Essays on the Principal-Expert Problem

by

Luis G. Zermeño Vallés

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author ....
Department of Economics
May 15, 2012

Certified by ....
Bengt Holmström
Paul A. Samuelson Professor of Economics
Thesis Supervisor

Certified by ....
Glenn Ellison
Gregory K. Palm Professor of Economics
Thesis Supervisor

Accepted by ....
Michael Greenstone
3M Professor of Environmental Economics

Essays on the Principal-Expert Problem

by

Luis G. Zermeño Vallés

Submitted to the Department of Economics
on May 15, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This dissertation studies the problem of motivating an expert to help a principal take a decision. The first chapter examines a principal-expert model in which the only source of friction is that the expert must be induced to acquire non-verifiable information relevant for the principal's decision. I show that contracts that specify a single transfer scheme are strictly dominated by contracts that specify a *menu* of transfer schemes from which the expert can choose. Optimal menu contracts often induce inefficient decision-making. Indeed, in environments where decisions affect the amount of information that is revealed ex-post, distorting decision-making in favor of decisions that reveal more information can help to provide better incentives for information acquisition. Without menus, there is an additional reason to distort decision-making. In this case, distorting decision-making is almost always optimal, and the distortions can favor decision that reveal *less* information ex-post.

The second chapter studies the role of authority in a more general version of the principal-expert model studied in chapter 1. Contracts specify a menu of transfer schemes from which the expert can choose. I consider three possible allocations of authority: 1) Full-commitment, under which the expert's choice from the menu also determines the decision to be taken. 2) Expert-authority, under which the expert can ultimately take any decision. 3) Principal-authority, under which the principal can ultimately take any decision. I provide conditions under which any Pareto-optimal outcome implementable under full-commitment can also be implemented when either one of the parties has authority.

The third chapter analyzes what happens if the expert is not motivated through a contract, but through his concern about his reputation. The expert can be a *charlatan* (and have no relevant information) or *informed*, and he privately knows his type. The principal makes inference about the expert's type based on the expert's report and on the outcome of the decision. I show that the expert's concern about his reputation coarsens the information that he can credibly transmit. As a result, decision-making is biased away from the status quo: the decision that the principal would take under the prior is taken too infrequently.

Thesis Supervisor: Bengt Holmström
Title: Paul A. Samuelson Professor of Economics

Thesis Supervisor: Glenn Ellison
Title: Gregory K. Palm Professor of Economics

To my parents

Acknowledgments

I would first like to thank Bengt Holmström and Glenn Ellison for their invaluable advice and support. Their insight and attention to detail have had a huge impact on my work. I would also like to thank Bob Gibbons for his guidance and support at the early stages of this process, and Juuso Toikka for many helpful conversations.

I have had the opportunity to discuss my work with several people. In particular, I would like to thank Ricardo Alonso, Luis Garicano, Pablo Kurlat, James Malcomson, Niko Matouschek, Parag Pathak, Heikki Rantakari, Jean Tirole and seminar participants at the MIT Organizational Economics Lunch, the MIT Theory Lunch and the ITAM Economics Seminar. Their comments on early drafts of this work have been extremely helpful.

I have had the fortune of sharing this experience with an incredible set of classmates and friends. In particular, I am grateful to Dan Barron, Alejandro Cornejo, Arun Chandrasekhar, Gabriel Carroll, Ben Golub, Sebastián Di Tella, Felipe Iachan, Anton Kolotilin, Horacio Larreguy, José Luis Montiel, Juan Passadore, Michael Powell, Felipe Severino, Alex Wolitzky and, especially, Juan Pablo Xandri. Their intellectual input and friendship were essential for the completion of this thesis.

This process would have been much harder without my family's continuous support. I am immensely grateful to my father for his timely wisdom, to my mother for making everything work, to María and Diego for being always there to remind me of what is important, and to Sofía for her love.

Cambridge, Massachusetts, USA

May 15, 2012

Contents

Chapter 1

A Principal-Expert Model and the Value of Menus

1.1 Introduction

The classical moral hazard literature has analyzed the problem of motivating an agent whose costly and non-verifiable effort is a direct input to the production process. The agent is induced to work through a contract that specifies a single transfer scheme contingent on some performance measure.[1] This model provides a good description for situations where the agent is a *worker*. However, there is another type of agent, an *expert*, whose effort does not affect output directly, but instead generates information relevant for the decisions that determine output. The problem of motivating experts has one fundamental difference from that of motivating workers: in the former, there is additional information arriving at an interim stage, after the expert exerts effort, but before outcomes are observed. This raises the possibility that the parties could do better by writing contracts that are qualitatively different from those used to motivate workers. A contract could specify a *menu* of transfer schemes from which the expert can choose at the interim stage.[2] The purpose of this paper

[1] See, for example, Mirrlees (1976), Holmström (1979), Harris and Raviv (1979), Shavell (1979), Grossman and Hart (1983), Holmström and Milgrom (1991) or Kim (1995).

[2] Menu contracts are common in the analysis of incentive problems that combine adverse selection with moral hazard (see Laffont and Tirole (1986), for example).

is to determine when and why menu contracts are valuable, and to study the features of optimal menu contracts.

I examine a model with two risk-neutral parties, a principal (she) and an expert (he). The principal is the residual claimant of output, which is determined by a decision and a *binary* state of nature. The expert has no direct interest in the decision and is protected by limited liability.[3] Both parties start with the same prior about the state of nature. After a contract is signed but before the state is realized, the expert can exert costly and non-verifiable effort to acquire non-verifiable information about the state. Formally, we may assume without loss of generality that each level of effort induces a distribution over *posteriors* about the state with expectation equal to the prior. Contracts specify a menu of pairs; each pair is formed by a decision and a transfer scheme, which may be contingent on everything that is observable at the end. In the case where transfer-scheme menus are not available, every decision must be paired with the same transfer scheme. At the interim stage, the expert privately observes a posterior drawn from the distribution corresponding to the effort that he exerted. Based on this observation, he then selects an element from the menu specified by the contract. The selection determines the decision to be taken and the transfer scheme for payments. Note that, since the parties are risk-neutral, the total surplus generated by the relationship is determined by the expert's effort and the decision rule, which specifies the decision that is taken after the expert observes each posterior. Thus, the *outcome* of the relationship can be fully described by a level of effort, a decision rule and a constant specifying the expert's net gain from the relationship.

The core of the analysis is concerned with determining how the set of outcomes that can be implemented depends on whether contracts specify transfer-scheme menus. In particular, the objective is to understand the interaction between the decision rule to be implemented and the level of effort that can be attained. Given a contract, the expert's incentives to acquire information are entirely determined by the expected payoffs that he will receive, given his optimal selections from the menu, after observing each posterior. These payoffs are described by the value function associated with the expert's choice problem at the interim

[3] For example, consultants or portfolio managers typically have no direct interest in their clients' choices.

stage.[4] This function has a central role in the analysis. Indeed, the first result in the paper is a characterization of implementable outcomes (with menus) expressed only in terms of the expert's interim value function.[5] This characterization serves two purposes: 1) it enables us to use the interim value function directly as the incentive instrument (replacing contracts); 2) it makes explicit the way in which the decision rule to be implemented restricts the shape that the interim value function can take (and thus, the amount of effort that can be induced).

With menus, decision rules add restrictions only when they prescribe decisions that do not reveal the state ex-post. As an illustration, consider the following examples:

Example 1: An expert is hired to advice a gambler on how to make a bet on the *Celtics* vs *Lakers* game. The set of states is $\{$*Celtics win, Lakers win*$\}$, the decision is how to bet, and the output corresponds to the gambler's profit. The parties observe at the end output, the decision and the state.

Example 2: An expert is hired to assess whether a project should be undertaken. The set of states is $\{$*Project succeeds, Project fails*$\}$. Output is equal to zero if the project is not undertaken, it equals some positive value if the project is undertaken and succeeds, and some negative value if the project is undertaken but fails. The parties observe the value of output and whether the project was undertaken.[6]

In Example 1, the realization of the state can eventually be observed regardless of the decision taken. We will see that, in this case, decision rules impose no restrictions on the shape that the interim value function can take, so there is no need to induce ex-post inefficient decision-making. Intuitively, with menus, a contract can give the expert choices that have different state-contingent payoffs even though they induce the same decision. This possibility

[4] This function depends only on posteriors.

[5] This result exploits the fact that, since the expert's observation is a posterior, his interim payoff, given any choice from the menu, is linear in his type. In this sense, the result is analogous to the standard characterizations of implementable outcomes in linear mechanism design environments (see, for example, Rochet (1987) or Jehiel and Moldovanu (2001)). The characterization in this paper has some unique features derived from the specific characteristics of this problem.

[6] This environment is studied in Lambert (1986), Levitt and Snyder (1997), and Inderst and Klein (2007).

enables the parties to *separate* the expert's incentives to acquire information from the way in which the decision is taken.

In Example 2, if the project is not undertaken, the state is never revealed (the counterfactual cannot be observed). We will see that, in this case, even though separation is still possible, decision rules do impose restrictions on the expert's interim value function. Intuitively, if the expert is to take a decision that does not reveal the realization of the state, the (optimal) choice from the menu that induces that decision must pay him the same amount regardless of the state. This fact restricts the shape that the interim value function can take. Such restrictions can only be relaxed by distorting decision-making in favor of decisions that reveal the state.

Without transfer-scheme menus it is no longer possible to separate the expert's incentives to exert effort from the decision taken; the expert can receive different state-contingent payoffs only if the decision taken changes. If the decision taken can be identified ex-post (as in the examples above), this is the only difference compared with the case where menus are available. The paper formalizes this claim by extending the characterization of implementable outcomes to the case without menus and where decisions are identifiable.[7]

The two characterizations illustrate how decision rules induce different restrictions on the expert's interim value function depending on whether contracts can rely on transfer-scheme menus. In order to assess the importance of these differences, we need to compare the Pareto-optimal outcomes that can be implemented when menus are available and when they are not. By working directly with the expert's interim value function, we identify conditions (regarding the information acquisition technology) under which "the first-order approach" is valid.[8] Under these conditions, we then characterize optimal contracts and Pareto-optimal outcomes (implementable with and without menus).[9] We show that optimal contracts (with

[7] If decisions are not identifiable ex-post, the set of implementable outcomes without transfer-scheme menus shrinks, making menus more valuable.

[8] The first-order approach replaces the expert's effort incentive compatibility constraints for a single local constraint. The conditions provided here are analogous to Rogerson (1985)'s conditions for the validity of the first-order approach in the standard moral hazard framework.

[9] The fact that the state of nature is binary provides significant tractability because the expert's type space (and the domain of the interim value function) becomes one-dimensional.

and without menus) specify at most three alternatives that induce different state-contingent payoffs for the expert. Thus, optimal menus are relatively simple. For instance, in Example 1, the optimal menu contract gives the expert a choice between two transfer schemes, one paying him only if the *Celtics* win, and the other paying him only if the *Lakers* win.

The comparison of the Pareto-optimal outcomes that are implementable depending on whether transfer-scheme menus are available gives the following conclusions: 1) Menus are usually valuable because they enable the parties to separate the expert's incentives to acquire information from the way in which the decision is taken. This improves decision-making and induces more information acquisition. Separation may require the expert to be given both the option to bet for and against the *same* decision. Indeed, in Example 1, the optimal menu typically includes an option in which the expert induces the principal to bet for one team and simultaneously chooses to be paid only if the other team wins. 2) The nature of optimal distortions in decision-making depends crucially on whether menus are available. With menus, optimal distortions are driven purely by differences in the amount of information that different decisions reveal about the state ex-post. Inducing the expert to exert more effort requires distorting decision-making in favor of decisions that reveal more information. Thus, in Example 1, the optimal decision rule is ex-post efficient. However, in Example 2, the optimal decision rule is distorted in favor of undertaking the project, which is the only decision that reveals the state. Without menus, the inability to separate the expert's incentives from the principal's decision introduces an additional motive to distort decision-making. As a result, ex-post efficient decision rules are optimal only in exceptional circumstances. Moreover, the direction of optimal distortions becomes ambiguous; it may also be optimal to distort decision-making in favor of decisions that do not reveal the state.

The previous literature recognized the presence of distortions in decision-making in principal-agent models where the agent exerts effort and, at the same time, helps take a decision in which he has no direct interest (Lambert (1986), Demski and Sappington (1987), Levitt and Snyder (1997), Diamond (1998), Athey and Roberts (2001), Inderst and Klein (2007), Malcomson (2009, 2011) and Chade and Kovrijnykh (2011)).[10] These studies either

[10] Osband (1989) and Prendergast (1993) analyze models with similar features, but take their analyses in

focus on contracts that specify a single transfer-scheme or analyze environments in which menus turn out to be not valuable (Levitt and Snyder (1997) and Inderst and Klein (2007)). By comparing the outcomes that can be attained depending on whether transfer-scheme menus are available, this paper makes two contributions to this literature. First, it uncovers the nature of optimal distortions to decision-making (with and without menus). Second it shows that, contrary to previous findings, menus are usually valuable.[11]

Another strand of the literature has studied the problem of inducing agents, who have a direct interest in decisions, to acquire non-verifiable information (Lewis and Sappington (1997), Aghion and Tirole (1997), Crémer, Khalil and Rochet (1998*a*), Crémer, Khalil and Rochet (1998*b*), Szalay (2005) and Szalay (2009)).[12] The incentive design problem analyzed in these papers is quite different from the one here, since now decision rules serve as the main incentive instrument.[13] Szalay (2005) is the most similar to this paper. It also has the feature that, even though the expert's preferences over decisions are ex-ante aligned with those of the principal, it is optimal to implement decision rules that are not ex-post efficient. Nevertheless, the nature of distortions there is very different from the one in this paper. In Szalay (2005), the expert has his own motive to acquire information, the problem is that he does not take into account the principal's gain when deciding how much to acquire. The expert can be induced to gather more information if the parties commit never to take the decisions that are efficient when the information is relatively inaccurate.

The rest of the paper is structured as follows. Section 1.2 describes the model and defines

different directions. In Osband (1989), the main source of friction is the fact that the principal does not know the expert's quality (how costly it is for him to acquire information). In Prendergast (1993), the principal receives a free (and private) informative signal, and the expert's compensation is based solely on how his report compares to the principal's signal. Dewatripont and Tirole (1999) and Gromb and Martimort (2007) are concerned with the question of when hiring two agents to acquire information may dominate hiring only one. The former has the distinguishing feature that the information acquired is verifiable.

[11] In Zermeño (2011*b*), I study the role of authority in a more general principal-expert model, where the state of nature can take a finite number of values and the expert's effort may shift the distribution of the state in addition to generating information.

[12] Other papers have studied information acquisition beyond principal-agent frameworks. For example, Bergemann and Välimäki (2002) study the implications of allowing information acquisition in a general mechanism design environment, and Persico (2000) analyzes the effect of information acquisition in auctions.

[13] In Aghion and Tirole (1997), for example, the allocation of authority effectively determines the decision rule to be implemented.

the main concepts that are used in the analysis. Section 1.3 derives the characterization of implementable outcomes (with menus) that lies at the heart of our analysis. Section 1.4 provides sufficient conditions for the validity of the first-order approach and characterizes the optimal menu contract. Section 1.5 extends the results to the case where transfer-scheme menus are not available and decisions are identifiable ex-post. Section 1.6 provides (implicit) characterizations of the Pareto-optimal outcomes that are implementable depending on whether menus are available. That section considers explicitly the relation between the decision rule to be implemented and the level of effort that can be induced, and introduces a class of information acquisition technologies for which this relation takes a particularly simple form. Section 1.7 works with these particular technologies to illustrate, in the context of Examples 1 and 2, when and why menus are valuable. Section 1.8 explores what could be the cost of motivating an expert with a contract designed for a worker. Section 1.9 concludes.

1.2 The model

There are two risk-neutral parties, a principal (she), and an expert (he). The principal has unlimited wealth, and is the residual claimant of output, $y(d, \theta)$, which is determined by a decision, $d \in D$, and a state of nature, θ. The state, $\theta \in \{\theta_1, \theta_2\}$, is *binary* and the set D is finite. The expert is protected by limited liability; his pledgeable income is ω dollars. Moreover, he does not have a direct interest in the decision, and is the only one capable of acquiring information regarding the state of nature.

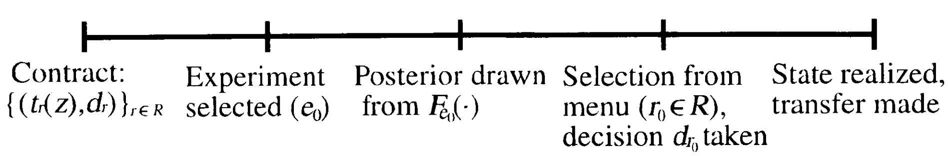
Both parties start with the same prior about the state of nature, $x_p \in (0, 1)$, where x_p is the probability that $\theta = \theta_2$. After the contract is signed, the expert chooses a non-verifiable effort cost normalized to $e \in [0, 1]$. Each level of effort produces an experiment, which is a joint distribution over a signal and the state of nature.[14] After effort is exerted, the expert privately observes the realization of the signal which results in a posterior over the state of nature. Note that the expert's effort induces a distribution over posteriors with

[14] The notion of experiments here corresponds to that in Blackwell (1953).

expectation equal to the prior (by the law of iterated expectations). Thus, the experiment corresponding to effort e_0 is fully described by a CDF over *posteriors*, $F_{e_0}(x)$, where x is the probability that $\theta = \theta_2$. In short, after incurring the effort cost, e_0, the expert privately observes a posterior, x_0, drawn from the distribution $F_{e_0}(x)$. This information is assumed to be *non-verifiable*, so the expert's *type*, in the mechanism design jargon, is $x_0 \in [0,1]$.[15] An *information acquisition technology* is a family of CDFs, $\{F_e(x)\}_{e\in[0,1]}$, where $\mathbb{E}_{F_{e_0}}[x] = x_p$, for all $e_0 \in [0,1]$.[16]

Transfer schemes are payments from the principal to the expert. They may be contingent on the contractible variable $z(d,\theta)$, which describes everything that will be observable after the state is realized and the decision is taken.[17] A *contract* specifies a menu of pairs, $\{(t_r(z), d_r)\}_{r\in R}$, where $r \in R$ is an index. The expert can choose any pair from the menu after acquiring information, but before the state is realized.[18] The selected pair, $r_0 \in R$, determines the transfer scheme through which the expert is compensated, $t_{r_0}(z)$, and the decision to be taken, $d_{r_0} \in D$.[19] The case in which contracts cannot specify transfer-scheme menus can be described by the additional constraint that $t_r(z) \equiv t_{r'}(z)$ for any $r, r' \in R$. Without loss of generality, we only consider contracts in which the expert contributes his whole pledgeable income at the beginning and always receives *non-negative* transfers.

To sum up, the principal and the expert have Von Neumann-Morgenstern preferences with Bernoulli utility functions $u_P \equiv y - t$, and $u_E \equiv t - e$ respectively (t is the transfer paid to the expert). The following timeline describes the sequence of events:



[15] The advantage of this formulation is that the expert's interim payoff becomes linear in his type.

[16] Throughout the paper, the subscripts after the expectation operator denote the distribution with respect to which the expectation is taken.

[17] The codomain of the function $z(d,\theta)$ is left unspecified because its natural specification depends on the environment under consideration. For instance, in Example 1 we have $z(d,\theta) \equiv (y(d,\theta), d, \theta)$, while in Example 2 we have $z(d,\theta) \equiv (y(d,\theta), d)$ (see the examples in section 1.1).

[18] Menus may have any number of elements; the set R is arbitrary.

[19] We restrict attention to contracts that specify deterministic decisions. Section 1.2.1 expands on this point.

1.2.1 Model discussion

There are several aspects of the model that deserve discussion. This is best done in the context of the two examples from the introduction. In these examples, the assumptions that the expert has no direct interest in decisions, and that the information that he acquires is non-verifiable are natural. In both cases, the expert is interested in the clients' choices only to the extent they may affect his compensation. The non-verifiability of information captures situations in which information is difficult to communicate unequivocally. For instance, the gambler's expert would struggle to *prove* that the *Celtics* will win with some probability.

Since the parties do not have any exogenous conflict of interest regarding decisions, the only source of friction in this model is the fact that the expert's effort is non-verifiable; if effort did not need to be motivated, decisions would always be taken efficiently given the information available. This feature allows us to isolate the interaction between the need to induce the expert to exert effort and the way decisions are taken. Even for problems where the expert might have a direct interest in decisions, the analysis in this paper serves as a useful benchmark.

The assumption that the set of states is binary provides significant tractability, since the interim mechanism design problem becomes one-dimensional (the expert's type, his posterior, is just a number). As Examples 1 and 2 illustrate, there are environments that are well described by two states. Moreover, many of the main insights that can be derived for the binary case also apply for the N-state case.[20]

In this model the parties operate under *full-commitment*. That is, the decision to be taken is determined by the contract. Thus, the present analysis establishes an upper bound for the set of outcomes that the parties could implement under alternative arrangements in which decisions were taken in some other way. Full-commitment, interpreted literally, may not be attainable in some situations. For instance, there are environments where the principal may have trouble committing not to overrule the expert's decision (Baker, Gibbons and Murphy (1999)). In Zermeño (2011*b*), I provide conditions that apply to this framework

[20] In Zermeño (2011*b*) I examine a principal-expert model with N states in which effort may affect output directly in addition to generating information.

under which any Pareto-optimal outcome implementable under full-commitment can also be implemented under the following two arrangements: 1) contracts specify a menu of transfer schemes (from which the expert can choose), and the expert has the right to irrevocably take *any* decision, $d \in D$; 2) the same, except that the principal keeps the right to overrule the expert's choice of $d \in D$. The conditions needed boil down to a restriction over the function $z(d, \theta)$: it must enable the party with the ultimate right to decide to be held accountable through transfers after unplanned decisions are taken. In particular, these conditions are met when transfer schemes may be contingent explicitly on decisions, as in the examples above.[21]

Finally, note that we have assumed that the decisions specified in contracts must be deterministic. In Zermeño (2011*b*), I show that, in general, the parties would benefit if contracts could specify lotteries over decisions. However, I show that outcomes in which stochastic decisions are taken are difficult to implement without full-commitment.[22]

1.2.2 Basic concepts

An *outcome* is a complete description of the variables that determine the size of the surplus generated by the relationship, and the way it is split between the parties. Formally,

Definition 1.2.1. *An outcome is a triplet,* $(e_0, d(x), \mathbb{T})$*, where* $e_0 \in [0,1]$ *is the effort exerted by the expert,* $d : [0,1] \to D$ *is a decision rule, mapping each posterior,* $x_0 \in [0,1]$*, to some decision,* $d(x_0) \in D$*, and* $\mathbb{T} \in \mathbb{R}$ *denotes the expert's (ex-ante) expected payment.*

Since the parties are risk neutral, e_0 and $d(x)$ fully determine the size of the surplus. Decision rules specify the decisions to be taken for all posteriors *observed* by the expert, not only those in the support of $F_{e_0}(x)$. The expert's ex-ante expected payment, $\mathbb{T}$, determines how the surplus is split between the parties.

[21] These results rely on the fact that the contracts considered in Zermeño (2011*b*) may specify transfer-scheme menus. Without menus, implementation under imperfect commitment may become more difficult.

[22] Formally, I show that any outcome with a stochastic decision rule that is implementable when the expert has the ultimate right to decide is Pareto-dominated by another outcome with a deterministic decision rule that is also implementable when the expert has the right to decide.

Since the parties may choose to opt-out of the relationship, outcomes must be *individually rational* (IR).

Definition 1.2.2. *An outcome,* $(e_0, d(x), \mathbb{T})$ *is individually rational if*

1. $\mathbb{T} - e_0 \geq \omega$.

2. $\omega + \mathbb{E}_{F_{e_0}}[E_x[y(d(x), \theta)]] - \mathbb{T} \geq 0$.

The first condition is the expert's individual rationality constraint. At the time of signing the contract, the expert's expected payoff must exceed the amount of money that he brings to the table. The second condition states that the principal must obtain a non-negative benefit from the relationship. It reflects the fact that she receives ω dollars from the expert at the beginning. The parties' outside options are normalized to zero.

Next we define *implementability.* We take into account the fact that effort and information are not verifiable, so a contract must induce the expert to willingly pick the planned choices. Note that, given a contract $\{(t_r(z), d_r)\}_{r\in R}$, in the interim stage the expert will always select an element from the menu that gives him the highest expected payoff conditional on his information. That is, for each $x_0 \in [0, 1]$, the contract actually pays the expert a conditional expected transfer, $T(x_0) \equiv \max_{r\in R} \mathbb{E}_{x_0}[t_r(z(d_r, \theta))]$. The function $T(x)$ denotes the *conditional expected payment* to the expert induced by the contract. This function will play a central role in the analysis. Then, we have:

Definition 1.2.3. *The IR outcome,* $(e_0, d(x), \mathbb{T})$*, is implementable if there exists a contract* $\{(t_r(z), d_r)\}_{r\in R}$ *(and its corresponding* $T(x)$*) such that:*

1. $e_0 \in \arg\max_{e\in[0,1]} \mathbb{E}_{F_e}[T(x)] - e$.

2. $\mathbb{E}_{F_{e_0}}[T(x)] = \mathbb{T}$.

3. *For all* $x_0 \in [0, 1]$*, there exists* $r_0 \in \arg\max_{r\in R} \mathbb{E}_{x_0}[t_r(z(d_r, \theta))]$ *s.t.* $d_{r_0} = d(x_0)$.[23]

[23] Note that we are imposing the requirement that contracts must be such that the expert's problem of choosing an element from the menu always has a solution (even for posteriors that are not in the support of $F_{e_0}(x)$). Potentially, it could be the case that the parties would be able to implement more outcomes if this constraint was relaxed by requiring that such solution exists only after observing posteriors in the support of $F_{e_0}(x)$. Zermeño (2011*b*) proves that this is actually not the case in a more general framework, so it is not the case here.

4. All transfers are non-negative.

The first condition states that the contract must induce the expert to exert the level of effort specified by the outcome. The second condition guarantees that the expert's ex-ante expected payment is indeed $\mathbb{T}$. The third condition ensures that the expert is always willing to select an element from the menu that is compatible with the decision rule specified by the outcome. The fourth condition is limited liability.

Definition 1.2.4. *A decision $d_0 \in D$ is revealing if $z(d_0, \theta_1) \neq z(d_0, \theta_2)$.*

Decisions are *revealing* when they allow the parties to identify the realization of the state ex-post. In Example 1, the parties are always able to observe who won regardless of the decision that was taken, so all decisions are revealing. In Example 2, by contrast, the only revealing decision is investment. If the project is not undertaken, the parties never observe the counterfactual. This is illustrated by Figure 1.2.1.

