B, \overline{\pi}] \cap C$ and either:*

1. $\int_0^{\overline{\pi}} \phi\left(\pi\right) \mathbf{1}_{\pi \in C} \,\mathrm{d}F\left(\pi \mid e_L\right) = 0$, *or*

2. $\int_0^{\overline{\pi}} \phi\left(\pi\right) \mathbf{1}_{\pi \in C} \,\mathrm{d}F\left(\pi \mid e_L\right) \leq 0$ *and $\phi\left(\pi\right)$ decreasing for $\pi \geq \pi_B$.*

Then, for any $e_H > e_L$ and any MLRP family F, we have that $\int_0^{\overline{\pi}} \phi\left(\pi\right) \mathbf{1}_{\pi \in C} \,\mathrm{d}F\left(\pi \mid e_L\right) \geq \int_0^{\overline{\pi}} \phi\left(\pi\right) \mathbf{1}_{\pi \in C} \,\mathrm{d}F\left(\pi \mid e_H\right)$.

Proof. Note that under case 1 we have:

$$\int_0^{\pi_B} \phi\left(\pi\right) \mathbf{1}_{\pi \in C} \,\mathrm{d}F\left(\pi \mid e_L\right) = -\int_{\pi_B}^{\overline{\pi}} \phi\left(\pi\right) \mathbf{1}_{\pi \in C} \,\mathrm{d}F\left(\pi \mid e_L\right) > 0.$$

[22] That is, C belongs to the Borel σ-algebra of the usual topology on $[0, \overline{\pi}]$.

Take any $e_H > e_L$ and consider:

$$\begin{aligned}
& \left(\int_0^{\overline{\pi}} \phi(\pi)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi \mid e_L) - \int_0^{\overline{\pi}} \phi(\pi)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi \mid e_H)\right) \int_0^{\pi_B} \phi(\pi_L)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi_L \mid e_L) \\
= & \int_0^{\pi_B} \phi(\pi_L)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi_L \mid e_L) \int_0^{\overline{\pi}} \phi(\pi)\, \mathbf{1}_{\pi\in C} \left(\frac{dF(\pi \mid e_L)}{dC(\pi)} - \frac{dF(\pi \mid e_H)}{dC(\pi)}\right) \mathrm{d}C(\pi) \\
= & \left(-\int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\, \mathbf{1}_{\pi\in C} \frac{dF(\pi_H \mid e_L)}{dC(\pi_H)}\, \mathrm{d}C(\pi_H)\right) \int_0^{\pi_B} \phi(\pi_L)\, \mathbf{1}_{\pi\in C} \left(\frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} - \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)}\right) \mathrm{d}C(\pi_L) \\
& + \left(\int_0^{\pi_B} \phi(\pi_L)\, \mathbf{1}_{\pi\in C} \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)}\, \mathrm{d}C(\pi_L)\right) \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\, \mathbf{1}_{\pi\in C} \left(\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} - \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}\right) \mathrm{d}C(\pi_H),
\end{aligned}$$

where we write $C(\pi)$ for $C(\pi \mid e_L, e_H)$. By Fubini's theorem (applies since the above are integrable and C is a probability measure and therefore σ-finite) the above equals:

$$\begin{aligned}
& -\int_0^{\pi_B} \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\, \phi(\pi_L)\, \mathbf{1}_{\pi\in C} \frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} \left(\frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} - \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)}\right) \mathrm{d}C(\pi_H)\ \mathrm{d}C(\pi_L) \\
& + \int_0^{\pi_B} \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\, \phi(\pi_L)\, \mathbf{1}_{\pi\in C} \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} \left(\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} - \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}\right) \mathrm{d}C(\pi_H)\ \mathrm{d}C(\pi_L) \\
= & \int_0^{\pi_B} \int_{\pi_B}^{\overline{\pi}} \phi(\pi_H)\, \phi(\pi_L)\, \mathbf{1}_{\pi\in C} \left[\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)} - \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}\right] \mathrm{d}C(\pi_H)\ \mathrm{d}C(\pi_L) \\
\geq & \ 0,
\end{aligned}$$

where the last inequality follows since $\phi(\pi_L) \geq 0$, $\phi(\pi_H) \leq 0$ and by the generalized MLRP (which also applies to $\pi \in C$ only, since conditional FOSD is equivalent to MLRP, thus we can condition on set C):

$$\frac{dF(\pi_H \mid e_L)}{dC(\pi_H)} \frac{dF(\pi_L \mid e_H)}{dC(\pi_L)} \leq \frac{dF(\pi_L \mid e_L)}{dC(\pi_L)} \frac{dF(\pi_H \mid e_H)}{dC(\pi_H)}.$$

Thus:

$$\left(\begin{array}{c} \int_0^{\overline{\pi}} \phi(\pi)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi \mid e_L) \\ - \int_0^{\overline{\pi}} \phi(\pi)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi \mid e_H) \end{array}\right) \int_0^{\pi_B} \phi(\pi_L)\, \mathbf{1}_{\pi\in C}\, \mathrm{d}F(\pi_L \mid e_L) \geq 0,$$

and since $\int_0^{\pi_B} \phi(\pi_L) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi_L \mid e_L) > 0$, we have that:

$$\int_0^{\overline{\pi}} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_L) \geq \int_0^{\overline{\pi}} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_H).$$

Case 2 follows similarly (using first-order stochastic dominance of $F(\cdot \mid e_H)$ over $F(\cdot \mid e_L)$ and the fact that ϕ is decreasing for $\pi \geq \pi_B$). May need to "split up" mass at π^*.

In particular, under case 2, there exists some π^* and $\alpha \in (0, 1]$ such that:

$$\begin{aligned} & \int_0^{\pi_B} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_L) \\ = & -\lim_{\pi' \to \pi^*_-} \int_{\pi_B}^{\pi'} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_L) - \alpha \left[F(\pi^* \mid e_L) - F\left(\pi^*_- \mid e_L\right)\right] \phi(\pi^*). \end{aligned}$$

We can then repeat the above, replacing $\overline{\pi}$ by π^*, with the alpha-mass adjustment. We have then shown that:

$$\begin{pmatrix} \int_0^{\pi^*} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_L) \\ -\int_0^{\pi^*} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_H) \\ -\alpha\phi(\pi^*) \left[f(\pi^* \mid e_L) - f(\pi^* \mid e_H)\right] \end{pmatrix} \int_0^{\pi_B} \phi(\pi_L) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi_L \mid e_L) \geq 0, \quad (1.A.5)$$

where

$$f(\pi^* \mid e_L) = F(\pi^* \mid e_L) - F\left(\pi^*_- \mid e_L\right).$$

Because $F(\pi \mid e_H)$ dominates $F(\pi \mid e_L)$ with respect to the monotone likelihood ratio order, it conditionally first-order stochastically dominates it. Conditioning on

$(\pi^*, \infty) \cap C$ and π^* implies that:

$$\int_{\pi^*_+}^{\overline{\pi}} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_L) + \alpha f(\pi^* \mid e_L) \phi(\pi^*) \tag{1.A.6}$$
$$\geq \int_{\pi^*_+}^{\overline{\pi}} \phi(\pi) \mathbf{1}_{\pi \in C} \, \mathrm{d}F(\pi \mid e_H) + \alpha f(\pi^* \mid e_H) \phi(\pi^*),$$

since $\phi(\pi)$ decreasing for $\pi \geq \pi_B$. Combining 1.A.5 and 1.A.6 we have the desired result. ■

1.A.8 Proof of Corollary 1.6.2

Corollary 1.A.11 *The principal's payoff from contract* $B(\pi) = \min\{\pi, z + \alpha\pi\}$ *is:*

$$R = \int_0^{\frac{z}{1-\alpha}} \pi \, \mathrm{d}G + \frac{\alpha V_A(B \mid F_0) + z\left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)}.$$

Proof. The principal's payoff from contract $B(\pi) = \min(\pi, z + \alpha\pi)$ is:

$$\begin{aligned} R &= \int_0^{\frac{z}{1-\alpha}} \pi \, \mathrm{d}G + \min_F \int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\alpha\pi + z) \, \mathrm{d}F \\ \text{s.t.} & \quad \int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\pi - \alpha\pi - z) \, \mathrm{d}F \geq V_A(B \mid F_0). \end{aligned}$$

For any F:

$$\begin{aligned} \int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\alpha\pi + z) \, \mathrm{d}F &= \alpha \int_{\frac{z}{1-\alpha}}^{\overline{\pi}} \pi \, \mathrm{d}F + z\left(F(\overline{\pi}) - F\left(\frac{z}{1-\alpha}\right)\right) \\ &= \alpha \int_{\frac{z}{1-\alpha}}^{\overline{\pi}} \pi \, \mathrm{d}F + z\left(1 - G\left(\frac{z}{1-\alpha}\right)\right), \end{aligned}$$

and hence:

$$\begin{aligned}\int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\pi - \alpha\pi - z)\,\mathrm{d}F &= (1-\alpha)\int_{\frac{z}{1-\alpha}}^{\overline{\pi}} \pi\,\mathrm{d}F - z\left(1 - G\left(\frac{z}{1-\alpha}\right)\right) \\ &\geq V_A\left(B \mid F_0\right).\end{aligned}$$

Thus the constraint in the principal minimization problem holds as an equality.

Solving the last equation for the integral we have that:

$$\int_{\frac{z}{1-\alpha}}^{\overline{\pi}} \pi\,\mathrm{d}F = \frac{V_A\left(B \mid F_0\right) + z\left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)},$$

and hence for any F:

$$\int_{\frac{z}{1-\alpha}}^{\overline{\pi}} (\alpha\pi + z)\,\mathrm{d}F = \frac{\alpha V_A\left(B \mid F_0\right) + z\left(1 - G\left(\frac{z}{1-\alpha}\right)\right)}{(1-\alpha)},$$

which gives the characterization in the corollary. ■

Chapter 2

Optimally Toothless Policies

This chapter is co-authored with Prof. Matias Iaryczower.[1]

2.1 Introduction

A central problem in economics is how to design institutions and organizations so that agents have incentives to share relevant information with the actors that are in charge of making decisions. The effort to cope with this informational asymmetry fundamentally defines the relationships between investors and asset managers, CEOs and lower level managers, and politicians and bureaucrats. The problem is particularly challenging when the principal cannot use transfers to alleviate incentive problems, as it is often the case in politics, and in interactions among economic agents within firms and non-profit organizations.

Some variants of this problem are well understood. This is the case, for example, when the policy space over which the principal contracts with the agent is unidimensional. In this situation, a principal who can choose among a large space of contracts will simply delegate decision-making authority to the agent over a set of

[1] Department of Politics, Princeton University, Princeton, NJ 08544, email: miaryc@princeton.edu

possible actions (Holmström (1977), Melmud and Shibano (1991), Alonso and Matouschek (2008)).[2] The principal's problem is then reduced to determining how much discretion to *delegate* to the agent.[3]

The situation in which the principal is so seriously handicapped, however, is extreme. In fact, a common feature in many agency relationships is that the principal can decide not only the content of policy but also its scope of applicability, or scale of implementation. In the investor/asset manager context, for example, the investor can choose the level of risk for the portfolio, but also how much to invest with the asset manager. In politics, elected politicians must not only determine how "harsh" they will allow the CIA's interrogation techniques to be, but also the scope of the so called black site operations (should they house only confirmed terrorist elites, or any suspect who might have relevant information).

A similar logic applies to resources dedicated to the monitoring and enforcement of a given policy. As chaotic traffic in many Latin American countries shows, strict traffic laws only matter if they are enforced. Similarly, environmental regulations and carbon pollution standards that aim to reduce greenhouse gas emissions can only be effective if the agency in charge of monitoring and enforcement (the EPA in the United States) is endowed with the resources necessary to accomplish these goals in the first place.

In this context, the principal can use the scale of the project in lieu of transfers, providing incentives by distorting both the scale and content of policy outcomes. This new distortion arises because the value of increasing the scale of the project is inexorably linked to the content of the underlying policy. Thus, differently to a

[2] Baron (2000) and Krishna and Morgan (2008) analyzed the problem in which the principal can contract over both policy and *money* transfers (where both principal and agent have quasilinear preferences over money), as in the standard screening problem (Baron and Myerson, 1982). With full commitment, the model here is a relatively standard screening model, except for a limited liability constraint on nonnegative transfers to the agent.

[3] The delegation solution has been studied extensively in political science (see Epstein and O'Halloran(1994), Huber and Shipan (2002), and Bendor and Meirowitz (2004)).

transfer of "money", the value of increasing the scale of the project will naturally be a function of how much each actor values the associated policy in each state of the world. If the asset manager uncovers an investment opportunity that can attain an extraordinarily high expected return with a larger risk, the investor will be willing to invest a larger fraction of its wealth in this high risk portfolio. If the CIA can in fact obtain actionable intelligence from detainees with an intense interrogation treatment, politicians will be more willing to allow the agency to apply these harsher techniques broadly.

In this paper we characterize the optimal institutional arrangement for the principal in this setting. While there has been an extensive literature on delegation, the solution to this class of problems has not yet been explored. Baron (2000) and Krishna and Morgan (2008) analyze the unidimensional policy space with transfers assuming quasilinear preferences and quadratic policy payoffs. The multidimensional case without transfers is less common in the literature; Koessler and Martimort (2012) study a two-dimensional policy space with separable quadratic payoffs, while Frankel (2014) considers an N-dimensional policy space with separable preferences and a non-Bayesian (max-min) principal. In our case, instead, the content and scope of policy are complements in the utility function of principal and agent. Thus, we have what Koessler and Martimort (2012) call "externalities across decisions".

Multidimensional mechanism design problems are generally difficult to solve because the order in which incentive constraints bind is often endogeneous to the problem (Rochet and Stole, 2003). We appeal to a generalized single crossing condition which abstracts from this difficulty, yet still the standard techniques to deal with screening problems do not readily apply. In particular, with non-separable preferences across policy dimensions, the common procedure of reparametrizing the problem in terms of information rents is not helpful.[4] In spite of this, we are able to make considerable progress. First, we solve the optimal contract in the two type case, and present a graphical analysis that makes the logic and results transparent. The graphical analysis also allows us to relate our results with the standard quasilinear setting easily. We then characterize the optimal separating contract in the continuum with a parametric assumption on payoffs (exponential payoffs).

Our analysis leads to a number of new insights. First, we show that whenever conflicts of interests are binding (always in the continuum; for sufficiently large bias in the two type case), the principal will overfund "low types" and underfund "high types". In our environmental policy example, this says that when climate change is indeed occurring at a fast pace, the resources dedicated to enforcing regulations curbing carbon emissions are too low relative to the first-best: it is in this sense that the principal optimally chooses a toothless policy.

The possibility of tinkering with the scale of implementation induces distortions in the content of policy that *can* be different from what would result in a comparable model with quasilinear payoffs (i.e., with transfers in lieu of project size). The qualitative nature of these distortions depends on the level of conflict between the agent and the principal relative to the "smallest" possible deviations. When the agent's bias is sufficiently large relative to the smallest possible deviation (which is always the case in the continuum) the distortions in policy direction in fact *are* different than what appear in a comparable model with quasilinear payoffs.

Indeed, both in the continuum and in the two type model with a large bias, the optimal separating contract partitions the state space in a "low" and "high" set of states, such that the principal overfunds and distorts towards the agent in low states, but underfunds and distorts against the agent in high states. In our EPA

[4] In general, we are dealing with a singular control problem of the type that does not admit the sort of straightforward bang-bang solutions commonly used in other contexts (e.g., the continuous-time literature often exploits bang-bang solutions).

example, this implies that the optimal contract sets overly stringent regulations that are heavily enforced when climate change is mild, and relatively weak regulations which are under-enforced in the states in which climate change is accelerating more heavily. Thus, the solution is ex-post inefficient, as both Congress and the EPA would both prefer to set more stringent, heavily enforced environmental regulations in the high states. This strong form of ex-post inefficiency in the optimal contract does not appear in standard quasi-linear models where utility is perfectly transferable between parties.

The solution for the continuum illustrates that in choosing the optimal policy function the principal faces a tradeoff between inducing distortions in the content and the scale of implementation of the policy. This is because in order to make the content of policy responsive to the state of the world (reducing distortions in policy direction), the scale of implementation needs to be responsive to the state as well, increasing distortions in project size.

The particular resolution of this tradeoff depends crucially on the *relative* sensitivity to policy loss of the agent vis a vis the principal. When this ratio is low enough (when the agent does not care too much about policy losses relative to the principal) it is relatively cheap for the principal to compensate the agent with changes in project size to obtain a policy that is very close to her first best, and the optimal contract is fully separating. When the agent is very sensitive to policy losses, on the other hand, attaining a policy close to the first best is very costly for the principal, and the principal would rather take a relatively unresponsive policy than introduce large distortions in project size. In fact, for sufficiently high sensitivity of the agent to policy loss, the principal will be better off with a pooling contract.

The rest of the paper is organized as follows. We review the related literature in Section 2.2, and describe the model in Section 2.3. The main results are in Section 2.4. We begin in Section 2.4.1 with the two type case, and consider the model with a continuum of states in Section 2.4.2. In Section 2.5 we explore in detail a version of our model with quasilinear payoffs for comparison. We conclude in Section 2.6. All proofs are in the Appendix.

2.2 Related Literature

This paper contributes to the optimal delegation literature initiated by Holmström (1977).[5] Holmström (1977) considers a problem in which an uninformed principal contracts with an informed agent over a unidimensional policy space, and the principal cannot use transfers. In this setting, the optimal mechanism for the principal makes policy either completely unresponsive to the agent's type, or equal to his ideal point. This outcome can be achieved by simply delegating decision-making power to the agent over an appropriately chosen set of policies. Melumad and Shibano (1991), and Alonso and Matouschek (2008) then fully characterize the solution to the delegation problem in the absence of restrictions on feasible delegation sets.[6] The optimal delegation set trades off the benefits of making the policy responsive to the state of the world against the loss of decision-making power to a biased agent. When the conflict of interest between principal and agent is sufficiently large the optimal contract is a pooling contract, in which the principal commits to a policy equal to her expected ideal point (delegation is not valuable). The optimal delegation literature presupposes that the principal can commit to a *mechanism* (or set of institutions) regulating her interaction with the agent. At the opposite extreme of the spectrum, Crawford and Sobel (1982) assume that the principal cannot commit to a policy choice. In this context, the principal will always choose her preferred

[5] At a more general level, our work builds on the classic mechanism design and screening literature (see Laffont and Martimort (2009) for a review). Unlike almost all of the literature, we study a two-dimensional screening problem with non-separable preferences. In particular, we deviate from the standard setting with quasilinear preferences, in which one of the policy dimensions is a transfer of money.

[6] See Ambrus and Egorov (2012) and Amador and Bagwell (2012, 2013) for a variant of this problem with money burning.

policy given the information provided by the agent, and as a result cannot reward the agent with policy concessions after the agent reveals her information. The delegation solution has been studied extensively in political science in the context of congressional control of the bureaucracy and executive/legislative relations (see Epstein and O'Halloran (1994), Huber and Shipan (2002)). Bendor and Meirowitz (2004)(Bendor and Meirowitz 2004) and Gailmard and Patty (2012) provide an overview of the theoretical literature in political science and a general framework for this family of models. In addition to these theoretical contributions, there is also a large empirical literature on congressional control of the bureaucracies, which often relies at least informally on the principal agent setup. See for example Weingast and Moran (1983), Wood and Anderson (1993), Wood and Waterman (1991), Carpenter (1996), Shipan (2004).

Baron (2000) and Krishna and Morgan (2008) analyze the problem in which the policy space is unidimensional, but the principal can use transfers to alleviate incentive constraints, assuming quasilinear preferences and quadratic policy payoffs (see also Walsh (1995)). Krishna and Morgan (2008) show that in the solution policy outcomes are systematically distorted to favor the agent's preferences. Thus, even if the principal can use transfers to fully align incentives, in the solution she will choose not to do this to the full extent. The unidimensional policy space with transfers is a natural benchmark for comparison with our model. We relegate this comparison to Section 2.5, where we develop two alternative versions of the model with transfers in the two type case that are directly comparable to our model (one introducing an individual rationality constraint as in the standard screening problems, the other with nonnegative transfers, as in Krishna and Morgan (2008)).