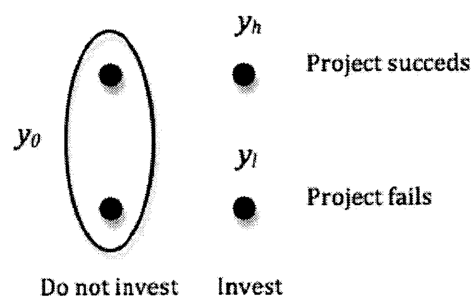


Figure 1.2.1: Example 2

1.3 Characterization of implementable outcomes

This section provides a characterization of implementable outcomes that is central to the analysis. One implication of this result is that, instead of contracts (which are relatively complicated objects), it is possible to use the expert's conditional expected payment function, $T(x)$, directly as the incentive instrument.[24] This observation simplifies the analysis considerably. Moreover, this characterization illustrates when and why it may be optimal to distort decision-making away from ex-post efficiency in environments where transfer-scheme menus are available.[25] We begin by stating the result:

Theorem 1.3.1. *The IR outcome $(e_0, d(x), \mathbb{T})$ is implementable if and only if there exists a function $T : [0,1] \to \mathbb{R}$ such that:*

1. $e_0 \in \arg\max_{e\in[0,1]} \mathbb{E}_{F_e}[T(x)] - e$.
2. $\mathbb{E}_{F_{e_0}}[T(x)] = \mathbb{T}$.
3. *$T(x)$ is convex and continuous (at the boundary).*
4. $T(0) + T'(0) \geq 0$.
5. $T(1) - T'(1) \geq 0$.[26]
6. *$T(x)$ reaches its minimum in the entire interval, $[\underline{x}(d(x)), \bar{x}(d(x))]$, where $\underline{x}(d(x)) \equiv \inf\{x_0 \in [0,1] \mid d(x_0)$ is unrevealing$\}$ and $\bar{x}(d(x)) \equiv \sup\{x_0 \in [0,1] \mid d(x_0)$ is unrevealing$\}$. This interval is empty if $d(x)$ is always revealing.*

Proof. See appendix 1.10.1. □

The key difference from the definition of implementability (definition 1.2.3) is that, in Theorem 1.3.1, the function $T(x)$ replaces contracts as the choice variable (from now on, we will refer to $T(x)$ as a contract). Before discussing the implications of this characterization, let us provide an intuition for why it is true.

Figure 1.3.1 illustrates the argument for necessity. Suppose an outcome, $(e_0, d(x), \mathbb{T})$, is implementable with some contract $\{(t_r(z), d_r)\}_{r\in R}$. The lines in Figure 1.3.1a describe the expert's expected payment conditional on his information given each particular selection

[24] This methodology corresponds to Rochet and Chone (1998)'s "dual approach", in which the agent's indirect utility function becomes the choice variable.
[25] A decision rule, $d(x)$, is *ex-post efficient* if $d(x_0) \in \arg\max_{d\in D} \mathbb{E}_{x_0}[y(d,\theta)]$ for all $x_0 \in [0,1]$.
[26] The expressions $T'(0)$ and $T'(1)$ denote the side derivatives of $T(x)$ at 0 and 1 respectively.

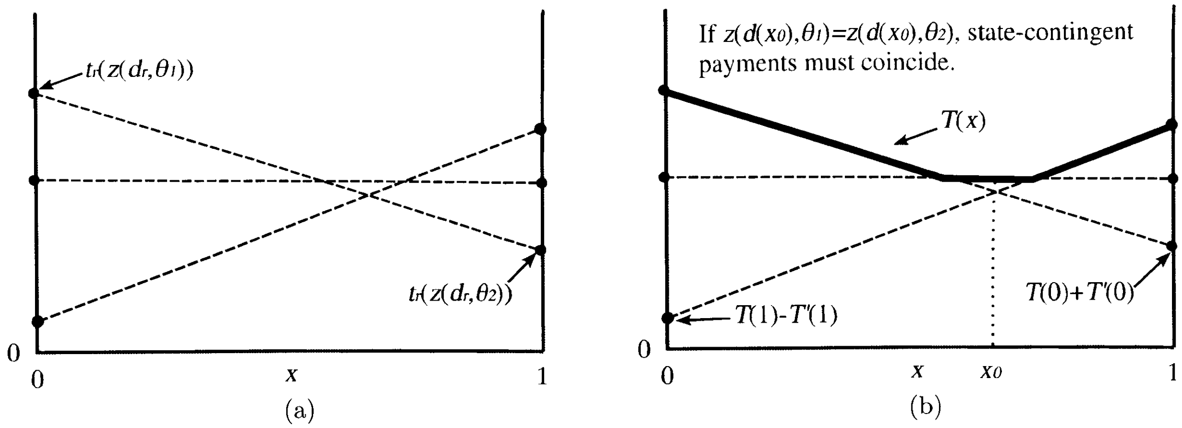


Figure 1.3.1: Theorem 1.3.1 - Necessity

from the menu. A contract boils down to a set of lines in this graph.[27] After observing a posterior, $x_0 \in [0, 1]$, the expert will select from the menu the element that corresponds to the highest line at x_0. Thus, the upper envelope, $T(x)$, describes the expert's actual expected payment conditional in his information. This function satisfies conditions (1)-(6) of Theorem 1.3.1. Conditions (1) and (2) follow directly from definition 1.2.3. Condition (3) is a consequence of the fact that $T(x)$ is the upper envelope of linear functions. Since the original contract satisfies limited liability, the points at the sides of Figure 1.3.1a must be non-negative. As Figure 1.3.1b illustrates, the expressions $T(0) + T'(0)$, and $T(1) - T'(1)$ correspond to the lowest point in each side. Thus, conditions (4) and (5) hold. Finally, condition (6) follows because, if the decision taken at some x_0 is unrevealing, the expert's payment cannot be contingent on the state. Thus, the highest line at this point must be flat (see Figure 1.3.1b).

Figure 1.3.2 illustrates the argument for sufficiency. Start with a function, $T(x)$, that satisfies conditions (1)-(6) of Theorem 1.3.1 given some outcome, $(e_0, d(x), \mathbb{T})$. As Figure 1.3.2a shows, for each $x_0 \in [0, 1]$, we can construct payments, $t_{x_0}(z(d(x_0), \theta_i))$, by using a tangent line to $T(x)$ at x_0. The payments specified by the transfer scheme, $t_{x_0}(z)$, for $z \neq z(d(x_0), \theta_i)$ can be any non-negative number. The contract $\{(t_{x_0}(z), d(x_0))\}_{x_0 \in [0,1]}$ built

[27] Although the contract depicted in the figure has only three elements, in general it may have an arbitrary number of elements.

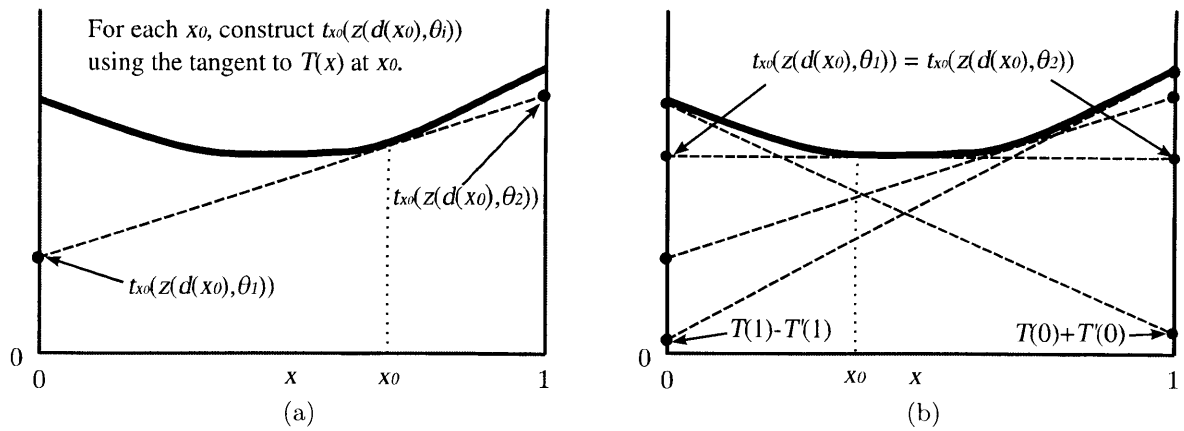


Figure 1.3.2: Theorem 1.3.1 - Sufficiency

in this way implements $(e, d(x), \mathbb{T})$. Condition (6) guarantees that these transfers are well defined. If, for some $x_0 \in [0,1]$, we have that $z(d(x_0), \theta_1) = z(d(x_0), \theta_2)$, condition (6) implies that there exists a flat tangent to $T(x)$ at x_0, so the payments constructed for this value of z actually coincide (see Figure 1.3.2b). Since $T(x)$ is convex, conditions (4) and (5) imply that all transfers are non-negative (see Figure 1.3.2b). In addition, convexity implies that condition (3) of definition 1.2.3 holds, and that the conditional expected payment function associated to the constructed contract coincides with our starting $T(x)$. Thus, conditions (1) and (2) here imply that the first two conditions of definition 1.2.3 are also satisfied.

Theorem 1.3.1 uncovers several insights about this problem. In particular, it illustrates that, when transfer-scheme menus are available, optimal distortions to decision-making are purely driven by differences in the amount of information about the state that different decisions reveal ex-post. The first condition in Theorem 1.3.1 establishes that the expert's incentives to acquire information are completely determined by the function $T(x)$. Thus, given a pair $(d(x), \mathbb{T})$, the parties need to choose $T(x)$ to induce their preferred level of effort subject to conditions (2)-(6) in the theorem. Note that the decision rule, $d(x)$, only restricts $T(x)$ through condition (6). In fact, if all decisions are revealing, $d(x)$ does not restrict $T(x)$ at all. When this is the case, any decision rule can be implemented without hindering the parties' ability to induce effort, so there is no need to distort decision-making.

The presence of unrevealing decisions, however, introduces a trade-off between the levels of effort that can be induced and the decision rule implemented. Indeed, as reflected by condition (6), taking unrevealing decisions restricts the shape that $T(x)$ can take, limiting the parties' ability to induce effort. Therefore, in this case, in order to foster information acquisition, it may be optimal to distort decision-making. Optimal distortions always take the form of revealing decisions being taken when unrevealing ones would be ex-post efficient. The following proposition formalizes this point:

Proposition 1.3.1. *If contracts may specify transfer-scheme menus, any implementable outcome, $(e_0, \tilde{d}(x), \mathbb{T})$, is Pareto-dominated (at least weakly) by another implementable outcome, $(e_0, d(x), \mathbb{T})$, in which $d(x)$ is such that, for some interval $[\underline{m}, \bar{m}]$,*

1. *$d(x_0) \in \arg\max_{d\in D} \mathbb{E}_{x_0}[y(d,\theta)]$ for all $x_0 \in [\underline{m}, \bar{m}]$.*

2. *$d(x_0) \in \arg\max_{d\in D^R} \mathbb{E}_{x_0}[y(d,\theta)]$ for all $x_0 \notin [\underline{m}, \bar{m}]$, where $D^R \subseteq D$ is the set of decisions that are revealing.*[28]

Proof. Suppose that the outcome, $(e_0, \tilde{d}(x), \mathbb{T})$, is implementable with a function $T(x)$. Since $T(x)$ is convex, we can define $[\underline{m}, \bar{m}] = \arg\min_{x\in[0,1]} T(x)$. Then $T(x)$ also implements $(e_0, d(x), \mathbb{T})$, where $d(x)$ is as described in the corollary given $[\underline{m}, \bar{m}]$ (by construction, $T(x)$ satisfies condition (6) in Theorem 1.3.1 given $d(x)$). This outcome makes both parties at least weakly better off. □

In Example 1 in section 1.1, the expert should always be induced to decide efficiently. In Example 2, however, the expert may optimally be induced to undertake the project too frequently; it can never be optimal to undertake the project too rarely.

Theorem 1.3.1 enables us to assess what is the cost of information being non-verifiable. If information were verifiable and transfer schemes could depend directly on posteriors, conditions (1) and (2) of Theorem 1.3.1 and the limited-liability constraint, $T(x) \geq 0$ for all

[28] I use $[\underline{m}, \bar{m}]$ as opposed to $[\underline{x}, \bar{x}]$, as in Theorem 1.3.1, because it may be optimal for a contract, $T(x)$, to reach its minimum in a strict superset of $[\underline{x}, \bar{x}]$.

$x \in [0, 1]$, would completely characterize implementable outcomes.[29] In this case, distorting decision-making away from ex-post efficiency would never be necessary. When information is non-verifiable (and transfers may only depend on variables observable ex-post), conditions (3)-(6) of Theorem 1.3.1 must be included to characterize implementable outcomes. Condition (3), convexity, reflects the intuitive observation that having more non-verifiable information can never be detrimental to the expert. Conditions (4)-(6) are needed because, here, the expert's compensation depends on his information only indirectly. Conditions (4) and (5) illustrate that now limited liability imposes stronger restrictions than in the case with verifiable information. The reason is that here transfers must be non-negative for every realization of the state, and not only in expectation. As discussed above, condition (6) introduces a reason to distort decision-making when information is non-verifiable.

The insights we have derived in this section are quite general. In Zermeño (2011*b*), I extend Theorem 1.3.1 to the case where there is a finite number of states, and where the expert's effort may affect output directly in addition to generating information. Furthermore, this result could also be generalized to environments in which the parties are expected utility maximizers with Bernoulli utility functions strictly increasing in money. Without risk-neutrality, the main difference would be that the surplus generated by the relationship could no longer be described by a decision rule and a level of effort; the structure of compensation would also affect total value.[30] Nevertheless, Theorem 1.3.1 relies on the fact that the function $T(x)$ contains all the relevant information available in a contract (as Figure 1.3.2a illustrates, the whole contract can be constructed from $T(x)$), and this is true regardless of preferences.[31]

[29] This situation would correspond to the standard moral hazard framework, as studied by Holmström (1979) or Kim (1995), for example. The agent's output (in this case information) is directly contractible.

[30] Differences in the parties' marginal utilities of income would make the *level* of compensation matter; differences in their attitudes towards risk would make the *variability* of compensation matter.

[31] Therefore, a similar characterization can be used to study the case where the principal is risk-neutral, the expert is risk-averse and there is no limited liability, as in Lambert (1986), Demski and Sappington (1987) and Malcomson (2009). The analysis is actually quite similar, since both, limited liability and risk-aversion, basically set limits to how convex $T(x)$ can be (with risk-aversion, steeper slopes mean more value lost).

1.4 Simple contracts

This section provides a sufficient condition under which any implementable outcome can be implemented with contracts that specify a significantly reduced set of alternatives for the expert. Formally, an *alternative* is a vector specifying the payment that the expert will receive after each of the states is realized. That is, when the expert makes a selection, $r_0 \in R$, from the menu, he is actually choosing the alternative, $(t_{r_0}(z(d_{r_0}, \theta_i)))_{i=1}^2$. Note that different selections from the menu may correspond to the same alternative, and different alternatives may be induced by selections that have the same transfer scheme. An alternative is *undominated* if it is strictly preferred by the expert after observing some posterior. The results in this section will be phrased in terms of the following notion of *simple* contracts:

Definition 1.4.1. *A contract is simple if it specifies at most three undominated alternatives. The third alternative (whenever it applies) pays the same amount regardless of the state.*

The term *simple* is used because, to be effective, a contract must provide at least two undominated alternatives. Here, we are only including the possibility of having a third riskless alternative. As an illustration, in the gambler's example a simple contract could give the expert the right to choose between two pre-specified bets, where each bet is a payment contingent solely on who won. Note that the function, $T(x) \equiv \max_{r \in R} \mathbb{E}_x[t_r(z(d_r, \theta))]$, induced by a simple contract may change its slope at most twice (see Figure 1.3.1).

Given the information acquisition technology, $\{F_e(x)\}_{e \in [0,1]}$, define $\mathcal{I}(x_0; e) \equiv \int_0^{x_0} F_e(x)dx$. Then, the following condition is sufficient for simple contracts to be optimal:

Assumption 1.4.1. *For all $x_0 \in [0, 1]$, $\mathcal{I}(x_0; e)$ is non-decreasing and concave in e.*

This condition is equivalent to requiring that, for any convex and continuous $T(x)$, the function, $\mathbb{E}_{F_e}[T(x)]$, is non-decreasing and concave in e. Indeed, if $\Delta(x_0)$ is a subderivative of a given convex function, $T(x)$, at x_0 for all $x_0 \in [0,1]$,[32] then

$$\mathbb{E}_{F_e}[T(x)] \equiv T(1) - (1 - x_p)\Delta(1) + \int_0^1 \mathcal{I}(x; e) d\Delta(x), ^{33} \tag{1.4.1}$$

where $\int_0^1 \mathcal{I}(x;e)d\Delta(x)$ denotes the Riemann-Stieltjes integral of $\mathcal{I}(x;e)$ with respect to $\Delta(x)$ over $[0,1]$ (it is well defined because $\mathcal{I}(x;e)$ is continuous in x and $\Delta(x)$ is non-decreasing).[34] The requirement that, for each $x_0 \in [0,1]$, $\mathcal{I}(x_0;e)$ is non-decreasing in e is intuitive. It is equivalent to asking that, if $e \geq e'$, then the experiment generated by e must be more informative (in Blackwell's sense) than the experiment generated by e'.[35] The concavity requirement corresponds to a strong notion of diminishing returns to information acquisition.[36] I shall return to Assumption 1.4.1 later on. Now, we can state the main result of this section:

Theorem 1.4.1. *Under Assumption 1.4.1, simple contracts minimize the expected transfer that needs to be paid to the expert in order to implement any given pair, $(e_0, d(x))$. Moreover, if $d(x)$ is always revealing, the contract that attains the minimum needs to specify only two undominated alternatives.*

Proof. See appendix 1.10.3. □

---

[32] The subderivative of a convex function $T : [0,1] \to \mathbb{R}$ at a point x_0 is a scalar, c, such that $T(x_0) + c(x - x_0) \leq T(x)$ for all $x \in [0,1]$.

[33] The equivalence follows from:

1. $\mathbb{E}_{F_e}[T(x)] = T(0)F_e(0) + \int_0^1 T(x)dF_e(x)$,
2. $\int_0^1 T(x)dF_e(x) + \int_0^1 F_e(x)\Delta(x)dx = T(1) - T(0)F_e(0)$,
3. $\int_0^1 F_e(x)\Delta(x)dx + \int_0^1 \mathcal{I}(x;e)d\Delta(x) = \mathcal{I}(1;e)\Delta(1) - \mathcal{I}(0;e)\Delta(0) = (1 - x_p)\Delta(1)$,

where all the integrals are Riemann-Stieltjes integrals.

[34] The argument for sufficiency is straight forward. For necessity note that, if there exists x_0 such that $\mathcal{I}(x_0;e)$ is not non-decreasing (or concave) in e, we can pick a convex $T(x)$ that changes slope only once at x_0. Then, $\mathbb{E}_{F_e}[T(x)] \equiv T(1) - (1 - x_p)\Delta(1) + \delta\mathcal{I}(x_0, e)$, where $\delta > 0$ is the increment in slope. This function is not non-decreasing (or concave) in e.

[35] The requirement that, for all $e \geq e'$ and $x_0 \in [0,1]$, $\int_0^{x_0} F_e(x)dx \geq \int_0^{x_0} F_{e'}(x)dx$ is equivalent to requiring that, for all $e \geq e'$, $F_e(x)$ is a mean-preserving spread of $F_{e'}(x)$ (see Rothschild and Stiglitz (1970)).

[36] Not all reasonable information acquisition technologies satisfy this requirement. Radner and Stiglitz (1984) provide sufficient conditions under which the marginal value of a small amount of information is zero (and, thus, the concavity requirement is violated). Chade and Schlee (2002) show that these conditions are rather restrictive, and provide examples where the concavity requirement is satisfied.

The rest of the section provides the intuition behind this result. Before we get there, however, the following corollary establishes its main implication.

Corollary 1.4.1. *Under Assumption 1.4.1, any implementable outcome is implementable with a simple contract. Moreover, if all decisions are revealing, contracts need to specify at most two undominated alternatives.*

Proof. Suppose the outcome, $(e_0, d(x), \mathbb{T})$, is implementable. By Theorem 1.4.1, there exists a simple contract, $\{(\hat{t}_r(z), d_r)\}_{r\in R}$, that implements $(e_0, d(x), \hat{\mathbb{T}})$, with $\hat{\mathbb{T}} \leq \mathbb{T}$. If $d(x)$ is always revealing, this simple contract specifies only two undominated alternatives. Then, the contract $\{(t_r(z), d_r)\}_{r\in R}$, where $t_r(z) \equiv \hat{t}_r(z) + \mathbb{T} - \hat{\mathbb{T}}$ for all $r \in R$, is also simple and implements $(e_0, d(x), \mathbb{T})$. □

Therefore, optimal menu contracts need not be complicated objects. Considering contracts that specify only a reduced set of alternatives is often enough.

Theorem 1.4.1 states that simple contracts solve Program 1.4.1 below. In this program we choose a function $T(x)$ to minimize the expert's expected payment subject to implementing a given level of effort, e_0, and decision rule, $d(x)$. That is, $T(x)$ must satisfy conditions (1) and (3)-(6) in Theorem 1.3.1, given e_0 and the interval, $[\underline{x}(d(x)), \bar{x}(d(x))]$.

Program 1.4.1 Cost-minimization problem given e_0 and $[\underline{x}, \bar{x}] = [\underline{x}(d(x)), \bar{x}(d(x))]$

$$\min_{T(x)} \ \mathbb{E}_{F_{e_0}}[T(x)],$$

subject to:

1. $T(x)$ is convex and continuous (at the boundary).
2. $\mathbb{E}_{F_{e_0}}[T(x)] - e_0 \geq \mathbb{E}_{F_e}[T(x)] - e$ for all $e \neq e_0$.
3. $T(0) + T'(0) \geq 0$.
4. $T(1) - T'(1) \geq 0$.
5. $T(x)$ reaches its minimum at every point in the (possibly empty) interval $[\underline{x}, \bar{x}]$.

This problem is non-standard in the sense that the choice variable is a convex function. Fortunately, it can be expressed in more familiar terms. We rely on the fact that a function, $T(x)$, is convex and continuous if and only if it can be expressed as $T(x_0) \equiv T(0) + \int_0^{x_0} \Delta(x)dx$, where for each $x_0 \in [0,1]$, $\Delta(x_0)$ is a subderivative of $T(x)$ at x_0 (note that $\Delta(x)$ must be non-decreasing). Then, as we will see below, it is possible to replace $T(x)$ with $\Delta(x)$ as the choice variable, and rewrite Program 1.4.1 in terms only of the *increment* of $\Delta(x)$, $d\Delta(x)$. The advantage of doing this is that the reformulated problem is linear in the increment, $d\Delta(x)$, and the convexity constraint turns into a non-negativity constraint on $d\Delta(x)$. Once here, using standard linear programing techniques (generalized to apply to this problem), it can be proved that, under Assumption 1.4.1, simple contracts, in which $d\Delta(x)$ is positive at most twice, solve this problem.

In what follows we will show how to express Program 1.4.1 in terms of only of $d\Delta(x)$, and discuss the role of Assumption 1.4.1. The first step is to note that any convex function $T(x)$ can be characterized by its (non-decreasing) subderivative, $\Delta(x)$, and its intercept, $T(1) \in \mathbb{R}$. Furthermore, by integrating by parts twice (as in equation 1.4.1), it is possible to rewrite Program 1.4.1 in terms of the *increment* of $\Delta(x)$, $d\Delta(x)$, and of two *levels*: the level of $T(x)$ (determined by $T(1)$), and the level of $\Delta(x)$ (determined by $\Delta(1)$).[37] Then, Program 1.4.1 becomes Program 1.4.2:

Program 1.4.2 Program 1.4.1 reformulated

$$\min_{\Delta(x),\ T(1)} \quad T(1) - (1 - x_p)\Delta(1) + \int_0^1 \mathcal{I}(x; e_0) d\Delta(x),$$

subject to:

1. $\Delta(x)$ is non-decreasing.
2. $\int_0^1 [\mathcal{I}(x; e_0) - \mathcal{I}(x; e)] d\Delta(x) \geq e_0 - e$ for all $e \neq e_0$.
3. $T(1) - \int_0^1 (1 - x) d\Delta(x) \geq 0$.[38]
4. $T(1) - \Delta(1) \geq 0$.
5. $\Delta(x) = 0$ in the (possibly empty) interval $[\underline{x}, \bar{x}]$.

[37] The term *level* is used to describe the "height" of a function. For example, if the function $\Delta(x)$ is fixed, the level of $T(x)$ is determined by its value at any point; in particular, by $T(1)$.

[38] This expression is equivalent to $T(0) + \Delta(0) \geq 0$. The equivalence follows from:

The variables, $d\Delta(x)$, $T(1)$ and $\Delta(1)$ can be chosen independently. For example, by shifting up the schedule $T(x)$, we can increase $T(1)$ without affecting $\Delta(x)$. Similarly, by shifting up $\Delta(x)$ we can increase $\Delta(1)$ without affecting $d\Delta(x)$ or $T(1)$. The next step is to pin down the levels, $T(1)$ and $\Delta(1)$ using the constraints, and express Program 1.4.1 only in terms of the increment, $d\Delta(x)$. For this task, there are four cases that must be considered: 1) the interval, $[\underline{x}, \bar{x}]$, is empty; 2) $0 = \underline{x} \leq \bar{x} < 1$; 3) $0 < \underline{x} \leq \bar{x} = 1$; 4) $0 < \underline{x} \leq \bar{x} < 1$.

In the first case constraints (3) and (4) of Program 1.4.2 pin down the optimal levels of $T(x)$ and $\Delta(x)$. Since increasing $\Delta(1)$ decreases the objective function, it is optimal to shift up the schedule $\Delta(x)$ to the point where $\Delta(1) = T(1)$. Substituting this value into the objective, we can see that, optimally, the schedule $T(x)$ should be shifted down to the point where $T(1) = \int_0^1 (1 - x) d\Delta(x)$. Thus, Program 1.4.2 becomes Program 1.4.3 below.

Program 1.4.3 Case 1: $[\underline{x}, \bar{x}]$ is empty

$$\min_{\Delta(x)} \int_0^1 [\mathcal{I}(x; e_0) + x_p(1 - x)] d\Delta(x),$$

subject to:

1. $\Delta(x)$ is non-decreasing.
2. $\int_0^1 [\mathcal{I}(x; e_0) - \mathcal{I}(x; e)] d\Delta(x) \geq e_0 - e$ for all $e \neq e_0$.

When the interval $[\underline{x}, \bar{x}]$ is not empty, the level of $\Delta(x)$ is pinned down by constraint (5) in Program 1.4.2. In particular, we must have $\Delta(1) = \int_{\bar{x}}^1 d\Delta(x)$. Substituting this value in Program 1.4.2, we can see that, in the second case, constraint (4) implies constraint (3). Thus, when $\underline{x} = 0$, it is optimal to have $T(1) = \int_{\bar{x}}^1 d\Delta(x)$, and Program 1.4.2 takes a very similar form to Program 1.4.3. The only differences are that the objective function becomes $\int_{\bar{x}}^1 [\mathcal{I}(x; e_0) + x_p] d\Delta(x)$, and that all the integrals run from $\bar{x}$ to 1. The third case is analogous to the second. Here constraint (3) in Program 1.4.2 implies constraint (4),

---

1. $\Delta(0) = \Delta(1) - \int_0^1 d\Delta(x)$,
2. $T(0) = T(1) - \int_0^1 \Delta(x) dx$,
3. $\int_0^1 \Delta(x) dx = -\int_0^1 x d\Delta(x) + \Delta(1)$.

and it is optimal to make $T(1) = \int_0^{\underline{x}}(1-x)d\Delta(x)$. Now, the objective function becomes $\int_0^{\underline{x}}[\mathcal{I}(x;e_0)+1-x]d\Delta(x)$, and all integrals run from 0 to $\underline{x}$.

The fourth case is qualitatively different. Here, there is no loss of generality in assuming that constraints (3) and (4) in Program 1.4.2 both hold with equality. The reason is that increasing only the value $\Delta(1)$ (while keeping the rest of the schedule $\Delta(x)$ fixed), only affects constraint (4), and decreasing the value of $\Delta(0)$ in a similar way only affects constraint (3).[39] Taking this into account, Program 1.4.2 becomes Program 1.4.4 below.

Program 1.4.4 Case 4: $0 < \underline{x} \le \bar{x} < 1$

$$\min_{\Delta(x)} \int_0^{\underline{x}} \mathcal{I}(x;e_0)d\Delta(x) + \int_{\bar{x}}^1 [\mathcal{I}(x;e_0)+x_p]d\Delta(x),$$

subject to:

1. $\Delta(x)$ is non-decreasing.
2. $\int_0^{\underline{x}}[\mathcal{I}(x;e_0)-\mathcal{I}(x;e)]d\Delta(x) + \int_{\bar{x}}^1[\mathcal{I}(x;e_0)-\mathcal{I}(x;e)]d\Delta(x) \ge e_0 - e$ for all $e \neq e_0$.
3. $\int_0^{\underline{x}}(1-x)d\Delta(x) = \int_{\bar{x}}^1 x d\Delta(x)$.

As we wanted to show, Programs 1.4.3 and 1.4.4 are linear and expressed only in terms of the increment, $d\Delta(x)$, which is restricted to be non-negative by constraint (1). Unfortunately, finding a general solution to these problems is difficult; the second constraint actually specifies a continuum of constraints from which the set that will bind depends on the specific information acquisition technology.[40] Assumption 1.4.1 guarantees that these constraints can be replaced by a single local condition. In this sense, this assumption is analogous to the conditions provided in Rogerson (1985) for the validity of the first-order approach in the classic moral hazard model. In Rogerson (1985), the agent's effort is a direct input to the production process and $x \in [0,1]$ is output. Rogerson's convexity of the distribution func-

[39] For example, consider what happens in Program 1.4.2 when increasing the value of $\Delta(1)$ by δ, while keeping the rest of $\Delta(x)$ and $T(1)$ fixed. The objective increases by $\delta(-(1-x_p)+\mathcal{I}(1;e_0)) = 0$. The first constraint is relaxed. The left hand side of the second constraint changes by $\delta(\mathcal{I}(1;e_0)-\mathcal{I}(1;e)) = 0$ for any $e \neq e_0$. The third constraint changes by $\delta(1-1)=0$. The fourth constraint becomes tighter by δ. Thus, we can always increase $\Delta(1)$ (without affecting the outcome) to the point where constraint (4) binds.

[40] In binary effort models, such as Inderst and Klein (2007), we do not have this problem, and simple contracts also solve Program 1.4.1.

tion condition (CDFC) is equivalent to the requirement that, for any non-decreasing $T(x)$, $\mathbb{E}_{F_e}[T(x)]$ is concave in e.[41] When x is output, the assumption that $F_e(x)$ satisfies the monotone likelihood ratio property (MLRP) is natural. This property, together with the CDFC, guarantees that optimal compensation schemes are actually non-decreasing. When effort is about acquiring information (and x is a posterior), we lose the MLRP, so we cannot require the function $T(x)$ to be non-decreasing. However, if information is non-verifiable, $T(x)$ must be convex and continuous, which is the fact that we exploit in Assumption 1.4.1.

Assumption 1.4.1 is not necessary for Theorem 1.4.1 to hold. Appendix 1.10.4 provides an example where it is not satisfied, but where simple contracts are still optimal. Whether Theorem 1.4.1 holds without Assumption 1.4.1 is an open question. I have not found an example where simple contracts are not optimal.

1.5 Single-transfer-scheme contracts with identifiable decisions

In this section we analyze the case in which contracts may only specify a single transfer scheme. That is, contracts are now menus of the form $\{(t(z), d_r)\}_{r\in R}$. The main insight that comes out of the analysis is that, without transfer-scheme menus, contracts can only induce different *alternatives* if the prescribed decision changes.[42] In fact, if $z(d, \theta)$ is such that the decisions taken can be identified at the end, this is the only difference between having and not having access to transfer-scheme menus. Proposition 1.5.1 provides a characterization of implementable outcomes that formalizes these observations. Moreover, we will see that the result that (under Assumption 1.4.1) simple contracts can implement any implementable outcome extends to the case where transfer-scheme menus are not available.

[41] The family of distributions $\{F_e(x)\}_{e\in[0,1]}$ satisfies the CDFC if for al $x_0 \in [0, 1]$, $F_e(x_0)$ is convex in e. Take any non-decreasing $T(x)$ for which the Riemann-Stieltjes integral with respect to $F_e(x)$ over [0,1] exists for any e. Then, integrating by parts, $\mathbb{E}_{F_e}[T(x)] \equiv T(1) - \int_0^1 F_e(x)dT(x)$. The equivalence follows from this equation.

[42] Recall that an *alternative* is a vector specifying the payment that the expert will receive after each realization of the state.

We will restrict attention to environments in which decisions can be identified after they are taken. In other words, the variable $z(d, \theta)$ must satisfy:

Assumption 1.5.1. *Decisions are identifiable. That is, $d \neq d'$ implies $z(d, \theta) \neq z(d', \theta')$ for any θ and θ'.*

This assumption is natural in many situations. For instance, it is satisfied in Examples 1 and 2 in section 1.1. In environments where decisions are identifiable, it is less likely that transfer-scheme menus are valuable, since the set of outcomes that are implementable without menus becomes larger. The specific role of Assumption 1.5.1 will be described in more detail after the following proposition:

Proposition 1.5.1. *Under Assumption 1.5.1, the IR outcome, $(e_0, d(x), \mathbb{T})$, is implementable with a single-transfer-scheme contract if and only if there exists $T : [0, 1] \rightarrow \mathbb{R}$ such that:*

1. $e_0 \in \arg\max_{e \in [0,1]} \mathbb{E}_{F_e}[T(x)] - e$.
2. $\mathbb{E}_{F_{e_0}}[T(x)] = \mathbb{T}$.
3. *$T(x)$ is convex and continuous (at the boundary).*
4. $T(0) + T'(0) \geq 0$.
5. $T(1) - T'(1) \geq 0$.
6. *$T(x)$ reaches its minimum in the entire interval, $[\underline{x}(d(x)), \bar{x}(d(x))]$, where $\underline{x}(d(x))$ and $\bar{x}(d(x))$ are as defined in Theorem 1.3.1.*
7. *If $d(x_0) = d(x')$, then $T(\alpha x' + (1 - \alpha)x_0) = \alpha T(x') + (1 - \alpha)T(x_0)$, for all $\alpha \in [0, 1]$.*

Proof. See appendix 1.10.5. □

The intuition behind this result is analogous to that of Theorem 1.3.1. The only innovation here with respect to our original characterization is condition (7). It reflects the fact that, when a contract specifies a single transfer scheme, each decision cannot correspond to more than one alternative. Indeed, suppose a contract, $\{(t(z), d_r)\}_{r\in R}$, implements an outcome $(e_0, d(x), \mathbb{T})$. If $d(x_0) = d(x')$, then we must have $z(d(x_0), \theta_i) = z(d(x'), \theta_i)$, and thus, $t(z(d(x_0), \theta_i)) = t(z(d(x'), \theta_i))$ for $i = 1, 2$. This implies that the same line must be tangent to $T(x) \equiv \max_{r\in R} \mathbb{E}_x[t(d_r, \theta)]$ at x_0 and x' (see Figure 1.3.1). Thus, by convexity, $T(x)$ must be linear between x_0 and x'.

If Assumption 1.5.1 were not satisfied, there would exist $d_0 \neq d'$ with $z(d_0, \theta_i) = z(d', \theta_j)$ for some $\theta_i, \theta_j \in \{\theta_1, \theta_2\}$. Suppose a decision rule had $d(x_0) = d_0$ and $d(x') = d'$. Then, in order to implement $d(x)$, a contract, $\{(t(z), d_r)\}_{r\in R}$, would need to induce best alternatives at x_0 and x', $r^*(x_0)$ and $r^*(x')$, satisfying $t(z(d_{r^*(x_0)}, \theta_i)) = t(z(d_{r^*(x')}, \theta_j))$.[43] This fact would impose additional constraint over the shape that $T(x) \equiv \max_{r\in R} \mathbb{E}_x[t(d_r, \theta)]$ could take. As we can see, in order to characterize implementable outcomes without Assumption 1.5.1, one would need to keep track of the relationship between each pair of decisions.

Condition (7) in Proposition 1.5.1 boils down to the requirement that the function $T(x)$ that implements an outcome, $(e_0, d(x), \mathbb{T})$, can only change its slope at posteriors where $d(x)$ specifies that the decision taken will change. Formally, each decision rule, $d(x)$, now induces a partition, $\mathcal{P}(d(x))$, on the set $[0, 1]$ given by:

$$\mathcal{P}(d(x)) \equiv \left\{ x_0 \in [0,1] \mid \{d(x') \in D \mid x' < x_0\} \bigcap \{d(x') \in D \mid x' > x_0\} = \varnothing \right\}.^{44}$$

Condition (7) states that the slope of $T(x)$ can only change at the points in $\mathcal{P}(d(x))$. Proposition 1.5.2 relies on this observation.

Proposition 1.5.2. *Under Assumptions 1.4.1 and 1.5.1, if contracts may only specify a single transfer scheme, any implementable outcome is implementable with a simple contract. Moreover, if all decisions are revealing, contracts need to specify at most two undominated alternatives.*

Proof. Here we need to solve Program 1.4.1 with one additional constraint: $T(x)$ can only change its slope at points in $\mathcal{P}(d(x))$. The proof that simple contracts solve this modified

[43] $r^*(x) \in \arg\max_{r\in R} \mathbb{E}_x[t(z(d_r, \theta))]$ for all $x \in [0, 1]$.
[44] Note that, since D is finite, $\mathcal{P}(d(x))$ must be finite.

program is analogous to the proof of Theorem 1.4.1. The modified version of Program 1.4.1 can still be expressed as a linear program only in terms of the increment of the subderivative of $T(x)$, $d\Delta(x)$. The only change is that now we only get to choose a finite number of variables (the size of the increment at the pre-specified points), as opposed to a continuum as before. Simple contracts still correspond to extreme points in the feasible set. $\square$

This result enables us to compare, in the following section, the Pareto-optimal outcomes that can be implemented depending on whether transfer-scheme menus are available.

1.6 Pareto-optimal outcomes

This section derives an (implicit) upper bound for the level of effort that the parties can implement given a decision rule, the expert's pledgeable income, ω, and the net expected utility that the expert will derive from the relationship. This upper bound enables us to (implicitly) characterize Pareto-optimal outcomes with and without menus, and to illustrate when and why menus are valuable.

The subsequent analysis relies on the results in the previous sections. Indeed, we will see that the value functions of the cost-minimization problems that we solved in Theorem 1.4.1 and Proposition 1.5.2 can be used to implicitly define the upper bound for implementable effort. Let $U_0 \in \mathbb{R}_+$ be the expert's net expected utility, and $V^j(d(x), e_0)$ denote the value function of the cost-minimization problem (given $d(x)$ and e_0), depending on whether transfer-scheme menus are available ($j = M$) or not ($j = S$). Note that outcomes can alternatively be defined in terms of the expert's net expected utility, U_0, instead of his ex-ante expected payment, $\mathbb{T}$ (given e_0, the expression $U_0 \equiv \mathbb{T} - e_0 - \omega$ defines a one-to-one mapping between U_0 and $\mathbb{T}$). Then, by the definition of the cost-minimization problems, the IR outcome, $(e_0, d(x), U_0)$, is implementable given $j \in \{M, S\}$ if and only if $U_0 + e_0 + \omega \geq V^j(d(x), e_0)$ (note that $\mathbb{T} \equiv U_0 + e_0 + \omega$). We will show that, under Assumptions 1.4.1 and 1.5.1, the inequality, $U_0 + e_0 + \omega \geq V^j(d(x), e_0)$, implicitly defines the highest level of effort that can be implemented given $d(x)$, U_0 and ω.

Under Assumptions 1.4.1 and 1.5.1, the expression $V^j(e_0, d(x))$ has already been derived (see appendix 1.10.3). Under these assumptions, $V^j(e_0, d(x))$, can be expressed in terms of the locations where the function, $T(x)$, that solves the cost-minimization problem (given e_0, $d(x)$ and $j \in \{M, S\}$) changes its slope. Recall that, if $j = M$ and $d(x)$ is always revealing, the slope of $T(x)$ optimally changes once at some point in $X^M(d(x)) \equiv [0, 1]$. If $j = M$, but $d(x)$ specifies unrevealing decisions, the slope of $T(x)$ may optimally change twice, once in $\underline{X}^M(d(x)) \equiv [0, \underline{x}(d(x))]$, and once in $\bar{X}^M(d(x)) \equiv [\bar{x}(d(x)), 1]$. If $j = S$, however, we must include the additional constraint that $T(x)$ can only change its slope at the points in $\mathcal{P}(d(x))$ (constraint (7) in Proposition 1.5.1). Thus, in this case, changes in the slope of the optimal $T(x)$ can only take place in $X^S(d(x)) \equiv X^M(d(x)) \bigcap \mathcal{P}(d(x))$ if all decisions are revealing, and in $\underline{X}^S(d(x)) \equiv \underline{X}^M(d(x)) \bigcap \mathcal{P}(d(x))$ and $\bar{X}^S(d(x)) \equiv \bar{X}^M(d(x)) \bigcap \mathcal{P}(d(x))$ if there are unrevealing decisions. Then, we have:

Proposition 1.6.1. *Let the index, $j \in \{M, S\}$, describe whether contracts may specify transfer-scheme menus (M), or only a single transfer scheme (S). Then, under Assumptions 1.4.1 and 1.5.1, the IR outcome $(e_0, d(x), U_0)$ is implementable (given j) if and only if:*

1. *If $d(x_0)$ is revealing for all $x_0 \in [0, 1]$,*

$$U_0 + \omega + e_0 \geq \left(\max_{x \in X^j(d(x))} \frac{\mathcal{I}_{e^-}(x; e_0)}{\mathcal{I}(x; e_0) + x_p(1 - x)} \right)^{-1} . \tag{1.6.1}$$

[45]

2. *If $d(x_0)$ is unrevealing for some $x_0 \in [0, 1]$,*

$$U_0 + \omega + e_0 \geq \left(\max_{\underline{m} \in \underline{X}^j(d(x)),\ \bar{m} \in \bar{X}^j(d(x))} \frac{\bar{m}\mathcal{I}_{e^-}(\underline{m}; e_0) + (1 - \underline{m})\mathcal{I}_{e^-}(\bar{m}; e_0)}{\bar{m}\mathcal{I}(\underline{m}; e_0) + (1 - \underline{m})(\mathcal{I}(\bar{m}; e_0) + x_p)} \right)^{-1} . \tag{1.6.2}$$

Proof. It follows from the previous discussion. Note that $U_0 + \omega + e_0 \equiv \mathbb{T}$, and the expressions on the right are the value functions of the cost-minimization problems. See appendix 1.10.3. □

45 The term $\mathcal{I}_{e^-}(x_0; e_0) \equiv \lim_{e \to e_0^-} \frac{\mathcal{I}(x_0; e_0) - \mathcal{I}(x_0; e)}{e_0 - e}$ is the left-derivative of $\mathcal{I}(x; e)$ with respect to e at (x_0, e_0), which is always well defined under Assumption 1.4.1. Thus, this characterization applies for outcomes with $e_0 > 0$. Any IR outcome with $e_0 = 0$ is implementable.

The maximizers in expressions 1.6.1 and 1.6.2 correspond to the locations where the cost-minimizing function, $T(x)$, changes its slope (given $d(x)$ and e_0). Note that the decision rule, $d(x)$, only affects these expressions by changing the set of points in which the slope of $T(x)$ is allowed to change. The following lemma demonstrates that expressions 1.6.1 and 1.6.2 implicitly define an upper bound for the level of effort that can be implemented given the expert's expected net utility, U_0, initial endowment, ω, and decision rule, $d(x)$.

Lemma 1.6.1. *Under Assumptions 1.4.1 and 1.5.1, if the outcome* $(e_0, d(x), U_0)$ *is implementable (given* $j \in \{M, S\}$*), then* $(e', d(x), U_0)$ *is also implementable (given* $j \in \{M, S\}$*) for any* $e' < e_0$.

Proof. See appendix 1.10.6. □

Moreover, under Assumption 1.6.1, this upper bound is attainable.[46]

Assumption 1.6.1. *The function* $\mathcal{I}_{e^-}(x_0; e_0) \equiv \lim_{e \to e_0^-} \frac{\mathcal{I}(x_0;e_0) - \mathcal{I}(x_0;e)}{e_0 - e}$ *is continuous in* (x_0, e_0).

Let $UB^j(d(x), U_0, \omega)$ denote the upper bound for the implementable levels of effort determined by expressions 1.6.1 and 1.6.2. The superscript $j \in \{M, S\}$ describes whether the parties have access to transfer-scheme menus. Then, under Assumptions 1.4.1, 1.5.1 and 1.6.1, *Pareto-optimal* outcomes solve Program 1.6.1 below. *Optimal* contracts are those that implement Pareto-optimal outcomes.

Program 1.6.1 Pareto frontier

$$\max_{e_0 \in [0,1], d(x)} \quad \mathbb{E}_{F_{e_0}}[\mathbb{E}_x[y(d(x), \theta)]] - e_0,$$

subject to:

1. $e_0 \leq UB^j(d(x), U_0, \omega)$.

There are several insights that can be derived from Program 1.6.1. For instance, consider the question of when transfer-scheme menus are valuable. Let $(e^M, d^M(x))$ solve Program

[46] By the maximum theorem, Assumption 1.6.1 guarantees that both sides of expressions 1.6.1 and 1.6.2 are continuous in e_0.

1.6.1 when transfer-scheme menus are available, and suppose that the constraint is binding. Then, transfer-scheme menus are valuable if the maximizer in expression 1.6.1 (or 1.6.2) when $e_0 = e^M$ is not in $\mathcal{P}(d^M(x))$.[47] If all decisions are revealing, this maximizer is determined solely by the information acquisition technology, whereas $d^M(x)$ is determined by the output function. Thus, in this case, the maximizer need not be in the set $\mathcal{P}(d^M(x))$, and trasnfer-scheme menus are generally valuable. When there are unrevealing decisions, however, menus may not be valuable in environments that are not exceptional. This point will be illustrated in section 1.7.

The expression, $UB^j(d(x), U_0, \omega)$, summarizes the trade-off between improving the incentives for information acquisition and inducing better ex-post decision-making. As discussed in section 1.3, when $j = M$, $d(x)$ only affects $UB^j(d(x), U_0, \omega)$ through the unrevealing decisions that it prescribes. The only way to increase this upper bound is to take such decisions less frequently.[48] When $j = S$, the set $\mathcal{P}(d(x))$ introduces an additional reason to distort decision-making. In this case, the analysis becomes more subtle because the direction of optimal distortions depends on what the maximizer in expression 1.6.1 (or 1.6.2) would be in the case with menus (without menus, effort may be increased by distorting $d(x)$ to approximate such maximizer with points in $\mathcal{P}(d(x))$). The maximizer in expression 1.6.1 (or 1.6.2) when $j = M$ is difficult to characterize in general, since it is jointly determined with the level of effort that is implemented. Section 1.6.1 introduces a class of information acquisition technologies for which this maximizer can be characterized. In section 1.7 we will work with this technology to illustrate how optimal distortions to decision-making work in the case without transfer-scheme menus.

Finally, note that the function $UB^j(d(x), U_0, \omega)$ is increasing in U_0 and ω.[49] That is, increases in the expert's bargaining power or pledgeable income lead to more information acquisition and better decision-making. This could explain why partners in private equity

[47] If this is the case, we would have $U^S(d^M(x), U_0, \omega) < U^M(d^M(x), U_0, \omega)$ and e^M would not be attainable.

[48] When unrevealing decisions are taken less frequently, the interval, $[\underline{x}(d(x)), \bar{x}(d(x)]$, becomes smaller, and the maximum problem embedded in expression 1.6.2 attains a greater value for any level of effort. This implies that inequality 1.6.2 is satisfied for higher levels of effort, so $U^M(d(x), U_0, \omega)$ increases.

[49] By in creasing ω or U_0, the inequalities in expressions 1.6.1 and 1.6.2 become easier to satisfy for any level of effort. Thus, the highest level of effort that will satisfy the inequalities becomes larger.

funds, for example, are required to invest a non-trivial amount of their own wealth in their respective funds.

1.6.1 Success-Failure information acquisition technologies

An information acquisition technology is *Success-Failure* if, when the expert exerts effort $e_0 \in [0,1]$: with probability $P(e_0)$, he observes a posterior drawn from some distribution $G(x)$ with mean equal to the prior (success); with probability $1 - P(e_0)$, the expert learns nothing and observes the prior (failure). We assume that the function, $P : [0,1] \to [0,1]$, is non-decreasing, continuously differentiable and concave. Then Success-Failure information acquisition technologies satisfy Assumptions 1.4.1 and 1.6.1.[50]

Success-failure information acquisition technologies have the attractive feature that the maximizers in expressions 1.6.1 and 1.6.2 can easily be characterized. Indeed, it is always optimal to choose x, $\underline{m}$, and $\bar{m}$ as close to x_p as possible in their respective problems. The following lemma formalizes these claims:

Proposition 1.6.2. *If the information acquisition technology is Success-Failure, then:*

1. *The expression* $\frac{\mathcal{I}_{e^-}(x;e_0)}{\mathcal{I}(x;e_0)+x_p(1-x)}$ *is non-decreaasing in x in the interval $[0,x_p]$, and non-increasing in x in the interval $[x_p,1]$.*

2. *For any $\underline{x} \leq \bar{x}$, the expression* $\frac{\bar{m}\mathcal{I}_{e^-}(\underline{m};e_0)+(1-\underline{m})\mathcal{I}_{e^-}(\bar{m};e_0)}{\bar{m}\mathcal{I}(\underline{m};e_0)+(1-\underline{m})(\mathcal{I}(\bar{m};e_0)+x_p)}$ *is non-decreasing in $\bar{m}$ in the interval $[\bar{x},1]\bigcap[0,x_p]$, and non-increasing in $\bar{m}$ in the interval $[\bar{x},1]\bigcap[x_p,1]$. Similarly, this expression is non-decreasing in $\underline{m}$ in the interval $[0,\underline{x}]\bigcap[0,x_p]$, and non-increasing in $\underline{m}$ in the interval $[0,\underline{x}]\bigcap[x_p,1]$.*

Proof. See appendix 1.10.7. □

Another way of reading Proposition 1.6.2 is: when the information acquisition technology is Success-Failure, the contract, $T(x)$, that minimizes the cost of implementing any pair,

[50] This class of information acquisition technologies has been extensively used in the literature. In particular, Lambert (1986), and Inderst and Klein (2007) work with discrete versions of this technology (effort is binary). In Lambert (1986) $G(x) \sim U[0,1]$, and in Inderst and Klein (2007) $G(x)$ is assumed to have continuous density.

$(e_0, d(x))$, is a simple contract that changes its slope at a point (or pair of points) as close to x_p as possible (given the restrictions imposed by $d(x)$). Since the optimal location of the change in slope is independent of e_0, it is also true that, given any pair, $(d(x), U_0)$, the contract that maximizes effort is a simple contract, $T(x)$, that changes its slope at a point (or pair of points) as close to x_p as possible. This is the main implication of Proposition 1.6.2.

1.7 The value of menus

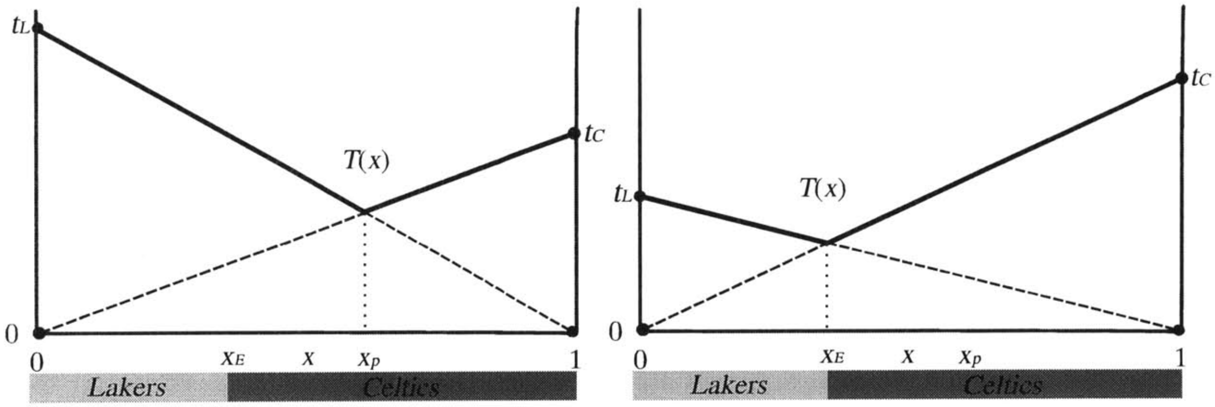
This section illustrates, through the examples described in section 1.1 (and in the context of Success-Failure information acquisition technologies), when and why transfer-scheme menus are valuable.