Koessler and Martimort (2012) consider a two-dimensional policy space with separable quadratic payoffs where the agent has the same ideal point in each dimension. They show that interval delegation sets are generally not optimal in this setting, as the optimal decisions on each dimension are never equal to the agent's ideal points.

Two results are particularly relevant for comparison with the results in our paper. First, in this setting pooling is never optimal for the principal. This is true in the two-type version of our model, but not in the continuum. Second, as in our model, the optimal contract can be ex-post inefficient. While in the optimal contract the distance between the policy outcomes in each dimension increases for lower types to induce information revelation, the spread between outcomes at low types can be ex-post too large for both principal and agent.

We are aware of only two papers in political science which consider a model of the interaction of "budget" and policy choice in a principal-agent context.[7] Both papers have fundamental differences with this paper. In particular, both of these papers posit a given sequence of play, and do not consider the optimal mechanism for the principal. In Ting (2001), the agency can choose a more right winged policy at a cost, which enters its quasilinear utility function as a transfer. Congress initially chooses a budget for the agency and, after observing a signal of the agency's choice of policy, an auditing level. In McCarty (2004), the agency needs resources to move policy away from the status quo. The President appoints the agent, while Congress chooses the agency's budget, and thus effectively a range of discretion for the agency around the status quo.

2.3 The Model

A principal is to contract with an agent who has private information about a payoff-relevant state variable. An outcome $(y, m) \in \mathbb{R} \times \mathbb{R}_+ =: X$, comprises a "policy" y, and a scale of implementation or scope m. The policy y is the result of an action $x \in \mathbb{R}$ and a random state variable $\omega \in \Omega$, for a compact set $\Omega \subset \mathbb{R}$. In particular, we let $y = x - \omega$. It is common knowledge that $\omega \sim F$, where we assume that

[7] Banks (1989) considers a model in which an agency has private information about the cost of providing a service, while Congress decides the agency's budget and whether to audit the agency or not.

$supp\,(F) = \Omega$. However, the realization of the random state variable ω is private information of the agent.

Let z_j denote j's ideal policy, $j \in \{P, A\}$. Without loss of generality, we fix $z_P = 0$ and $z_A = b > 0$. We say that b is the *bias* of the agent relative to the principal. The principal and agent have state-contingent preferences. For any action x, state ω, and program size m, the principal's payoff is $U^p\,(x, m|\omega) := u^p(\ell^P(x, \omega), m) - \gamma(m)$, and the agent's payoff is $U^a\,(x, m|\omega) = u^a(\ell^A(x, \omega), m)$, where $\ell^j(x, \omega) := (x - \omega - z_j)^2$. We assume that $\gamma(\cdot)$ is increasing and convex, and that for $j = P, A$, (i) $u^j_\ell \leq 0$ and $u^j_{\ell\ell} \leq 0$, (ii) $u^j_m \geq 0$ and $u^j_{mm} \leq 0$, and (iii) $u^j_{m\ell} \leq 0$. This last assumption says that the value for player j of an extra dollar invested in the program is decreasing in ℓ^j. Our assumptions imply that for each state $\omega \in \Omega$ and player $j \in \{P, A\}$, the "better than" sets $B^j\,(u|\omega) := \{(x, m) : U^j(x, m|\omega) \geq u\}$ are convex. Moreover, they imply that the agent's preferences satisfy the generalized single crossing condition (SCC):

$$\frac{\partial}{\partial \omega}\left(\frac{U^a_x(x, m|\omega)}{U^a_m(x, m|\omega)}\right) \geq 0.$$

We consider the problem of maximizing the principal's payoff by choosing a state-contingent contract with the agent. To allow a rich contract space, we take a mechanism design approach. Without loss of generality, we consider direct truthful mechanisms, in which the principal proposes a menu of contracts $\{(x(\omega), m(\omega))\}_{\omega\in\Omega}$ to the agent, and is committed to implementing the policy $(x(\hat{\omega}), m(\hat{\omega}))$ if the agent announces that the realized state is $\hat{\omega} \in \Omega$. By the revelation principle, any equilibrium outcome of a contract with arbitrary communication protocols between the principal and the agent is implementable by a truthful direct mechanism. Thus, while we will not recover the particular protocol that principal and agent might be using, the solution will capture the equilibrium relation between states and outcomes. Throughout, we will restrict to deterministic mechanisms. We discuss this issue in our concluding remarks.

2.4 Optimal Delegation with Strings Attached

In this section we present our main results. We begin with the binary state space $\Omega = \{0, 1\}$, which presents our results in a highly tractable setting. In Section 2.4.2 we extend our analysis to allow for a continuum of states, i.e., $\Omega = [0, 1]$.

2.4.1 Two Types

Our first order of business is to characterize the first best policy for the principal in state ω, $(\widehat{x}_\omega, \widehat{m}_\omega)$. This is straightforward. Since the principal wants policy to match the state of the world, $\widehat{x}_\omega = \omega$. The optimal scale of the project with full information, on the other hand, is such that the marginal benefit of project expansion *at the ideal policy for the principal* equals the marginal cost; i.e., $u^p_m(0, \widehat{m}_\omega) = \gamma_m(\widehat{m}_\omega)$. Thus $\widehat{m}_0 = \widehat{m}_1 =: \widehat{m}$.

When the agent is privately informed about the realization of the state the principal's problem is to choose (x_0, m_0) and (x_1, m_1) to maximize

$$\sum_{\omega \in \{0,1\}} f(\omega) U^p(x_\omega, m_\omega | \omega)$$

subject to the incentive compatibility (IC) constraints:

$$U^a(x_\omega, m_\omega|\omega) \geq U^a(x_{\omega'}, m_{\omega'}|\omega) \quad \text{for } \omega, \omega' \in \{0, 1\}.$$

The nature of the solution depends on the level of conflict of interests between the principal and the agent. First, as usual in these type of problems, if the conflict of interests between the principal and the agent is sufficiently low ($b \leq 1/2$ in our case),

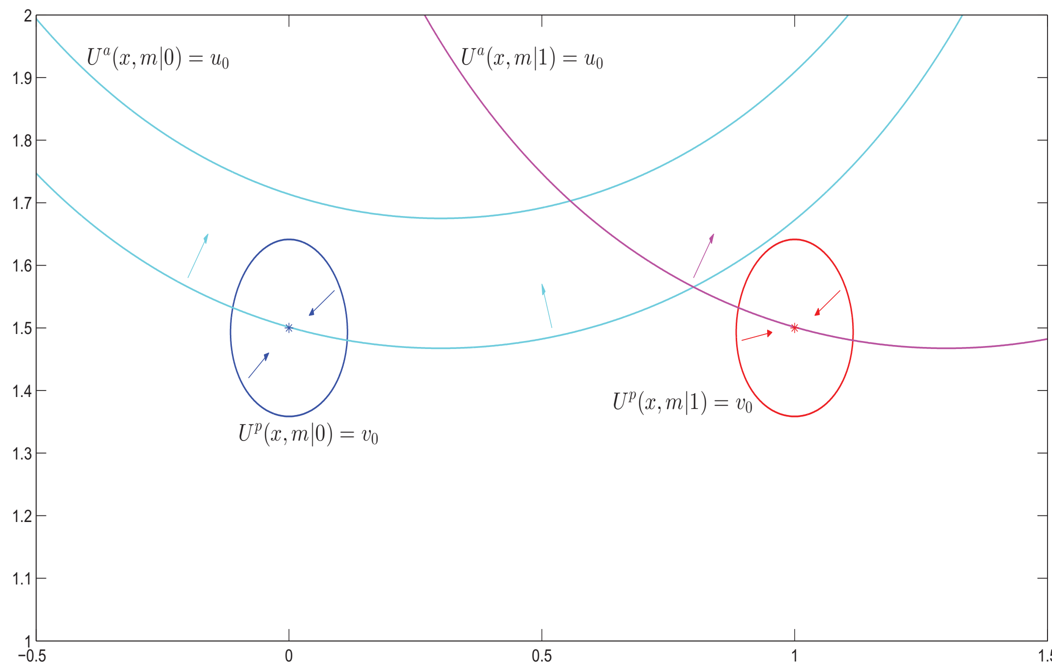


Figure 2.1: First-best achieved with low bias.

the incentive constraints will not be binding in the solution and the principal will be able to achieve her first-best policy in each state (see Lemma 2.A.1). Since the agent has an upward policy bias ($b > 0$), the state 1 incentive constraint is trivially not binding at the first best. And given $b < 1/2$, $(\hat{x}_0, \hat{m})$ is preferred to $(\hat{x}_1, \hat{m})$ for the agent in state 0, since $|\hat{x}_0 - b| < |\hat{x}_1 - b|$ (see Figure 1).

When $b > 1/2$, instead, the principal will not be able to implement the first-best. In this case, the optimal solution for the principal implies trading-off losses in the two states to achieve a policy function that is incentive compatible for the agent. We begin by showing that *generically*, it is optimal for the principal to give the agent some discretion over policy outcomes.[8]

[8]We consider the topological notion of genericity — where a property is generic if it is satisfied in an open dense set; and not generic if it is satisfied only in a closed nowhere dense set. We show that the utility functions which satisfy a necessary condition for pooling are nowhere dense in the appropriate Sobolev space, $W^{1,p}(X)$.

Theorem 2.4.1 (No Pooling) *A pooling contract (x_p^*, m_p^*) is generically suboptimal for the principal.*

The intuition for the result is illustrated in Figure 2.4.1. The principal's indifference curves in state 0 and 1 are depicted in blue and red, respectively. The set of points where the indifference curves in the two states are tangent to one another is shown by the green line. Note that if an optimal pooling contract (x_p^*, m_p^*) is proposed, it will be somewhere on this line, for otherwise we can improve the principal's utility by proposing a pooling contract in this set. (In particular, the optimal pooling contract for $f(0) = 2/5$ is shown by the black circle.) Note however that if (x_p^*, m_p^*) is an optimal contract, it must be that the agent's indifference curve in state 0 is also tangent to the principal's indifference curves at this point. Otherwise utility can be improved by moving "inside" the principal's better-than sets in each state, as shown by the black triangles in the figure. It follows that a pooling contract (x_p^*, m_p^*) can only be optimal if a triple tangency of indifference curves is satisfied, a property that only holds in a closed nowhere dense set of utility functions.

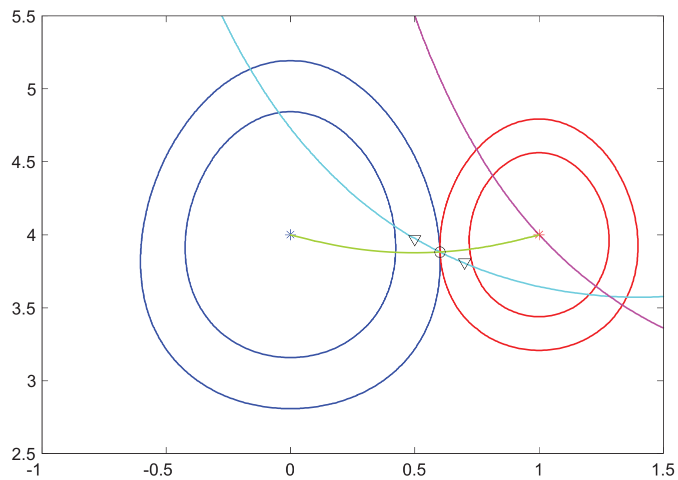


Figure 2.2: Agency Discretion: Pooling Contracts are Not Optimal

Given that the second-best solution involves granting the agent discretion, the principal has to design the policy function to ensure that the agent has incentives to report truthfully. Achieving incentive compatibility in this setup will necessarily imply policy distortions, which are themselves costly to the principal. The principal will therefore shape the policy function to achieve its objective in the least costly manner. Given that the binding incentive constraint is that of state 0, this entails making policy in state 1 *less* attractive to the agent and/or policy in state 0 *more* attractive to the agent in the least costly manner for the principal.

Formally, we define the state-contingent contract curves, a *reward curve* in state 0, $CC(0)$, and a *discipline curve* in state 1, $CC(1)$. Let $V := [U^a(\widehat{x}_0, \widehat{m}_0|0), U^a(\widehat{x}_1, \widehat{m}_1|0)]$. Then

Definition 2 *The reward curve is the set of points* $CC\,(0) := \{(\widetilde{x}^0(u), \widetilde{m}^0(u)) : u \in V\}$*, where*

$$(\widetilde{x}^0(u), \widetilde{m}^0(u)) := \arg\max_{(x,m)} U^p(x, m|0) \text{ s.t. } U^a(x, m|0) \geq u.$$

The discipline curve is the set of points $CC\,(1) := \{(\widetilde{x}^1(u), \widetilde{m}^1(u)) : u \in V\}$*, where*

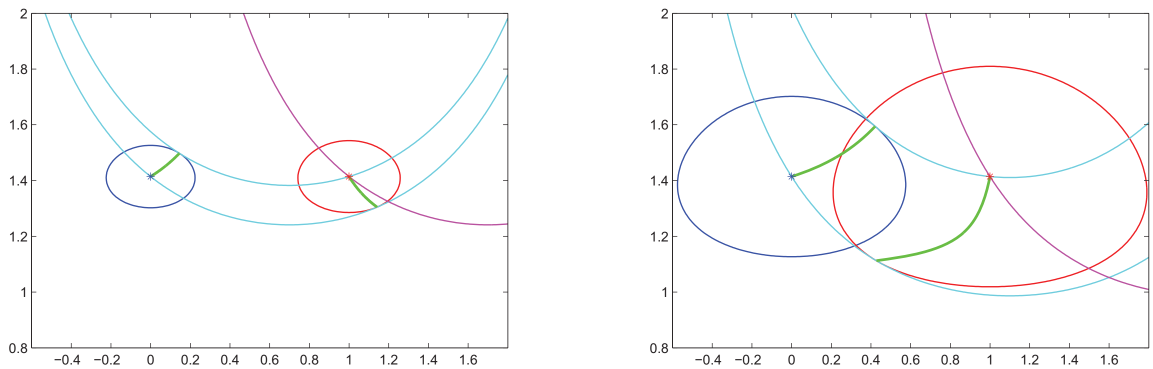
$$(\widetilde{x}^1(u), \widetilde{m}^1(u)) := \arg\max_{(x,m)} U^p(x, m|1) \text{ s.t. } U^a(x, m|0) \leq u.$$

Because policies on the contract curves reward the agent in state 1 and discipline the agent in state 0 efficiently, a policy lying anywhere outside the contract curves can be improved with an alternative policy that preserves incentives and increases the principal's utility. As a result,

Lemma 2.4.2 *The optimal incentive compatible policy for the principal lies on the contract curves; i.e.,* $(x^*_\omega, m^*_\omega) \in CC\,(\omega)$*, and thus*

$$\frac{U^p_x(x^*_\omega, m^*_\omega|\omega)}{U^p_m(x^*_\omega, m^*_\omega|\omega)} = \frac{U^a_x(x^*_\omega, m^*_\omega|0)}{U^a_m(x^*_\omega, m^*_\omega|0)} \quad \text{for } \omega = 0, 1. \tag{2.4.1}$$

Lemma 2.4.2 allows us to characterize the nature of the distortions in the scope and direction of policy through the shape of the contract curves. To do this it is useful to distinguish two cases. We say that the agent is a *moderate* if his ideal policy in state 0 is below the first best policy for the principal in state 1; i.e., if $b < 1$. We say that the agent is a *zealot* if $b > 1$.



Moderate ($1/2 < b < 1$) Zealot ($b > 1$)

Figure 2.3: Contract curves reflect the nature of distortions

The left panel of Figure 2.4.2 plots representative contract curves for a moderate agent. As the figure illustrates, the reward curve is increasing, and the discipline curve is decreasing. This means that policy will be distorted in the direction of the agent's bias in both states, and that the scale of implementation will be *larger* than the first best for the principal in state $\omega = 0$, but *smaller* than the first best for the principal in state $\omega = 1$.

The intuition for the result can be seen graphically in Figure 2.4.1. When $b < 1$, the state 1 indifference curves for the principal and the state 0 indifference curves for the agent (blue) are tangent below and to the right of the principal's ideal point in state 1. Because the ideal policy of the agent in state 0 is still lower than the ideal policy of the principal in state 1, the least costly way to leave the agent at some utility level u below than what he would obtain at $(\hat{x}_1, \hat{m}_1)$ is to reduce the scale of the program m and increase the policy direction x (achieving this payoff for the agent with some point $x < 1$ would be more costly to the principal, as any point on the blue indifference curve with $x < 1$ is in a lower state 1 indifference curve for the principal).

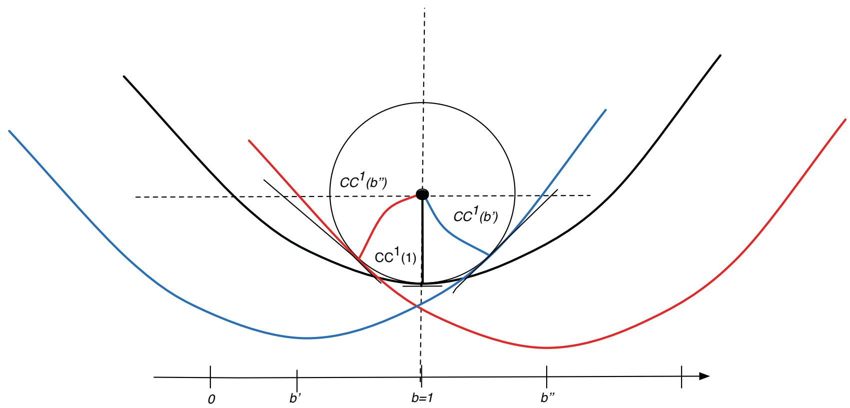


Figure 2.4: Recoil of the Optimal Policy

How much does policy need to adjust relative to project size depends on the strength of the agent's bias. Consider a point in the disciplining curve $CC(1)$ for an agent with bias $b' < 1$, as illustrated in Figure 2.4.1. At any such point, the relative value of changes in policy direction and scale in state 1 for a state-0 agent is given by the agent's state-0 MRS between policy and project size at this point. Now suppose that we increase the agent's bias to $\tilde{b} \in (b', 1)$. This agent would have a flatter indifference curve through the point. Because in state 0 the $\tilde{b}$ bias agent's preferred policy is closer to the ideal policy of the principal in state 1, this agent is willing to give away a larger loss in policy direction to get a given amount of additional project size. Thus, the contract curve for an agent $\tilde{b} \in (b', 1)$ will be steeper than for b'. In the extreme, for $b = 1$, the ideal policy of the agent in state 0 coincides with the ideal policy of the principal in state 1. Thus, the most efficient way to punish the (state-0) agent in state 1 is to reduce the implementation scale without changing policy.

When the agent is a zealot, instead, the nature of the optimal policy changes. In this case, *both* the reward curve $CC(0)$ and the discipline curve $CC(1)$ are increasing (right panel of Figure 2.4.2). This means that while the implementation scale of the policy in state $\omega = 1$ will be smaller than in the first-best as in the previous case, the direction of the policy outcome in state $\omega = 1$ will now be distorted *against* the direction of the agent's bias. The reasoning is symmetric to the previous case. When $b > 1$, as for b'' in the figure, the ideal policy of a state 0 agent is larger than the ideal policy of the principal in state 1. Thus, the least costly way to leave the agent at some utility level u below what he would obtain at $(\hat{x}_1, \hat{m}_1)$ is now to *decrease* policy direction x and reduce the implementation scale m as before. Furthermore, as before (in logic if not in direction), as we continue to increase the bias of the agent above $b = 1$, increases in the value of the state 1 implementation scale become less valuable for the state 0 agent relative to gains in policy, and the most efficient way to punish this agent is through small reductions in project size and sharp distortions in policy (a flatter discipline curve).

The next theorem summarizes the previous discussion.