In order to induce the expert to acquire information, a contract must create a decision problem for him for which the information is relevant. We will see that menus are valuable because they enable the parties to separate this problem from the way in which the principal's decision is taken. Moreover, we will see that attaining separation requires menu contracts to take a form that may seem counterintuitive at first sight. Indeed, the value of menus stems from the fact that they introduce the possibility of giving the expert the option to bet for or against the same decision. Thus, in the optimal menu contract, the expert often has the opportunity to choose a decision for the principal and get rewarded when that decision fails.

1.7.1 Example 1: Betting on a game

We begin by considering an example in which all decisions are revealing. Let θ_1 =*Lakers win*, θ_2 =*Celtics win*, and suppose that there are only two decisions: *bet on the Lakers* and *bet on the Celtics*. Fix U_0 and ω small enough so that the first best is not attainable when contracts may specify menus (this implies that constraint (1) in Program 1.6.1 is binding). The information acquisition technology is Success-Failure.

We will see that, in this case, the only reason to distort decision-making comes from the inability to separate the principal's decision from the expert's incentives to acquire information. Figure 1.7.1 depicts the contracts that implement the maximum level of effort conditional on inducing ex-post efficient decision-making, depending on whether menus are available (Figure 1.7.1a) or not (Figure 1.7.1b). The yellow (green) region includes posteriors for which betting on the *Lakers* (*Celtics*) maximizes conditional expected output.



(a) With transfer-scheme menus

(b) Without transfer-scheme menus

Figure 1.7.1: Ex-post efficient contracts

Recall from Theorem 1.3.1 that, if contracts may specify transfer-scheme menus and all decision are revealing, decision rules impose no restrictions over the shape that $T(x)$ can take. Thus, by Proposition 1.6.2, the contract that maximizes effort conditional on efficient ex-post decision-making is a *simple* contract that changes its slope at x_p (see Figure 1.7.1a). In fact, this contract maximizes the level of effort that can be induced for any decision rule (given U_0 and ω), so it is optimal.[51] The optimal function $T(x)$ has two different slopes in the region where the decision is to bet on the *Celtics*. Thus, the contract that implements the optimal outcome must have three elements: 1) bet on the *Lakers* and get paid t_L only when the *Lakers* win; 2) bet on the *Celtics* and get paid t_C only when the *Celtics* win; 3) bet on the *Celtics* and get paid t_L only when the *Lakers* win. The expert will choose the first option after observing posteriors in $[0, x_E]$, the second option after observing posteriors

[51] The fact that the optimal contract changes its slope exactly at x_p is a specific feature of Success-Failure information acquisition technologies.

in $[x_p, 1]$, and the third option after observing posteriors in $[x_E, x_p]$. Note that, when the expert chooses the third option, he is making his client bet on the *Celtics*, but he is getting rewarded only when the *Lakers* win. Indeed, it is the possibility of letting the expert bet for or against the same decision (in this case, bet on the *Celtics*) that enables menus to separate the expert's incentives to acquire information from the way in which decisions are taken.[52]

Without menus, it is impossible to attain separation; in this case, the function $T(x)$ can only have different slopes at points where different decisions are taken (condition (7) in Proposition 1.5.1). Therefore, as illustrated in Figure 1.7.1b, implementing the ex-post efficient decision rule requires $T(x)$ to change its slope at x_E. By Proposition 1.6.2, this contract implements a lower level of effort than the contract in Figure 1.7.1a. Inducing the expert to acquire more information requires distorting decision-making. Indeed, in order to enable $T(x)$ to change its slope at a point closer to x_p, the bet must be placed on the *Lakers* more frequently than what would be efficient. Doing this actually leads to a Pareto-improvement.[53]

This example illustrates a more general insight: when the parties do not have access to transfer-scheme menus, distorting decision-making is generally optimal.[54] In particular, this implies that, in environments where all decisions are revealing, menus are usually valuable. Moreover, although in this example distortions favor betting on the *Lakers*, by modifying the output function, we could have constructed an example where distortions would go in the opposite direction. That is, without transfer-scheme menus, optimal distortions to decision-making do not follow a general pattern (as they do when menus are available). In fact, without menus, the intuitive result that distortions should always favor decisions that are more revealing breaks down. Below we provide an example where, in the absence of menus, it is optimal to distort decision-making in favor of the unrevealing decision.

[52] The same logic can be applied in any environment where the transfers may depend explicitly on state of nature. Portfolio managers, for example, could be compensated with a menu of transfer-schemes contingent only on stock prices. This would motivate them to acquire information relevant for their clients' choices without encouraging them to take inefficient decisions on their behalf.

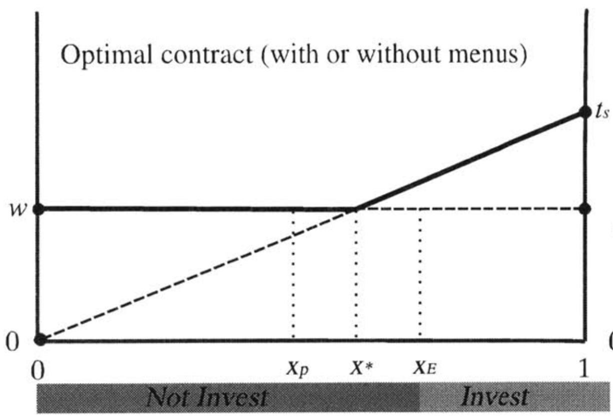
[53] Modifying the contract in Figure 1.7.1b slightly to induce more effort generates a second-order loss of value due to betting in the *Lakers* too frequently, but a first-order gain due to the additional effort.

[54] It is not optimal only in the (unlikely) event that both decisions generate the same expected output after observing the posterior in which $T(x)$ should change its slope to maximize effort.

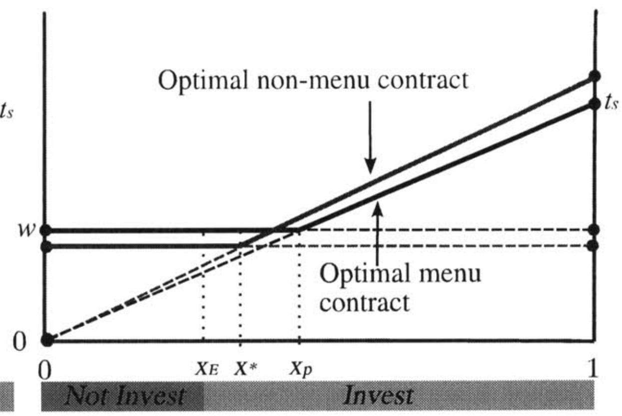
1.7.2 Example 2: Undertaking a project

Here the decisions are *invest* or *not invest*, θ_1 =*project fails* and θ_2=*project succeeds*.[55] The only revealing decision is to invest in the project. We will see that, when there are unrevealing decisions: 1) there is a reason to distort decision-making even when transfer-scheme menus are available and the expert's incentives can be separated from the principal's decision; 2) menus may be not valuable under circumstances that are not exceptional; 3) menus are still often valuable (indeed, without menus, it may be optimal to distort decision-making in favor of the unrevealing decision).

Figure 1.7.2 depicts optimal contracts given two different output functions (everything else is kept unchanged). The red (green) regions include posteriors for which not investing (investing) is the ex-post efficient decision.



(a) Transfer-scheme menus are not valuable



(b) Transfer-scheme menus are valuable

Figure 1.7.2: Optimal contracts

Figure 1.7.2a illustrates an example in which menus are *not* valuable. In this case, the decision that maximizes expected output under the prior is to not invest (Inderst and Klein (2007) focus on this case). Start by considering the optimal menu contract. If the parties wanted to implement an ex-post efficient decision rule, the function $T(x)$ would have to be flat in the interval $[0, x_E]$ (condition (6) in Theorem 1.3.1). However, by Proposition 1.6.2, the parties would be able to increase information acquisition by having $T(x)$ change its slope

[55] This is the environment studied in Lambert (1986) and Inderst and Klein (2007).

at a point closer to x_p. This can only be achieved by investing in the project more frequently. Thus, the optimal menu contract changes its slope at a point $x^* \in [x_p, x_E]$.[56] In order to induce a function $T(x)$ with these features, a contract could give the expert two alternatives: 1) not invest and get a fixed wage, w; 2) invest and get paid t_s only if the project succeeds. Note that, in each of the two alternatives, a different decision is taken. Therefore, the same arrangement could be replicated with a single-transfer-scheme contract. In this example, distortions to decision-making (the project is undertaken too frequently) are purely driven by the fact that the decision not to invest is unrevealing.

Figure 1.7.2b illustrates an example where menus are valuable again. The only difference with respect to the previous case is that now investing in the project is ex-post efficient under the prior. Consider the contracts that implement the highest possible level of effort conditional on ex-post efficient decision-making (given U_0 and ω). By Proposition 1.6.2, the function $T(x)$ (that is restricted to be flat in the interval $[0, x_E]$) that maximizes effort (given U_0) is actually flat in the (larger) interval, $[0, x_p]$.[57] In fact, if menus are available, this contract is optimal.[58] Therefore, in this example, the optimal menu contract has three elements: 1) not invest and get w (chosen by the expert after observing posteriors in $[0, x_E]$); 2) invest and get w (chosen by the expert after observing posteriors in $[x_E, x_p]$); 3) invest and get paid t_s only if the project succeeds (chosen by the expert after observing posteriors in $[x_p, 1]$). Again, the principal's decision is separated from the expert's incentives by letting the expert bet for or against the same decision. In the second alternative, the expert chooses to invest, but does not let his own compensation depend on the investment's outcome. In

[56] Distorting the decision rule slightly in favor of investing at points near x_E creates a second-order loss of value due to inefficient decision-making, but a first-order gain due to the additional effort. If the decision rule were distorted at points near $x = 0$, this would also generate a first-order gain due to increased effort. However, in this case such distortions would also generate a first-order loss of value, since at $x = 0$, not investing is strictly more efficient than investing. We assume that distorting the decision rule at points close to $x = 0$ is not worthwhile.

[57] Recall that, under Success-Failure information acquisition technologies, in order to maximize effort (given U_0), $T(x)$ should change its slope at a point as close as possible to x_p given the constraints imposed by the decision rule.

[58] This follows because, in order to be able to induce more effort, decision-making would have to be distorted at $x = 0$. Deciding to invest at points near $x = 0$ would generate a first-order loss of value, which (we assume) would not be compensated by the increase in effort.

this case, it is optimal to implement the ex-post efficient decision rule.

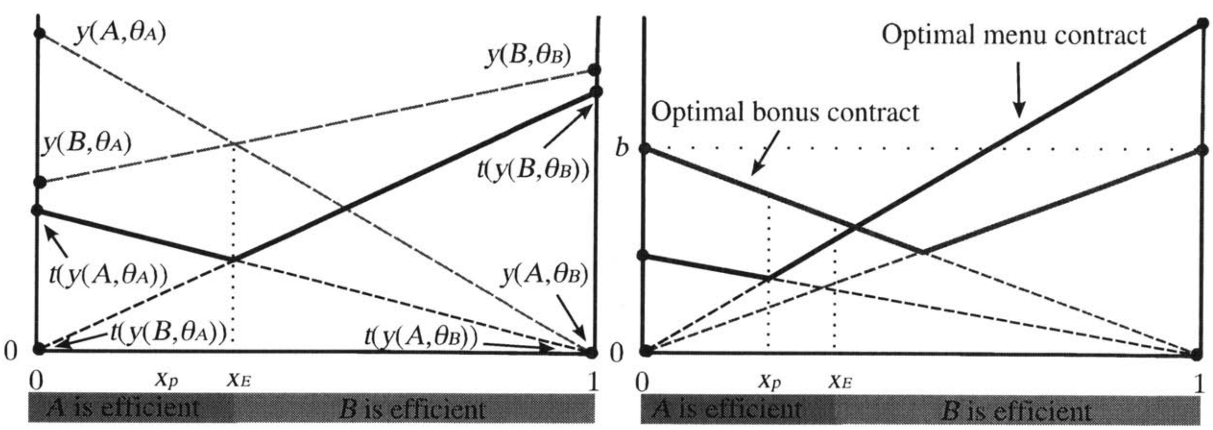
Without menus the story is quite different. Now implementing the ex-post efficient decision rule would require $T(x)$ to change its slope at x_E. Now, by Proposition 1.6.2, the parties could do better by increasing information acquisition at the expense of distorting decision-making in favor of not undertaking the project (in this way, the slope of $T(x)$ could change at a point closer to x_p, as in Figure 1.7.2b). Therefore, in this case, it is optimal to distort decision-making in favor of the *unrevealing* decision!

Lambert (1986), Demski and Sappington (1987) and Malcomson (2009) analyze the Principal-Expert problem considering contracts that specify a single output-contingent transfer scheme. These studies interpret the nature of distortions to decision-making in terms of the differences in the variability of output induced by different decisions. We can use our previous analysis to reassess this interpretation. The examples illustrate that the output function may affect optimal distortions to decision-making through two different channels: 1) it determines the set of posteriors for which each decision is efficient; 2) it may affect which decisions are revealing. Note that the riskiness associated with each decision only affects distortions indirectly, since decisions that induce more variable output (invest, in the project example) are also those that are revealing.

1.8 Can we treat experts like workers?

Consider the problem of motivating an agent whose contribution to output comes, to a large extent, through his influence on decisions (take, for example, a CEO or a portfolio manager). Can we analyze this problem (in theory and in practice) through the lens of the standard moral hazard framework? In particular, if we only consider contracts that specify a single output-contingent transfer scheme, should the contract used to motivate the CEO resemble the contracts that could be optimal in the standard moral hazard model? This section provides an example to illustrate two points. First, there is little that can be said in general about the shape of optimal output-contingent contracts when the agent is an expert. Indeed, it is easy to construct examples in which the information acquisition technology is natural and the optimal contract is non-monotone in output. Second, contracts that may be optimal in the moral-hazard framework can easily lead to poor performance when used with experts. In particular, we will see that a contract with two levels of compensation (a bonus contract) can easily induce poor decision-making without fostering information acquisition.

Suppose that $d \in \{A, B\}$ and $\Theta = \{\theta_A, \theta_B\}$. Let the output function be such that $y(A, \theta_A) > y(B, \theta_A)$, $y(B, \theta_B) > y(A, \theta_B)$ and $y(d, \theta) \neq y(d', \theta')$ for any $(d, \theta) \neq (d', \theta')$. Suppose contracts may only specify a single transfer-schemes contingent in output, and that the information acquisition technology is Success-Failure. Figure 1.8.1a illustrates an example in which the optimal transfer scheme is non-monotone in output. The figure depicts the output function and the transfer scheme that implements the highest possible level of effort, conditional on the decision-rule being ex-post efficient (given U_0). Note that, since $x_E < \frac{1}{2}$, it must be the case that $t(y(B, \theta_B)) > t(y(A, \theta_A))$, while $y(A, \theta_A) > y(B, \theta_B)$. The same would be true for any simple contract in which the limited-liability constraint is binding, and where the slope of $T(x)$ changes at a point in $[x_p, x_E]$. Thus, it is true for the optimal contract.[59]



(a) Non-monotone transfer scheme

(b) Bonus contracts

Figure 1.8.1: Output-contingent contracts

This result contrasts with the conclusion in Diamond (1998) that, as output becomes

[59] By Proposition 1.6.2, the induced level of effort here can only be increased by letting $T(x)$ change its slope at a point closer to x_p.

more important relative to the expert's cost of effort, the expert's optimal compensation converges to a transfer-scheme linear in output.[60] This conclusion relies on two especial features of Diamond (1998)'s model. First, there, transfer schemes implement the ex-post efficient decision rule if and only if they are linear in output. In the example above, the contract that maximizes the implemented level of effort conditional on the decision rule being ex-post efficient is actually non-monotone. The difference is that, in Diamond (1998), the output function has $y(A, \theta_B) = y(B, \theta_A) = 0$. Second, in Diamond (1998)'s model, as output becomes more important relative to the cost of effort, it is optimal for the distortions to decision-making to tend to zero (this implies that linear contracts are asymptotically optimal). This feature does not seem general. It is true that, as output becomes more important, the marginal cost of distorting decision-making increases. However, the marginal benefit of acquiring more information also becomes larger, making the net effect over optimal distortions ambiguous in general.

Figure 1.8.1b, illustrates what would happen in this example if the expert was rewarded with a contract with two levels of compensation (a bonus contract).[61] In order to induce effort, the bonus would have to be paid if and only if output is greater than or equal to $y(B, \theta_B)$. Moreover, to maximize effort (given U_0), the baseline wage would have to be zero. Note that any bonus contract necessarily induces the expert to take decision A if and only if his posterior is smaller than $\frac{1}{2}$. Thus, A is taken too frequently. Furthermore, the non-linear contract in Figure 1.8.1a induces ex-post efficient decision-making and, by Proposition 1.6.2, implements a higher level of effort than the optimal bonus contract. That is, in this example, bonus contracts distort decision-making without fostering information acquisition.

1.9 Concluding remarks

This paper has analyzed the problem of motivating an expert to help a principal take a decision. In order to induce the expert to exert costly and non-verifiable effort to acquire

[60] Diamond (1998) analyzes a Principal-Expert model in which contracts may specify a single output-contingent transfer scheme.

[61] These contracts can be optimal in the standard moral hazard framework.

information, a contract must create a choice problem for him for which the information is relevant. Contracts that specify menus of transfer schemes are valuable because they enable the parties to separate the expert's choice problem from the way the principal's decision is taken. Indeed, when menus are available and all decisions reveal the same amount of information about the state ex-post, optimal contracts always induce efficient decision-making conditional on the information available. However, in environments where different decisions reveal different amounts of information about the state, it may be optimal to distort decision-making even when separation is possible. Fostering information acquisition requires distorting decision rules in favor of decisions that reveal more information. Intuitively, when a decision that is not revealing is taken, the expert's payment cannot depend on the realization of the state, limiting his incentives to acquire information.

When contracts may only specify a single transfer scheme, the expert's choice problem after acquiring information is inevitably linked to the way the principal's decision is taken; different choices at the interim stage can only induce different state-contingent payoffs if they lead to different decisions. This fact introduces an additional motive to distort decision-making. As a result, in this case it is optimal to implement the ex-post efficient decision rule only in exceptional circumstances. Moreover, the intuition for menus may be overturned: without menus, it may be optimal to distort decision-making in favor of decisions that reveal less information about the state, rather than more.

We have shown that, under reasonable conditions, optimal contracts have at most three elements that induce distinct state-contingent payoffs for the expert (with and without transfer-scheme menus). Moreover, we have illustrated how, in order to separate the expert's incentives for effort from the principal's decision, menu contracts must include choices that induce different transfer schemes but lead to the same decision. This is often valuable, but has the counterintuitive implication that optimal menu contracts typically include choices in which the expert induces a decision and is simultaneously rewarded when that decision leads to low output.

These results are relevant for the study of optimal incentives for agents whose contribution to output comes, to a large extent, through their influence on decision-making (CEOs, consultants or portfolio managers, for example). The present analysis suggests that the standard moral hazard model may not provide a good description for such incentive problems. In particular, menu contracts, which are qualitatively different from the contracts considered in the standard model, are able to improve performance. A priori, there is no reason to rule out menus. Understanding to what extent menus are actually used in practice or why they might not be used is an interesting avenue for future research.

The main insights derived in this paper are quite general. The heart of the analysis is the characterization of implementable outcomes provided by Theorem 1.3.1 (and its extension to the case without menus provided by Proposition 1.5.1). In Zermeño (2011*b*), I extend this result to the case where there is a finite number of states of nature and where the expert's effort may affect output directly in addition to generating information.

The result that simple menu contracts are optimal is more delicate. We have only provided sufficient conditions under which the result is true, and illustrated why the conditions are sufficient but not necessary. Characterizing optimal menu contracts without Assumption 1.4.1 or in the case where the state of nature is not binary could be interesting. Such result would help us understand how the number of alternatives that a menu needs to provide is related to the number of states of nature in the environment.

1.10 Appendix

1.10.1 Proof of Theorem 1.3.1

We begin with necessity. Suppose that an outcome $(e_0, d(x), \mathbb{T})$ is implementable with some contract, $\{(t_r(z), d_r)\}_{r\in R}$. For all $x_0 \in [0,1]$, let $T(x_0) = \max_{r\in R} \mathbb{E}_{x_0}[t_r(z(d_r,\theta))]$, and $r^*(x_0) \in \arg\max_{r\in R} \mathbb{E}_{x_0}[t_r(z(d_r,\theta))]$. Then $T(x)$ satisfies conditions (1)-(6) of Theorem 1.3.1. Conditions (1)-(2) must be satisfied by definition 1.2.3. Condition (3) is satisfied because $T(x)$ is the upper envelope of linear functions of x, so it must be convex and continuos.

Moreover, by the envelope theorem (see Milgrom and Segal (2002)), we must have:

$$T(x_0) \equiv T(0) + \int_0^{x_0} \Delta(x)dx,$$

where $\Delta(x) \equiv t_{r^*(x)}(z(d_{r^*(x)}, \theta_2)) - t_{r^*(x)}(z(d_{r^*(x)}, \theta_1))$. By the convexity of $T(x)$, $\Delta(x)$ is non-decreasing. Note that condition (6) follows by the convexity of $T(x)$ and because, if $d(x_0)$ is unrevealing, $\Delta(x_0) = 0$, and $T(x)$ must reach its minimum at x_0.

Finally, conditions (4) and (5) are satisfied because

$$T(0)+T'(0) = t_{r^*(0)}(z(d_{r^*(0)}, \theta_1))+\Delta(0^+) \geq t_{r^*(0)}(z(d_{r^*(0)}, \theta_1))+\Delta(0) = t_{r^*(0)}(z(d_{r^*(0)}, \theta_2)) \geq 0,$$

$$T(1)-T'(1) = t_{r^*(1)}(z(d_{r^*(1)}, \theta_2))-\Delta(1^-) \geq t_{r^*(1)}(z(d_{r^*(1)}, \theta_2))-\Delta(1) = t_{r^*(1)}(z(d_{r^*(1)}, \theta_1)) \geq 0.$$ [62]

Now we prove sufficiency. Suppose that a function $T(x)$ satisfies conditions (1)-(6) on Theorem 1.3.1 given the outcome, $(e_0, d(x), \mathbb{T})$. For each $x_0 \in [0, 1]$, let $\Delta(x_0)$ be a subderivative of $T(x)$ at x_0 with the following properties: $\Delta(x_0) = 0$ if $T(x)$ reaches its minimum at $x_0 \in [0, 1]$; if $T(x)$ does not reach its minimum at $x = 0$, $\Delta(0) = \lim_{x \to 0^+} \Delta(x)$; if $T(x)$ does not reach its minimum at $x = 1$, $\Delta(1) = \lim_{x \to 1^-} \Delta(x)$.[63] For each $x_0 \in [0, 1]$, define:

$$t_{x_0}(z(d(x_0), \theta_1)) = T(x_0) - x_0\Delta(x_0),$$

$$t_{x_0}(z(d(x_0), \theta_2)) = T(x_0) + (1 - x_0)\Delta(x_0),$$

and $t_{x_0}(z) = 0$ for any other z. Since $T(x)$ satisfies condition (6), these transfers are well defined. The contract $\{(t_x(z), d(x))\}_{x \in [0,1]}$ implements $(e_0, d(x), \mathbb{T})$. First note that limited liability is satisfied. We have,

$$t_{x_0}(z(d(x_0), \theta_1)) = T(x_0) - x_0\Delta(x_0) \geq T(1) - T'(1) \geq 0, \ \forall \ x_0 \in [0, 1).$$

62 The expressions $\Delta(0^+)$ and $\Delta(1^-)$ denote the limits of $\Delta(x)$ as x approaches 0 and 1 respectively.
63 The values $\Delta(0)$ and $\Delta(1)$ must be bounded because $T(x)$ satisfies conditions (4) and (5).

$$t_{x_0}(z(d(x_0),\theta_2)) = T(x_0) + (1-x_0)\Delta(x_0) \geq T(0) + T'(0) \geq 0, \ \forall \ x_0 \in (0,1].$$

The first inequalities follow because the expressions $T(x) - x\Delta(x)$ and $T(x) + (1-x)\Delta(x)$ are non-increasing and non-decreasing in x respectively.[64] The second inequalities follow from conditions (4)-(5) in Theorem 1.3.1. When $T(x)$ does not reach its minimum at 0 or 1, all the inequalities still hold for $x_0 \in \{0,1\}$. If $T(x)$ reaches its minimum at $\hat{x} \in \{0,1\}$, $t_{\hat{x}}(z(d(\hat{x}),\theta_1)) = t_{\hat{x}}(z(d(\hat{x}),\theta_2)) \geq 0$ by the inequalities above.

Finally, note that the expert's best response after observing x_0 is to report x_0, so condition (3) of definition 1.2.3 is satified. If the expert reported x' instead, his payment would be

$$T(x') - (1-x_0)x'\Delta(x') + x_0(1-x')\Delta(x') = T(x') + (x_0 - x')\Delta(x') \leq T(x_0),$$

which is what he would get by reporting x_0. The inequality follows because $\Delta(x')$ is a subderivative of $T(x)$ at x'. Thus, for each $x_0 \in [0,1]$, $T(x_0) = \max_{x\in[0,1]} \mathbb{E}_{x_0}[t_x(z(d(x),\theta))]$, so conditions (1) and (2) of definition 1.2.3 are implied by conditions (1)-(2) of Theorem 1.3.1.

1.10.2 Weak duality

Theorem 1.4.1 relies extensively on Lemma 1.10.1, which is stated and proved here. This lemma is an adaptation of the standard weak duality result used in linear programming theory.

Given the real-valued functions $a(x)$, $k(e)$, $l(x)$, and $b(x;e)$, defined over $x \in [0,1]$, and $e \in [0,1]$, consider the following *primal* problem and its corresponding *dual*:

[64] Take $x' > x_0$. Then $T(x') - x'\Delta(x') - (T(x_0) - x_0\Delta(x_0)) = \int_{x_0}^{x'} \Delta(x)dx - x'\Delta(x') + x_0\Delta(x_0) \leq (x'-x_0)\Delta(x') - x'\Delta(x') + x_0\Delta(x_0) = x_0(\Delta(x_0) - \Delta(x')) \leq 0$, where the inequalities follow because, by condition (3) of Theorem 1.3.1, $\Delta(x)$ must be non-decreasing. Similarly, take $x' > x_0$. Then $T(x') + (1 - x')\Delta(x') - T(x_0) - (1-x_0)\Delta(x_0) = \int_{x_0}^{x'} \Delta(x)dx + (1-x')\Delta(x') - (1-x_0)\Delta(x_0) \geq (1-x')(\Delta(x') - \Delta(x_0)) \geq 0$.

Program 1.10.1 Primal

$$\min_{\Delta(x)} \int_0^1 a(x)d\Delta(x),$$

subject to:

1. $\Delta(x)$ is non-decreasing and right-continuous.
2. For all $e \in [0,1]$, $a(\cdot), l(\cdot)$ and $b(\cdot; e)$ are integrable with respect to the Lebesgue-Stieltjes measure and σ-algebra derived form $\Delta(x)$.
3. $\int_0^1 b(x;e)d\Delta(x) \geq k(e)$, for all $e \in [0,1]$.
4. $\int_0^1 l(x)d\Delta(x) = K$.