Theorem 2.4.3 *Suppose $b > 1/2$. Then the optimal incentive compatible solution entails distortions in both states; i.e., $(x^*_\omega, m^*_\omega) \neq (\widehat{x}_\omega, \widehat{m}_\omega)$ for $\omega = 0, 1$. Moreover,*

1. *The principal overfunds the agent relative to first best in state 0 ($m_0 > \hat{m}_0$) and underfunds the agent in state 1 ($m_1 < \hat{m}_1$).*

2. *The optimal contract for a moderate agent distorts policy towards the agent in both states; i.e., $x_\omega > \omega$ for all $\omega \in \{0, 1\}$. The optimal contract for a zealot distorts policy in favor of the agent in state 0 but against the agent in state 1; i.e., $0 < x_0 < x_1 < 1$. When $b = 1$, $x_1 = 1$.*

3. *In the optimal contract for a moderate (a zealot) the distortion in state 1 project size increases (decreases) continuously with the agent's bias b and the distortion in policy direction decreases (increases) continuously with b.*

In the context of our environmental policy example, for instance, Theorem 2.4.3 says that Congress underfunds the EPA relative to the first best level precisely when climate change is occurring rapidly, and overfunds the agency relative to the first best if climate change turns out not to be a grave concern. Overfunding in the "low" state always comes together with an environmental policy that is overly aggressive for the median legislator. The distortions in policy in the "high" state, however, depend on the extent of conflict of interests between Congress and the agency: when the EPA is only moderately biased relative to the median legislator, environmental policy does more to curb emissions than what Congress would want, but when the conflict of interests between the EPA and Congress is high, the optimal incentive compatible plan sets a lax environmental policy when climate change is accelerating. This implies that when climate change is indeed occurring at a fast pace, *both regulations and resources* dedicated to curbing carbon emissions are too low relative to Congress' first best policy. Thus, ex-post, in these cases both Congress and the agency would favor more stringent regulations and an increase of resources to the EPA.

Theorem 2.4.3 characterizes the qualitative nature of the distortions and was entirely independent of the principal's prior, f. *How much* the principal distorts policy in each state depends on the likelihood of each state. Note that since the principal chooses between pairs of points on the contract curves, we can rewrite the principal's problem as:

$$\max_{u \in V} \; f(0)U^p(\widetilde{x}^0(u), \widetilde{m}^0(u)|0) + f(1)U^p(\widetilde{x}^1(u), \widetilde{m}^1(u)|1) \tag{2.4.2}$$

In particular, the first order conditions at the optimal level u^* (interior by Theorem 2.4.3) imply that:

$$\frac{f(0)}{1-f(0)} = -\frac{\partial U^p(\widetilde{x}^1(u), \widetilde{m}^1(u)|1)/\partial u}{\partial U^p(\widetilde{x}^0(u), \widetilde{m}^0(u)|0)/\partial u}.$$

Note that the optimal trade-off between distortions in state $\omega = 1$ and state $\omega = 0$ depends on the likelihood of each state. In fact, since $\partial U^p(\widetilde{x}^0(u), \widetilde{m}^0(u))/\partial u < 0$ and $\partial U^p(\widetilde{x}^1(u), \widetilde{m}^1(u))/\partial u > 0$, as $f(0)$ increases we "move down the contract curves" in Figure 2.4.2, reducing the size of the distortion in state $\omega = 0$ in exchange for an increased policy distortion in state $\omega = 1$. Therefore, as state 0 becomes more probable, the magnitude of the distortions in the direction and implementation scale of policy in state 1 will be more severe.

2.4.2 Continuum of Types

We now extend our analysis to the case in which there is a continuum of states. In this context, the principal offers the agent a menu of incentive compatible contracts $\{x(\omega), m(\omega)\}_{\omega \in [0,1]}$. Letting $\mathcal{U}^a(\hat{\omega}, \omega) := U^a(x(\hat{\omega}), m(\hat{\omega})|\omega)$, the principal's problem is:

$$\max_{\{x(\omega), m(\omega)\}} \int_0^1 U^p(x(\omega), m(\omega)|\omega) f(\omega) d\omega \tag{PP}$$

subject to:

$$\mathcal{U}^a(\omega, \omega) \geq \mathcal{U}^a(\hat{\omega}, \omega) \text{ for all } \omega, \widehat{\omega} \in [0,1].$$

Our main goal is to establish whether the nature of the distortions in policy we obtained in the two-type model extend naturally to the case in which there are multiple states. With this goal in mind, we will focus on characterizing the optimal fully separating contract, which we assume to be differentiable.

For our richer results, we will assume that principal and agent have *exponential payoffs*;[9] i.e., that:

$$U^a(x, m|\omega) = m \exp(-\beta|x - \omega - b|),$$

and:

$$U^p(x, m|\omega) = m \exp(-\eta/2(x - \omega)^2) - \frac{\gamma}{2}m^2.$$

With this assumption, we will be able to characterize the optimal menu of contracts in sufficiently rich detail so as to compare the results with the two-type case. We will also show that in this case the optimal contract is continuous and piecewise differentiable. Thus, the original assumption of differentiability only rules out kinks in the optimal contract. We will then also provide conditions under which the fully separating contract dominates any pooling contract in this context.

Our first step is to reduce the continuum of incentive compatibility constraints in (PP) in the usual way. We show that as in the standard quasilinear model, as long as the policy function has the property that $x(\cdot)$ is increasing in ω, only *local* deviations are relevant.[10] This argument has two parts. The local incentive compatibility constraint for type ω ensures that type ω can not gain by announcing to be a type arbitrarily close to ω. A necessary condition for no profitable local deviations at ω is that:

$$\left.\frac{\partial \mathcal{U}^a(\hat{\omega}, \omega)}{\partial \hat{\omega}}\right|_{\hat{\omega}=\omega} = 0,$$

[9]Assuming a specific utility function simplifies the analysis considerably and is standard in the literature. Baron (2000) and Krishna and Morgan (2008) assume quadratic policy payoffs in a unidimensional policy space with separability of transfers (i.e., a quasilinear utility function). Melumad and Shibano (1991) assume quadratic payoffs in a unidimensional policy space with no transfers. In the same context, Alonso and Matouschek (2008) assume quadratic payoffs for the principal, and a single-peaked symmetric utility function for the agent. Koessler and Martimort (2012)(Koessler and Martimort 2012) assume that payoffs are quadratic in each dimension and separable across dimensions. We deviate from the quadratic payoffs assumption that is prevalent in the literature because of the non-separability of payoffs that is at the core of our problem.

[10]Part of the complexity of multidimensional mechanism design problems is that the order in which incentive constraints bind is typically endogeneous to the mechanism (e.g., see the review by Rochet and Stole (2003). The fact that the agent's preferences in our model satisfy a generalized single crossing condition allows us to abstract from these complexities. However we still have to deal with technical issues arising from the lack of quasilinearity.

or equivalently:

$$m'(\omega) = -\underbrace{\frac{U_x^a(x(\omega), m(\omega)|\omega)}{U_m^a(x(\omega), m(\omega)|\omega)}}_{MRS_{xm}^a(\omega)} x'(\omega). \tag{2.4.3}$$

Thus, incentive compatibility implies that at any point ω, the rate of change in the scope of policy in the optimal contract must be proportional to the rate of change of policy direction by a factor given by the agent's marginal rate of substitution in that state. Condition (2.4.3) is also sufficient to assure no profitable local deviations if $x(\cdot)$ is nondecreasing (see Lemma 2.A.2). In fact, because of the single crossing condition, if $x(\cdot)$ is nondecreasing, (2.4.3) is necessary and sufficient to rule out both local and global deviations (see Lemma 2.A.3).

We can then write the principal's problem (PP) as:

$$\max_{\{x(\omega), m(\omega)\}} \int_0^1 U^p(x, m|\omega) f(\omega) d\omega$$

subject to the law of motion (2.4.3) and the constraints $x'(\omega) \geq 0$, $m(\omega) \geq 0$.

The law of motion defines a functional equation for the project size m in terms of x, the only control variable in the above problem is x. We can write the above problem in Bolza form by introducing a new function y, which will be a "stand-in" for x'. This, of course, requires an additional constraint reflecting that relationship. The optimal control problem[11] is therefore:

$$\max_{y(\omega)} \int_0^1 U^p(x, m|\omega) f(\omega) d\omega$$

subject to:

$$\begin{aligned} m'(\omega) &= -\frac{U_x^a(x(\omega), m(\omega)|\omega)}{U_m^a(x(\omega), m(\omega)|\omega)} y(\omega), \\ x'(\omega) &= y(\omega), \; m(\omega) \geq 0. \end{aligned}$$

[11] Ignoring the $x'(\omega) \geq 0$ constraint for now.

The Hamiltonian for this problem is then:

$$\mathcal{H} = U^p(x, m|\omega)f(\omega) - \lambda_1 \frac{U_x^a(x, m|\omega)}{U_m^a(x, m|\omega)} y + \lambda_2 y + \mu m.$$

The associated Euler-Lagrange conditions for a fully separating solution are characterized in Remark 3 in the Appendix.

We can now prove our second substantive result. We have already established that the policy content $x(\cdot)$ is nondecreasing in type. (In fact, in a separating equilibrium $x(\cdot)$ must be strictly increasing, for if $x' = 0$ in an interval $[a, b] \subset [0, 1]$, then (2.4.3) implies that $m' = 0$ in $[a, b]$, which implies pooling.) We next show that the implementation scale $m(\cdot)$ similarly must be nonincreasing in type. Thus, as in the binary state environment, the scale of implementation of policy decreases with the direction of policy x. To show this result, we must first establish an intermediate step.

Lemma 2.4.4 *In the solution to (PP), $x(\omega) \leq \omega + b$ for all $\omega \in [0, 1]$.*

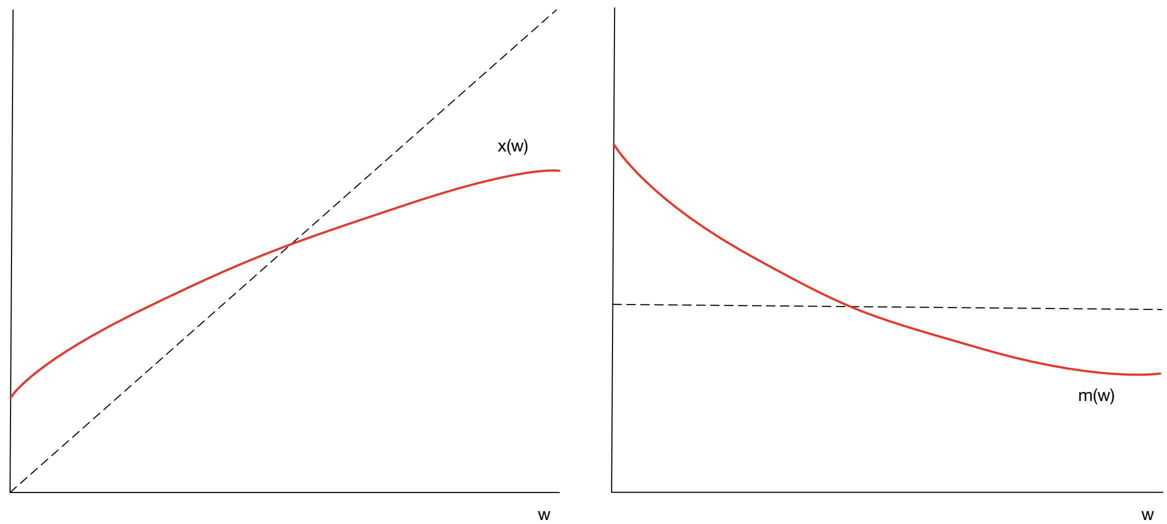
Thus, in the optimal separating contract, the direction of policy is never larger than the ideal policy of the agent. Our result now follows immediately using Lemma 2.4.4. Note that given that $x(\cdot)$ is nondecreasing and $U_m^a(\cdot) > 0$, the truth telling condition (2.4.3) says that $m(\cdot)$ will be weakly decreasing at ω if and only if $U_x^a(\cdot) \geq 0$. But this happens if and only if $x(\omega) \leq \omega + b$. We then have

Corollary 2.4.5 *In the optimal incentive compatible plan (x, m)*

$$\frac{dm}{dx}(\omega) = -\frac{U_x^a(x(\omega), m(\omega)|\omega)}{U_m^a(x(\omega), m(\omega)|\omega)} \leq 0 \quad \text{for all } \omega \in [0, 1],$$

with strict inequality whenever $x(\omega) < \omega + b$.

A second implication of Lemma 2.4.4 is that the equilibrium payoff of the agent is decreasing in type. Thus, the lowest type (type 0) makes the largest informational



$x(\cdot)$ $m(\cdot)$

Figure 2.5: In the optimal separating contract, policy $x(\cdot)$ is a strictly increasing function of the state ω and project size is a weakly decreasing function of the state ω (strictly decreasing whenever $x(\omega) < \omega + b$).

rent. To see this, note that by the envelope theorem (or substituting eq. 2.4.3), $U_x^a(x(\omega), m(\omega)|\omega)x'(\omega) + U_m^a(x(\omega), m(\omega)|\omega)m'(\omega) = 0$. Then

$$\frac{d}{d\omega}U^a(x(\omega), m(\omega)|\omega) = U_\omega^a(x(\omega), m(\omega)|\omega) < 0,$$

where the inequality follows from the fact that $x(\omega) \leq \omega + b$.

Providing a more detailed characterization of the solution at this level of generality is difficult. However, in order to evaluate whether the lessons we learned in the two-type case extend to this environment, we need to characterize the nature of the distortions relative to the first best. As a first step in this direction, we will assume hereafter that principal and agent have exponential payoffs. A key property of this payoff specification is that from the truth telling condition (2.4.3) we can obtain a closed form for the project size function $m(\cdot)$ as a function of the policy $x(\cdot)$ and the state ω itself. This, in turn, allows us to take a direct approach to solve PP without solving for the multipliers $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ in the constrained optimization formulation of the Euler-Lagrange conditions (2.A.12-2.A.17).[12]

Note that since in the solution $x(\omega) \leq \omega + b$ for all ω by Lemma 2.4.4, we can write $U^a(x, m|\omega) = m \exp(\beta[x - \omega - b])$. Then (2.4.3) becomes $\frac{m'(\omega)}{m(\omega)} = -\beta x'(\omega)$, and thus

$$m(\omega) = m_0 \exp\left(-\beta[x(\omega) - x_0]\right) \tag{2.4.4}$$

It follows that the optimal project size $m(\cdot)$ is a strictly decreasing function of ω, which decreases faster the steeper $x(\omega)$ is and the more responsive the agent is to policy loss, as measured by β.

Given that expression (2.4.4) incorporates the IC constraints (2.4.3), we can now directly substitute (2.4.4) into the objective function of the principal. Since $m(\omega) > 0$ for all ω, the constraint $m(\omega) \geq 0$ is not binding. In addition, as it is common in the literature, we will ignore the constraint that $x'(\omega) \geq 0$ and check that the solution satisfies the constraint ex-post. With these remarks, we can rewrite the principal's problem as

$$\max_{x_0, m_0, x(\cdot)} J(x_0, m_0, x(\cdot)) = \int_0^1 U^p(x(\omega), m_0 \exp(-\beta[x(\omega) - x_0])|\omega) f(\omega) d\omega \quad \text{s.t. } x(0) = x_0.$$

The necessary first order condition with respect to $x(\omega)$ then gives

$$MRS^p_{xm}(\omega) \equiv \frac{U^p_x(x(\omega), m(\omega)|\omega)}{U^p_m(x(\omega), m(\omega)|\omega)} = \beta m(\omega) = MRS^a_{xm}(\omega) \quad \forall \omega \in [0, 1]. \tag{2.4.5}$$

Note that since $m(\cdot)$ is strictly decreasing, this implies that the principal's marginal rate of substitution in the optimal contract is decreasing in ω, so that increases in policy are relatively less valuable for the principal in higher states, while

[12] This simplification does come at a cost in generality. In particular, we want to point out that in this case the single crossing condition is only satisfied weakly, as $\frac{\partial}{\partial \omega}\left(\frac{U^a_x(x,m|\omega)}{U^a_m(x,m|\omega)}\right) = 0$. As a result, the optimal contract will remove the incentive for the agent to misrepresent his type leaving him indifferent over sending alternative reports.

increases in implementation scale are relatively more valuable for the principal in higher states.

The optimality condition (2.4.5) leads to two immediate implications. First, note that since $U_x^p(\cdot) \leq 0$ if and only if $x \geq \omega$, the optimal contract will overfund any project that distorts policy towards the agent and underfund projects that distort policy against the agent.

Theorem 2.4.6 *In the optimal separating solution,* $m(\omega) \geq \hat{m} = 1/\gamma \Leftrightarrow x(\omega) \geq \omega$.

Second, using this result we can show that the optimal separating solution is continuous and piecewise differentiable. Thus, in particular, the optimal differentiable contract that we characterize here dominates any solution with discontinuities, and the original assumption only rules out kinks in the optimal contract.

Theorem 2.4.7 *Suppose principal and agent have exponential payoffs. Then fully separating solutions to the principal's problem are continuous and piecewise differentiable.*

Theorem 2.4.7 builds on Lemma 2.A.4 in the Appendix, which shows that in general (for any utility functions satisfying the assumptions of the model) the equilibrium payoffs of principal $U^p\left(x\left(\omega\right), m\left(\omega\right)|\omega\right)$ and agent $U^a\left(x\left(\omega\right), m\left(\omega\right)|\omega\right)$ are continuous in ω. This rules out all but a specific kind of discontinuity, which we can then rule out with (2.4.5).

It is important to note that the equality of the marginal rates of substitution of the principal and the agent in (2.4.5) does *not* imply that the optimal incentive compatible contract is efficient. Consider for example Figure 2.4.2, which plots a possible solution $(x(\cdot), m(\cdot))$ in the (x, m) space. Note that at each point in the curve, the marginal rates of substitution of the principal and the agent are equal. However, the contract underfunds and distorts policy against the agent in "high" states (states $\omega > \hat{\omega}$), as in the zealot solution of the two type model. In these high states, the indifference curve of the agent is tangent to the indifference curve of the principal from below, and both the principal and the agent would prefer to increase funding and choose a higher policy.

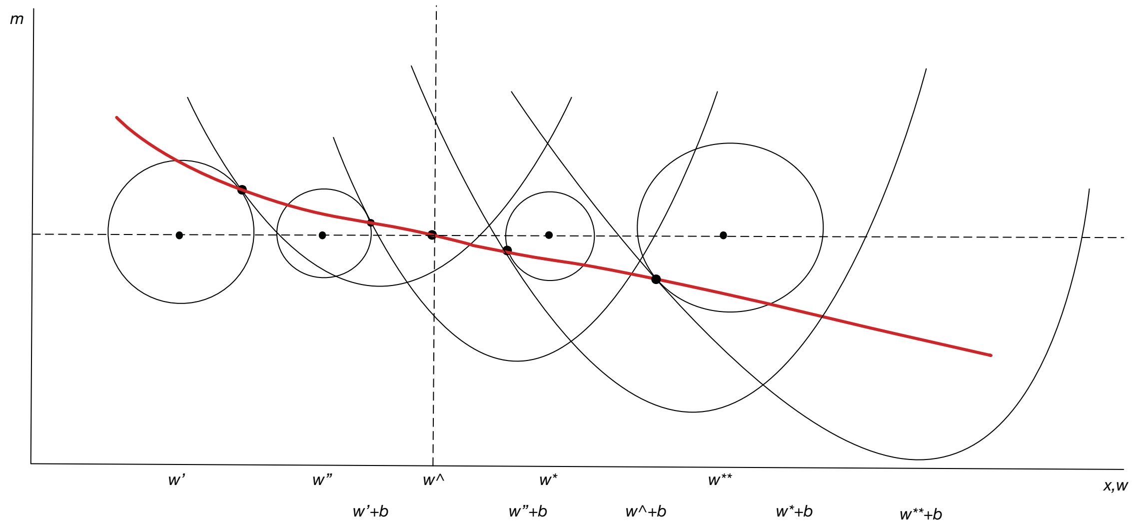


Figure 2.6: Optimal Contract with Exponential Payoffs.