Program 1.10.2 Dual

$$\max_{\lambda(e),\mu} \mu K + \int_0^1 k(e)d\lambda(e),$$

subject to:

1. $\lambda(e)$ is non-decreasing and right-continuous.
2. For all $x \in [0,1]$, $k(\cdot)$ and $b(x; \cdot)$ are integrable with respect to the Lebesgue-Stieltjes measure and σ-algebra derived from $\lambda(e)$.
3. $\mu l(x) + \int_0^1 b(x;e)d\lambda(e) \leq a(x)$, for all $x \in [0,1]$.

Lemma 1.10.1. *Let $\Delta(x)$, and $(\lambda(e), \mu)$ be feasible in Programs 1.10.1 and 1.10.2 respectively, and such that $\int_0^1 \left(\int_0^1 b(x;e)d\lambda(e)\right) d\Delta(x) = \int_0^1 \left(\int_0^1 b(x;e)d\Delta(x)\right) d\lambda(e)$. Then,*

$$\int_0^1 a(x)d\Delta(x) \geq \mu K + \int_0^1 k(e)d\lambda(e).$$

Proof. Suppose that $\Delta(x)$, and $(\lambda(e), \mu)$ satisfy all the conditions in the lemma. Then,

$$\int_0^1 a(x)d\Delta(x) \geq \int_0^1 \left(\mu l(x) + \int_0^1 b(x;e)d\lambda(e)\right) d\Delta(x) = \mu \int_0^1 l(x)d\Delta(x) +$$

$$+ \int_0^1 \left(\int_0^1 b(x;e)d\Delta(x)\right) d\lambda(e) \geq \mu K + \int_0^1 k(e)d\lambda(e),$$

where the first inequality follows because $(\lambda(q), \mu)$ satisfy condition (3) in Program 1.10.2. The second inequality comes from changing the order of integration, which can be done by assumption. The third inequality follows because $\Delta(x)$ satisfies conditions (3) and (4) in Program 1.10.1. □

If $b(x;e)$ is bounded and measurable (with respect to the product space) for any feasible $\Delta(x)$ and $\lambda(e)$, then, by Fubini's theorem, the order-of-integration condition always holds. In this case, Lemma 1.10.1 implies that, if $\Delta^*(x)$ and $(\lambda^*(e), \mu^*)$ are feasible and attain the same value in their respective programs, they are optimal. The proof of Theorem 1.4.1 exploits this fact.

1.10.3 Proof of Theorem 1.4.1

We begin with two preliminary observations that will be used extensively:

Observation 1: Define $\mathcal{I}_{e^-}(x; e_0) \equiv \lim_{e \to e_0^-} \frac{\mathcal{I}(x;e_0) - \mathcal{I}(x;e)}{e_0 - e}$. Then, under Assumption 1.4.1, if the constraint set of Program 1.4.1 is non-empty, there must exist $x_0 \notin (\underline{x}, \bar{x})$ such that $\mathcal{I}_{e^-}(x_0; e_0) > 0$.

Note that the second constraint in Program 1.4.1 (as expressed in Program 1.4.2) can be rewritten as:

2a. $\inf_{e<e_0} \int_0^1 \frac{\mathcal{I}(x;e_0) - \mathcal{I}(x;e)}{e_0 - e} d\Delta(x) \geq 1$;

2b. $\sup_{e>e_0} \int_0^1 \frac{\mathcal{I}(x;e_0) - \mathcal{I}(x;e)}{e_0 - e} d\Delta(x) \leq 1$.

Under Assumption 1.4.1, for each $x_0 \in [0, 1]$, the expression $\frac{\mathcal{I}(x_0;e_0) - \mathcal{I}(x_0;e)}{e_0 - e}$ is non-negative and non-increasing in e. Thus,

$$\inf_{e<e_0} \int_0^1 \frac{\mathcal{I}(x; e_0) - \mathcal{I}(x; e)}{e_0 - e} d\Delta(x) = \lim_{e \to e_0^-} \int_0^1 \frac{\mathcal{I}(x; e_0) - \mathcal{I}(x; e)}{e_0 - e} d\Delta(x).$$

If we had $\mathcal{I}_{e^-}(x_0; e_0) = 0$, for all $x_0 \notin (\underline{x}, \bar{x})$, then it would hold that

$$\lim_{e \to e_0^-} \int_0^1 \frac{\mathcal{I}(x; e_0) - \mathcal{I}(x; e)}{e_0 - e} d\Delta(x) = \int_0^1 \lim_{e \to e_0^-} \frac{\mathcal{I}(x; e_0) - \mathcal{I}(x; e)}{e_0 - e} d\Delta(x) = 0,$$

where the equality follows by Dini's theorem (a decreasing sequence of continuous functions defined over a compact set converging pointwise to a continuous function converges uniformly). Since this would hold for any non-decreasing $\Delta(x)$, the constraint set of Program 1.4.1 would be empty.

Observation 2: Without loss of generality, for the purposes of this proof, we can change the notion of integration from Riemann-Stieltjes to Lebesgue-Stieltjes. This is necessary because we will use Lebesgue's dominated convergence theorem.

Program 1.4.2 is expressed in terms of Riemann-Stieltjes integrals, which are always well defined because $\Delta(x)$ is required to be non-decreasing and the integrand functions are continuous. The notion of Lebesgue-Stieltjes integration requires $\Delta(x)$ to be right-continuous. However, since this function is endogenous, it need not be right-continuous. To deal with this issue we will actually solve a modified version of Program 1.4.2. There are three differences. First, the function $\Delta(x)$ is now defined over the reals and is required to be right-continuous. Second, the closed interval in constraint (5) of Program 1.4.2 is replaced by the interval $[\underline{x}, \bar{x})$. Third, the notion of integration is now Lebesgue-Stieltjes integration.

The modified program is equivalent to Program 1.4.2 in the sense that a value is attainable in the modified program if and only if it is also attainable in Program 1.4.2. To see why take $\Delta(x)$ feasible in Program 1.4.2. Let $\hat{\Delta}(x_0) = \Delta(0)$ for $x_0 < 0$, $\hat{\Delta}(x_0) = \Delta(1)$ for $x_0 > 1$, and $\hat{\Delta}(x_0) = \lim_{x \to x_0^+} \Delta(x)$ for $x_0 \in [0, 1]$. The function $\hat{\Delta}(x)$ is right-continuous. Moreover, for any continuous $h(x)$, $\mathcal{RS} \int_0^1 h(x) d\Delta(x) = \mathcal{LS} \int_0^1 h(x) d\hat{\Delta}(x)$.[65] Thus, $\hat{\Delta}(x)$ is feasible in the modified program and attains the original value. Similarly, if $\hat{\Delta}(x)$ is feasible in the modified program, let $\Delta(0) = \lim_{x \to 0} \hat{\Delta}(x)$, $\Delta(\bar{x}) = 0$, and $\Delta(x_0) = \hat{\Delta}(x_0)$ for any other $x_0 \in [0, 1]$. Then, for any continuous $h(x)$, $\mathcal{RS} \int_0^1 h(x) d\Delta(x) = \mathcal{LS} \int_0^1 h(x) d\hat{\Delta}(x)$, so $\Delta(x)$ is feasible in Program 1.4.2, and attains the same value.[66]

Body of the proof: There are four cases that need to be considered when solving

[65] The notation $\mathcal{RS} \int$ denotes Riemann-Stieltjes integrals, and $\mathcal{LS} \int$ denotes Lebesgue-Stieltjes integrals.

[66] If $\underline{x} = \bar{x}$, the necessary modification to Program 1.4.2 becomes more involved. The difference is that now we must choose *two* non-decreasing and right-continuous functions, $\underline{\Delta}(x)$, and $\bar{\Delta}(x)$ defined over the reals. We require that $\bar{\Delta}(x_0) = 0$ for all $x_0 < \bar{x}$, and $\underline{\Delta}(x_0) = 0$ for all $x_0 \geq \underline{x}$. Finally, any integral, $\mathcal{RS} \int_0^1 h(x)\Delta(x)$, in Program 1.4.2 is replaced by $\mathcal{LS} \int_0^{\underline{x}} h(x) d\underline{\Delta}(x) + \mathcal{LS} \int_{\bar{x}}^1 h(x) d\bar{\Delta}(x)$. This modified program is equivalent to Program 1.4.2 when $\underline{x} = \bar{x}$.

Program 1.4.1: 1) the interval $[\underline{x}, \bar{x}]$ is empty; 2) $\underline{x}$=0; 3) $\bar{x} = 1$; 4) $0 < \underline{x} \leq \bar{x} < 1$. As argued in section 1.4, in the first case, Program 1.4.1 can be rewritten as Program 1.4.3, and in the fourth case it can be rewritten as Program 1.4.4. The second and third cases are analogous to the first one in the sense that Program 1.4.1 can be rewritten in a very similar way to Program 1.4.3. Thus, the same arguments used for the first case still apply to the second and third, which will not be discussed here.

The proof proceeds in three steps. First, we show that in Programs 1.4.3 and 1.4.4 there exists an optimal contract within the class of contracts that are feasible *and* simple. Second, for each case we compute the optimal simple contract. Third, using Lemma 1.10.1, we show that the derived simple contracts are optimal (within the class of *all* feasible contracts) in their respective programs.

Step 1: Programs 1.10.3 and 1.10.4 below summarize the problem when only simlpe contract are allowed. In this case, the function $\Delta(x)$ may only increase once in Program 1.4.3, and twice in Program 1.4.4 (once in each side of $[\underline{x}, \bar{x}]$). Thus, Program 1.4.3 becomes a problem with two variables: the location of the change (say $\hat{x}$), and its size (say δ); and Program 1.4.4 becomes a problem with four variables: the locations of the changes (say $\underline{m}$ and $\bar{m}$), and their corresponding sizes (say $\underline{\delta}$ and $\bar{\delta}$).

Program 1.10.3 Program 1.4.3 when only simple contracts are allowed

$$\min_{\hat{x}\in[0,1],\delta\geq 0} \quad \delta[\mathcal{I}(\hat{x}; e_0) + x_p(1 - \hat{x})],$$

subject to:

1. $\delta[\mathcal{I}(\hat{x}; e_0) - \mathcal{I}(\hat{x}; e)] \geq e_0 - e$ for all $e \neq e_0$.

Note that the objective function in Programs 1.10.3 and 1.10.4 are continuous in all the variables. Moreover, the constraint sets are closed. Thus, in order to show that a solution exists, it is enough to prove that δ, $\underline{\delta}$ and $\bar{\delta}$ can be bounded from above (this would guarantee that the constraint sets are compact). In order to find an upper bound for δ in Program 1.10.3, take any x_0 such that $\mathcal{I}_{e^-}(x_0; e_0) > 0$, and let $\delta_0 = (\mathcal{I}_{e^-}(x_0; e_0))^{-1}$. By Assumption 1.4.1, the pair (x_0, δ_0) is feasible, and is attains some positive value V_0. Note

Program 1.10.4 Program 1.4.4 when only simple contracts are allowed

$$\min_{\underline{m}\in[0,\underline{x}],\bar{m}\in[\bar{x},1],\underline{\delta}\geq 0,\bar{\delta}\geq 0} \quad \underline{\delta}\mathcal{I}(\underline{m};e)+\bar{\delta}[\mathcal{I}(\bar{m};e_0)+x_p],$$

subject to:

1. $\underline{\delta}[\mathcal{I}(\underline{m};e_0)-\mathcal{I}(\underline{m};e)]+\bar{\delta}[\mathcal{I}(\bar{m};e_0)-\mathcal{I}(\bar{m};e)]\geq e_0-e$ for all $e\neq e_0$.
2. $(1-\underline{m})\underline{\delta}=\bar{m}\bar{\delta}$.

---

that $\min_{x\in[0,1]}\{\mathcal{I}(x;e)+x_p(1-x)\}>0$. Thus, if $\delta>V_0\left(\min_{x\in[0,1]}\{\mathcal{I}(x;e)+x_p(1-x)\}\right)^{-1}$, it will attain a strictly higher value than V_0, and cannot be optimal. Upper bounds for $\underline{\delta}$ and $\bar{\delta}$ in Program 1.10.4 can be found in a similar way.

Step 2: We start with the solution to Program 1.10.3. Fix $\hat{x}$ and consider the optimal choice of δ. A necessary condition for feasibility is that $\delta\mathcal{I}_{e^-}(\hat{x};e_0)\geq 1$. Moreover, by Assumption 1.4.1, if this inequality holds with equality, then $(\delta,\hat{x})$ is feasible. Since the objective is non-negative, the inequality will optimally hold with equality. Thus, substituting δ in the reciprocal of the objective, our problem becomes:

$$\max_{\hat{x}\in[0,1]}\frac{\mathcal{I}_{e^-}(\hat{x};e_0)}{\mathcal{I}(\hat{x};e_0)+x_p(1-\hat{x})}. \tag{1.10.1}$$

We know that a solution exists (even though $\mathcal{I}_{e^-}(\hat{x};e_0)$ cannot be guaranteed to be continuous in $\hat{x}$) because we showed that there exists an optimal contract within the class of simple contracts.

Similarly, we can find the solution to Program 1.10.4. Using the same arguments as before, we can see that, in this case, given the cutoffs $\underline{m}$ and $\bar{m}$, having $\underline{\delta}$ satisfy $\underline{\delta}[\bar{m}\mathcal{I}_{e^-}(\underline{m};e_0)+(1-\underline{m})\mathcal{I}_{e^-}(\bar{m};e_0)]=\bar{m}$ is optimal and feasible. Substituting $\underline{\delta}$ (and the implied $\bar{\delta}$) in the reciprocal of the objective, our problem becomes:

$$\max_{\underline{m}\in[0,\underline{x}],\bar{m}\in[\bar{x},1]}\frac{\bar{m}\mathcal{I}_{e^-}(\underline{m};e_0)+(1-\underline{m})\mathcal{I}_{e^-}(\bar{m};e_0)}{\bar{m}\mathcal{I}(\underline{m};e_0)+(1-\underline{m})(\mathcal{I}(\bar{m};e_0)+x_p)}. \tag{1.10.2}$$

Again, we know that a solution exists because we proved that there exists an optimal contract within the class of simple contracts.

Step 3: Let $\hat{x}^*$ maximize equation 1.10.1, and $\delta^* = \left(\mathcal{I}_{e^-}(\hat{x}^*; e_0)\right)^{-1}$. Similarly, let $(\underline{m}^*, \bar{m}^*)$ maximize equation 1.10.2, and define $\underline{\delta}^* = \frac{\bar{m}^*}{\bar{m}^*\mathcal{I}_{e^-}(\underline{m}^*;e_0)+(1-\underline{m}^*)\mathcal{I}_{e^-}(\bar{m}^*;e_0)}$ and $\bar{\delta}^* = \frac{1-\underline{m}^*}{\bar{m}^*\mathcal{I}_{e^-}(\underline{m}^*;e_0)+(1-\underline{m}^*)\mathcal{I}_{e^-}(\bar{m}^*;e_0)}$. We will show that the simple contracts characterized by these values are optimal in Programs 1.4.3 and 1.4.4 respectively. To do so we will rely essentially on Lemma 1.10.1.

We begin by rewriting Programs 1.4.3 and 1.4.4 (the modified versions) in terms of our canonical primal program (Program 1.10.1). Define $b(x;e) \equiv \frac{\mathcal{I}(x;e_0)-\mathcal{I}(x;e)}{e_0-e}$ for all $e < e_0$, $b(x;e) \equiv -\frac{\mathcal{I}(x;e_0)-\mathcal{I}(x;e)}{e_0-e}$ for all $e > e_0$, and $b(x;e_0) \equiv \mathcal{I}_{e^-}(x;e_0)$. Note that, by Assumption 1.4.1, $|b(x;e| \leq max_{x\in[0,1]}b(x;0)$, so $b(x;e)$ is bounded. Moreover, $b(x;e)$ is measurable with respect to the product space generated by any feasible $\Delta(x)$ and $\lambda(e)$.[67] This implies that the order-of-integration condition in Lemma 1.10.1 is satisfied for any feasible $\Delta(x)$ and $\lambda(e)$.

Note that, for any $e \neq e_0$, and $\Delta(x)$ non-decreasing and right-continuous, $b(x;e)$ is integrable with respect to the Lebesgue-Stieltjes measure derived from $\Delta(x)$ ($b(x;e)$ is continuous in x). Then, by Assumption 1.4.1 and Lebesgue's dominated convergence theorem, $b(x;e_0)$ is also integrable and

$$\int_0^1 b(x;e_0)d\Delta(x) = \lim_{e\to e_0^-}\int_0^1 b(x;e)d\Delta(x) = \inf_{e<e_0}\int_0^1 b(x;e)d\Delta(x).$$

Therefore, constraint (2) in Programs 1.4.3 and 1.4.4 implies that $\int_0^1 b(x;e_0)d\Delta(x) \geq 1$, and so it can be rewritten as:

$$\int_0^1 b(x;e)d\Delta(x) \geq k(e), \quad \forall \enspace e \in [0,1],$$

where $k(e) = 1$ for $e \leq e_0$, and $k(e) = -1$ for $e > e_0$. Once constraint (2) is written in

[67] To see this take any open subset of $\mathbb{R}$ (say O). Then

$$b^{-1}(O) = \left(b^{-1}(O)\bigcap[0,1]\times\{e\in[0,1]\mid e\neq e_0\}\right)\bigcup\left(b^{-1}(O)\bigcap[0,1]\times\{e_0\}\right).$$

The first term is in the product σ-algebra because, since $b(x;e)$ is continuous at any point with $e \neq e_0$, it is an open set. Note also that $b(x;e')$ is measurable, as a function of x, for all $e' \neq e_0$ (it is continuous). Thus, since $b(x;e_0) \equiv inf_{e<e_0}\{b(x;e)\}$, it is also measurable as a function of x. Therefore, the second term is also in the product σ-algebra (it is a measurable rectangle).

this way, Programs 1.4.3 and 1.4.4 have the exact same form as Program 1.10.1.[68] Programs 1.10.5 and 1.10.6 are their respective duals.

Program 1.10.5 Dual for program 1.4.3

$$\max_{\lambda(e)} \int_0^1 k(e)d\lambda(e),$$

subject to:

1. $\lambda(e)$ is non-decreasing and right-continuous.
2. For all $x \in [0,1]$, $k(\cdot)$ and $b(x;\cdot)$ are integrable with respect to the Lebesgue-Stieltjes measure and σ-algebra derived form $\lambda(e)$.
3. $\int_0^1 b(x;e)d\lambda(e) \leq \mathcal{I}(x;e) + x_p(1-x)$, for all $x \in [0,1]$.

Program 1.10.6 Dual for program 1.4.4

$$\max_{\lambda(e),\mu} \int_0^1 k(e)d\lambda(e),$$

subject to:

1. $\lambda(e)$ is non-decreasing and right-continuous.
2. For all $x \in [0,1]$, $k(\cdot)$ and $b(x;\cdot)$ are integrable with respect to the Lebesgue-Stieltjes measure and σ-algebra derived form $\lambda(e)$.

3a. $(1-x)\mu + \int_0^1 b(x;e)d\lambda(e) \leq \mathcal{I}(x;e_0)$, for all $x \in [0,\underline{x}]$.

3b. $-x\mu + \int_0^1 b(x;e)d\lambda(e) \leq \mathcal{I}(x;e_0) + x_p$, for all $x \in [\bar{x},1]$.

In order to show that the simple contracts above are optimal, we must exhibit multipliers that are feasible in their respective programs and that attain the same value as the optimal simple contracts (the reciprocal of expressions 1.10.1 and 1.10.2). Note that the functions $k(\cdot)$ and $b(x;\cdot)$ are left-continuous for any $x \in [0,1]$. Therefore, for any non-decreasing and right-continuous $\lambda(e)$, they are Riemann-Stieltjes integrable, and thus Lebesgue-Stieltjes integrable.

[68] The third constraint in Program 1.4.4 can be rewritten as $\int_0^1 l(x)d\Delta(x) = 0$, where $l(x) = 1-x$ for $x \leq \underline{x}$, and $l(x) = -x$ otherwise.

We begin by finding the multiplier for Program 1.10.5. Let $\lambda^*(e) = 0$ for $e < e_0$, and $\lambda^*(e) = \frac{\mathcal{I}(\hat{x}^*;e_0)+x_p(1-\hat{x}^*)}{\mathcal{I}_{e^-}(\hat{x}^*;e_0)}$ otherwise. Clearly, $\lambda^*(e)$ attains the same value as the simple contract characterized by $(\hat{x}^*, \delta^*)$ (see equation (1.10.1)). To see that it is feasible, note that constraint (3) of Program 1.10.5 becomes $b(x;e_0) \leq [\mathcal{I}(x;e_0) + x_p(1-x)]\frac{\mathcal{I}_{e^-}(\hat{x}^*;e_0)}{\mathcal{I}(\hat{x}^*;e_0)+x_p(1-\hat{x}^*)}$, for all $x \in [0,1]$. It holds because $b(x;e_0) \equiv \mathcal{I}_{e^-}(x;e_0)$, and $\mathcal{I}(x;e_0) + x_p(1-x) > 0$.

Now we proceed to finding suitable multiplier for Program 1.10.6. Let

$$\lambda^*(e) = \begin{cases} 0 & if\ e < e_0 \\ \frac{\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)+(1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)}{\bar{m}^*\mathcal{I}_{e^-}(\underline{m}^*;e_0)+(1-\underline{m}^*)\mathcal{I}_{e^-}(\bar{m}^*;e_0)} & if e \geq e_0 \end{cases},$$

and

$$\mu^* = \frac{\mathcal{I}(\underline{m}^*;e_0)b(\bar{m}^*;e_0) - b(\underline{m}^*;e_0)[\mathcal{I}(\bar{m}^*;e_0)+x_p]}{\bar{m}^*b(\underline{m}^*;e_0) + (1-\underline{m}^*)b(\bar{m}^*;e_0)}.$$

Again, it is straight forward to verify that these multipliers attain the same value as the simple contract above (see equation 1.10.2). To check feasibility take any $x \in [\bar{x}, 1]$. Then,

$$-x\mu^* + b(x;e_0)\frac{\bar{m}^*\mathcal{I}(\underline{m}^*;e_0) + (1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)}{\bar{m}^*b(\underline{m}^*;e_0) + (1-\underline{m}^*)b(\bar{m}^*;e_0)} \leq \mathcal{I}(x;e_0) + x_p$$

holds if and only if:

$$\begin{aligned} - &\quad x[\mathcal{I}(\underline{m}^*;e_0)b(\bar{m}^*;e_0) - b(\underline{m}^*;e_0)(\mathcal{I}(\bar{m}^*;e_0)+x_p)] + b(x;e_0)[\bar{m}^*\mathcal{I}(\underline{m}^*;e_0) + \\ + &\quad (1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)] \leq [\mathcal{I}(x;e_0)+x_p][\bar{m}^*b(\underline{m}^*;e_0) + (1-\underline{m}^*)b(\bar{m}^*;e_0)]. \end{aligned}$$

Rearranging and multiplying by $(1-\underline{m}^*)$ on each side we obtain

$$\begin{aligned} &(1-\underline{m}^*)[\mathcal{I}(\bar{m}^*;e_0)+x_p][xb(\underline{m}^*;e_0) + (1-\underline{m}^*)b(x;e_0)] + (1-\underline{m}^*)b(x;e_0)\bar{m}^*\mathcal{I}(\underline{m}^*;e_0) \leq \\ &[x\mathcal{I}(\underline{m}^*;e_0) + (1-\underline{m}^*)(\mathcal{I}(x;e_0)+x_p)]b(\bar{m}^*;e_0)(1-\underline{m}^*) + [\mathcal{I}(x;e_0)+x_p]\bar{m}^*b(\underline{m}^*;e_0)(1-\underline{m}^*). \end{aligned}$$

Adding $x\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)b(\underline{m}^*;e_0)$ on each side,

$$[(1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)+\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)][xb(\underline{m}^*;e_0)+(1-\underline{m}^*)b(x;e_0)] \leq$$

$$[x\mathcal{I}(\underline{m}^*;e_0)+(1-\underline{m}^*)(\mathcal{I}(x;e_0)+x_p)][b(\bar{m}^*;e_0)(1-\underline{m}^*)+\bar{m}^*b(\underline{m}^*;e_0)],$$

which holds by equation 1.10.2.

Similarly, take any $x \in [0, \underline{x}]$. Then

$$(1-x)\mu^* + b(x;e_0)\frac{\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)+(1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)}{\bar{m}^*b(\underline{m}^*;e_0)+(1-\underline{m}^*)b(\bar{m}^*;e_0)} \leq \mathcal{I}(x;e_0)$$

holds if and only if:

$$(1-x)[\mathcal{I}(\underline{m}^*;e_0)b(\bar{m}^*;e_0)-b(\underline{m}^*;e_0)(\mathcal{I}(\bar{m}^*;e_0)+x_p)]+b(x;e_0)[\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)+$$

$$+\quad (1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)] \leq \mathcal{I}(x;e_0)[\bar{m}^*b(\underline{m}^*;e_0)+(1-\underline{m}^*)b(\bar{m}^*;e_0)].$$

Rearranging and multiplying by $\bar{m}^*$ on each side we obtain:

$$\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)[(1-x)b(\bar{m}^*)+\bar{m}^*b(x;e_0)]+\bar{m}^*(1-\underline{m}^*)b(x;e_0)(\mathcal{I}(\bar{m}^*;e_0)+x_p) \leq$$

$$\bar{m}^*b(\underline{m}^*;e_0)[(1-x)(\mathcal{I}(\bar{m}^*;e_0)+x_p)+\bar{m}^*\mathcal{I}(x;e_0)]+\bar{m}^*(1-\underline{m}^*)\mathcal{I}(x;e_0)b(\bar{m}^*;e_0).$$

Adding $(1-x)(1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)b(\bar{m}^*;e_0)$ on each side,

$$[\bar{m}^*\mathcal{I}(\underline{m}^*;e_0)+(1-\underline{m}^*)(\mathcal{I}(\bar{m}^*;e_0)+x_p)][(1-x)b(\bar{m}^*;e_0)+\bar{m}^*b(x;e_0)] \leq$$

$$[(1-\underline{m}^*)b(\bar{m}^*;e_0)+\bar{m}^*b(\underline{m}^*;e_0)][(1-x)(\mathcal{I}(\bar{m}^*;e_0)+x_p)+\bar{m}^*\mathcal{I}(x;e_0)],$$

which holds by equation 1.10.2.