The result sketched in the figure is interesting because it represents the natural generalization of the results for the two-type model. Since in the continuum local deviations are "small" relative to the size of the bias, the agent is always a "zealot", and thus, based on the logic of the two-type model, we would expect that the optimal contract will overfund and distort policy in favor of the agent for "low types" and underfund and distort policy against the agent for "high types", leading to the ex-post inefficiency.[13]

[13]Consider first a finite state space with typical element ω_t, $t = 1, \ldots, T$, such that $\omega_{t+1} - \omega_t = \Delta$ for some $\Delta > 0$. A direct extrapolation of the two-type results to the finite state case would imply that (i) $x(\omega_t) = \omega_t$ for all t whenever $b < \Delta/2$, as in the "low bias" case, (ii) that $x(\omega_t) > \omega_t$ whenever $\Delta/2 < b < \Delta$, as in the "moderate" case, and that (iii) when $b > \Delta$ there exists a $\bar{t}$ such that $x(\omega_t) > \omega_t$ for $t \leq \bar{t}$ and $x(\omega_t) < \omega_t$ for $t > \bar{t}$, so that policy distorts in favor of the agent for "low types" and against the agent for "high types", as in the "zealot" case. Now as $\Delta \to 0$, eventually $b > \Delta$, and only the "zealot" case has bite in the continuum.

Our main goal is to confirm whether in fact the solution has this strong inefficiency ex-post, or if instead the optimal incentive compatible contract eliminates all gains from trade (as is the case for "low" states in the contract depicted in the figure). The result is stated in Theorem 2.4.8. For this it will be convenient to define $r(\omega) \equiv (\eta/2)[(x(\omega)-\omega)^2 - x_0^2] - \beta[x(\omega) - x_0]$.

Theorem 2.4.8 *The optimal fully separating incentive compatible policy* $(x(\cdot), m(\cdot))$ *is such that there exists a* $\hat{\omega} \in (0,1)$ *such that* $x(\omega) > \omega$, *and* $m(\omega) > \hat{m}$ *for* $\omega \in [0, \hat{\omega})$ *and* $x(\omega) < \omega$, *and* $m(\omega) < \hat{m}$ *for* $\omega \in (\hat{\omega}, 1]$. *Moreover, for all* $\omega \in [0,1]$,

$$x'(\omega) = \frac{(x(\omega)-\omega)\exp(r(\omega)) - \frac{1}{(\beta+\eta x_0)}}{\left[(x(\omega)-\omega) - \frac{\beta}{\eta}\right]\exp(r(\omega)) - \frac{1}{(\beta+\eta x_0)}}. \tag{2.4.6}$$

The proof of Theorem 2.4.8 builds on two key facts. First, we show from the transversality condition in the Euler-Lagrange equations that *(i)* the optimal policy $x(\cdot)$ cannot be always above or always below ω. Second, a direct examination of (2.4.6) shows that $x'(\omega) < 1$ whenever $x(\omega) < \omega$. Thus, *(ii)* if $x(\omega) < \omega$ at some $\omega \in [0,1]$, then $x(\omega') < \omega'$ for all $\omega' \geq \omega$. This in turn implies that $x_0 > 0$, for otherwise x would always be below ω, which contradicts *(i)*. So $x(\omega)$ starts above ω and then must cross ω at least once. But by *(ii)*, if x is ever below ω, it will not go back up. Thus it must cross ω only once.

Fact (i) above captures the resolution of a tradeoff for the principal. Recall that in the first best the policy matches the state, $x(\omega) = \omega$ for all $\omega \in [0,1]$, and the project size $m(\cdot)$ is flat at $\hat{m} = 1/\gamma$. Now, in order to provide incentives without using transfers, the principal needs to introduce distortions in policy and/or implementation scale. In the exponential case, the tradeoff between these distortions is given by the expression $m'(\omega) = -\beta m(\omega) x'(\omega)$. This reflects that in order to make $x(\cdot)$ responsive to the state (reducing distortions in policy), $m(\cdot)$ needs to be responsive to the state as well (increasing distortions in implementation scale). Thus, in choosing how close $x(\cdot)$ can trace ω, the principal faces a tradeoff between inducing distortions in policy direction versus inducing distortions in implementation scale.[14]

The particular resolution of the tradeoff between distortions in policy and project size depends crucially on the *relative* sensitivity to policy loss of the agent vis a vis the principal, captured here by the ratio β/η. Note that the denominator of (2.4.6) is equal to the numerator minus the term $\frac{\beta}{\eta}\exp(r(\omega))$. It follows that $x'(\omega) \to 1$ for all ω as $\beta/\eta \to 0$. In this situation the agent does not care too much about policy losses relative to the principal. Thus, it is relatively cheap for the principal to compensate the agent with changes in project size to obtain a policy that is very close to her first best. In particular, if $\beta \to 0$, the agent does not care much about policy losses (not only relative to the principal) and thus is cheap to compensate with small changes in project size. As we can see from (2.4.4), in this case $m(\cdot)$ will be relatively flat. This of course is excellent for the principal, who can achieve an outcome close to her first best.

Now, recall that we solved the principal's problem PP assuming that in the solution $x(\cdot)$ would be nondecreasing. The previous discussion makes clear that when the agent is sufficiently insensitive to policy loss relative to the principal in fact $x'(\omega) > 0$ for all $\omega \in [0,1]$, and thus the optimal incentive compatible contract will be fully separating.

Theorem 2.4.9 *There exists $\delta > 0$ such that if $\beta/\eta < \delta$, the optimal incentive compatible contract is a fully separating contract.*

[14]This simple logic is due in part to the exponential/linear payoffs. In this case, the agent's MRS does not depend on the distance between policy $x(\omega)$ and the agent's preferred policy, $\omega + b$. Thus, if $x(\cdot)$ is always below or always above ω the principal can do better in all states by just shifting $x(\cdot)$ above or below. Now, in general, the agent's MRS can be decreasing as x gets closer to $\omega + b$. When this is the case, the principal has two competing ways of reducing distortions in project size: by making $x(\cdot)$ flatter, and by making $x(\cdot)$ close to $\omega + b$.

When the agent is very sensitive to policy losses, on the other hand (β large), attaining a policy $x(\cdot)$ close to the first best is very costly in terms of project size distortions, and the principal would rather take a relatively unresponsive policy than introduce large distortions in project size. In fact, for sufficiently high β, the principal will be better off with a pooling contract. To see this, note that with $\beta \to \infty$ the numerator of (2.4.6) goes to $x(\omega) - \omega$ and the denominator to $-\infty$. Since $x(\omega) > \omega$ for some ω, this implies that $x' < 0$, which violates the monotonicity constraint. When the agent is much more sensitive to policy loss than the principal the agent is too costly to compensate, and it is better for the principal to pool types. It is worth noting here that while this result is in line with many papers in the literature, it is in contrast with Koessler and Martimort (2012), who show that in a model with separability across policy dimensions pooling is never optimal.[15]

2.5 Discussion: The Quasilinear Model

To put our results in the context of the previous literature with transferable utility, we consider a version of the model in which "project size" enters into the utility of principal and agent simply as transfers in a quasilinear utility function. For simplicity, we do this here for the two type model. For any action x, state ω, and program size m, the principal's payoff is now $U^p\left(x, m|\omega\right) := v^p(\ell^p(x, \omega)) - m$, and the agent's payoff is $U^a\left(x, m|\omega\right) = v^a(\ell^a(x, \omega)) + m$. As before, we assume that for $j = P, A$, (i) $v^j_\ell < 0$ and (ii) $v^j_{\ell\ell} \leq 0$.

In order for this problem to be well defined, we need to ensure that it is bounded in some way (otherwise the principal could ask for infinite transfers from the agent). There are two possible ways to do this, and the choice makes a large difference to the qualitative results. The first approach is to introduce the standard individual

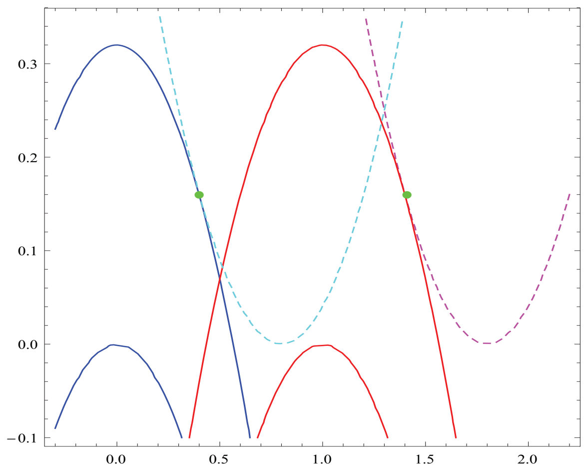
[15]The result holds for any level of conflict of interest (bias) between principal and agent, provided the conflict of interest between the principal and the agent are different on each dimension (Martimort and Semenov, 2006).

rationality constraint in principal-agent models, where the contract must assure the agent a minimum utility level determined by an outside option. The second is to impose the constraint that transfers are non-negative, which puts a hard lower-bound on what can be achieved. This corresponds to the model of delegation with transfers in Krishna and Morgan (2008), Section 3.

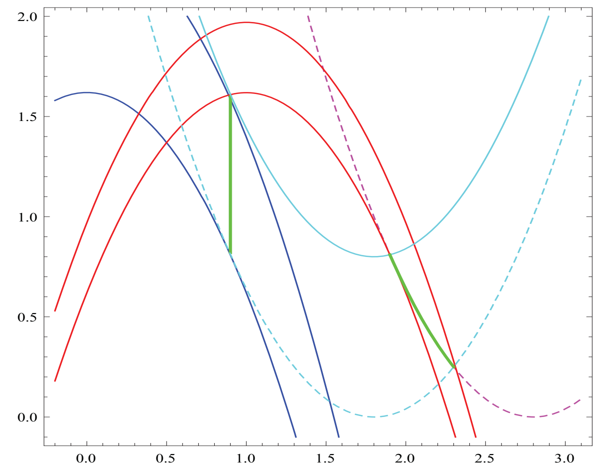
In both cases the Principal chooses (x_0, m_0) and (x_1, m_1) to maximize expected utility $E[v^p(\ell^p(x_\omega, \omega)) - m_\omega]$ subject to the incentive compatibility constraints that the plan (x_ω, m_ω) is optimal for the agent in state ω; i.e., that $v^a(\ell^a(x_\omega, \omega)) + m_\omega \geq v^a(\ell^a(x_{\omega'}, \omega)) + m_{\omega'}$ for $\omega, \omega' \in \{0, 1\}$. In the agent individual rationality version, we close the model by including the individual rationality (IR) constraints:

$$v^a(\ell^a(x_\omega, \omega)) + m_\omega \geq 0. \tag{2.5.1}$$

In this formulation, the problem boils down to a standard screening problem. When bias is "low" ($b \leq \hat{b}$ for some $\hat{b}$), the first-best is incentive compatible (left panel of Figure 2.5). When $b > \hat{b}$, the first-best is unattainable, and the solution has many of the features of the textbook screening problem,. This is illustrated in the right panel of Figure 2.5. The contract curves (in green) characterize the efficient frontier of a tradeoff between extracting more surplus in state 0 or state 1 with incentive compatible contracts. Points to the south and southeast of the contract curves in state $\omega = 0$ and $\omega = 1$ lead to a higher payoff for the principal in state 0, and points to the north and northwest of the contract curves leading to a higher payoff for the principal in state 1. Since utility is perfectly transferable between the principal and agent, both x_0 and x_1 are increasing in the agent's bias b, so unlike in our model, there is no recoil effect. As in our model, though, in the quasilinear model with an IR constraint pooling is generically suboptimal.



$b < \hat{b}$ (First Best)



$b > \hat{b}$ (Constrained Optimal)

Figure 2.7: Quasilinear Model with an IR constraint. Policy x on the horizontal axis and transfers m on the vertical axis. The principal's indifference curves are blue in state 0 and red in state 1. The agent's indifference curves in states 0 and 1 are light blue and light red, respectively. Indifference curves satisfying the IR constraint strictly are dashed.

Consider now the case of non-negative transfers, as in Krishna and Morgan (2008). Now we replace the IR constraint (2.5.1) with

$$m_\omega \geq 0, \text{ for } \omega \in \{0, 1\}\,. \tag{2.5.2}$$

If the agent's bias is sufficiently low ($b \leq 1/2$) we can implement the first best, so assume $b > 1/2$. In a separating solution, the IC constraint in state $\omega = 1$ will not bind, and the IC constraint in state $\omega = 0$ will hold with equality. From here we can obtain $m_0 = m_1 + [v^a(\ell^a(x_1, 0)) - v^a(\ell^a(x_0, 0))]$. Substituting in the objective function and noting that in the solution $m_1^* = 0$, we can write the principal's problem as

$$\max_{x_0,x_1} f(0)[v^p(\ell^p(x_0,0)) - v^a(\ell^a(x_1,0)) + v^a(\ell^a(x_0,0))] + f(1)v^p(\ell^p(x_1,1))$$

From the first order condition with respect to x_0 we obtain $U_x^p(x_0^*, m_0^*|0) = -U_x^a(x_0^*, m_0^*|0)$. Since $U_m^p(x,m|\omega) = -1$ and $U_m^a(x,m|\omega) = 1$, this implies that $(x_0^*, m_0^*) \in CC(0)$. From the state 1 FOC, however, $f(1)U^p(x_1^*, m_1^*|1) = f(0)U^a(x_1^*, m_1^*|0)$, so $(x_1^*, m_1^*) \notin CC(1)$: the contract curve in state 1 is constrained so that it has to be on the m-axis, i.e., $m = 0$.

The solution is illustrated in Figure 2.5. The left panel depicts the case of moderate bias, $1/2 < b < 1$. The points A and B in the figure lie on the state 0 indifference curve for the agent that goes through the principal's ideal point in state 1. Point A includes positive transfers to the agent, and is on the state 0 contract curve, at the tangency of the agent's indifference curve with a state 0 indifference curve for the principal. Point B includes zero transfers, and is in the constrained contract curve. The contract (A, B) maximizes the principal's payoff in state 1 among incentive compatible contracts but implies large transfers to the agent in state 0. If the likelihood of state 0 is high, the principal is better off moving to a point like (A', B'), trading transfers in state 0 for *upward* distortions in policy in state 1. For a sufficiently large likelihood on state 0, the contract must be in the constrained contract curve in both states, moving to a point like (A'', B''), with no transfers and distortions in the policy in both states.

The right panel illustrates the case of $b \in (1, 2)$.[16] Note that distorting policy in state 1 upwards is never optimal for the principal. Here the tradeoff is achieved distorting policy *downwards* in state 1 (against the agent's preferred policy) and

[16] If $b \geq 2$ we get pooling at the prior as the solution to the problem.

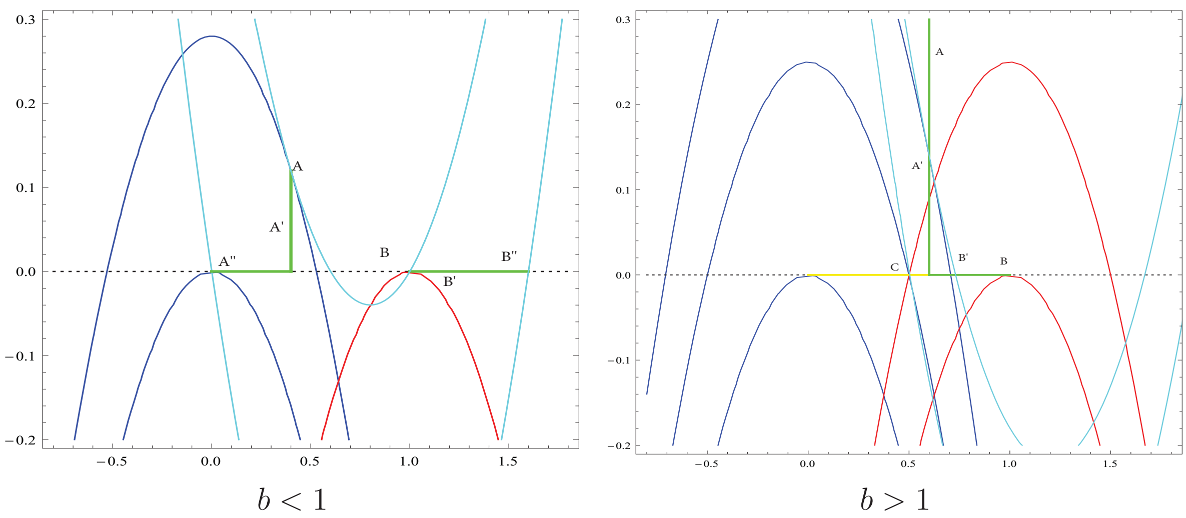


Figure 2.8: Quasilinear Model with Non-negative Transfers. The figure shows the principal's indifference curves in state 0 (in blue) and state 1 (in red), and the agent's indifference curves in state 0 (in light blue).

reducing transfers in state 0, as in a point like (A', B'). Thus, differently than the IR constraint version of the model, the quasilinear model with non-negative transfers does have a recoil effect. Differently than in our model, though, this recoil effect is discontinuous in the bias b.

For a sufficiently high likelihood of state 0, no separating contract exists, and the principal is reduced to a pooling contract along the yellow constrained contract curve. To see why pooling obtains in this context, consider point C on the yellow curve. Note that the triple tangency condition argument against pooling in our model (Theorem 2.4.1) does not apply here, because given the nonnegative constraint on transfers, the principal's indifference curves will not be tangent at the optimal pooling point. Thus pooling is optimal if the agent's state 0 indifference curve runs through the non-trivial gap between the principal's indifference curves.

2.6 Conclusion

A common feature in many agency relationships is that the principal can decide both the direction *and* the scope or implementation scale of a policy. In such cases, there is a natural complementarity between these dimensions of policy, as the value of expanding the scale of implementation increases for both principal and agent the closer the implemented policy is to their preferred policy. In this paper we characterize the optimal contract for the principal in this environment when she cannot count on transfers to alleviate incentive problems.

Because of the non-separability across policy dimensions that is at the core of our problem, the common procedure of re-parametrizing the problem in terms of information rents is not helpful. In general we are dealing with a singular control problem of the type that does not admit the sort of straightforward bang-bang solutions commonly used in other contexts. However, we are able to make considerable progress. First, we solve the optimal contract in the two type case, and present a graphical analysis that makes the logic and results transparent. We then characterize the optimal separating contract in the continuum with a parametric assumption on payoffs (exponential payoffs).

We show that the optimal separating contract is equivalent to delegation "with strings attached": an agent with an upward policy bias can only choose higher policies by reducing the scale of the project. The possibility of tinkering with the scale of implementation induces distortions in the content of policy that *can* be different from what would result in a comparable model with quasilinear payoffs, and that *are* in fact different when the agent's bias is sufficiently large relative to the smallest possible deviation (which is always the case in the continuum). Indeed, in this case the optimal separating contract partitions the state space in a "low" and "high" set of states, such that the principal overfunds and distorts towards the agent in low states, but underfunds and distorts against the agent in high states.

This strong form of ex-post inefficiency in the optimal contract does not appear in the standard model with quasilinear payoffs where utility is perfectly transferable between parties, and leads to new insights in applications. In the environmental regulation case, for example, this implies that the optimal contract sets overly stringent regulations that are heavily enforced when climate change is mild, and relatively weak regulations which are under-enforced in the states in which climate change is accelerating more heavily. Thus, both Congress and the EPA would both prefer to set more stringent, heavily enforced environmental regulations in the high states.

While we have made significant progress in analyzing this problem, much work remains. As much of the literature before us, we have restricted the principal to choose among deterministic mechanisms. A key direction for future work would be to extend the analysis in this paper to stochastic mechanisms (e.g., Strausz (2006), Kováč and Mylovanov (2009)). For example, if the principal is less risk averse than the agent this kind of contract can be useful for the principal in this context, allowing her to relax incentive constraints without utility loss. Alternatively, if the agent's risk attitudes change throughout the policy space these differences can be exploited (e.g., if the agent is risk loving near the principal's state 0 optimal point, but risk averse near the principal's state 1 optimal point the principal may be able to reward and/or punish through random mechanisms depending on the principal's preferences).