1.10.4 Binary-signal information acquisition technologies

Binary-signal information acquisition technologies generate experiments with signals that have two possible realizations. Thus, given a level of effort e, the expert observes the posteriors $x_h(e) \equiv x_p + h(e)$ with probability $\hat{x}(e)$, and $x_l(e) \equiv x_p - l(e)$ with probability $1 - \hat{x}(e)$, where $\hat{x}(e)h(e) - (1-\hat{x}(e))l(e) \equiv 0$. The functions $h : [0,1] \to [0, 1-x_p]$, and $l : [0,1] \to [0, x_p]$ are assumed to be non-decreasing and continuous. It is straight-forward to verify that this class of technologies does not satisfy Assumption 1.4.1. However, the following proposition implies that simple contracts are still optimal.

Proposition 1.10.1. *With binary-signal information acquisition technologies, if the outcome* $(e_0, d(x), \mathbb{T})$ *is implementable, then it is implementable with a simple contract.*

Proof. Suppose that $T(x)$ implements $(e_0, d(x), \mathbb{T})$. If $d(x)$ prescribes unrevealing decisions, let $[\underline{x}, \bar{x}]$ be the interval over which $T(x)$ must be minimized, and $[\underline{m}, \bar{m}]$ the interval over which it is actually minimized ($[\underline{x}, \bar{x}] \subseteq [\underline{m}, \bar{m}]$). Figure 1.10.1 illustrates how we can build a simple contract, $T^s(x)$, that implements the same outcome.

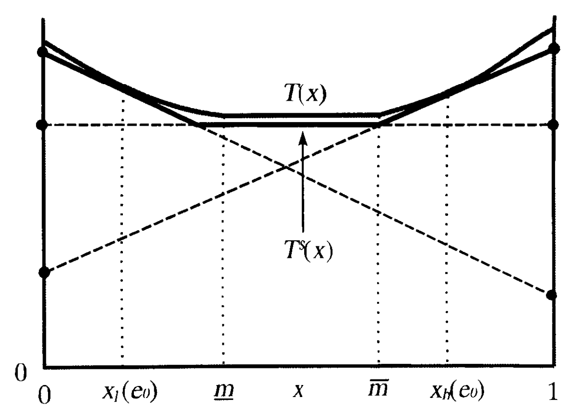


Figure 1.10.1: Proposition 1.10.1

The function $T^s(x)$ is the upper envelope of three lines. The first two are tangent lines to $T(x)$ at $x_l(e_0)$ and $x_h(e_0)$. The third one is the lowest flat line that guarantees that $T^s(x)$

is flat in the interval $[\underline{m}, \bar{m}]$. Note that, given the outcome, $(e_0, d(x), \mathbb{T})$, $T^s(x)$ satisfies conditions (2)-(6) in Theorem 1.3.1. Condition (1) is also satisfied because, by the convexity of $T(x)$, $T^s(x) \leq T(x)$ for all $x \in [0,1]$, with equality in $x_l(e_0)$ and $x_h(e_0)$. Thus, for all $e \neq e_0$,

$$\mathbb{E}_{F_{e_0}}[T^s(x)] - e_0 = \mathbb{E}_{F_{e_0}}[T(x)] - e_0 \geq \mathbb{E}_{F_e}[T(x)] - e \geq \mathbb{E}_{F_e}[T^s(x)] - e.$$

□

1.10.5 Proof of Proposition 1.5.1

The argument is analogous to that of Theorem 1.3.1, so Figures 1.3.1 and 1.3.2 are helpful to provide intuition. We start with necessity. Suppose an outcome $(e_0, d(x), \mathbb{T})$ is implementable with a contract $\{(t(z), d_r)\}_{r\in R}$. Let $T(x) \equiv max_{r\in R}\mathbb{E}_x[t(z(d_r, \theta))]$. Theorem 1.3.1 implies that conditions (1)-(6) must hold. Condition (7) must be satisfied because, if $d(x_0) = d(x')$, then $t(z(d(x_0), \theta_i)) = t(z(d(x'), \theta_i))$ for $i = 1, 2$. Thus, for any $\alpha \in [0,1]$, we have

$$\begin{aligned} T(\alpha x_0 + (1-\alpha)x') &\geq [1 - (\alpha x_0 + (1-\alpha)x')]t(z(d(x_0), \theta_1)) + [\alpha x_0 + (1-\alpha)x']t(z(d(x_0), \theta_2)) \\ &= [(1-\alpha)(1-x') + \alpha(1-x_0)]t(z(d(x_0), \theta_1)) + [\alpha x_0 + (1-\alpha)x']t(z(d(x_0)\theta_2)) \\ &= \alpha T(x_0) + (1-\alpha)T(x'), \end{aligned}$$

where the last line follows because $t(z(d(x_0), \theta_i)) = t(z(d(x'), \theta_i))$ for $i = 1, 2$. The reverse inequality follows from the convexity of $T(x)$.

To show that conditions (1) to (7) are sufficient, take $T(x)$ that satisfies them given the outcome $(e_0, d(x), \mathbb{T})$. Transfer schemes can be constructed as in Figure 1.3.2. For each $x_0 \in [0,1]$, let $\Delta(x_0)$ be a subderivative of $T(x)$ at x_0. If $d(x_0) = d(x')$, pick $\Delta(x_0) = \Delta(x') = \frac{T(x_0)-T(x')}{x_0-x'}$, which are subderivatives of $T(x)$ at x_0 and x' by condition (7). Also, let

$\Delta(0) = \inf_{x>0} \Delta(x)$, and $\Delta(1) = \sup_{x<1} \Delta(x)$. Then, for each $x_0 \in [0,1]$, define

$$t(z(d(x_0), \theta_1)) = T(x_0) - x_0\Delta(x_0),$$

and

$$t(z(d(x_0), \theta_2)) = T(x_0) + (1 - x_0)\Delta(x_0).$$

Note that, if $d(x_0) = d(x')$,

$$\begin{aligned} t(z(d(x_0), \theta_1)) &= T(x_0) - x_0\Delta(x_0) = T(x') + (x_0 - x')\Delta(x_0) - x_0\Delta(x_0) = T(x') - x'\Delta(x') \\ &= t(z(d(x'), \theta_1)), \end{aligned}$$

where the second equality follows from condition (7). Similarly, $t(z(d(x_0), \theta_2)) = t(z(d(x'), \theta_2))$. Moreover, by Assumption 1.5.1, if $d(x_0) \neq d(x')$, $z(d(x_0), \theta) \neq z(d(x'), \theta')$ for any θ, θ'. Therefore, these transfers are well defined (for each z_0, only one value is assigned to $t(z_0)$).

The argument to show that the the contract $\{(t(z), d(x))\}_{x\in[0,1]}$ built in this way implements the outcome $(e_0, d(x), \mathbb{T})$ is analogous to the one used in Theorem 1.3.1.

1.10.6 Proof of Lemma 1.6.1

Note that inequalities 1.6.1 and 1.6.2 are satisfied if and only if:

1. If $d(x_0)$ is revealing for all $x_0 \in [0,1]$,

$$\max_{x \in X^j(d(x))} \{(U_0 + \omega + e_0)\mathcal{I}_{e^-}(x; e_0) - [\mathcal{I}(x; e_0) + x_p(1-x)]\} \geq 0. \quad (1.10.3)$$

2. If $d(x_0)$ is unrevealing for some $x_0 \in [0,1]$,

$$\max_{\underline{m} \in \underline{X}_j(d(x)),\ \bar{m} \in \bar{X}^j(d(x))} \{(U_0 + \omega + e_0)[\bar{m}\mathcal{I}_{e^-}(\underline{m}; e_0) + (1 - \underline{m})\mathcal{I}_{e^-}(\bar{m}; e_0)] \\ -[\bar{m}\mathcal{I}(\underline{m}; e_0) + (1 - \underline{m})(\mathcal{I}(\bar{m}; e_0) + x_p)]\} \geq 0. \tag{1.10.4}$$

To see why suppose $d(x_0)$ is always revealing. If inequality 1.6.1 is satisfied, then there exists $x^* \in X^j(d(x))$ such that

$$U_0 + \omega + e_0 \geq \left(\frac{\mathcal{I}_{e^-}(x^*; e_0)}{\mathcal{I}(x^*; e_0) + x_p(1 - x^*)} \right)^{-1},$$

which hold if and only if

$$(U_0 + \omega + e_0)\mathcal{I}_{e^-}(x^*; e_0) - [\mathcal{I}(x^*; e_0) + x_p(1 - x^*)] \geq 0.$$

Thus, inequality 1.10.3 is also satisfied.

Similarly, if inequality 1.10.3 is satisfied, there exists $x^* \in X^j(d(x))$ such that:

$$(U_0 + \omega + e_0)\mathcal{I}_{e^-}(x^*; e_0) - [\mathcal{I}(x^*; e_0) + x_p(1 - x^*)] \geq 0,$$

which holds if and only if

$$U_0 + \omega + e_0 \geq \left(\frac{\mathcal{I}_{e^-}(x^*; e_0)}{\mathcal{I}(x^*; e_0) + x_p(1 - x^*)} \right)^{-1}.$$

Thus, inequality 1.6.1 is also satisfied.

The case where $d(x)$ prescribes unrevealing decisions is analogous.

Now, under Assumption 1.4.1, for any $x \in X^j(d(x))$, and $(\underline{m}, \bar{m}) \in \underline{X}^j(d(x)) \times \bar{X}^j(d(x))$, the expressions

$$(U_0 + \omega + e_0)\mathcal{I}_{e^-}(x; e_0) - [\mathcal{I}(x; e_0) + x_p(1 - x)],$$

and

$$(U_0+\omega+e_0)[\bar{m}\mathcal{I}_{e^-}(\underline{m};e_0)+(1-\underline{m})\mathcal{I}_{e^-}(\bar{m};e_0)]-[\bar{m}\mathcal{I}(\underline{m};e_0)+(1-\underline{m})(\mathcal{I}(\bar{m};e_0)+x_p)]$$

are decreasing in e_0. To see why take $e_0 > e'$. Then,

$$(U_0+\omega+e_0)\mathcal{I}_{e^-}(x;e_0)-[\mathcal{I}(x;e_0)+x_p(1-x)]-[(U_0+\omega+e')\mathcal{I}_{e^-}(x;e')-[\mathcal{I}(x;e')+x_p(1-x)]]$$

$$\leq e_0\mathcal{I}_{e^-}(x;e_0)-e'\mathcal{I}_{e^-}(x;e')-(\mathcal{I}(x;e_0)-\mathcal{I}(x;e'))$$

$$=(e_0-e')\mathcal{I}_{e^-}(x;e_0)+e'(\mathcal{I}_{e^-}(x;e_0)-\mathcal{I}_{e^-}(x;e'))-(\mathcal{I}(x;e_0)-\mathcal{I}(x;e'))$$

$$\leq e'(\mathcal{I}_{e^-}(x;e_0)-\mathcal{I}_{e^-}(x;e'))\leq 0,$$

where the next-to-last inequality follows because, by Assumption 1.4.1,

$$(e_0-e')\mathcal{I}_{e^-}(x;e_0)-(\mathcal{I}(x;e_0)-\mathcal{I}(x;e'))\leq 0.$$

The other case is analogous.

1.10.7 Proof of Proposition 1.6.2

Success-Failure information acquisition technologies have

$$\mathcal{I}(x_0;e_0)\equiv\int_0^{x_0}\mathbb{I}_{[x_p,1]}(x)dx+P(e_0)\int_0^{x_0}[G(x)-\mathbb{I}_{[x_p,1]}(x)]dx,$$

and $\mathcal{I}_{e^-}(x_0;e_0)\equiv P'(e_0)\int_0^{x_0}[G(x)-\mathbb{I}_{[x_p,1]}(x)]dx$. Note that both expressions in the proposition (as functions of $x\in[0,1]$, and $(\underline{m},\bar{m})\in[0,\underline{x}]\times[\bar{x},1]$ respectively) are continuous.

Moreover, their left partial derivatives are well defined, since

$$\mathcal{I}_{x^-}(x_0;e_0) \equiv \lim_{x\to x_0^-} \frac{\mathcal{I}(x_0;e_0)-\mathcal{I}(x;e_0)}{x_0-x} \equiv \mathbb{I}_{[x_p,1]}(x_0^-) + P(e_0)(G(x_0^-) - \mathbb{I}_{[x_p,1]}(x_0^-)),$$

and

$$\mathcal{I}_{e^-x^-}(x_0;e_0) \equiv \lim_{x\to x_0^-} \frac{\mathcal{I}_{e^-}(x_0;e_0)-\mathcal{I}_{e^-}(x;e_0)}{x_0-x} \equiv P'(e_0)(G(x_0^-) - \mathbb{I}_{[x_p,1]}(x_0^-)),$$

where, for any function $h(x)$, $h(x_0^-)$ denotes $\lim_{x\to x_0^-} h(x)$.

In order to complete the proof, we will see that the left-derivative of the first expression with respect to x is positive for all $x_0 \leq x_p$, and negative for all $x_0 > x_p$. Similarly, we will see that, fixing any $\underline{m} \in [0, \underline{x}]$, the left-derivative of the second expression with respect to $\bar{m}$ is positive if $\bar{m}_0 \leq x_p$, and negative if $\bar{m}_0 > x_p$; and, fixing any $\bar{m} \in [\bar{x}, 1]$, the left-derivatve of the second expression with respect to $\underline{m}$ is positive if $\underline{m}_0 \leq x_p$, and negative if $\underline{m}_0 > x_p$.

To simplify the algebra, note that $\mathcal{I}_{e^-}(x_0;e_0) \equiv \frac{P'(e_0)}{P(e_0)}\left(\mathcal{I}(x_0;e_0) - \int_0^{x_0} \mathbb{I}_{[x_p,1]}(x)dx\right)$. We start with the first expression. The numerator of the left-derivative with respect to x at x_0 is proportional to:

$$P(e_0)[G(x_0^-) - \mathbb{I}_{[x_p,1]}(x_0^-)][\mathcal{I}(x_0;e_0) + x_p(1-x_0)]$$

$$-[P(e_0)G(x_0^-) + (1-P(e_0))\mathbb{I}_{[x_p,1]}(x_0^-) - x_p]\left[\mathcal{I}(x_0;e_0) - \int_0^{x_0} \mathbb{I}_{[x_p,1]}(x)dx\right] \qquad (1.10.5)$$

If $x_0 \leq x_p$, expression 1.10.5 becomes:

$$P(e_0)G(x_0^-)x_p(1-x_0) + x_p\mathcal{I}(x_0;e_0) \geq 0.$$

If $x_0 > x_p$, expression 1.10.5 becomes:

$$P(e_0)(G(x_0^-)-1)x_0(1-x_p) - (1-x_p)(\mathcal{I}(x_0;e_0) - x_0 + x_p) \leq 0,$$

where the inequality follows because $\mathcal{I}(x_0;e_0) - x_0 + x_p \geq \mathcal{I}(1;e_0) - 1 + x_p = 0$.

Now for the second expression. The numerator of the left partial derivative with respect to $\bar{m}$ at $(\underline{m}_0, \bar{m}_0)$ is proportional to:

$$\left[\mathcal{I}(\underline{m}_0; e_0) - \int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx + (1 - \underline{m}_0)P(e_0)(G(\bar{m}_0^-) - \mathbb{I}_{[x_p,1]}(\bar{m}_0^-))\right]$$

$$[\bar{m}_0\mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)(\mathcal{I}(\bar{m}_0; e_0) + x_p)]$$

$$-[\mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)[P(e_0)G(\bar{m}_0^-) + (1 - P(e_0))\mathbb{I}_{[x_p,1]}(\bar{m}_0^-)]]$$

$$\left[\bar{m}_0\left(\mathcal{I}(\underline{m}_0; e_0) - \int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right) + (1 - \underline{m}_0)\left(\mathcal{I}(\bar{m}_0; e_0) - \int_0^{\bar{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right)\right] \tag{1.10.6}$$

If $\bar{m}_0 \leq x_p$, expression 1.10.6 becomes:

$$\left[\mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)P(e_0)G(\bar{m}_0^-)\right](1 - \underline{m}_0)x_p \geq 0.$$

If $\bar{m}_0 > x_p$, expression 1.10.6 becomes:

$$\left[\mathcal{I}(\underline{m}_0; e_0) - \int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx + (1 - \underline{m}_0)P(e_0)(G(\bar{m}_0^-) - 1)\right]$$

$$[\bar{m}_0 \mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)(\mathcal{I}(\bar{m}_0; e_0) + x_p)]$$

$$-[\mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)[P(e_0)(G(\bar{m}_0^-) - 1) + 1]] \quad .$$

$$\left[\bar{m}_0 \left(\mathcal{I}(\underline{m}_0; e_0) - \int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right) + (1 - \underline{m}_0)\left(\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p\right)\right]$$

$$=(1 - \underline{m}_0)P(e_0)(G(\bar{m}_0^-) - 1)\left[\bar{m}_0 \int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx + \bar{m}_0(1 - \underline{m}_0)\right]$$

$$-(1 - \underline{m}_0)(\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p)\left[\int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx + (1 - \underline{m}_0)\right] \leq 0,$$

where the last inequality follows because $\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p \geq \mathcal{I}(1; e_0) - 1 + x_p = 0$.

Finally, the numerator of the left partial derivative with respect to $\underline{m}$ at $(\underline{m}_0, \bar{m}_0)$ is proportional to:

$$\left[\bar{m}_0 P(e_0)[G(\underline{m}_0^-) - \mathbb{I}_{[x_p,1]}(\underline{m}_0^-)] - \left(\mathcal{I}(\bar{m}_0; e_0) - \int_0^{\bar{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right)\right]$$

$$[\bar{m}_0 \mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)(\mathcal{I}(\bar{m}_0; e_0) + x_p)]$$

$$-[\bar{m}_0[P(e_0)G(\underline{m}_0^-) + (1 - P(e_0))\mathbb{I}_{[x_p,1]}(\underline{m}_0^-)] - (\mathcal{I}(\bar{m}_0; e_0) + x_p)]$$

$$\left[\bar{m}_0 \left(\mathcal{I}(\underline{m}_0; e_0) - \int_0^{\underline{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right) + (1 - \underline{m}_0)\left(\mathcal{I}(\bar{m}_0; e_0) - \int_0^{\bar{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right)\right] \tag{1.10.7}$$

If $\underline{m}_0 \leq x_p$, expression 1.10.7 becomes:

$$\left[\bar{m}_0 P(e_0)G(\underline{m}_0^-) - \left(\mathcal{I}(\bar{m}_0; e_0) - \int_0^{\bar{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right)\right] [\bar{m}_0 \mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)(\mathcal{I}(\bar{m}_0; e_0) + x_p)]$$

$$-[\bar{m}_0 P(e_0)G(\underline{m}_0^-) - (\mathcal{I}(\bar{m}_0; e_0) + x_p)] \left[\bar{m}_0 \mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)\left(\mathcal{I}(\bar{m}_0; e_0) - \int_0^{\bar{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right)\right]$$

$$= [\bar{m}_0 P(e_0)G(\underline{m}_0^-)(1 - \underline{m}_0) + \bar{m}_0 \mathcal{I}(\underline{m}_0; e_0)] \left(x_p + \int_0^{\bar{m}_0} \mathbb{I}_{[x_p,1]}(x)dx\right) \geq 0.$$

If $\underline{m}_0 > x_p$, expression 1.10.7 becomes:

$$\left[\bar{m}_0 P(e_0)(G(\underline{m}_0^-) - 1) - (\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p)\right] \left[\bar{m}_0 \mathcal{I}(\underline{m}_0; e_0) + (1 - \underline{m}_0)(\mathcal{I}(\bar{m}_0; e_0) + x_p)\right]$$

$$-\left[\bar{m}_0 P(e_0)(G(\underline{m}_0^-) - 1) - (\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p)\right]$$

$$\left[\bar{m}_0 \left(\mathcal{I}(\underline{m}_0; e_0) - \underline{m}_0 + x_p\right) + (1 - \underline{m}_0)\left(\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p\right)\right]$$

$$= -\left[(\mathcal{I}(\bar{m}_0; e_0) - \bar{m}_0 + x_p) + \bar{m}_0 P(e_0)(1 - G(\underline{m}_0^-))\right] \bar{m}_0(1 - x_p) \leq 0.$$

Chapter 2

The Role of Authority in a General Principal-Expert Model

2.1 Introduction

Since the seminal contributions of Crawford and Sobel (1982) and Holmström (1977, 1984), the literature on decision-making in organizations has relied extensively on an exogenous conflict of interest between the parties regarding the decisions that are to be taken.[1] There are situations, however, where the agents involved in the decision-making process do not have a direct interest in decisions. For instance, portfolio managers or consultants typically are not directly concerned with their clients' choices; they become interested only to the extent that these choices affect their own compensation. In these situations, the core of the incentive problem is not the exogenous disagreement regarding decisions, but the need to motivate experts to acquire costly information relevant for decision-making.

This paper examines a general principal-expert model in which the only source of friction is the fact that the expert must be induced to exert effort. The principal (she) is the residual claimant of output, which is determined by a decision and the state of nature. The expert (he)

[1] See, for example, Aghion and Tirole (1997), Baker, Gibbons and Murphy (1999), Dessein (2002), Krishna and Morgan (2008), Alonso and Matouschek (2008), Rantakari (2008) and Alonso, Dessein and Matouschek (2008).

has no direct interest in the decision and is protected by limited liability. By exerting costly and non-verifiable effort, the expert may affect output directly (by shifting the distribution of the state), or indirectly (by generating non-verifiable information about the state). To induce effort, the expert is paid through transfer schemes, which can depend on everything that is observable after the decision is taken and the state is realized.[2] The analysis focuses on the following questions. How does the allocation of authority (who has the right to take the decision) affect the set of outcomes that can be implemented? How should compensation schemes respond to changes in the allocation of authority? Can optimal incentive design lead to an endogenous conflict of interest regarding decisions? If so, what is the nature of this conflict? What kind of biases can systematically emerge?

The first set of results is concerned with the first two questions. We start by considering the set of outcomes that can be implemented under *full-commitment*. In this case, contracts specify a menu of pairs, where each pair is formed by a decision and by a transfer scheme. After exerting effort but before the state is realized, the expert can select a pair from the menu. The selection determines the decision to be taken and the transfer scheme for payments. The full-commitment case establishes an upper bound for the set of outcomes that can be implemented when decisions are determined in some other way. This paper provides conditions under which any Pareto-optimal outcome implementable under full-commitment can also be implemented under each of the following two arrangements:[3] 1) Expert-authority, under which contracts specify a menu of transfer schemes from which the expert can choose, and the expert can ultimately take *any* decision; 2) Principal-authority, under which contracts specify a menu of transfer schemes from which the expert can choose, and the principal can ultimately take *any* decision after observing the expert's selection from

[2] As in the standard moral hazard model (see, for example, Mirrlees (1976), Holmström (1979), Harris and Raviv (1979), Shavell (1979), Grossman and Hart (1983) or Kim (1995)), in this framework, if effort were verifiable, the first-best would be attainable. Nevertheless, in contrast to the standard model, here there is additional information arriving after the expert exerts effort but before the state is realized. Thus, as in Laffont and Tirole (1986), in addition to inducing effort, the parties must solve an adverse-selection problem: the expert must be induced to use his private information adequately. Moreover, here output is not only determined by the expert's effort, but also by the way decisions are taken given the information available. Thus, incentives designed to induce effort must take into account their influence on decision making.

[3] An outcome is *Pareto-optimal* if there does not exists any other implementable outcome that makes both parties better off.

the menu and recommendation.

Optimal full-commitment contracts establish a *plan* that specifies the decision that should be taken as a function of the transfer scheme selected by the expert. Under expert-authority or principal-autority, the party in control can deviate from the plan and take any alternative decision. The conditions identified in the analysis guarantee that it is always possible to make such deviations unappealing to the party in control, while keeping the expert's incentives to acquire information the same as in the full-commitment benchmark. We require that, for any pair of decisions, d and d', if what the parties observe ex-post depends on whether d or d' was taken (given the realization of some state of nature), then the parties must be able to recognize that d was not taken when d' was taken.[4] If this condition is met, the allocation of authority is irrelevant, in the sense that any Pareto-optimal outcome implementable under full-commitment can also be implemented regardless of who has authority.

This result emphasizes the value of *accounatability*, understood as the possibility of setting limits to authority by making the parties responsible for their decisions. In environments where accountability is possible, compensation schemes can be adjusted to changes in the allocation of authority, muting the effects found in studies where the allocation of authority is the only incentive instrument (see Aghion and Tirole (1997), Dessein (2002), Alonso, Dessein and Matouschek (2008) and Rantakari (2008, 2010), for example). If accountability is not possible, however, the set of implementable outcomes depends in general on the allocation of authority and is a proper subset of the set of implementable outcomes under full-commitment.

In this model, optimal compensation schemes respond to changes in the allocation of authority. Indeed, the party ultimately taking decisions is the one who is held accountable after unplanned decisions are detected; optimal contracts penalize the party in control even if such decisions lead to a positive outcome.

The second part of the analysis derives a characterization of implementable outcomes in the full-commitment case.[5] This result illustrates when and why, in order to generate

[4] In particular, if the parties can observe ex-post the decision taken or the realization of the state, this condition is satisfied.

[5] This result extends the characterization provided in Zermeño (2011*a*) to this more general framework. In Zermeño (2011*a*), the state of nature is binary and the expert's effort cannot affect output directly.

better incentives to exert effort, it may be optimal to distort decision-making away from what would be efficient given the information available. When a contract induces inefficient decision-making, the parties' preferred decisions given the information available do not always coincide. Thus, such contracts create an endogenous conflict of interest.

It has long been recognized in the literature that, in models where agents must be induced simultaneously to exert effort and to participate in decision-making, optimal contracts often lead to inefficient decision-making given the information available (Lambert (1986), Demski and Sappington (1987), Diamond (1998), Prendergast (2002), Athey and Roberts (2001), Malcomson (2009, 2011), Rantakari (2011), Inderst and Klein (2007), Levitt and Snyder (1997)).[6] These studies focus on contracts that can only specify a single transfer scheme contingent on what the parties can observe ex-post. With such contracts, it is impossible to separate the agents' dual tasks of exerting effort and participating in decision-making (see Zermeño (2011*a*)).[7] This problem does not arise if contracts can specify *menus* of transfer schemes. In this case, separation can be attained because the same decision can be associated with different transfer schemes. Nevertheless, we will see that it can still be optimal to distort decision-making. With menus, optimal distortions are purely driven by differences in the amounts of information that different decisions reveal about the state ex-post.

In environments where all decisions reveal the same amount of information ex-post, optimal contacts always induce efficient decision-making. For example, portfolio managers are

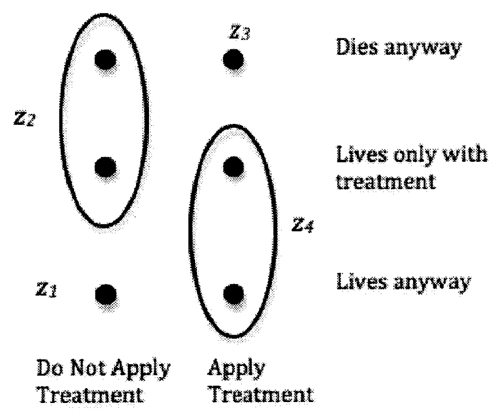
[6] These papers study models with three common features: 1) the agents' private information is non-verifiable; 2) the agents do not have a direct interest in decisions; 3) the agents are induced to work through compensation schemes based on information revealed ex-post. Szalay (2005) analyzes a principal-expert model where the expert preferences over decisions are also aligned with those of the principal. The main differences with respect to the present analysis are that, in Szalay (2005), the expert also has a direct interest in decisions and, since there are no transfer schemes, decision rules are the only incentive instrument. In Prendergast (1993), the principal also receives a (free) private informative signal, and the expert's compensation depends solely on how his report compares to the principal's information. Dewatripont and Tirole (1999) study an information acquisition model where information is verifiable.

Other related studies include: Osband (1989), who focuses on the friction created when the principal does not know the expert's quality at the time of offering a contract; Gromb and Martimort (2007) and Krishna and Morgan (2001) who analyze whether a principal should hire one or two experts; and Lewis and Sappington (1997), Crémer, Khalil and Rochet (1998*a*), Crémer, Khalil and Rochet (1998*b*) and Szalay (2009), who study a Baron-Myerson setup where the agent has the possibility of acquiring information.

[7] A possible solution proposed in Athey and Roberts (2001) is to separate the tasks by assigning them to different agents. However, if the agent's effort is about acquiring non-verifiable information relevant for decision-making, the same agent must inevitably perform both tasks.

hired to take investment decisions based on their assessment of what stock prices are going to be. The decisions that they take do not affect the parties' ability to observe the actual prices ex-post, so optimal contracts induce efficient investment decisions.

There are environments, however, where different decisions reveal different amounts of information ex-post. To illustrate, suppose that a doctor is hired to assess whether a patient should get a treatment. There are three states of nature: *the patient will die anyway, the patient will live only if she gets treatment* and *the patient will live anyway.* Transfer schemes can depend on what is observable ex-post: whether the patient received treatment and whether she survived. This situation is illustrated in the diagram below:



If the patient gets treatment and she survives (z_4), it is not possible to distinguish whether she lived because of the treatment or because she would live anyway. If she gets treatment and dies (z_3), the parties learn that she would have died anyway. On the other hand, if the patient does not get treatment and dies (z_2), the parties cannot recognize between the states *the patient will die anyway* and *the patient will live only if she gets treatment.* If she survives (z_1), they learn that she was going to live anyway.

In environments where decisions can be ranked in terms of the amount of information that they reveal ex-post, inducing more effort requires distorting decision-making in favor of decisions that reveal more information. Intuitively, in order to motivate the expert to exert effort, his compensation must depend on the realization of the state.[8] Taking a decision that does not enable the parties to recognize between two states, forces the expert to receive the same payment after each of the two states is realized, limiting the parties' ability to induce effort.

When decisions cannot be ranked in terms of the amount of information that they reveal ex-post (as in the example above), we cannot say in general how optimal decision rules are going to be distorted. This depends on the features of the information acquisition technology. In the example, if information acquisition is mainly about distinguishing between the states *the patient will die anyway* or *the patient will live only if she gets treatment*, then it is optimal to over-treat the patient, since the decision to apply the treatment reveals which of these two states was realized. However, if information acquisition is mainly about distinguishing between the states *the patient will live anyway* or *the patient will live only if she gets treatment*, then it is optimal to under-treat the patient.[9]

The rest of the paper is structured as follows: section 2.2 introduces the model, and defines the basic concepts that will be used throughout the paper. Section 2.3 contains the results regarding the role of authority and accountability in this incentive problem. Section 2.4 provides a characterization of implementable outcomes in the full-commitment case, which illustrates when and why it may be optimal to distort decision-making. Section 2.5 provides an example that shows how this characterization can be used to derive the shape of optimal contracts. Moreover, it illustrates how the set of implementable outcomes may depend on the allocation of authority when accountability is not possible. Finally, section 2.6 concludes.

[8] This is true regardless of whether effort influences the realization of the state or purely generates information about what the realization is going to be.

[9] A situation in which information acquisition is mainly about distinguishing between the states *the patient will die anyway* or *the patient will live anyway* is unlikely to be relevant in this environment because, since the efficient decision is the same in both cases (do not treat the patient), acquiring this information would not be valuable.

2.2 The model

Consider an environment with two risk-neutral parties, a principal (she) and an expert (he). The principal has unlimited wealth, and is the residual claimant of output, $y(d, \theta)$, which is determined by a decision, $d \in D$, and the state of nature, $\theta \in \Theta$. Assume that $\Theta = \{\theta_1, ..., \theta_N\}$ and D are finite sets. The expert does not have a direct interest in the decision, and is protected by limited liability; he starts with a pledgeable income of ω dollars.[10] After being hired, the expert may exert non-observable effort, $e_0 \in E$, at private cost $c(e_0)$.[11] The principal and the expert have Von Neumann-Morgenstern preferences with Bernoulli utility functions $u_P = y - t$ and $u_E = t - c(e)$ respectively, where t is a transfer made from the principal to the expert.

The expert's effort may have two effects. First, it may affect output directly by shifting the distribution of the state of nature. This effect is captured by the fact that the parties' (common) prior over the state of nature may be contingent on effort; let $x^p(e_0) \in int(X)$, where X is the set of all possible distributions over Θ (the *simplex*), denote the prior.[12] Second, effort may generate information about the future realization of the state of nature; each level of effort creates an *experiment* (that is, a joint probability measure over a signal and the state of nature).[13] After an experiment is created, the expert privately observes the realization of the signal, and, through Bayes rule, generates a posterior, $x^0 \in X$, over the states of nature (x_i^0 denotes that probability that $\theta = \theta_i$). Note that the experiment created with effort, $e_0 \in E$, generates a probability measure over *posteriors* denoted by $P_{e_0}(\cdot)$. Thus, an *information acquisition technology* can be characterized by a cost function, $c : E \rightarrow \mathbb{R}$, and a set of probability measures over posteriors, $\{P_{e_0}(\cdot)\}_{e_0 \in E}$, with the property that $\mathbb{E}_{P_{e_0}}[x] = x^p(e_0)$ for all $e_0 \in E$ (by the law of iterated expectations, the expected value of the posterior must be the prior). Summarizing, after exerting effort, $e_0 \in E$, the expert

[10] The assumptions of risk-neutrality and limited liability are not essential for the results, but make the exposition cleaner. The same results would still be obtained in an environment with risk-averse parties.

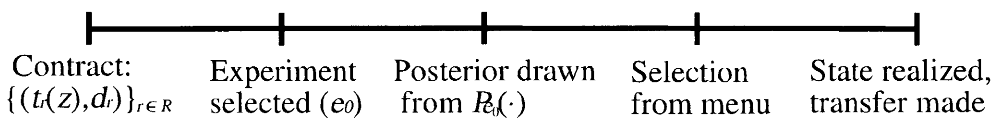
[11] The set E could potentially be multidimensional, so the model embeds multi-tasking framework such as Holmström and Milgrom (1991).

[12] The statement that the parties share the same prior is a statement about the whole function $x^p(e)$.

[13] The notion of *experiment* here is the same as in Blackwell (1953).

privately observes a posterior, $x^0 \in X$, drawn from the probability measure, $P_{e_0}(\cdot)$. The vector, $x^0 \in X$, is the expert's *type* in the mechanism design jargon, so the expert's payoff at the interim stage is *linear* in his type.

Transfer schemes are payments from the principal to the expert. They can be contingent on the contractible variable, $z(d, \theta)$, which describes everything that is observable after the decision is taken and the state is realized.[14] A *full-commitment contract* specifies a menu of pairs, $\{(t_r(z), d_r)\}_{r \in R}$, where $r \in R$ is an index. After exerting effort, but before the state is realized, the expert can choose a pair from the menu. The expert's selection, $r_0 \in R$, determines the transfer scheme, $t_{r_0}(z)$, through which payments are made, and the decision to be taken, $d_{r_0} \in D$. Without loss of generality, we consider contracts in which the expert forfeits to the principal his whole pledgeable income (ω) at the beginning, and receives in return *non-negative* transfers. The following timeline summarizes the sequence of events:



First, a contract is signed by the parties. Then, the expert exerts effort, e_0, at cost $c(e_0)$, and observes a posterior drawn from $P_{e_0}(\cdot)$. At this point, the expert selects an element from the menu, determining his compensation scheme and the decision to be taken. Finally, the state is realized and the payment is made.

2.2.1 Model discussion

This formulation of the principal-agent problem is quite general; it embeds as special cases the standard moral hazard framework, where the expert's effort affects output *directly*, and the case where the agent is a pure expert, in the sense that his effort only generates information. The *state of nature* plays a crucial role in this formulation. It must satisfy two properties: 1) its distribution must be independent from the decision taken; 2) in conjunction with the

[14] The codomain of this function is left unspecified because its natural specification depends on the environment under consideration.

decisions, it must account for all possible values that output can take, and must describe everything that is observable at the time payments are made. For example, the standard moral hazard model corresponds to the case where $D = \{d_0\}$, $\Theta = V \times P$, $y(d, \theta) = v$ with $v \in V$, $z(d, \theta) = p$ with $p \in P$, and $P(\cdot \mid e_0)$ assigns probability one to $x^p(e_0)$. The case where effort is purely about information acquisition has $x^p(e) \equiv x^p$. The main results of the paper hold for any information acquisition technology. The assumption that the spaces of decisions and states are finite is made for technical convenience.

The model does have a lot of structure in some important respects. First, as is standard in the literature, the technology is common knowledge. Moreover, at the time of signing, the parties have the same information, since they share the same prior about the state of nature. The analysis if this assumption were relaxed seems challenging; contracts would have to solve a double screening problem because the expert would still acquire new private information after signing.

The expert has two special features. First, he does not have an intrinsic interest in decisions. This assumption has been made because one of the purposes of this study is to assess what kind of biases regarding decision-making may emerge endogenously. Although it is natural in some environments, it may be strong in others. For example, a doctor might care about his patients, or different decisions might have different private implementation costs for a manager. Even in these cases, the present analysis serves as a useful benchmark. Second, the information that he acquires is non-verifiable. This assumption captures situations in which the expert's findings may be difficult to communicate unequivocally to the principal (maybe) because of her lack of time or expertise. The structure of the problem would change drastically if information were verifiable.

We have imposed one important restriction: full-commitment contracts specify only *deterministic* decisions. If full-commitment contracts could specify lotteries over decisions instead, the parties would generally be able to improve. However, committing to such decision rules is more challenging. Appendix 2.7.2 expands on these points.

2.2.2 Basic concepts

An outcome is a complete description of the variables that determine the size of the surplus generated by a relationship, and the way it is split between the parties.

Definition 2.2.1. *An outcome is given by the triplet* $(e_0, d(x), \mathbb{T})$, *where* $e_0 \in E$ *is the effort exerted by the expert,* $d(x)$ *denotes the decision rule, which maps each posterior,* $x^0 \in X$, *to some decision,* $d(x^0) \in D$, *and* $\mathbb{T} \in \mathbb{R}_+$ *is the expert's ex-ante expected payment.*

Since both parties are risk-neutral, the size of the surplus depends solely on the effort exerted by the expert and on the way decisions are taken. If the parties were risk-averse, then the surplus would also depend on the way the risk is shared. It is this feature that motivated the assumption of risk-neutrality; it makes the exposition cleaner. The expert's expected payment, $\mathbb{T}$, determines how the surplus is split between the parties.

Since both parties must be willing to sign a contract, outcomes must be *individually rational* (IR).

Definition 2.2.2. *The outcome* $(e_0, d(x), \mathbb{T})$ *is said to be individually rational (IR) if:*

1. $\mathbb{T} - c(e_0) \geq \omega$.

2. $\omega + \mathbb{E}_{P_{e_0}}[\mathbb{E}_x[y(d(x), \theta)]] - \mathbb{T} \geq 0$.

The first condition states that the benefit derived by the expert from the relationship must exceed the amount of money that he brings to the table. The second condition is the principal's IR constraint, and reflects the fact that she receives ω from the expert at the beginning. Both parties' outside options are normalized to zero.

Next we define *implementability*. We take into account that the effort exerted by the expert and the information that he receives in the interim stage are non-verifiable. Thus, a contract must induce the expert to act in accordance with the outcome to be implemented. Note that, in the interim stage, the expert will choose one of the elements of the menu, $\{(t_r(z), d_r)\}_{r \in R}$, that gives him the highest expected payment conditional on the information that he received. Therefore, each contract induces a *conditional expected payment function*, $T(x) \equiv \max_{r \in R} \mathbb{E}_x[t_r(z(d_r, \theta))]$. Then we have:

Definition 2.2.3. *The IR outcome* $(e_0, d(x), \mathbb{T})$ *is implementable if there exists a contract,* $\{(t_r(z), d_r)\}_{r \in R}$ *(with its induced conditional expected payment function,* $T(x)$*), such that:*

1. $\mathbb{E}_{P_{e_0}}[T(x)] = \mathbb{T}$.

2. $e_0 \in \arg\max_{e \in E} \ \{\mathbb{E}_{P_e}[T(x)] - c(e)\}$.

3. *For all* $x^0 \in X$*, there exists* $r_0 \in \arg\max_{r \in R} \ \mathbb{E}_{x^0}[t_r(z(d_r, \theta))]$ *such that* $d_{r_0} = d(x^0)$.

4. *All transfers are non-negative.*

The first condition states that the expert's expected payment must indeed be $\mathbb{T}$. The second condition reflects the fact that effort is not observable, so the expert must be willing to exert e_0. The third condition establishes that, at the interim stage, the expert's optimal choice from the menu must be consistent with the decision rule to be implemented. The fourth condition is the limited-liability constraint.

Note that definition 2.2.3 requires the expert to have a best response at the interim stage for *every* possible posterior. In principle, it could be the case that more outcomes could be implemented if such requirement were relaxed to hold *only* for posteriors in the support of $P_{e_0}(\cdot)$. Appendix 2.7.1 shows that this is actually not the case.

Pareto-optimal outcomes are those observed when the parties design their contracts optimally. Formally,

Definition 2.2.4. *The IR and implementable outcome* $(e_0, d(x), \mathbb{T})$ *is Pareto-dominated by the implementable outcome* $(e', d'(x), \mathbb{T}')$ *if either*

1. $\mathbb{T}' - c(e') \geq \mathbb{T} - c(e_0)$;

2. $\mathbb{E}_{P_{e'}}[\mathbb{E}_x[y(d'(x), \theta)]] - \mathbb{T}' \geq \mathbb{E}_{P_{e_0}}[\mathbb{E}_x[y(d(x), \theta)]] - \mathbb{T}$;

with at least one strict inequality; or

1. $\mathbb{T}' - c(e') = \mathbb{T} - c(e_0)$;

2. $\mathbb{E}_{P_{e'}}[\mathbb{E}_x[y(d'(x), \theta)]] - \mathbb{T}' = \mathbb{E}_{P_{e_0}}[\mathbb{E}_x[y(d(x), \theta)]] - \mathbb{T}$;

3. $\mathbb{E}_{x^0}[y(d'(x^0), \theta)] \geq \mathbb{E}_{x^0}[y(d(x^0), \theta)]$ *for all* $x^0 \in X$, *with strict inequality for at least one* $x^0 \in X$.

The first part of the definition states that an outcome is Pareto-dominated if there exists another implementable outcome that makes both parties better off at the time of signing the contract. The second part establishes the assumption that, if two outcomes generate the same expected payoffs, the parties will prefer the one with the most efficient decision rule. We say that an implementable outcome is *Pareto-optimal* if it is not Pareto-dominated by any other outcome. Finally,

Definition 2.2.5. *A decision* d_0 *is more revealing than a decision* d' *if* $z(d_0, \theta) = z(d_0, \theta')$ *implies* $z(d', \theta) = z(d', \theta')$ *for any* $\theta, \theta' \in \Theta$. *That is,* $z(d_0, \theta)$ *induces a finer partition over the set* Θ *than* $z(d', \theta)$.

Note that it may not be possible to rank decisions in terms of how revealing they are.

2.3 The role of authority and accountability

The notion of implementability established in definition 2.2.3 (the full-commitment case) relies on contracts, $\{(t_r(z), d_r)\}_{r \in R}$, that effectively give the expert *restricted authority* over decisions. That is, after choosing a transfer scheme, say $t_0(z) \in \{t_r(z)\}_{r \in R}$, the expert has the right to choose any decision from the set $\{d_r \in D \mid r \in R,\ t_r(z) \equiv t_0(z)\}$. There are environments, however, where committing to such arrangements may be problematic. In some situations authority may only reside at the top (Baker, Gibbons and Murphy (1999)), and the principal may have trouble committing not to overrule the expert's choice. In other cases, where the principal may credibly commit not to intervene, the expert's discretion may be hard to restrict.

This section provides conditions under which Pareto-optimal outcomes, implementable under full-commitment, can also be implemented when one of the parties has the right to take *any* decision after observing the expert's choice of a transfer scheme and his recommendation.

We begin by introducing two stronger notions of implementability than the one established in definition 2.2.3.

Definition 2.3.1. *The IR outcome* $(e_0, d(x), \mathbb{T})$ *is implementable under expert-authority if there exists a contract* $\{t_r(z)\}_{r \in R}$ *such that:*

1. $\mathbb{E}_{P_{e_0}}[T(x)] = \mathbb{T}$;
2. $e_0 \in arg \max_{e \in E} \{\mathbb{E}_{P_e}[T(x)] - c(e)\}$;
3. *for all* $x^0 \in X$, *there exists* $r_0 \in R$ *s.t.* $(r_0, d(x^0)) \in arg \max_{r \in R, d \in D} \mathbb{E}_{x^0}[t_r(z(d, \theta))]$;
4. *all transfers are non-negative;*

where $T(x) \equiv \max_{r \in R, d \in D}\{\mathbb{E}_x[t_r(z(d, \theta))]\}$.

The only difference with respect to definition 2.2.3 is that here contracts specify only a menu of transfer schemes, and the expert may take any decision, $d \in D$, after making a selection from the menu.

Definition 2.3.2. *The IR outcome* $(e_0, d(x), \mathbb{T})$ *is implementable under principal-authority if there exists a contract,* $\{t_r(z)\}_{r \in R}$, *a message space,* M, *strategies,* $r^* : X \to R$, $m^* : X \to M$, *and* $d^* : R \times M \to D$, *and beliefs,* $x^* : R \times M \to X$, *such that:*

1. $\mathbb{E}_{P_{e_0}}[T(x)] = \mathbb{T}$;
2. $e_0 \in arg \max_{e \in E} \{\mathbb{E}_{P_e}[T(x)] - c(e)\}$;
3. *for all* $x^0 \in X$, $d^*(r^*(x^0), m^*(x^0)) = d(x^0)$;
4. *all transfers are non-negative;*
5. *for all* $x^0 \in X$, $(r^*(x^0), m^*(x^0)) \in arg \max_{r \in R, m \in M} \mathbb{E}_{x^0}[t_r(z(d^*(r, m), \theta))]$;
6. *for all* $(r_0, m_0) \in R \times M$, $x^*(r_0, m_0) = \mathbb{E}_{P_{e_0}}[x \mid r^*(x) = r_0, m^*(x) = m_0]$ *if* $r^{*-1}(r_0) \bigcap m^{*-1}(m_0) \neq \varnothing$, *and can be anything otherwise;*

7. for all $(r_0, m_0) \in R \times M$, $d^*(r_0, m_0) \in arg\max_{d \in D} \mathbb{E}_{x^*(r_0,m_0)}[y(d, \theta) - t_{r_0}(z(d, \theta))]$;

where $T(x) \equiv \max_{r \in R, m \in M} \mathbb{E}_x[t_r(z(d^*(r, m), \theta))]$.

That is, an outcome is implementable under principal-authority if it is induced by a Perfect Bayesian Equilibrium of the game where, given a contract $\{t_r(z)\}_{r \in R}$, the principal may choose any decision after observing the expert's choice from the menu and recommendation.[15] Conditions (1)-(4) are analogous to those in definition 2.2.3. Condition (5) states that the expert's strategy, $(r^*(\cdot), m^*(\cdot))$, is optimal given the principal's strategy, $d^*(\cdot)$. Condition (6) guarantees that the principal's beliefs are consistent given the expert's optimal behavior. Condition (7) states that the principal's strategy, $d^*(\cdot)$ is optimal given her beliefs. Note that the principal must take into account the effect that decisions have on output.

Following Aghion and Tirole (1997), we say that an outcome, $(e_0, d(x), \mathbb{T})$, is implementable under principal-authority with a *rubber-stamping scheme* if the conditions in definition 2.3.2 are met with $M = D$ and $m^*(x^0) = d(x^0)$, for all $x^0 \in X$. Rubber-stamping schemes have the attractive feature that they minimize the amount of communication that needs to take place between the expert and the principal in order to implement a given decision rule. The expert only needs to be able to describe the decision that is to be taken (or the principal needs to be able to observe the decision that the expert intends to take); a detailed description of the information acquired (which may be impossible) is not required. They also have a natural interpretation: the expert is taking the decisions under the principal's supervision. He knows that if he takes an unplanned decision, he will be overruled, so, along the equilibrium path, his decisions are always carried out.

Evidently, outcomes implementable under expert-authority or principal-authority can also be implemented under full-commitment. Here we provide conditions under which the converse is also true:

Assumption 2.3.1. *For all* $d, d' \in D$ *and* $\theta_i \in \Theta$, *if* $z(d', \theta_i) \neq z(d, \theta_i)$, *then* $z(d', \theta_i) \notin \{z(d, \theta) \mid \theta \in \Theta\}$.

[15] See Fudenberg and Tirole (1991) for a formal definition of Perfect Bayesian Equilibrium.

In words, suppose that, given a benchmark decision, d, there exists a *significant* deviation (i.e. a decision $d' \neq d$ such that, for some $\theta_i \in \Theta$, $z(d', \theta_i) \neq z(d, \theta_i)$). Then, when the decision d' is taken and θ_i is realized, it is possible to recognize that d was not taken. Note that, if transfers may depend explicitly on the state of nature or on decisions, then assumption 2.3.1 is satisfied; the converse is not true. Then, we have:

Proposition 2.3.1. *If assumption 2.3.1 is satisfied,*

(a) Any outcome implementable under full-commitment can also be implemented under expert-authority. Implementation requires the expert to be penalized when unplanned decisions are detected.[16]

(b) Any Pareto-optimal outcome implementable under full-commitment can also be implemented under principal-authority with a rubber-stamping scheme. Implementation requires the principal to be penalized when unplanned decisions are detected.

Proof. See appendices 2.7.3 and 2.7.4. □

In other words, under assumption 2.3.1, the allocation of authority is irrelevant in terms of the outcomes that are implementable. However, it has a substantial effect on the way compensation schemes should be structured. It is the party who ultimately takes the decision the one that should be held accountable if unplanned decisions are taken. Note that, under principal-authority, we require outcomes to be Pareto-optimal. This restriction is needed because, when the principal is considering alternative decisions, she also takes into account their effect on output.[17]

Proposition 2.3.1 illustrates the basic need in organizational design, identified in Arrow (1974), to set limits to authority through *accountability*.[18] Assumption 2.3.1 guarantees that

[16] The full-commitment contract that implements the outcome under consideration establishes a *plan* that determines the decision that is to be taken as a function of the transfer scheme selected by the expert. *Unplanned* decisions correspond to deviations from this plan.

[17] There could be a deviation, $d' \neq d(x^0)$, such that $z(d', \theta) = z(d(x^0), \theta)$ for all $\theta \in \Theta$. If such deviation was profitable for the principal, the original outcome would be Pareto-dominated by an outcome with a decision rule specifying that d' is to be taken after x^0 is observed. Thus, Pareto-optimality rules out such deviations.

[18] Arrow (1974) uses the term *responsibility* instead.

it is always possible to detect significant deviations in decision-making, enabling the party responsible for such deviations to be penalized through transfers. In this framework, these penalties suffice to induce the desired decision rule.[19]

Although, in this model, accountability can only be achieved through transfers, in reality it may may be achieved through other channels. For example, the decisions taken may affect a manager's chances of promotion or of getting fired, or a consultant's reputation. Baker, Gibbons and Murphy (1999) analyze a model where the principal always keeps the right to ultimately take any decision. There, accountability is possible because the parties interact repeatedly. The principal's choices are disciplined by the threat that the expert may brake the relationship if he is overruled.

If assumption 2.3.1 is not satisfied, (perfect) accountability is unattainable. In this case, the set of implementable outcomes when full-commitment is not possible generally depends on who is in control. Moreover, it is a subset of the set of implementable outcomes under full-commitment. Section 2.5 provides an example.

Studies where the allocation of authority affects the set of implementable outcomes assume away (implicitly or explicitly) the possibility of accountability. In Aghion and Tirole (1997), Dessein (2002), Alonso, Dessein and Matouschek (2008) and Rantakari (2008), for example, transfers are ruled out. Athey and Roberts (2001) and Rantakari (2011) restrict attention to transfer schemes which are linear in the performance measure. These restrictions may not be innocuous when considering situations where deviations in decision-making could be detected ex-post, making accountability possible. In such situations, the effect of changes in the allocation of authority may be muted by adjusting the compensation schemes.

[19] Although the expert is protected by limited liability, he has no intrinsic interest in decisions. Thus, a payment of zero is always punishment enough. The principal, on the other hand, has a direct interest in output. However, since she is assumed to have unlimited wealth, her penalty if she deviates may be arbitrarily large.

2.4 A characterization of implementable outcomes; when and how will decision-making be distorted?

Given a full-commitment contract, $\{(t_r(z), d_r)\}_{r\in R}$, the expert's incentives for effort are entirely determined by the conditional expected payment function, $T(x) \equiv \max_{r\in R} \mathbb{E}_x[t_r(z(d_r, \theta))]$, induced by the contract. Indeed, the expert's choice of effort can be viewed as a choice of a probability measure over posteriors, and the function, $T(x)$, specifies the expert's actual expected payment as a function of the posterior observed. This section presents a characterization of implementable outcomes that expresses this incentive problem only in terms of the function $T(x)$.[20] The value of this result is twofold. First, it simplifies the problem considerably, since functions $T(x)$ effectively replace contracts as the incentive instrument; it is *as if* we could choose a payment scheme for the expert contingent directly on posteriors.[21] Second, the characterization makes explicit what restrictions are imposed on the functions $T(x)$ that can be chosen due to the fact that posteriors are not actually directly contractible. In particular, it illustrates how the decision rule to be implemented restricts the functions $T(x)$ that can be chosen and, thus, the incentives that can be provided for the expert to acquire information. In other words, the characterization uncovers a tradeoff that is central in this incentive problem; it allows us to understand when and why it may be optimal to distort decision making away from what would be efficient (conditional on the information available) in order to improve the expert's incentives to acquire information.