To see this, consider the binary state model. For incentive compatibility, the principal needs to make the policy in state $\omega = 0$ more attractive, and the policy in state $\omega = 1$ less attractive to the state 0 agent. Now suppose, for simplicity, that the principal is risk neutral, while the agent is risk averse. Then the principal can gain substituting the state 1 policy in the optimal deterministic contract (x_1^*, m_1^*) with a lottery that plays a point (x_1', m_1') with a probability $\mu' \in (0, 1)$ and a point (x_1'', m_1'') with probability $(1 - \mu)$ in state $\omega = 1$.

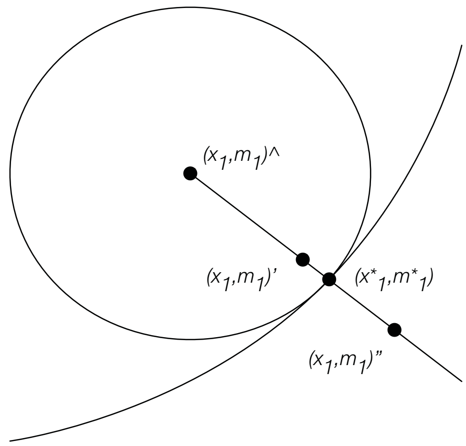


Figure 2.9: A stochastic mechanism.

This is illustrated in Figure 2.9. Choose a point (x_1', m_1') that is a convex combination between (x_1^*, m_1^*) and the principal's state 1 ideal point, $(\hat{x}_1, \hat{m}_1)$, choose a point (x_1'', m_1'') and a probability $\mu \in (0, 1)$ such that $(x_1^*, m_1^*) = \mu(x_1', m_1') + (1-\mu)(x_1'', m_1'')$. Note that the risk neutral principal is indifferent between (x_1^*, m_1^*) and the lottery $\mu[(x_1', m_1')] + (1-\mu)[(x_1'', m_1'')]$, but the state 0 agent is strictly worse-off. Then letting $\mu' = \mu + \varepsilon$ for $\varepsilon > 0$ small, the agent is still strictly worse-off but the principal is better-off in $\mu'[(x_1', m_1')] + (1-\mu')[(x_1'', m_1'')]$ than in (x_1^*, m_1^*).

This simple example illustrates that if the principal is less risk averse than the agent, there is space to improve outcomes by considering stochastic mechanisms. This also suggests that deterministic mechanisms can be optimal when the principal is at least as risk averse as the agent. This, we believe, is at the heart of the comment in Koessler and Martimort (2012) regarding the optimality of deterministic mechanisms in that context. A full analysis of contracting with stochastic mechanisms in our setup is beyond the scope of this paper, and is left for future work.

2.A Omitted Proofs

2.A.1 Principal gets first-best when $b < 1/2$

Lemma 2.A.1 *$(x^*_\omega, m^*_\omega) = (\widehat{x}_\omega, \widehat{m}_\omega)$ for $\omega \in \{0,1\}$ if and only if $b \le 1/2$.*

Proof of Lemma 2.A.1. The Lagrangian for the principal is:

$$\sum_\omega U^p(x_\omega, m_\omega|\omega) f(\omega) + \lambda_0 \left[U^a(x_0, m_0|0) - U^a(x_1, m_1|0)\right] + \lambda_1 \left[U^a(x_1, m_1|1) - U^a(x_0, m_0|1)\right]$$

The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial x_0} = U^p_x(x_0, m_0|0) f(0) + \lambda_0 U^a_x(x_0, m_0|0) - \lambda_1 U^a_x(x_0, m_0|1) = 0, \tag{2.A.1}$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = U^p_x(x_1, m_1|1) f(1) - \lambda_0 U^a_x(x_1, m_1|0) + \lambda_1 U^a_x(x_1, m_1|1) = 0, \tag{2.A.2}$$

$$\frac{\partial \mathcal{L}}{\partial m_0} = U^p_m(x_0, m_0|0) f(0) + \lambda_0 U^a_m(x_0, m_0|0) - \lambda_1 U^a_m(x_0, m_0|1) = 0, \tag{2.A.3}$$

$$\frac{\partial \mathcal{L}}{\partial m_1} = U^p_m(x_1, m_1|1) f(1) - \lambda_0 U^a_m(x_1, m_1|0) + \lambda_1 U^a_m(x_1, m_1|1) = 0, \tag{2.A.4}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_0} = U^a(x_0, m_0|0) - U^a(x_1, m_1|0) \ge 0, \quad \lambda_0 \ge 0, \quad \lambda_0 \frac{\partial \mathcal{L}}{\partial \lambda_0} = 0, \tag{2.A.5}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = U^a(x_1, m_1|1) - U^a(x_0, m_0|1) \ge 0, \quad \lambda_1 \ge 0, \quad \lambda_1 \frac{\partial \mathcal{L}}{\partial \lambda_1} = 0, \tag{2.A.6}$$

Suppose that neither constraint is binding. Then we have $\lambda^*_0 = \lambda^*_1 = 0$. Thus (2.A.1) becomes $U^p_x(x_0, m_0|0) = 0 \Leftrightarrow x^*_0 = \widehat{x}_0 = 0$, and (2.A.2) becomes $U^p_x(x_1, m_1|1) = 0 \Leftrightarrow x^*_1 = \widehat{x}_1 = 1$. Then (2.A.3) becomes $U^p_m(x_0, m_0|0) = 0$, and since $x_0 = 0$, then $U^p_m(0, m_0|0) = 0$. Therefore $u^p_m(0, m_0) = \gamma_m(m_0)$, or $m_0 = \hat{m}$. Similarly, from (2.A.4), $m_1 = \hat{m}$. Then (2.A.5) and (2.A.6) are reduced to:

$$u^a\left(b^2, \widehat{m}\right) \ge u^a\left((1-b)^2, \widehat{m}\right) \quad \Leftrightarrow b \le \frac{1}{2},$$

and

$$u^a\left(b^2,\widehat{m}\right)\geq u^a\left((1+b)^2,\widehat{m}\right)\quad\Leftrightarrow b\geq -\frac{1}{2}.$$

This concludes the proof of the lemma. ■

2.A.2 Proof of Theorem 2.4.1

Proof of Theorem 2.4.1. We will show that the utility functions which satisfy a necessary condition for pooling are nowhere dense in the appropriate Sobolev space, $W^{1,p}(X)$. In particular, we show that if a pooling contract x_o^*, m_o^* is optimal for the Principal, then the following triple tangency condition must be satisfied:

$$\gamma_{p,0}'\left(t_{p,0}^*\right)=\pm\gamma_{p,1}'\left(t_{p,1}^*\right)=\pm\gamma_{a,0}'\left(t_{a,0}^*\right),$$

where $\gamma_{j,\omega}: I\to\mathbb{R}^2$ is the parametrization by arclength of $IC^j\left(x_o^*,m_o^*|\omega\right)$, $\gamma_{j,\omega}\left(t_{j,\omega}^*\right)=(x_o^*,m_o^*)$ and $I\subset\mathbb{R}$ is a non-empty interval.

First, the Principal's indifference curves have to be tangent, for otherwise (x_o^*,m_o^*) does not solve the optimal pooling problem,

$$\max_{(x_o,m_o)}\sum_{\omega\in\{0,1\}}f(\omega)U^p(x_o,m_o|\omega).$$

To see this, note that the first-order condition of the above problem implies:

$$\nabla U^p(x_o^*,m_o^*|0)=\frac{-f_1}{f_0}\nabla U^p(x_o^*,m_o^*|1).$$

It follows that for any (x,m):

$$\begin{aligned}\nabla U^p(x_o^*,m_o^*|0)\cdot[(x,m)-(x_o^*,m_o^*)] &= 0\\ \Leftrightarrow \nabla U^p(x_o^*,m_o^*|1)\cdot[(x,m)-(x_o^*,m_o^*)] &= 0.\end{aligned}$$

Now, by definition of $\gamma_{j,\omega}$:

$$\begin{aligned} \nabla U^p(x_o^*, m_o^*|0) \cdot \left[\gamma'_{p,0}\left(t^*_{p,0}\right) - (x_o^*, m_o^*)\right] &= 0, \text{ and} \\ \nabla U^p(x_o^*, m_o^*|1) \cdot \left[\gamma'_{p,1}\left(t^*_{p,1}\right) - (x_o^*, m_o^*)\right] &= 0, \end{aligned}$$

and since $\left\|\gamma'_{p,1}\left(t^*_{p,1}\right)\right\| = \left\|\gamma'_{p,0}\left(t^*_{p,0}\right)\right\| = 1$ (because $\gamma_{j,\omega}$ is the natural parametrization), it follows that $\gamma'_{p,0}\left(t^*_{p,0}\right) = \pm\gamma'_{p,1}\left(t^*_{p,1}\right)$.[17] This proves the first equality.

We will show the second equality by contradiction. Assume that $\gamma'_{p,1}\left(t^*_{p,1}\right) \neq \pm\gamma'_{a,0}\left(t^*_{a,0}\right)$. Note that for any ε, the menu $\left(\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right), -\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right)\right)$ is incentive compatible[18], since both contracts are on $IC^a\left(x_o^*, m_o^*|0\right)$ [alternatively, could move along IC curve by proposing menu $\left(\gamma_{a,0}\left(t^*_{a,0} - \varepsilon\right), \gamma_{a,0}\left(t^*_{a,0} + \varepsilon\right)\right)$]. Because $\gamma'_{p,1}\left(t^*_{p,1}\right) \neq \pm\gamma'_{a,0}\left(t^*_{a,0}\right)$, either

$$\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right) \in intB^p\left(x_o^*, m_o^*|0\right), -\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right) \in intB^p\left(x_o^*, m_o^*|1\right)$$

or

$$\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right) \in intB^p\left(x_o^*, m_o^*|1\right), -\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right) \in intB^p\left(x_o^*, m_o^*|0\right).$$

Without loss of generality, assume the first holds. This implies that $U^p(\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right)|0) > U^p(x_o^*, m_o^*|0)$ and $U^p(-\varepsilon\gamma'_{a,0}\left(t^*_{a,0}\right)|1) > U^p(x_o^*, m_o^*|1)$, thus the separating menu dominates the pooling menu state by state, which contradicts the optimality of pooling.

■

[17] Note that the sign depends on the direction of the parametrization; reversing the direction would change the sign.

[18] Note that we are using the convention in differential geometry that $\gamma'_{j,\omega}(t)$ is a vector with base point at $\gamma_{j,\omega}(t)$ and that $\varepsilon\gamma'_{j,\omega}(t)$ is a tangent vector length ε, instead of the unit tangent vector (this is a slight abuse of notation).

2.A.3 Proof of Lemma 2.4.2

Proof of Lemma 2.4.2. Suppose in the solution of the Principal's optimization problem $\lambda_0^* > 0$ and $\lambda_1^* = 0$. Then equations (2.A.1) and (2.A.3) boil down to:

$$\lambda_0 = -f(0)\frac{U_x^p(x_0, m_0|0)}{U_x^a(x_0, m_0|0)} = -f(0)\frac{U_m^p(x_0, m_0|0)}{U_m^a(x_0, m_0|0)} \Rightarrow \frac{U_x^p(x_0, m_0|0)}{U_m^p(x_0, m_0|0)} = \frac{U_x^a(x_0, m_0|0)}{U_m^a(x_0, m_0|0)} \tag{2.A.7}$$

so that $(x_0^*, m_0^*) \in CC\,(0)$, and equations (2.A.2) and (2.A.4) boil down to:

$$\lambda_0 = f(1)\frac{U_x^p(x_1, m_1|1)}{U_x^a(x_1, m_1|0)} = f(1)\frac{U_m^p(x_1, m_1|1)}{U_m^a(x_1, m_1|0)} \Rightarrow \frac{U_x^p(x_1, m_1|1)}{U_m^p(x_1, m_1|1)} = \frac{U_x^a(x_1, m_1|0)}{U_m^a(x_1, m_1|0)} \tag{2.A.8}$$

and thus $(x_1^*, m_1^*) \in CC\,(1)$. ■

2.A.4 Proof of Theorem 2.4.3

Proof of Theorem 2.4.3. Consider equation (2.A.7). Note that since $\lambda_0 > 0$ and $U_m^a(x_0, m_0|0) > 0$, then (2.A.7) implies that $U_m^p(x_0, m_0|0) < 0$. Thus, there is overfunding in state 0; i.e., $m_0 > \hat{m}$. Also, since $U_x^p(x_0, m_0|0) \geq 0$ iff $x_0 \leq 0$ and $U_x^a(x_0, m_0|0) \geq 0$ iff $x_0 \leq b$, (2.A.7) implies that $x_0 \in (0, b)$, so the optimal policy in state 0 distorts in favor of the agent. Consider next expression (2.A.8). Note that $\lambda_0 > 0$ and $U_m^a(x_1, m_1|0) > 0$ in (2.A.8) imply that $U_m^p(x_1, m_1|1) > 0$. Thus there is underfunding in state 1; i.e., $m_1 < \hat{m}$. And from the first equality, we have that $U_x^p(x_1, m_1|1)$ and $U_x^a(x_1, m_1|0)$ have to have the same sign, so either $x_1 < \min\{1, b\}$ or $x_1 > \max\{1, b\}$. So suppose first that $b < 1$. Then either $x_1 < b$ or $x_1 > 1$. However, it cannot be that $x_1 < b$. To see this, note that in this case the symmetric point about the ideal point of the agent $(2b - x_1, m_1)$ would give the agent the same payoff but would increase the utility of the principal. Thus such $(x_1, m_1) \notin CC(1)$. It follows that if $b < 1$, then $x_1 > 1$. Suppose next that $b > 1$. Then either $x_1 < 1$ or $x_1 > b$, but by a similar argument as before, it must be that $x_1 < 1$.

Finally, we show that if $f(0) \in (0,1)$, the optimal incentive compatible solution entails distortions in both states: $(x^*_\omega, m^*_\omega) \neq (\widehat{x}_\omega, \widehat{m}_\omega)$ for $\omega = 0, 1$. Equivalently, we need to show that if $f(0) \in (0,1)$, the solution to Problem 2.4.2 satisfies $u^* \in (U^a(\widehat{x}_0, \widehat{m}_0|0), U^a(\widehat{x}_1, \widehat{m}_1|0))$. We will show that if $f(0) \neq 0$ then $u < U^a(\widehat{x}_1, \widehat{m}_1|0)$. A similar argument proves the opposite direction. We have that:

$$\frac{\partial}{\partial u} U^p(\widetilde{x}^\omega(u), \widetilde{m}^\omega(u)|\omega) = u^p_\ell(\ell^P(\widetilde{x}^\omega(u), \omega), \widetilde{m}^\omega(u)) 2\left(\widetilde{x}^\omega(u) - \omega\right) \widetilde{x}^\omega_u(u)$$
$$+ \left[u^p_m(\ell^P(\widetilde{x}^\omega(u), \omega), \widetilde{m}^\omega(u)) - \gamma_m(\widetilde{m}^\omega(u))\right] \widetilde{m}^\omega_u(u),$$

and

$$\left.\frac{\partial}{\partial u} U^p(\widetilde{x}^1(u), \widetilde{m}^1(u)|1)\right|_{U^a(\widehat{x}_1, \widehat{m}_1|0)} = \left[u^p_m(0, \widehat{m}_1) - \gamma_m(\widehat{m}_1)\right] \widetilde{m}^0_u(U^a(\widehat{x}_1, \widehat{m}_1|0)) = 0,$$

where the last part follows by the definition of $\widehat{m}_1$, i.e., the FOC that this first-best has to satisfy is $u^p_m(0, \widehat{m}_1) - \gamma_m(\widehat{m}_1) = 0$. Next, note that:

$$\left.\frac{\partial}{\partial u} U^p(\widetilde{x}^0(u), \widetilde{m}^0(u)|0)\right|_{U^a(\widehat{x}_1, \widehat{m}_1|0)} = 2u^p_\ell(1, \widehat{m}_1)\widetilde{x}^0_u(U^a(\widehat{x}_1, \widehat{m}_1|0))$$
$$+ \left[u^p_m(1, \widehat{m}_1) - \gamma_m(\widehat{m}_1)\right] \widetilde{m}^\omega_u(U^a(\widehat{x}_1, \widehat{m}_1|0)).$$

We have that $2u^p_\ell(1, \widehat{m}_1) < 0$, and since we are overfunding always in state 0, $u^p_m(1, \widehat{m}_1) - \gamma_m(\widehat{m}_1) < 0$. Furthermore, since the indifference curve moves in the north-east direction, we have that $\widetilde{x}^0_u(U^a(\widehat{x}_1, \widehat{m}_1|0)) > 0$ and $\widetilde{m}^0_u(U^a(\widehat{x}_1, \widehat{m}_1|0)) > 0$. All of this implies that:

$$\left.\frac{\partial}{\partial u} U^p(\widetilde{x}^0(u), \widetilde{m}^0(u)|0)\right|_{U^a(\widehat{x}_1, \widehat{m}_1|0)} < 0,$$

which means utility can be improved by decreasing u if $f(0) \neq 0$; thus $u^* < U^a(\widehat{x}_1, \widehat{m}_1|0)$. ■

2.A.5 Incentive Compatibility in Continuum Model

Lemma 2.A.2 *If a policy function $q(\cdot) = (x(\cdot), m(\cdot))$ is implementable, $x(\cdot)$ is non-decreasing.*

Proof of Lemma 2.A.2. The proof follows the standard line, and is included here for completeness. The first-order condition for truth-telling is:

$$\left.\frac{\partial \mathcal{U}^a(\hat{\omega}, \omega)}{\partial \hat{\omega}}\right|_{\hat{\omega}=\omega} = U_x^a(x(\hat{\omega}), m(\hat{\omega})|\omega)x'(\hat{\omega}) + U_m^a(x(\hat{\omega}), m(\hat{\omega})|\omega)m'(\hat{\omega})\Big|_{\hat{\omega}=\omega} =: 0, \tag{2.A.9}$$

or equivalently,

$$m'(\omega) = -\frac{U_x^a(x(\omega), m(\omega)|\omega)}{U_m^a(x(\omega), m(\omega)|\omega)}x'(\omega). \tag{2.A.10}$$

The second-order condition for no (local) profitable deviations is:

$$\left.\frac{\partial^2 \mathcal{U}(\hat{\omega}, \omega)}{\partial \hat{\omega}^2}\right|_{\hat{\omega}=\omega} \leq 0 \tag{2.A.11}$$

Differentiating (2.A.9) gives

$$\left[\frac{\partial^2 \mathcal{U}(\omega, \omega)}{\partial \hat{\omega}^2} + \frac{\partial^2 \mathcal{U}(\omega, \omega)}{\partial \hat{\omega} \partial \omega}\right] d\omega = 0 \Rightarrow \frac{\partial^2 \mathcal{U}(\omega, \omega)}{\partial \hat{\omega} \partial \omega} = -\frac{\partial^2 \mathcal{U}(\omega, \omega)}{\partial \hat{\omega}^2}$$

so that (2.A.11) is

$$\frac{\partial^2 \mathcal{U}(\omega, \omega)}{\partial \hat{\omega} \partial \omega} \geq 0$$

From (2.A.9), this is

$$U_{x\omega}^a(x(\omega), m(\omega)|\omega)x_{m\omega}'^a(x(\omega), m(\omega)|\omega)m'(\omega) \geq 0$$

Substituting $m'(\omega)$ from (2.A.10), this is

$$x'(\omega)\left[U^a_{x\omega}(x(\omega),m(\omega)|\omega)-U^a_{m\omega}(x(\omega),m(\omega)|\omega)\frac{U^a_x(x(\omega),m(\omega)|\omega)}{U^a_m(x(\omega),m(\omega)|\omega)}\right]\geq 0$$

Since the bracket is nonnegative from the SCC $\frac{\partial}{\partial\omega}\left(\frac{U^a_x(x,m|\omega)}{U^a_m(x,m|\omega)}\right)\geq 0$, then $x'(\omega)\geq 0$.

■

Lemma 2.A.3 *If $x(\cdot)$ is nondecreasing and (2.4.3) holds for all $\omega\in\Omega$,*

$$\mathcal{U}^a(\omega,\omega)\geq\mathcal{U}^a(\hat{\omega},\omega)\quad \textit{for all}\quad \omega,\widehat{\omega}\in[0,1]\,.$$

Proof of Lemma 2.A.3. We want to show that:

$$0\geq\frac{\partial\mathcal{U}^a(\hat{\omega},\omega)}{\partial\hat{\omega}}=U^a_x(x(\hat{\omega}),m(\hat{\omega})|\omega)x'(\hat{\omega})+U^a_m(x(\hat{\omega}),m(\hat{\omega})|\omega)m'(\hat{\omega})\quad\forall\omega,\omega'\in[0,1]$$

Dividing and multiplying by $U^a_m(x(\hat{\omega}),m(\hat{\omega})|\omega)$, we have

$$\frac{\partial\mathcal{U}^a(\hat{\omega},\omega)}{\partial\hat{\omega}}=U^a_m(x(\hat{\omega}),m(\hat{\omega})|\omega)\left[\frac{U^a_x(x(\hat{\omega}),m(\hat{\omega})|\omega)}{U^a_m(x(\hat{\omega}),m(\hat{\omega})|\omega)}x'(\hat{\omega})+m'(\hat{\omega})\right]$$

By the SCC, if $\widehat{\omega}>\omega$,

$$\frac{\partial\mathcal{U}^a(\hat{\omega},\omega)}{\partial\hat{\omega}}\leq U^a_m(x(\hat{\omega}),m(\hat{\omega})|\omega)\underbrace{\left[\frac{U^a_x(x(\hat{\omega}),m(\hat{\omega})|\hat{\omega})}{U^a_m(x(\hat{\omega}),m(\hat{\omega})|\hat{\omega})}x'(\hat{\omega})+m'(\hat{\omega})\right]}_{=0\text{ by (2.4.3)}}=0$$

A similar argument holds for $\hat{\omega}<\omega$. ■

Remark 3 *The Hamiltonian for problem (PP) is*

$$\mathcal{H}=U^p(x,m|\omega)f(\omega)-\lambda_1\frac{U^a_x(x,m|\omega)}{U^a_m(x,m|\omega)}y+\lambda_2 y$$

The necessary and sufficient conditions for a fully separating solution are that there exist $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ such that:

$$m' = \mathcal{H}_{\lambda_1} = -\frac{U_x^a(x,m|\omega)}{U_m^a(x,m|\omega)}y \tag{2.A.12}$$

$$x' = \mathcal{H}_{\lambda_2} = y \tag{2.A.13}$$

$$0 = \mathcal{H}_y = -\lambda_1 \frac{U_x^a(x,m|\omega)}{U_m^a(x,m|\omega)} + \lambda_2 \tag{2.A.14}$$

$$\lambda_1' = -\mathcal{H}_m = -U_m^p(x,m|\omega)f(\omega) + \lambda_1 y \frac{\partial}{\partial m}\left(\frac{U_x^a(x,m|\omega)}{U_m^a(x,m|\omega)}\right) \tag{2.A.15}$$

$$\lambda_2' = -\mathcal{H}_x = -U_x^p(x,m|\omega)f(\omega) + \lambda_1 y \frac{\partial}{\partial x}\left(\frac{U_x^a(x,m|\omega)}{U_m^a(x,m|\omega)}\right), \tag{2.A.16}$$

$$0 = \mu m, \tag{2.A.17}$$

with initial conditions $m(0) = m_0$ and $x(0) = x_0$ and transversality conditions $\lambda_1(1) = 0$ and $\lambda_2(1) = 0$, and $\lambda_1(0) = 0$ and $\lambda_2(0) = 0$. From the Pontryagin Maximum Principle (for example, see Zeidler (1985) Theorem 48.C), any optimum for the Principal satisfies the Euler-Lagrange equations above. Moreover, the optimal control problem in equations (2.A.12-2.A.17) satisfies the weak Mangasarian sufficient condition for a maximum (the problem is in general weakly concave), and thus a solution to (2.A.12-2.A.17) is a global maximizer.

2.A.6 Proof of Lemma 2.4.4

Proof of Lemma 2.4.4. The proof follows a similar argument in Krishna and Morgan (2008). Suppose, to the contrary, that there exists an ω such that $x(\omega) > b + \omega$. Consider (2.A.14). Since $x(\omega) > b + \omega$, we have $U_x^a(x,m|\omega) < 0$. Suppose first that $\lambda_1 > 0$. Since $U_m^a(x,m|\omega) > 0$ and $\lambda_2 \geq 0$, we have

$$-\lambda_1 \frac{U_x^a(x,m|\omega)}{U_m^a(x,m|\omega)} + \lambda_2 > 0$$

which is a contradiction (this expression has to equal zero by (2.A.14)). Suppose then that $\lambda_1 = 0$. Then from (2.A.16)

$$\lambda_2' = -U_x^p(x, m|\omega)f(\omega) > 0 \Rightarrow \lambda_2(\omega) > 0,$$

which again contradicts (2.A.14). ■

Lemma 2.A.4 *The payoffs of principal and agent in the solution* $\{x(\cdot), m(\cdot)\}$ *to the principal's problem,* $U^p(x(\omega), m(\omega)|\omega)$ *and* $U^a(x(\omega), m(\omega)|\omega)$*, are continuous in* ω*.*

Proof of Lemma 2.A.4. To see that $U^a(x(\omega), m(\omega)|\omega)$ is continuous, assume by way of contradiction that there exists some $\varepsilon > 0$, such that for all $\delta > 0$ sufficiently small, $|U^a(x(\omega), m(\omega)|\omega) - U^a(x(\omega-\delta), m(\omega-\delta)|\omega-\delta)| > \varepsilon$. Then since $U^a(x, m|\omega)$ is continuous in ω, we have that for $\delta > 0$ sufficiently small:

$$\begin{aligned} 0 &\leq U^a(x(\omega), m(\omega)|\omega) - U^a(x(\omega), m(\omega)|\omega-\delta) < \varepsilon, \\ 0 &\leq U^a(x(\omega-\delta), m(\omega-\delta)|\omega-\delta) - U^a(x(\omega-\delta), m(\omega-\delta)|\omega) < \varepsilon, \end{aligned}$$

where the absolute values are not needed by truth-telling. But then, if $U^a(x(\omega), m(\omega)|\omega) - U^a(x(\omega-\delta), m(\omega-\delta)|\omega-\delta) > \varepsilon$:

$$\begin{aligned} U^a(x(\omega-\delta), m(\omega-\delta)|\omega-\delta) &< U^a(x(\omega), m(\omega)|\omega) + \varepsilon \\ &< U^a(x(\omega), m(\omega)|\omega-\delta), \end{aligned}$$

which is a contradiction since truth-telling fails for $\omega - \delta$. Similarly, if

$$U^a(x(\omega-\delta), m(\omega-\delta)|\omega-\delta) - U^a(x(\omega), m(\omega)|\omega) > \varepsilon$$

then truth-telling will fail for type ω.

To see that $U^p(x(\omega), m(\omega)|\omega)$ is continuous in ω, assume by way of contradiction that there exists some $\varepsilon > 0$, such that for all $\delta > 0$,

$$U^p(x(\omega), m(\omega)|\omega) - U^p(x(\omega-\delta), m(\omega-\delta)|\omega-\delta) > \varepsilon$$

(the other case follows similarly). Clearly this implies that either x or m or both are discontinuous at ω. Since $U^a(x(\omega'), m(\omega')|\omega)$ is continuous in ω' around $\omega' = \omega$, for all $\eta > 0$, there exists a $\delta > 0$ such that for any $\omega' \in (\omega-\delta, \omega)$ there exists some $\widetilde{x}(\omega'), \widetilde{m}(\omega')$ such that:

$$\begin{aligned} U^a(x(\omega'), m(\omega')|\omega') &= U^a(\widetilde{x}(\omega'), \widetilde{m}(\omega')|\omega'), \text{ and} \\ |\widetilde{x}(\omega') - x(\omega)| &< \eta, \\ |\widetilde{m}(\omega') - m(\omega)| &< \eta, \end{aligned}$$

and which preserves local IC. Since local IC implies global IC by the single-crossing property changing x, m to $\widetilde{x}, \widetilde{m}$ respects incentive constraints. Furthermore, by the continuity of $U^p(x, m|\omega)$ in x, m we have that for some $\delta' > 0$, and $\omega' \in (\omega, \omega-\delta')$ we have:

$$\begin{aligned} |U^p(x(\omega), m(\omega)|\omega) - U^p(\widetilde{x}(\omega'), \widetilde{m}(\omega')|\omega)| &< \frac{\varepsilon}{2}, \text{ and} \\ |U^p(\widetilde{x}(\omega'), \widetilde{m}(\omega')|\omega') - U^p(\widetilde{x}(\omega'), \widetilde{m}(\omega')|\omega)| &< \frac{\varepsilon}{2}, \end{aligned}$$

so that by the triangle inequality:

$$|U^p(x(\omega), m(\omega)|\omega) - U^p(\widetilde{x}(\omega'), \widetilde{m}(\omega')|\omega')| < \varepsilon.$$

But then for all $\omega' \in (\omega, \omega - \delta')$:

$$\begin{aligned} U^p\left(\widetilde{x}\left(\omega'\right), \widetilde{m}\left(\omega'\right)|\omega'\right) &> U^p\left(x\left(\omega\right), m\left(\omega\right)|\omega\right) - \varepsilon \\ &> U^p\left(x\left(\omega'\right), m\left(\omega'\right)|\omega'\right), \end{aligned}$$

thus since f has full support the proposed x, m cannot be optimal. ■

2.A.7 Proof of Lemma 2.4.7

Proof of Lemma 2.4.7. Let x,m be a solution to the principal's problem. By monotonicity, we know that x and m have to be differentiable and hence continuous almost everywhere. At points of differentiability, we have that:

$$MRS^p_{xm}(\omega) = \frac{U^p_x(x\left(\omega\right), m\left(\omega\right)|\omega)}{U^p_m(x\left(\omega\right), m\left(\omega\right)|\omega)} = \frac{U^a_x(x\left(\omega\right), m\left(\omega\right)|\omega)}{U^a_m(x\left(\omega\right), m\left(\omega\right)|\omega)} = MRS^a_{xm}(\omega),$$

which is a tangency condition between the indifference curves of the principal and agent. Note that for our specified utility functions generically there is at most one point of tangency of indifference curves. This is also generically true for other utility functions. To be precise, note that for the indifference curve $\upsilon_a = U^a(x\left(\omega\right), m\left(\omega\right)|\omega)$ we have:

$$m = \upsilon_a \exp\left(-\beta\left(x - \omega - b\right)\right),$$

where we are restricting attention to the relevant range, i.e., $x < \omega + b$. Thus on this indifference curve:

$$\frac{dm}{dx} = -\beta\upsilon_a \exp\left(-\beta\left(x - \omega - b\right)\right).$$

Similarly, for the indifference curve of the principal $\upsilon_p = U^p(x\left(\omega\right), m\left(\omega\right)|\omega)$, we have:

$$\frac{dm}{dx} = \frac{\eta\left(x - \omega\right)\exp(-\frac{\eta}{2}(x - \omega)^2)\left(\pm 1 - \sqrt{1 - 2\gamma\upsilon_p \exp(\eta(x - \omega)^2)}\right)}{\gamma\sqrt{1 - 2\gamma\upsilon_p \exp(\eta(x - \omega)^2)}}.$$

It is clear that equality of these expressions cannot hold for multiple x when υ_a and υ_p are fixed, unless very special choices are made for utility function parameters, e.g., $\beta = 0$, $\eta = 0$. Thus generically, there is at most one point of tangency. We further note that the tangency condition is differentiable and hence continuous.

Since the tangency condition is continuous, and x, m is differentiable in a neighborhood above and below $\widehat{\omega}$:

$$\begin{aligned}
\frac{U_x^p(x(\widehat{\omega}_-), m(\widehat{\omega}_-)|\widehat{\omega}_-)}{U_m^p(x(\widehat{\omega}_-), m(\widehat{\omega}_-)|\widehat{\omega}_-)} &= \frac{U_x^a(x(\widehat{\omega}_-), m(\widehat{\omega}_-)|\widehat{\omega}_-)}{U_m^a(x(\widehat{\omega}_-), m(\widehat{\omega}_-)|\widehat{\omega}_-)}, \\
\frac{U_x^p(x(\widehat{\omega}_+), m(\widehat{\omega}_+)|\widehat{\omega}_+)}{U_m^p(x(\widehat{\omega}_+), m(\widehat{\omega}_+)|\widehat{\omega}_+)} &= \frac{U_x^a(x(\widehat{\omega}_+), m(\widehat{\omega}_+)|\widehat{\omega}_+)}{U_m^a(x(\widehat{\omega}_+), m(\widehat{\omega}_+)|\widehat{\omega}_+)}.
\end{aligned}$$

But $U^a(x(\omega), m(\omega)|\omega)$ and $U^p(x(\omega), m(\omega)|\omega)$ are continuous in ω by lemma 2.A.4, we have:

$$\begin{aligned}
U^a(x(\widehat{\omega}_+), m(\widehat{\omega}_+)|\widehat{\omega}_+) &= U^a(x(\widehat{\omega}_-), m(\widehat{\omega}_-)|\widehat{\omega}_-), \text{ and} \\
U^p(x(\widehat{\omega}_+), m(\widehat{\omega}_+)|\widehat{\omega}_+) &= U^p(x(\widehat{\omega}_-), m(\widehat{\omega}_-)|\widehat{\omega}_-),
\end{aligned}$$

thus the tangency would have to occur on the same indifference curves in the limit. But since there is a unique tangency point for type $\widehat{\omega}$, we have that $x(\widehat{\omega}_-) = x(\widehat{\omega}_+)$ and $m(\widehat{\omega}_-) = m(\widehat{\omega}_+)$. ■

2.A.8 Proof of Theorem 2.4.8

Proof of Theorem 2.4.8. Part 1. First we show that in the solution we cannot have either $x(\omega) > \omega$ for all $\omega \in [0,1]$ or $x(\omega) < \omega$ for all $\omega \in [0,1]$. In the exponential case, the Euler-Lagrange equation (2.A.16) becomes

$$\lambda_2' = \eta(x-\omega)m\exp\left(-\frac{\eta}{2}(x-\omega)^2\right)f(\omega) \tag{2.A.18}$$

Substituting (2.4.4) in (2.A.18) we get

$$\lambda_2' = \eta(x-\omega)m_0 \exp\left(-\frac{\eta}{2}(x-\omega)^2 - \beta[x(\omega)-x_0]\right) f(\omega),$$

so that

$$\lambda_2(\omega) = \eta m_0 \int_0^{\omega} (x-\upsilon) \exp\left(-\frac{\eta}{2}(x-\upsilon)^2 - \beta[x-x_0]\right) f(\upsilon) dz \qquad (2.A.19)$$

Note then that the transversality condition $\lambda_2(1) = 0$ gives

$$\int_0^1 (x-\upsilon) \exp\left(-\frac{\eta}{2}(x-\upsilon)^2 - \beta[x-x_0]\right) f(\upsilon) d\upsilon = 0 \qquad (2.A.20)$$

and the result follows since $\exp(\cdot) > 0$.

<u>Part 2</u>. Next, we characterize properties of the optimal separating contract and derive (2.4.6). Let $r(\omega) \equiv (\eta/2)[(x(\omega)-\omega)^2 - x_0^2] - \beta[x(\omega)-x_0]$ and let $\tilde{r}(\omega) = r(\omega) + (\eta/2)x_0^2$. Note that we can write (2.4.5) as

$$\exp(\tilde{r}(\omega)) = \frac{\beta + \eta(x(\omega)-\omega)}{\beta\gamma m_0} \quad \forall \omega \in [0,1]. \qquad (2.A.21)$$

Imposing the constraint that $x(0) = x_0$ in (2.A.21), we obtain

$$m_0 = [1 + (\eta/\beta)x_0] \frac{1}{\gamma} \exp\left(-(\eta/2)x_0^2\right) \qquad (2.A.22)$$

Substituting back in (2.A.21), we have

$$\exp(r(\omega)) = \frac{[\beta + \eta(x(\omega)-\omega)]}{(\beta + \eta x_0)} \quad \forall \omega \in [0,1]. \qquad (2.A.23)$$

Equation (2.A.23) completely characterizes the optimal policy $x(\cdot)$ as a function of the initial value x_0. Differentiating (2.A.23), we obtain (2.4.6). To obtain the optimal x_0, note that we can now rewrite the principal's problem as

$$\max_{x_0} J(x_0) = \int_0^1 U^p \left(x(\omega), \frac{(\beta + \eta x_0)}{\beta\gamma} \exp\left(-\beta[x(\omega) - x_0] - \left(\frac{\eta}{2}\right) x_0^2 \right) |\omega \right) f(\omega) d\omega$$

subject to (2.A.23), so at the optimum

$$\frac{\partial J}{\partial x_0}(x_0) = 0. \tag{2.A.24}$$

Equations (2.4.4), (2.A.22), (2.A.23), and (2.A.24) completely characterize the optimal fully separating incentive compatible contract for the principal, provided the solution exists.

Part 3. Recall that by (2.4.6),

$$x'(\omega) = \frac{(x(\omega) - \omega) \exp(r(\omega)) - \frac{1}{(\beta + \eta x_0)}}{\left[(x(\omega) - \omega) - \frac{\beta}{\eta}\right] \exp(r(\omega)) - \frac{1}{(\beta + \eta x_0)}}$$

Note that if $x(\omega) < \omega$, then the numerator and denominator of (2.4.6) are negative. Then $x'(\omega) < 1$ if and only if $\frac{\beta}{\eta} \exp(r(\omega)) > 0$, which is always the case. It follows that if $x(\omega') < \omega'$ for some $\omega' \in [0, 1]$, then $x(\omega) < \omega$ for all $\omega \in [\omega', 1]$. Then it must be that $x_0 > 0$, for otherwise $x(\omega) < \omega$ for all $\omega \in [0, 1]$, contradicting (2.A.20). So $x(\omega)$ starts above ω and then must cross ω at least once. But note that if it crosses once, it will not go back up. This concludes the proof. ■

Chapter 3

Depreciation in Monotone Games: A Folk Theorem

This chapter is co-authored with Prof. Marco Battaglini.[1]

3.1 Introduction

Starting with Gale (2001), an important literature has been dedicated to the study of monotone contribution games. Monotone contribution games (henceforth MCGs) are dynamic games in which players' payoffs are nondecreasing in their own and other players' actions, and in which actions are assumed to be irreversible: in no period can an agent choose an action that is lower then the action chosen in the previous period. These games have been applied to study variety of environments: public good contribution games in which the public good is the sum of private contributions and is durable;[2] models of investment and adoption of new technologies;[3] models of

[1]Department of Economics, Cornell University, Ithaca, NY 14850, email: battaglini@cornell.edu

[2]Gale (2001), Lockwood and Thomas (2002), Yildirim (2006), Battaglini et al. (2015), among others.

[3]Gale (1995).

holdup;[4] market games;[5] models of bargaining in which players can make monotonic concessions.[6]

The key result in this literature is an antifolk theorem. Studying MCG with no discounting and in which players maximize long term payoffs, Gale (2001) has been the first to observe that the set of subgame perfect equilibria (hence SPE) is a strict subset of the set of feasible strictly individually rational payoffs. Extending the result to games with discounting, Lockwood and Thomas (2002) have shown that an efficient allocation is not feasible in a SPE for any discount factor $\delta < 1$ when payoffs are differentiable; and that all equilibria are characterized by inefficiently slow convergence to a steady state, a phenomenon they call gradualism. These insights have recently been greatly generalized by Matthews (2013) who has provided necessary conditions for SPE for a very general environment and who fully characterizes the set of SPE in more specific environments.[7]

In some of the applications presented above, the assumption of strict irreversibility is not ideal. Consider the case of public good contributions: while it is natural to assume some form of irreversibility and durability of the public good, it is less natural to assume that the rate of depreciation of the cumulative stock is exactly zero (as implied by strict irreversibility). Similarly, in models of technological adoption it is plausible to assume that the relative efficiency of an adoption would decrease over time; and so on so forth in other applications.[8]

In this paper we generalize the standard model of contribution with irreversibility to allow for depreciation on the players' contributions. We show that when depreci-

[4]Pitchford and Snyder (2004)

[5]Gale (2001)

[6]Compte and Jehiel (2004)

[7]Important results are also provided, among others, by Admati and Perry (1991), Compte and Jehiel (2004) who, studying environments with indivisibilities, characterize conditions in which there is a unique SPE.

[8]We will develop this point in Section 3 were we return on application of bargaining with concessions.

ation is allowed, even an arbitrarily small level of depreciation, a standard version of the folk theorem holds.

3.2 Model

Consider a dynamic contribution game with N players, $i \in I = \{1, 2, ..., N\}$. In the stage game, Γ, players simultaneously choose an action $a_i \in A_i = \mathbb{R}_+$ and player i's stage game payoff is given by $u_i(a_i, a_{-i})$, when action profile $a \in A$ is chosen. Assume that u_i is bounded, concave, continuously differentiable and satisfies **positive spillovers**, so that $u_i(a_i, \cdot)$ is **strictly** increasing in a_j for $j \neq i$. For technical reasons we will also need the following transversality condition, that $\lim_{a_i \to \infty} u_i(a_i, a_{-i}) = -\infty$ for all a_{-i}.

The stage game will be played in periods $t \in \{0, 1, 2, ...\}$. Before providing a formal description of the dynamic game we give a few preliminary definitions, which are standard in the repeated games literature[9] and will be applicable to our model.

Let $\mathcal{H}^t = A^t$ be the set of t-period histories, where $A^0 = \{\emptyset\}$ and note that a typical element $h^t \in \mathcal{H}^t$ is a list of t-period action profiles $(a_i, a_{-i}) \in A = A_1 \times ... \times A_N$. The set of complete histories is $\mathcal{H} = \cup_{t=0}^{\infty} \mathcal{H}^t$. A pure strategy for player i is $\sigma_i \colon \mathcal{H} \to A_i$ and a continuation strategy after history h^t, given σ_i, is $\sigma_{i|h^t}(h^s) = \sigma_i(h^t, h^s)$, where $(h^t, h^s) \in \mathcal{H}$ is the concatenation of h^t followed by h^s.

An outcome (or outcome path) of an infinitely repeated game is $\mathbf{a} = (a^0, a^1, ...) \in A^\infty$, where each $a^t \in A$. Denote by $\mathbf{a}^t \in \mathcal{H}^t$ the partial history which matches outcome $\mathbf{a}$ up to period t. Note that a pure strategy profile σ induces an outcome

[9] Our notation is consistent with Mailath and Samuelson (2006).

$\mathbf{a}(\sigma) = (a^0(\sigma), a^1(\sigma), ...)$ which is defined recursively as follows:

$$\begin{aligned} a^0(\sigma) &= (\sigma_i(\emptyset))_{i=1}^N, \\ a^1(\sigma) &= \left(\sigma_i\left(a^0(\sigma)\right)\right)_{i=1}^N, \\ a^2(\sigma) &= \left(\sigma_i\left(a^0(\sigma)\right), \sigma_i\left(a^1(\sigma)\right)\right)_{i=1}^N, \text{ etc.} \end{aligned}$$

For the analysis in this paper, there is no loss of generality to focus on pure strategies; so henceforth we will focus on pure strategies.

In the repeated game, denoted by Γ_δ, players maximize average discounted payoffs in each period t, i.e., given any σ_{-i} player i solves:

$$\max_{\sigma_i} U(\sigma_i, \sigma_{-i}) = \max_{\sigma_i} (1-\delta) \sum_{t=0}^{\infty} \delta^t u\left(a^t(\sigma)\right).$$

Note that players have a common discount factor $\delta \in (0,1)$. In period t, player i's continuation payoff, given outcome path $\mathbf{a}$, is denoted by:

$$U^t(\mathbf{a}) = (1-\delta) \sum_{\tau=t}^{\infty} \delta^\tau u\left(a_i^t, a_{-i}^t\right).$$

A subgame perfect Nash equilibrium σ is symmetric if for any t, $a^t(\sigma) = (a_t, ..., a_t)$ for some $a_t \in A$, i.e., the outcome path is such that all players play the same action in every period. Define:

$$\mathcal{S}(\Gamma_\delta) = \left\{(a_t)_{t=1}^{\infty} : a_t = a^t(\sigma) \text{ for all } t, \text{ for some symmetric SPNE } \sigma\right\}$$

to be the set of outcome paths of symmetric SPNE of Γ_δ. Along with the literature on dynamic contribution games, we will use subgame perfect equilibrium as the solution concept.

Up to this point, the game described above is a standard repeated game. A monotone game differs from a repeated game because the players' strategies are restricted to be monotonic non decreasing.

Definition 3 (monotonicity) *A pure strategy profile σ is monotonic if $\sigma(h^t) \geq a^{t-1}$ for any $h^{t-1} \in \mathcal{H}^{t-1}$ and a^{t-1} such that $h^t = (h^{t-1}, a^{t-1}) \in \mathcal{H}^t$.*

The first economic problem to which MCG have been applied is the entry game studied by Gale (1995), in which a_i's can be interpreted as private investments in a new technology with positive spillovers. In this case we may, for example, assume $u(a^t) = f(a^t) - c(a_i^t)$, where $f(a^t)$ is the production function and $c(a_i^t)$ is the opportunity cost of the idividual contribution. An application often cited to motivate the irreversibility assumption is that of games of accumulation of durable public goods (Gale, Matthews, Loockwod and Thomas, etc). In these games a_i is interpreted as the contribution to the public good by agent i. Assuming standard quasilinear preferences, the agent's i per period utility in period t is $U(a_t^{\Sigma}) - c_i^t$, where $a_t^{\Sigma} = \sum_j a_j^t$ is the level of public good and $c_i^t = a_i^t - a_i^{t-1} \geq 0$ is the individual contribution at time t. It can be easily shown that this game is strategically equivalent to a monotone contribution game in which agent i's utility function is $u(a^t) = \widetilde{u}(\sum_j a_j^t) - (1-\varepsilon)a_i^t$, and so fits in the framework described above.

In many applications it is natural to assume imperfect irreversibility. In the technological investment application, for example, it is natural to assume that as time passes the level of effectiveness of the technology depreciate at some positive rate ε: if a firms invests a_i^{t-1} at at time $t-1$ and investments are irreversible, then at time t the firm can choose any level $a_i^t \geq (1-\varepsilon)a_i^{t-1}$. Similarly, in the public good game described above, it may be natural to assume that contributions depreciate at a rate $\varepsilon > 0$, so $a_i^t = c_i^t + (1-\varepsilon)\sum_j a_j^{t-1}$ where c_i^t is the contribution at time t. In this case, it is natural to require non negative contributions, so $a_i^t \geq (1-\varepsilon)\sum_j a_j^{t-1}$.

To deal with this case, we propose the following weaker definition of monotonicity.

Definition 4 (ε-monotonicity) *A pure strategy profile σ is ε-monotonic if $\sigma(h^t) \geq (1-\varepsilon)a^{t-1}$ for any $h^{t-1} \in \mathcal{H}^{t-1}$ and a^{t-1} such that $h^t = (h^{t-1}, a^{t-1}) \in \mathcal{H}^t$.*

Monotonicity can be considered a special case of ε-monotonicity in which $\varepsilon = 0$; by allowing $\varepsilon > 0$, however, it allows to study the extensions of standard MCG described above; since ε can be arbitrarily small, moreover, ε-monotonicity allows to condsider cases that are just qualitatively different from cases in which monotonicity holds.

Let $\Gamma_\delta^\varepsilon$ be the dynamic (or extensive form) game given by Γ_δ with the added assumption that players' strategies must satisfy ε-monotonicity. We extend the above definitions to this environment in the natural way; for example, $\mathcal{S}(\Gamma_\delta^\varepsilon)$ denotes the set of outcome paths of symmetric SPNE of $\Gamma_\delta^\varepsilon$. Note that Γ_δ^1 is a standard repeated game.

3.2.1 Folk Theorem for Dynamic Contribution Games

Matthews (2013) studies Γ_δ^0 and characterizes the set of achievable outcomes as the "undercore", generally a strict subset of feasible, individually-rational payoff profiles. In particular, theorem 2 in Matthews (2013) shows that for any $\delta < 1$, if there are strictly positive spillovers, then any achievable outcomes is inefficient. [This result doesn't rely on the additional prisoner's dilemma assumption, but it holds with it as well.]

We show that a "folk theorem" holds in the game $\Gamma_\delta^\varepsilon$ for any $\varepsilon > 0$. In particular, for any individually rational, feasible payoff profile, v, and any $\varepsilon > 0$, there exists a $\underline{\delta}(\varepsilon) < 1$ such that for any $\delta > \underline{\delta}(\varepsilon)$ we have a subgame perfect equilibrium of $\Gamma_\delta^\varepsilon$ which delivers an average payoff of v.

For each player i define the **minmax** payoff as $\underline{v}_i = \min_{a_{-i}} \max_{a_i} u_i(a_i, a_{-i}) = \max_{a_i} u_i(a_i, 0)$. Let:

$$\mathcal{V} = \left\{ (u_i(a_i, a_{-i}))_{i=1}^N : a \in \mathbb{R}_+^N \text{ and } u_i(a_i, a_{-i}) \geq \underline{v}_i \text{ for all } i \right\},$$

be the set of feasible, individually rational payoffs.

In general we need to assume that $\mathcal{V}$ is convex and $\mathcal{V}^{\circ} \neq \emptyset$. However we can use the concavity of u and positive spillovers to prove the following lemma.

Denote the plane at the minmax utility level for player i by $\mathcal{K}_i := \{v \in \mathcal{V} : v_i = \underline{v}_i\}$, the corresponding half-space by $\mathcal{H}_i := \{v \in \mathcal{V} : v_i \geq \underline{v}_i\}$, the Pareto frontier by $PF(\mathcal{V}) := \{v \in \mathcal{V} : \nexists v' \in \mathcal{V} \text{ such that } v' \gg v\}$ and let

$$PF_{-}(\mathcal{V}) := \left\{w \in \mathbb{R}^N : \exists v \in PF(\mathcal{V}) \text{ such that } v \geq w\right\}$$

be the set majorized by the Pareto frontier.

Lemma 3.2.1 *$\mathcal{V}$ is convex and $\mathcal{V}^{\circ} \neq \emptyset$.*

Proof. To prove that $\mathcal{V}$ is convex, we show that (1) $PF_{-}(\mathcal{V})$ is convex, (2) for every i, j, $(v_{-ij}, v_j, \underline{v}_i) \in \mathcal{K}_i$, there is some v_j^* such that $\left(v_{-ij}, v_j^*, \underline{v}_i\right) \in PF(\mathcal{V}) \cap \mathcal{K}_i$, (3) $\mathcal{V} = \left(PF_{-}(\mathcal{V}) \cap \left\{\cap_{i=1}^N \mathcal{H}_i\right\}\right)$ on a dense set. Thus, (1) and (2) imply $\partial\mathcal{V} = \partial\left(PF_{-}(\mathcal{V}) \cap \left\{\cap_{i=1}^N \mathcal{H}_i\right\}\right)$, while (3) shows that no open balls are missing and then continuity of the utility functions shows $\mathcal{V} = \left(PF_{-}(\mathcal{V}) \cap \left\{\cap_{i=1}^N \mathcal{H}_i\right\}\right)$ and this is indeed convex since it is a finite intersection of conves sets.

(1) Take any $w, w' \in PF_{-}(\mathcal{V})$ and $\lambda \in (0,1)$. Let $v, v' \in PF(\mathcal{V})$ be such that $v \geq w$, $v' \geq w'$, and a, a' be action profiles for which $v = u(a)$ and $v' = u(a')$. Note that for each i:

$$u_i(\lambda a + (1-\lambda) a') \geq \lambda v_i + (1-\lambda) v_i' \geq \lambda w_i + (1-\lambda) w_i',$$

thus $\lambda w + (1-\lambda) w' \in PF_{-}(\mathcal{V})$, since $u(\lambda a + (1-\lambda) a') \in \mathcal{V}$.

(2) Assume by way of contradiction that there is an i,j,$(v_{-ij}, v_j, \underline{v}_i) \in \mathcal{K}_i$ such that $(v_{-ij}, v_j, \underline{v}_i) \notin PF(\mathcal{V})$ for all v_j such that $(v_{-ij}, v_j, \underline{v}_i) \in \mathcal{K}_i$. Let v_j^* be the largest v_j such that $\left(v_{-ij}, v_j^*, \underline{v}_i\right) \in \mathcal{K}_i$. By contradiction there exists $v' \in \mathcal{V}$ such

that $v' = u(a') > \left(v_{-ij}, v_j^*, \underline{v}_i\right)$. But then consider asking i and $-ij$ to make larger contributions, $a_{-j} \geq a'_{-j}$ so that $u_{-ij}\left(a'_j, a_{-j}\right) = v_{-ij}$ and $u_i\left(a'_j, a_{-j}\right) = \underline{v}_i$ (we can always do this by concavity, transversality, which implies utility is eventually decreasing in your own action, and positive spillovers). But then clearly $v_j^* \leq u_j(a') < u_j\left(a'_j, a_{-j}\right)$, which contradicts that v_j^* was the largest v_j such that $\left(v_{-ij}, v_j^*, \underline{v}_i\right) \in \mathcal{K}_i$.

Together (1) and (2) imply that $\mathcal{V}$ cannot have convexity issues at the boundaries. Thus it remains to show that the interior is convex.

(3) In this part of the proof we focus on $N = 2$, the more general case is analogous. Assume that there is an open set $U \subset \left(PF_{-}(\mathcal{V}) \cap \left\{\cap_{i=1}^{N} \mathcal{H}_i\right\}\right)$ such that $U \cap \mathcal{V} = \emptyset$. Note that we can consider U such that $\partial U \subset \mathcal{V}$, since by the closed map lemma[10] u is a closed map and thus $\mathcal{V}$ is closed.

Either there exists a $v \in \partial U$ such that an open ball $B \subset U$ has v on the north-west, south-east or south-west boundaries (since an open set must have some width). The case that v is on the north-west or south-east boundary is illustrated in the figure below. If $v = u(a)$, note that one of the players, i, can be asked to contribute a little

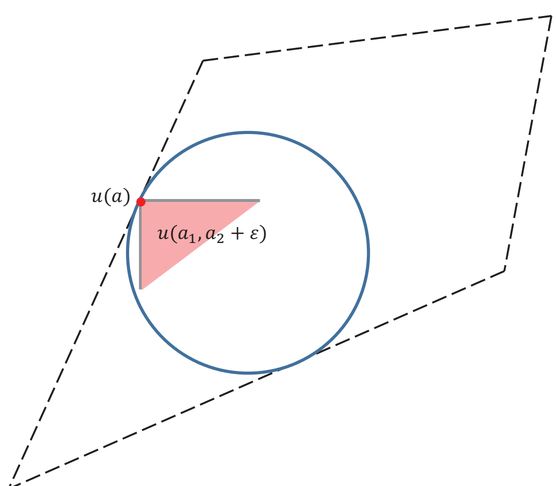


Figure 3.1: Proof idea when v is NW or SE

[10]This applies since by the transversality condition we can treat $A = [0, K]^N$, for some very large K.

more. By concavity and the transversality condition, we can take a such that i's utility is decreasing in his contribution. We therefore have that $u_i(a_i + \varepsilon, a_{-i}) < u_i(a)$ and by positive spillovers $u_{-i}(a_i + \varepsilon, a_{-i}) > u_{-i}(a)$ which would have to be in the shaded triangle, by continuity of u, for small enough ε. This gives a contradiction.

Alternatively, we can be on the south-west quadrant, which is illustrated in the figure below. In this case concavity and continuity of u_i implies that for λ suffi-

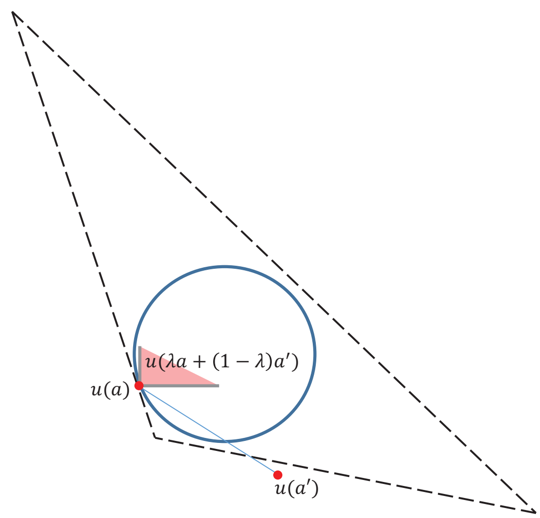


Figure 3.2: Proof idea when v is SW

ciently close to 1, we would have $u(\lambda a + (1-\lambda)a')$ in the shaded triangle, which is a contradiction.

Note that for any convergent sequence of a^t such that $u(a^t) \in \mathcal{V}$, by continuity of u we have that $\lim_{t\to\infty} u(a^t) \in \mathcal{V}$. Since $\mathcal{V}$ is not missing any open sets, we can approach any point in $\left(PF_-(\mathcal{V}) \cap \left\{\cap_{i=1}^N \mathcal{H}_i\right\}\right)$ by some sequence in $\mathcal{V}$ and thus $\mathcal{V} = \left(PF_-(\mathcal{V}) \cap \left\{\cap_{i=1}^N \mathcal{H}_i\right\}\right)$.

To prove that $\mathcal{V}^\circ \neq \emptyset$, note that from (2) above we have that $(v_{-i}, \underline{v}_i) \in PF(\mathcal{V})$ for each i where $v_{-i} \gg \underline{v}_{-i}$. Thus we have N linearly indepenedent vectors constructed from (if we were to draw the origin at $\underline{v}$) for each i, and hence $\mathcal{V}$ is N-dimensional. The claim follows since a convex set of full dimension has a non-empty interior. ■

A *passive* strategy for a player is not to contribute more than the minimum amount given by the ε-monotonicity constraint, i.e., to let the player's contribution depreciate by ε in each period.

Theorem 3.2.2 *For any $\varepsilon > 0$ and $v \in \{\widehat{v} \in \mathcal{V} \colon \widehat{v} \gg v' \text{ for some } v' \in \mathcal{V}\}$, there exists a $\underline{\delta}(v, \varepsilon) < 1$ such that for any $\delta \in (\underline{\delta}(v, \varepsilon), 1)$ there exists a SPNE σ of the game $\Gamma_\delta^\varepsilon$ whose average payoff profile is v.*

The proof is a variation of the classic pure-strategy folk theorems, with the caveat being that we have to respect the non-monotonicity property. We assume that $\mathcal{V}$ is convex in what follows and hope to be able to prove it (see lemma above for an attempt which is not too far off).

Proof. Let the pure action profile which achieves v be denoted by a^0. This is possible since $\mathcal{V}$ is convex. Note that since $\mathcal{V}$ is convex and has non-empty interior, there exists a $v' \in int(\mathcal{V})$ such that $v \gg v'$. Let $\eta > 0$ be such that $B_{N^{1/2}\eta}(v') \subset int(\mathcal{V})$. Define a "punishment of player i phase" (played after minmaxing) as the pure action profile a^i, which achieves the following payoff:

$$\left(v_1' + \eta, ..., v_{i-1}' + \eta, v_i', v_{i+1}' + \eta, ..., v_N' + \eta\right) \in int(\mathcal{V}).$$

We propose that the following strategy constitutes a subgame perfect equilibrium of the dynamic game which attains an average payoff close to v:

- If there has not been a deviation converge to profile a^0 and play this indefinitely
- If player i deviates, minmax player i for L periods (L to be computed shortly) and then converge to profile a^i
    - Deviations during the punishment phase of player i by another player j, result in punishment of j according to the above

- Simultaneous deviations by two or more players are ignored

By PS, minmaxing player i involves the other players pursuing a passive strategy, i.e., letting their contributions depreciate to 0. Thus, an upper bound on the benefit from deviating, t periods after the deviation took place is:

$$b_i^t = \max_{a_i} u_i \left(a_i, (1-\varepsilon)^t a_{-i}^0\right).$$

Here we are assuming that at the time of the initial deviation from the path converging to v, or at any time afterwards, the deviating player does not have to respect any monotonicity constraint, thus it is certainly an upper bound on the benefit from deviation.

Note that, $b_i^t \to \underline{v}_i < v_i$ and thus b_i^t is eventually in a γ-neighborhood of $\underline{v}_i$, where $\gamma = \frac{1}{2}(v_i' - \underline{v}_i) > 0$. That is, there exists a $T(\varepsilon)$ such that for all $t \geq T(\varepsilon)$, $b_i^t < \underline{v}_i + \gamma = \frac{1}{2}v_i' + \frac{1}{2}\underline{v}_i < \frac{1}{2}v_i + \frac{1}{2}\underline{v}_i$.

We follow a standard construction for folk theorems, where the deviating player is minmaxed for L periods and then the punishers are rewarded.

Note that with sufficiently high δ, player i has no incentive to deviate since:

$$\begin{aligned}
& (1-\delta)\left[\sum_{t=0}^{T(\varepsilon)-1} \delta^t b_i^t + \sum_{t=T(\varepsilon)}^{L-1} \delta^t b_i^t + \sum_{t=L}^{\infty} \delta^t v_{i,t}'\right] \\
< \; & (1-\delta)\left[\sum_{t=0}^{T(\varepsilon)-1} \delta^t b_0 + \sum_{t=T(\varepsilon)}^{L-1} \delta^t (v_i - \gamma) + \sum_{t=L}^{\infty} \delta^t v_{i,t}'\right] \\
= \; & v_i + (1-\delta)\left[\sum_{t=0}^{T(\varepsilon)-1} \delta^t (b_0 - v_i) - \sum_{t=T(\varepsilon)}^{L-1} \delta^t \gamma + \sum_{t=L}^{\infty} \delta^t \left(v_{i,t}' - v_i\right)\right] \\
\leq \; & v_i + \left(1-\delta^{T(\varepsilon)}\right)(b_0 - v_i) - \left(\delta^{T(\varepsilon)} - \delta^L\right)\gamma + \delta^L (v_i' - v_i) \\
\leq \; & v_i + \left(1-\delta^{T(\varepsilon)}\right)(b_0 - v_i) - \left(\delta^{T(\varepsilon)} - \delta^L\right)\gamma,
\end{aligned}$$

where $v'_{i,t}$ is the efficient way to converge to v'_i (this happens after a finite number of periods) and $v'_i \geq v'_{i,t}$ by construction. The last inequality follows since $v \gg v'$. Let $L = T(\varepsilon)\, k(\delta)$ with $k(\delta) = \log_\delta \frac{1}{2}$, then the above is less than v_i (and thus playing a^0 is an equilibrium) if:

$$\begin{aligned} \left(1-\delta^{T(\varepsilon)}\right)(b_0 - v_i) - \left(\delta^{T(\varepsilon)} - \delta^{T(\varepsilon)k}\right)\gamma &< 0 \\ \left(1-\delta^{T(\varepsilon)}\right)(b_0 - v_i) &< \delta^{T(\varepsilon)}\left(1-\delta^k\right)\gamma = \delta^{T(\varepsilon)}\frac{1}{2}\gamma \end{aligned}$$

or if:

$$\delta^{T(\varepsilon)} > \frac{b_0 - v_i}{b_0 - v_i + \frac{1}{2}\gamma}.$$

Thus, let $\underline{\delta} = \left(\frac{b_0 - v_i}{b_0 - v_i + \frac{1}{2}\gamma}\right)^{1/T(\varepsilon)} \in (0,1)$ since $(b_0 - v_i) < \infty$ (by transversality and concavity) and note that deviations off the equilibrium path are not profitable for any $\delta \in (\underline{\delta}, 1)$.

We are left to ensure that the punishments are subgame perfect. This follows by a similar modification of the standard argument to the one illustrated above and results in another lowerbound for δ, say $\underline{\delta}'$. We can then choose $\underline{\delta}(\varepsilon, v) = \max\{\underline{\delta}, \underline{\delta}'\}$.

■

Thus a significantly larger set than the undercore of Matthews (2013) is immediatebly achievable in $\Gamma^\varepsilon_\delta$ (note in Lockwood and Thomas and Matthews these are not achievable even in the limit).

It is possible to extend this result in a number of ways. One possibility is to consider the convex hull of $\mathcal{V}$ (in the case that it is not convex). To make that proof simpler, we could assume the existence of a public correlating device. This assumption would not be essential, but without it we would need a nonstationary sequence of actions to achieve v (and possibly a higher minimum δ). This type of argument is standard in the repeated games literature (see Mailath and Samuelson, 2006, p.69).

3.3 Generalized Model

We shall now assume a slightly more general structure than the preceding literature on MCGs. In particular, motivated by public good games, we assume that the flow utility from the public good is a function of the aggregate contribution level at present, but that the cost is simply your marginal (or current period) contribution. Thus, we want to consider a MCG with ε-monotonicity where player i's utility is given by:

$$u_i(a^t \mid a^{t-1}) = \widetilde{u}_i\left(\sum_j a_j^t\right) - c_i\left(a_i^t - (1-\varepsilon)\, a_i^{t-1}\right).$$

Note that $a_i^t - (1-\varepsilon)\, a_i^{t-1} \geq 0$, by the ε-monotonicity assumption. We further assume that $\widetilde{u}_i$ and c_i are continuous, $\widetilde{u}_i$ is concave and c_i is convex (maybe some more assumptions on these to guarantee convexity of $\mathcal{V}$).

This can therefore be modeled a dynamic game with action set for player i, $A_i = \mathbb{R}_+$ and states of the world $S = A$, where $A = A_1 \times ... \times A_N$. Thus, we have that:

$$u_i : A \times S \to \mathbb{R},$$

is continuous and furthermore, the state transition function, $q : A \times S \cup \{\emptyset\} \to S$, is deterministic, with:

$$q : (a, s) \mapsto a.$$

We shall denote the stage game with state s by $\Gamma(s)$. Note that there is no need to change our notation for histories, since they already capture all the relevant information about the state of the world (since this is just the previous period action profile). We will denote the dynamic game starting from state s with discount factor δ by $\Gamma_\delta(s)$ and the the ε-monotonic game by $\Gamma_\delta^\varepsilon(s)$.

Note that since both A and S are infinite, the folk theorem of Dutta (1995) does not apply. We shall use the structure of our problem to prove a more general folk theorem. It is not clear that a theorem of the level of generality of Dutta's would hold in this setting. In particular, we note that in our setting it is natural to consider a deterministic q, while Dutta (1995) allows for random state transitions.

To make proofs neater, we add an assumption on the utility function of the players, which is sometimes used in the MCG literature. We assume that u_i satisfies the **prisoner's dilemma** (PD) assumption, i.e., $\widetilde{u}_i'(0) < c_i'(0)$ for all i. Given that u_i already satisfies positive spillover's, this means that we can use simpler Nash-reversion arguments in proving the folk theorem, as opposed to the more intricate constructions above. The proof extends to the more general setting in the obvious way, as long as we are careful to define minmax payoffs appropriately (see the definitions in Dutta (1995) for example).

We also assume that it is efficient to provide some public good, i.e., the sum of players' utilities exceeds $c_i'(0)$ for all i, to ensure non-triviality of the problem (this also guarantees that the set $\mathcal{V}$ is non-empty).

3.3.1 Impossibility Result

The impossibility results of Matthews (2013) applies to this setting. In particular, Matthews shows that in the monotone game there are inefficient contributions even in the limit. This implies that the discounted average payoffs are also inefficient and thus a failure of the folk theorem for any discount factor δ when $\varepsilon = 0$.

3.3.2 Folk Theorem

We prove a "folk theorem" for $\Gamma_\delta^\varepsilon(s)$ for any $\varepsilon > 0$. In particular, for any individually rational, feasible payoff profile, v, and any $\varepsilon > 0$, there exists a $\underline{\delta}(\varepsilon, v) < 1$ such that for any $\delta > \underline{\delta}(\varepsilon, v)$ we have a subgame perfect equilibrium of $\Gamma_\delta^\varepsilon$ which delivers an average payoff of v.

Let:

$$\mathcal{V}=\left\{(u_i\left(a\mid(1-\varepsilon)a\right))_{i=1}^N : a\in\mathbb{R}_+^N \text{ and } u_i\left(a^t\mid a^{t-1}\right)\geq \underline{v}_i \text{ for all } i\right\},$$

be the set of feasible, individually rational payoffs. Assume this is convex and non-empty (follows from non-triviality).

Note that with positive spillovers (PS) and the prisoner's dillemma (PD) assumptions, the natural minmax payoff for each player is $u_i(\mathbf{0}\mid\mathbf{0})=\underline{v}_i$. This also has the benefit of not being state-dependent.

Theorem 3.3.1 *For any $\varepsilon>0$ and $v\in\{\widehat{v}\in\mathcal{V}\colon \widehat{v}\gg v' \text{ for some } v'\in\mathcal{V}\}$, there exists a $\underline{\delta}(v,\varepsilon)<1$ such that for any $\delta\in(\underline{\delta}(v,\varepsilon),1)$ there exists a SPNE σ of the game $\Gamma_\delta^\varepsilon(\mathbf{0})$ whose average payoff profile is arbitrarily close to v.*

While this seems rather complicated, the structure imposed allows for a much simpler proof than that in Dutta [1995]. We assume that $\mathcal{V}$ is convex in what follows and hope to be able to prove it (see lemma above for an attempt which is not too far off).

Proof. Let $v_t\to v$, be the efficient way to converge to v (note that it could be prohibitively costly to jump to a^0 immediately, but it can be achieved in $K<\infty$ periods). In the normal phase player i receives average payoff:

$$(1-\delta)\sum_{t=0}^{K-1}\delta^t v_{t,i}+\delta^K v_i,$$

which can be made arbitrarily close to v_i if K is sufficiently high.

Define a "normal" phase where players play an action profile which achieves v, say a^0 (again we may need to converge to this, which is why we can only get arbitrarily close to v).

Following a deviation, all players revert to playing the passive strategy and letting their contributions depreciate. This is equivalent to (constrained) repetition of the stage-game Nash equilibrium and therefore subgame perfect.

By PD, an upper bound on the benefit player i can get from deviating in state s is:

$$b_i^t = u_i\left(0, (1-\varepsilon)^{t+1} s \mid (1-\varepsilon)^t s\right).$$

Now, $b_i^t \to \underline{v}_i < v_i$ and thus b_i^t is eventually in a γ-neighborhood of $\underline{v}_i$, where $\gamma = \frac{1}{2}(v_i - \underline{v}_i) > 0$. That is, there exists a $T(\varepsilon)$ such that for all $t \geq T(\varepsilon)$, $b_t < v_i - \gamma = \frac{1}{2}v_i + \frac{1}{2}\underline{v}_i$. Now:

$$\begin{aligned}
(1-\delta)\left[\sum_{t=0}^{T(\varepsilon)-1} \delta^t b_i^t + \sum_{t=T(\varepsilon)}^{\infty} \delta^t b_i^t\right] &\leq (1-\delta)\sum_{t=0}^{T(\varepsilon)-1} \delta^t b_0 + (1-\delta)\sum_{t=T(\varepsilon)}^{\infty} \delta^t (v_i - \gamma) \\
&= v_i + (1-\delta)\left[\sum_{t=0}^{T(\varepsilon)-1} \delta^t (b_0 - v_i) - \sum_{t=T(\varepsilon)}^{\infty} \delta^t \gamma\right] \\
&\leq v_i + \left(1 - \delta^{T(\varepsilon)}\right)(b_0 - v_i) - \delta^{T(\varepsilon)}\gamma,
\end{aligned}$$

which is less than v_i if $\delta^{T(\varepsilon)} > \frac{b_0 - v_i}{b_0 - v_i + \gamma}$. Thus, let $\underline{\delta}(v, \varepsilon) = \left(\frac{b_0 - v_i}{b_0 - v_i + \gamma}\right)^{1/T(\varepsilon)}$ and note that $\underline{\delta}(\varepsilon) \in (0,1)$, since $(b_0 - v_i) < \infty$. Thus, for $\delta > \underline{\delta}(v, \varepsilon)$ we have that punishment by passive strategies (constrained reversion to Nash) is sufficient to discourage deviations. ■

3.4 Concluding Remarks

This paper has shown that the anti-folk theorem results in the literature on dynamic contribution games are not robust when some depreciation is added. In particular, we have shown that a positive rate of depreciation of contributions restores the folk theorem. However, this does require the discount factor to get arbitrarily close to 1 while keeping the rate of depreciation fixed. Continuity between the present result and the preceding literature can be achieved if we vary the deprecation rate with the discount rate. A related area for future work is to characterize the set of achievable payoffs for a fixed discount factor and depreciation rate, in the style of Abreu et al. (1990).

Bibliography

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica: Journal of the Econometric Society*, pp. 1041–1063.

Alonso, R., and N. Matouschek (2008): "Optimal delegation," *The Review of Economic Studies*, 75(1), 259–293.

Amador, M., and K. Bagwell (2012): "Tariff revenue and tariff caps," *The American Economic Review*, 102(3), 459–465.

——— (2013): "The theory of optimal delegation with an application to tariff caps," *Econometrica*, 81(4), 1541–1599.

Ambrus, A., and G. Egorov (2012): "Delegation and Nonmonetary Incentives," Typeset, Northwestern University.

Athey, S. (2002): "Monotone comparative statics under uncertainty," *The Quarterly Journal of Economics*, 117(1), 187–223.

Axelson, U. (2007): "Security design with investor private information," *The Journal of Finance*, 62(6), 2587–2632.

Banks, J. S. (1989): "Agency budgets, cost information, and auditing," *American Journal of Political Science*, pp. 670–699.

Baron, D. P. (2000): "Legislative Organization with Informational Committees," *American Journal of Political Science*, 44, 485–505.

Baron, D. P., and R. B. Myerson (1982): "Regulating a monopolist with unknown costs," *Econometrica: Journal of the Econometric Society*, pp. 911–930.

Bendor, J., and A. Meirowitz (2004): "Spatial models of delegation," *American Political Science Review*, 98(2), 293–310.

Biais, B., and T. Mariotti (2005): "Strategic liquidity supply and security design," *The Review of Economic Studies*, 72(3), 615–649.

Brooks, B. (2014): "Surveying and selling: Belief and surplus extraction in auctions," *Working Paper*.

Carpenter, D. P. (1996): "Adaptive signal processing, hierarchy, and budgetary control in federal regulation," *American Political Science Review*, pp. 283–302.

Carroll, G. (2014): "Robustness and Linear Contracts," *American Economic Review, forthcoming*.

Chassang, S. (2013): "Calibrated incentive contracts," *Econometrica*, 81(5), 1935–1971.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

Dang, T. V., G. Gorton, and B. Hölmstrom (2012): "Ignorance, Debt and Financial Crises," *Yale School of Management Working Paper*.

DeMarzo, P., and D. Duffie (1999): "A liquidity-based model of security design," *Econometrica*, 67(1), 65–99.

DeMarzo, P. M. (2005): "The pooling and tranching of securities: A model of informed intermediation," *Review of Financial Studies*, 18(1), 1–35.

DeMarzo, P. M., I. Kremer, and A. Skrzypacz (2005): "Bidding with Securities - Auctions and Security Design," *American Economic Review*, 95(4), 936–959.

Dutta, P. K. (1995): "A folk theorem for stochastic games," *Journal of Economic Theory*, 66(1), 1–32.

Epstein, D., and S. O'Halloran (1994): "Administrative procedures, information, and agency discretion," *American Journal of Political Science*, pp. 697–722.

Frankel, A. (2014): "Aligned delegation," *The American Economic Review*, 104(1), 66–83.

Gailmard, S., and J. W. Patty (2012): "Formal models of bureaucracy," *Annual Review of Political Science*, 15, 353–377.

Holmström, B. (1977): "On Incentives and Control in Organizations," Ph.D. thesis, Stanford University.

——— (1979): "Moral hazard and observability," *The Bell Journal of Economics*, pp. 74–91.

Huber, J. D., and C. R. Shipan (2002): *Deliberate discretion?: The institutional foundations of bureaucratic autonomy*. Cambridge University Press.

Inderst, R., and H. M. Mueller (2006): "Informed lending and security design," *The Journal of Finance*, 61(5), 2137–2162.

Innes, R. D. (1990): "Limited Liability and Incentive Contracting with ex-ante Action Choices," *Journal of Economic Theory*, 52(1), 45–67.

Kaplan, S. N., and P. Strömberg (2003): "Financial contracting theory meets the real world: An empirical analysis of venture capital contracts," *The Review of Economic Studies*, 70(2), 281–315.

Koessler, F., and D. Martimort (2012): "Optimal delegation with multi-dimensional decisions," *Journal of Economic Theory*, 147(5), 1850–1881.

Kováč, E., and T. Mylovanov (2009): "Stochastic mechanisms in settings without monetary transfers: The regular case," *Journal of Economic Theory*, 144(4), 1373–1395.

Krishna, V., and J. Morgan (2008): "Contracting for information under imperfect commitment," *The RAND Journal of Economics*, 39(4), 905–925.

Laffont, J.-J., and D. Martimort (2009): *The theory of incentives: the principal-agent model*. Princeton University Pres, New York.

Madarász, K., and A. Prat (2014): "Sellers with Misspecified Models," *Working Paper*.

Mailath, G. J., and L. Samuelson (2006): *Repeated Games and Reputations: Long-run Relationships*. Oxford University Press, New York.

Martimort, D., and A. Semenov (2006): "Continuity in mechanism design without transfers," *Economics Letters*, 93(2), 182–189.

Matthews, S. A. (2001): "Renegotiating moral hazard contracts under limited liability and monotonicity," *Journal of Economic Theory*, 97(1), 1–29.

McCarty, N. (2004): "The appointments dilemma," *American Journal of Political Science*, 48(3), 413–428.

Melumad, N. D., and T. Shibano (1991): "Communication in settings with no transfers," *The RAND Journal of Economics*, pp. 173–198.

Myerson, R. B. (1983): "Mechanism design by an informed principal," *Econometrica*, pp. 1767–1797.

Penot, J. (2013): *Calculus without derivatives*. Springer, New York.

Poblete, J., and D. Spulber (2012): "The form of incentive contracts: agency with moral hazard, risk neutrality, and limited liability," *The RAND Journal of Economics*, 43(2), 215–234.

Rochet, J.-C., and L. A. Stole (2003): "The economics of multidimensional screening," *Econometric Society Monographs*, 35, 150–197.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press, Princeton.

Shavell, S. (1979): "On moral hazard and insurance," *The Quarterly Journal of Economics*, pp. 541–562.

Shipan, C. R. (2004): "Regulatory regimes, agency actions, and the conditional nature of congressional influence," *American Political Science Review*, 98, 467–480.

Strausz, R. (2006): "Deterministic versus stochastic mechanisms in principal–agent models," *Journal of Economic Theory*, 128(1), 306–314.

Ting, M. M. (2001): "The"Power of the Purse"and its Implications for Bureaucratic Policy-Making," *Public Choice*, 106(3-4), 243–274.

Walsh, C. E. (1995): "Optimal contracts for central bankers," *The American Economic Review*, pp. 150–167.

Weingast, B. R., and M. J. Moran (1983): "Bureaucratic discretion or congressional control? Regulatory policymaking by the Federal Trade Commission," *The Journal of Political Economy*, pp. 765–800.

Wood, B. D., and J. E. Anderson (1993): "The politics of US antitrust regulation," *American Journal of Political Science*, pp. 1–39.

Wood, B. D., and R. W. Waterman (1991): "The dynamics of political control of the bureaucracy," *The American Political Science Review*, pp. 801–828.

Yang, M. (2013): "Optimality of Debt under Flexible Information Acquisition," *SSRN Working Paper*.

Yang, M., and Y. Zeng (2014): "Security Design in a Production Economy with Flexible Information Acquisition," *SSRN Working Paper*.

Zeidler, E. (1985): *Nonlinear Functional Analysis and its Applications III: Variational Methods and Optimization*. Springer-Verlag, New York.