Let $\bar{X} = \{x \in \mathbb{R}^N \mid \sum_{i=1}^N x_i = 1\}$ be the *extended* simplex. For all $i \in \{1, ..., N\}$, let $x^i \in X$ be the posterior that assigns probability one to state θ_i, and X^{-i} be the set of posteriors that assign probability zero to state θ_i. Finally, for any vector $v \in \mathbb{R}^N$ with $\sum_{i=1}^N v_i = 0$, and convex function $T : \bar{X} \to \mathbb{R}$, let

$$T'(x^0 \mid v) = \lim_{h\to 0^+} \frac{T(x^0 + hv) - T(x^0)}{h}$$

[20] This result extends the characterization in Zermeño (2011*a*) to this more general framework.

[21] In Zermeño (2011*a*) I use this characterization to solve for optimal contracts in the case where the state of nature is binary and where the expert's effort is only about generating information.

be the right directional derivative of $T(x)$ at x^0 in direction v. Now we can state the result:

Theorem 2.4.1. *The IR outcome* $(e_0, d(x), \mathbb{T})$ *is implementable if and only if there exists* $T : \bar{X} \to \mathbb{R}$ *such that:*

1. $\mathbb{E}_{P_{e_0}}[T(x)] = \mathbb{T}$.

2. $e_0 \in \arg\max_{e \in E} \mathbb{E}_{P_e}[T(x)] - c(e)$.

3. $T(x)$ *is convex.*

4. $T(x^{-i}) + T'(x^{-i} \mid x^i - x^{-i}) \geq 0 \;\; \forall \;\; i = 1, ..., N, \quad x^{-i} \in X^{-i}$.

5. *For any* $x^0 \in X$, *and* $v \in span\left(\{x^i - x^j \mid z(d(x^0), \theta_i) = z(d(x^0), \theta_j)\}\right)$,

$$T\left(x^0\right) \leq T(x^0 + v).$$

Proof. See Appendix 2.7.5. □

That is, an outcome is implementable if and only if there exists a function, $T(x)$, that satisfies five conditions. The main difference with respect the definition of implementability (definition 2.2.3) is that functions, $T(x)$, replace contracts as the incentive instrument. The first condition states that the expert's ex-ante expected payment must indeed be the one specified by the outcome. The second condition is the expert's effort incentive compatibility constraint. These two conditions would be present even if we could offer the expert a payment scheme contingent on the posterior that he observes. Conditions (3)-(5) must be included because posteriors are actually not directly verifiable. In order to understand these conditions, next we will go over the intuition behind the result.

2.4.1 The intuition behind Theorem 2.4.1

When the expert makes a selection, $r_0 \in R$, from a full-commitment contract, $\{(t_r(z), d_r)\}_{r \in R}$, the selection determines a set of state-contingent payments, $\{t_{r_0}(z(d_{r_0}, \theta_i))\}_{i=1}^{N}$, or a *lot-*

tery. In the binary-state case, each lottery corresponds to a line in figure 2.4.1a. Indeed, the horizontal axis describes the posterior observed by the expert, and each point in the line describes the expert's expected payment conditional on observing a posterior after making that particular selection.[22] In general, a selection, $r_0 \in R$, from a full-

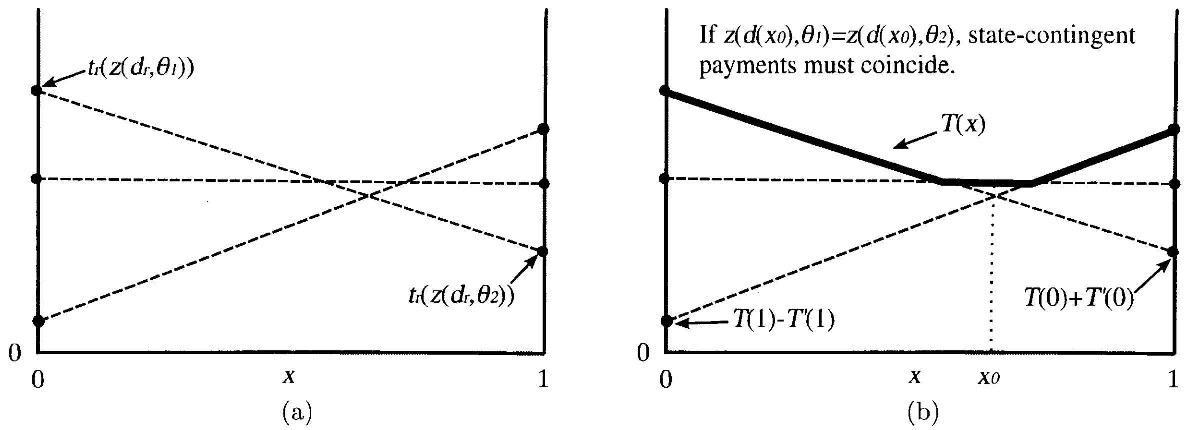


Figure 2.4.1: Theorem 2.4.1 - Necessity

commitment contract, $\{(t_r(z), d_r)\}_{r \in R}$, corresponds to the hyperplane in $\mathbb{R}^{N+1}$ given by $\{(x, w) \in \mathbb{R}^N \times \mathbb{R} \mid w = \sum_{i=1}^N x_i t_{r_0}(z(d_{r_0}, \theta_i))\}$. Then, from the expert's perspective, a contract boils down to a set of hyperplanes. After observing a posterior, $x^0 \in X$, it is in the expert's best interest to choose the element form the contract that corresponds to the highest hyperplane at x^0. Thus, the expert's actual expected payment as a function of his information is given by the upper envelope of the hyperplanes induced by the contract (the function $T(x)$ in figure 2.4.1b).

To prove that conditions (1)-(5) in Theorem 2.4.1 are necessary, one can show that they are satisfied by the upper envelope, $T(x)$, induced by the contract that implements the outcome.[23] Conditions (1) and (2) follow directly from definition 2.2.3. Condition (3) is a consequence of the fact that $T(x)$ is the upper envelope of hyperplanes.[24] This condition

[22] With two states, a posterior can be describes by a number in the interval $[0, 1]$, where $x = 1$ means that the expert is sure that $\theta = \theta_2$.

[23] The upper envelope can be defined over the extended simplex as $T(x) = \sup_{r \in R} \sum_{i=1}^N x_i t_r(z(d_r, \theta_i))$ for all $x \in \bar{X}$.

[24] The convexity of the agents' indirect utility function is a central feature of other characterizations of

has a clear economic interpretation. The constraint that the expert's expected payment must be convex in his posterior is saying that the expert cannot be worse off by having more non-verifiable information.[25] Condition (4) follows from limited liability. Figure 2.4.1b illustrates the argument in the case where the state of nature is binary. Since the contract that implements the outcome satisfies limited liability, all the points on the sides of the figure must be non-negative. The lowest point on the right-hand side can be expressed as $T(0) + T'(0)$ and the lowest point on the left-hand side as $T(1) - T'(1)$. Condition (4) states that these expressions must be non-negative.[26] The intuition in the general case is analogous. Finally, condition (5) must be satisfied. Figure 2.4.1b illustrates the argument for the binary-state case. If the decision taken at x^0 has $z(d(x^0), \theta_1) = z(d(x^0), \theta_2)$ then the highest lottery at that point must pay the same amount regardless of the realization of the state. This implies that the upper envelope of the lotteries must reach its minimum at x^0.[27]

Conversely, it can be shown that if we start with a function $T(x)$ that satisfies conditions (1)-(5) given an outcome, $(e_0, d(x), \mathbb{T})$, then we can construct a contract that implements the outcome. Figure 2.4.2a illustrates how to construct the transfer schemes in the binary-

implementable outcomes in mechanism design environments where the agents' utility is linear in their type (e.g. Rochet (1987), Jehiel and Moldovanu (2001) and Müller, Perea and Wolf (2007)). More generally, Krishna and Maenner (2001) show that the agents' indirect utility function is convex in their type in any mechanism design environment where the agents' utility is convex in their type.

[25] Indeed, the experiment characterized by the probability measure over posteriors, $P_e(\cdot)$, is more informative in Blackwell's sense than the experiment characterized by $P_{e'}(\cdot)$ if and only if $P_e(\cdot)$ is a mean-preserving spread of $P_{e'}(\cdot)$. Requiring $T(x)$ to be convex, is equivalent to requiring that the expert's ex-ante expected utility must be higher under $P_e(\cdot)$ than under $P_{e'}(\cdot)$ whenever $P_e(\cdot)$ is a mean-preserving spread of $P_{e'}(\cdot)$.

[26] Figure 2.4.1b contains an abuse of notation. Indeed, strictly speaking, a posterior in the binary-state case should be a two-dimensional vector, whereas in the figure we are representing posteriors in terms of a scalar specifying the probability that $\theta = \theta_2$. Thus, the expression $T(0) + T'(0)$ corresponds to

$$T(x^1) + T'(x^1 \mid x^2 - x^1) = T(x^1) + \lim_{h \to 0^+} \frac{T(x^1(1-h) + hx^2) - T(x^1)}{h}.$$

Similarly, the expression $T(1) - T'(1)$ corresponds to

$$T(x^2) + T'(x^2 \mid x^1 - x^2) = T(x^2) + \lim_{h \to 0^+} \frac{T(x^2(1-h) + hx^1) - T(x^2)}{h}.$$

[27] More generally, take any $x^0 \in X$ and $v \in span\left(\{x^i - x^j \mid z(d(x^0), \theta_i) = z(d(x^0), \theta_j)\}\right)$. The highest hyperplane at x^0 cannot change its hight as we reshuffle weight between states that induce the same payment. Thus, the height of this hyperplane at x^0 must be the same as its height at $x^0 + v$ (and equal to $T(x^0)$). Since $T(x)$ is the upper envelope of all the hyperplanes, this implies that $T(x^0 + v) \geq T(x^0)$.

state case. For each $x^0 \in X$, $t_{x^0}(z(d(x^0), \theta_i))$ for $i \in \{1, ..., N\}$ can be constructed using a

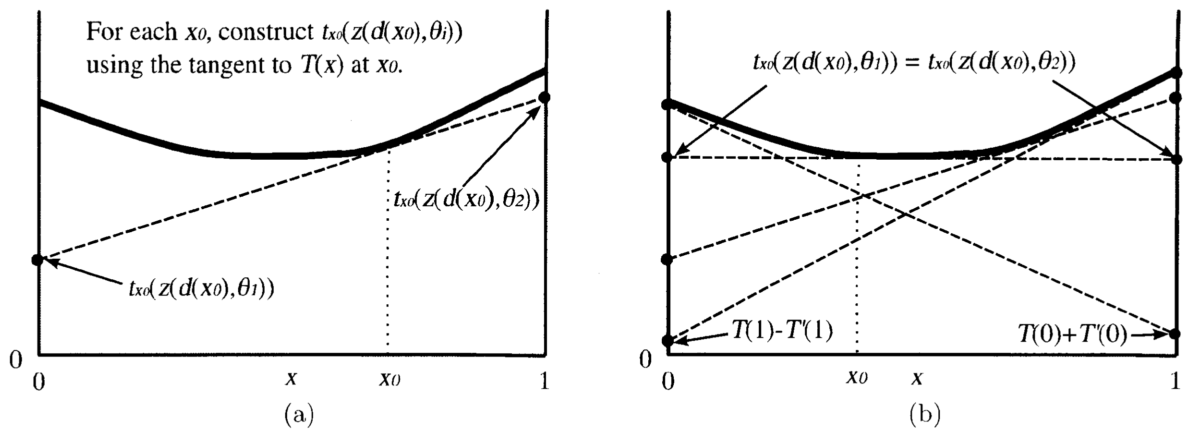


Figure 2.4.2: Theorem 2.4.1 - Sufficiency

supporting hyperplane of the graph of $T(x)$ at x^0 ($t_{x^0}(z)$ can be any non-negative number for any other value of z). Condition (5) guarantees that there always exists a supporting hyperplane such that the transfer schemes constructed in this way are well-defined. That is, if $z(d(x^0), \theta_i) = z(d(x^0), \theta_j)$, then $t_{x^0}(z(d(x^0), \theta_i)) = t_{x^0}(z(d(x^0), \theta_j))$ (see figure 2.4.2b). Then the contract, $\{(t_x(z), d(x))\}_{x \in X}$, can be shown to implement the outcome. Condition (3) guarantees that the original $T(x)$ is the upper envelope of the lotteries induced by the contract, so, by construction, the contract implements the level $\mathbb{T}$, the level of effort e_0 and the decision rule, $d(x)$. Finally, condition (4) guarantees that the constructed transfer schemes satisfy limited liability (see figure 2.4.2b). The fact that $T(x)$ is defined over the extended simplex comes into play to make sure that the supporting hyperplanes at the points in the boundary of the simplex are not vertical, so that we can use them to construct the transfers at these points.

2.4.2 The quality of decision making vs the quality of information

Theorem 2.4.1 uncovers a tradeoff that is central in this problem. This tradeoff is summarized by condition (5), as it illustrates the way the decision rule to be implemented restricts the set of functions, $T(x)$, that can be chosen. Since $T(x)$ is our instrument to induce the expert to exert effort, this condition determines when and why it may be optimal to distort decision-making (conditional on the information available) to induce more information acquisition.

There are environments where the decision taken does not affect the amount of information about the state that is revealed ex-post. For example, a portfolio manager's decisions do not affect our ability to observe stock prices ex-post. Formally, in this example we have that for all $d \in D$ and $\theta \neq \theta'$, $z(d, \theta) \neq z(d, \theta')$. In this case, condition (5) disappears and the decision rule to be implemented imposes no restrictions on the functions $T(x)$ that can be chosen. Thus, here, the decision rule to be implemented can be chosen independently from the level of effort to be induced, so there is no reason to distort decision making. More generally, we can see that, when all decisions reveal the same amount of information about the state ex-post, there is nothing to be gained by implementing decision rules that are not efficient. Indeed, when this is the case, changing the decision rule does not alter the restrictions imposed by condition (5) on $T(x)$.

Nevertheless, there are situations where the decision taken does affect the amount of information about the state that is revealed ex-post. For instance, the decision could be whether or not to invest in a project, and the state could represent all the possible outcomes of the project. In this case, the state is only revealed ex-post if the decision is to invest in the project (otherwise, the parties never observe the counterfactual).[28] Theorem 2.4.1 implies that, in environments where decisions can be ranked in terms of how revealing they are, the parties could improve the incentives to acquire information by distorting decision making in favor decisions that reveal more information ex-post. Indeed, condition (5) states that for each $x^0 \in X$, $T(x^0)$ must be the minimum of $T(x)$ when we restrict its domain to $\{x \in \bar{X} \mid x = x^0 + v,\ v \in span(\{x^i - x^j \mid z(d(x^0), \theta_i) = z(d(x^0), \theta_j)\})\}$. Since $T(x)$ must be convex, if the decision $d(x^0)$ is taken at any two points x, x' within this set, then it must be the case that $T(\alpha x + (1 - \alpha)x') = T(x^0)$ for all $\alpha \in [0, 1]$. When information acquisition generates dispersion of posteriors along this line, allowing $T(x)$ to actually vary would improve the incentives for effort. However, if decisions can be ranked in terms of how

[28] The models analyzed in Lambert (1986), Levitt and Snyder (1997) and Inderst and Klein (2007) have this structure.

revealing they are, this variation can only be achieved by taking a more revealing decision in some points of the line.

When decisions cannot be ranked in terms of how revealing they are, the direction of optimal distortions to decision making becomes ambiguous. It depends on the features of the information acquisition technology. Consider the medical example discussed in the introduction as an illustration. If information acquisition was mainly about distinguishing between the states *dies anyway* or *lives only with treatment*, then it would be optimal to over-treat the patient, since the decision to not apply the treatment would not enables us to tell apart these two states ex-post. However, if information acquisition was mainly about distinguishing between the states *lives anyway* or *lives only with treatment*, then it would be optimal to under-treat the patient. Indeed, treating the patient would not allow us to pay the doctor different amounts after these two states are realized.[29]

2.4.3 The role of limited liability

Condition (4) of Theorem 2.4.1 summarizes the role of limited liability in this problem. In a nutshell, limited liability restricts how convex $T(x)$ can be, in the sense that it limits its steepness at the points in the boundary of the simplex. Formally,

$$T'\left(x^{-i} \mid x^{i} - x^{-i}\right) \geq -T\left(x^{-i}\right) \quad \forall \;\; i = 1, ..., N, \;\; x^{-i} \in X^{-i}.$$

This restriction is important because it is precisely the convexity of $T(x)$ that makes information valuable for the expert, and thus induces him to exert effort (see footnote 25).

Increasing the expert's pledgeable income, ω, relaxes the constraints imposed by limited liability. Indeed, definition 2.2.2 implies that an outcome, $(e_0, d(x), \mathbb{T})$, is individually rational if and only if

$$\omega + c(e_0) \leq \mathbb{T} \leq \omega + \mathbb{E}_{P_{e_0}}[\mathbb{E}_x[y(d(x), \theta)]].$$

[29] A situation in which information acquisition is mainly about distinguishing between the states *dies anyway* or *lives anyway* is unlikely to be relevant in this environment because, since the efficient decision is the same in both cases (do not treat the patient), acquiring this information would not be valuable.

Thus, given any e_0 and $d(x)$, the set of values for $\mathbb{T}$ that make the outcome $(e_0, d(x), \mathbb{T})$ individually rational is increasing in ω. This means that, if we fix the share of the surplus that goes to each party, increases in ω shift the schedule $T(x)$ upwards, relaxing the restrictions imposed by limited liability, as specified in condition (4) of Theorem 2.4.1. In this way, increasing the expert's pledgeable income helps to provide better incentives for information acquisition and decision making. This could explain why partners in private equity funds, for example, are asked to invest a non-trivial fraction of their own wealth in the funds they manage.

2.5 The role of authority without accountability

The goal in this section is to show that, if assumption 2.3.1 is not satisfied, the set of Pareto-optimal outcomes in general depends on the allocation of authority. In addition, we will see how Theorem 2.4.1 can be used to solve for optimal contracts.

Throughout the section we restrict attention to environments with two states of nature and where every decision reveals the state ex-post (i.e. $z(d, \theta_1) \neq z(d, \theta_2)$ for all $d \in D$). Moreover, we will work with a particular class of information acquisition technologies, *Success-Failure* experiments, which makes solving for optimal contracts considerably simpler.[30] Formally, assume that $E = [0, 1]$, and that $c(e)$ is strictly convex, continuously differentiable, and such that $c'(0) = 0$ and $\lim_{e \to 1} c'(e) = \infty$. If the expert exerts effort e, the experiment *fails* with probability $1 - e$, and nothing is learnt (the expert observes the prior, x^p). With probability e, the experiment *succeeds*, and the expert observes an informative signal. The signal generates a distribution over posteriors characterized by the CDF over posteriors, $G(x)$, where x is the probability that $\theta = \theta_2$. By the law of iterated expectations, it must be the case that $\mathbb{E}_G[x] = x^p$. Since effort does not affect the prior, this is a pure information acquisition environment.

We begin by showing that, under these conditions, any implementable outcome can be

[30] Because of their tractability, this class of technologies has been widely used in the literature (see Aghion and Tirole (1997) or Inderst and Klein (2007), for example).

implemented with contracts that have a very simple form.

Proposition 2.5.1. *Suppose that $N = 2$, that every decision reveals the state ex-post and that the information acquisition technology is Success-Failure. Then any implementable outcome can be implemented with a contract that induces only two different state-contingent payments (or lotteries) for the expert. Moreover, the expert is indifferent between the two lotteries exactly when $x = x^p$, and both lotteries have the same low payment.*

In Zermeño (2011*a*), I prove a more general version of this proposition.[31] Figure 2.5.1 illustrates the argument when the information acquisition technology is Success-Failure.

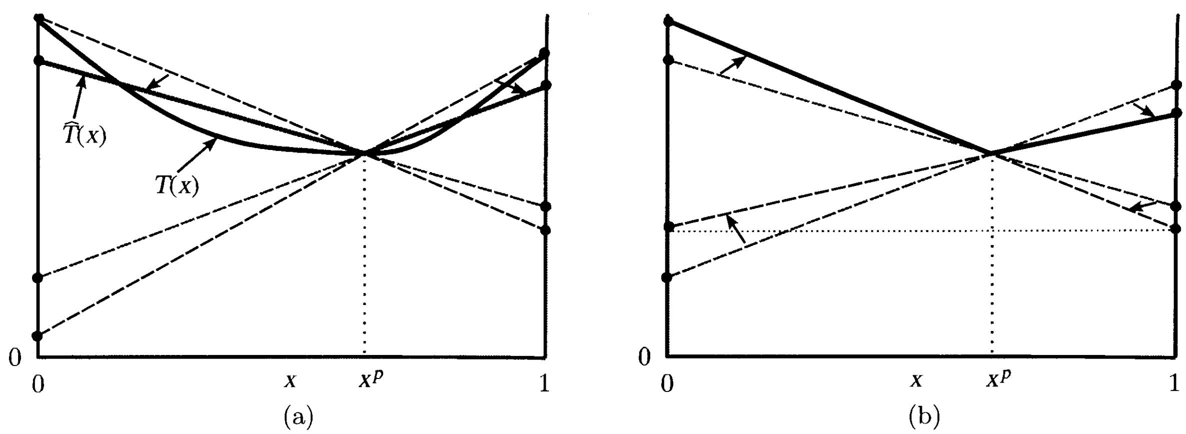


Figure 2.5.1: Proposition 2.5.1: Intuition

Figure 2.5.1a shows how any implementable outcome can be implemented with a contract that induces only two lotteries that make the expert indifferent at the prior. Suppose $T(x)$ satisfies the conditions in Theorem 2.4.1 given the outcome $(e_0, d(x), \mathbb{T})$. Consider the two lotteries that cross at the point $(x^p, T(x^p))$ and that have high payments $T(0)$ and $T(1)$ respectively (the dashed black lines in the figure). By the convexity of $T(x)$, the upper envelope of these lotteries is always above $T(x)$ and satisfies condition (4) of Theorem 2.4.1. Now, as illustrated in Figure 2.5.1a, rotate the two lotteries around the point $(x^p, T(x^p))$

[31] Specifically, I show that in the binary-state case, if the "first-order approach" is valid, then any implementable outcome can be implemented with a a contract that induces at most three different state-contingent payments for the expert. The fact that in Proposition 2.5.1 the expert is indifferent between the two lotteries precisely at the prior is specific to Success-Failure information acquisition technologies.

until their upper envelope, $\hat{T}(x)$, is such that $\mathbb{E}_G[\hat{T}(x)] = \mathbb{E}_G[T(x)]$. The function $\hat{T}(x)$ is the upper envelope of two lotteries that make the expert indifferent at the prior, and satisfies all the conditions in Theorem 2.4.1 given $(e_0, d(x), \mathbb{T})$. Indeed, since $\hat{T}(x^p) = T(x^p)$ and $\mathbb{E}_G[\hat{T}(x)] = \mathbb{E}_G[T(x)]$, conditions (1) and (2) are satisfied.[32] Condition (3) is trivially satisfied, as well as condition (5) (all decisions reveal the state). Finally, condition (4) is satisfied because it was satisfied by the upper envelope of the lotteries before we rotated them.

Figure 2.5.1b illustrates how we can implement the same outcome with a contract that induces two lotteries that make the expert indifferent at x^p *and* that have the same low payment. This follows because, if we rotate the two lotteries constructed in Figure 2.5.1a around the point $(x^p, T(x^p))$, keeping the change in slope of their upper envelope unaltered, the two resulting lotteries induce the same ex-ante expected payment for the expert and the same incentives for effort.[33] Thus, we can rotate the two original lotteries to the point where their low payments coincide, without changing the implemented outcome.

In the rest of the section, we provide an example in which Assumption 2.3.1 is not satisfied and where Pareto-optimal outcomes implementable under full-commitment can be implemented under Principal-authority, but not under Expert-authority. Suppose there are two decisions, d_1 and d_2, which may succeed (lead to high output, $\bar{y}$) or fail (lead to low output, $\underline{y}$). Decision d_i succeeds if and only if $\theta = \theta_i$, and transfer schemes can only depend on whether the decision taken succeeded or failed. Note that both decisions reveal the state ex-post, but Assumption 2.3.1 is not satisfied ($z(d_1, \theta_1) = z(d_2, \theta_2)$ and $z(d_1, \theta_2) = z(d_2, \theta_1)$).

[32] To see this, note that $\mathbb{E}_{P_{e_0}}[\hat{T}(x)] = e_0 E_G[\hat{T}(x)] + (1 - e_0)\hat{T}(x^p) = e_0 \mathbb{E}_G[T(x)] + (1 - e_0)T(x^p) = \mathbb{T}$. Condition (2) is satisfied because, with Success-Failure information acquisition technologies, it is equivalent to $\mathbb{E}_G[T(x)] - T(x^p) = c'(e_0)$.

[33] To see this let $T(x)$ be a convex function that changes slope once and at the prior. Then

$$E_G[T(x)] = G(0)T(0) + \int_0^1 T(x)dG(x) = T(1) - \int_0^1 \Delta(x)G(x)dx = T(x^p) + (1 - x^p)\Delta(1) - \int_0^1 \Delta(x)G(x)dx$$
$$= T(x^p) + (\Delta(1) - \Delta(0)) \int_0^{x^p} G(x)dx,$$

where $\Delta(x)$ is a subderivative of $T(x)$ and all integrals are Riemann-Stieltjes integrals. The second equality is derived by integrating by parts, and the fourth equality follows because $\int_{x^p}^1 G(x)dx = (1 - x^p) - \int_0^{x^p} G(x)dx$.

We will see that, in this example, Principal-authority dominates Expert-authority.

We begin by finding the optimal contract under full-commitment. Let the expert's expected net utility derived from an outcome, $(e_0, d(x), \mathbb{T})$, be denoted by $U_0 \equiv \mathbb{T} - c(e_0) - \omega$. Recall that, since both decisions reveal the state, every Pareto-optimal outcome implements the efficient decision rule. Thus, given U_0, our problem is to find the contract that implements the level of effort that generates the highest expected surplus conditional on efficient decision-making. In fact, for ω and U_0 small enough so that the first-best level of effort is not attainable, our problem is to find the contract that implements the highest possible level of effort. In what follows we focus on this case.

Figure 2.5.2 illustrates how to find the full-commitment contract that maximizes effort given the expert's expected net utility U_0. Proposition 2.5.1 implies that we can restrict our

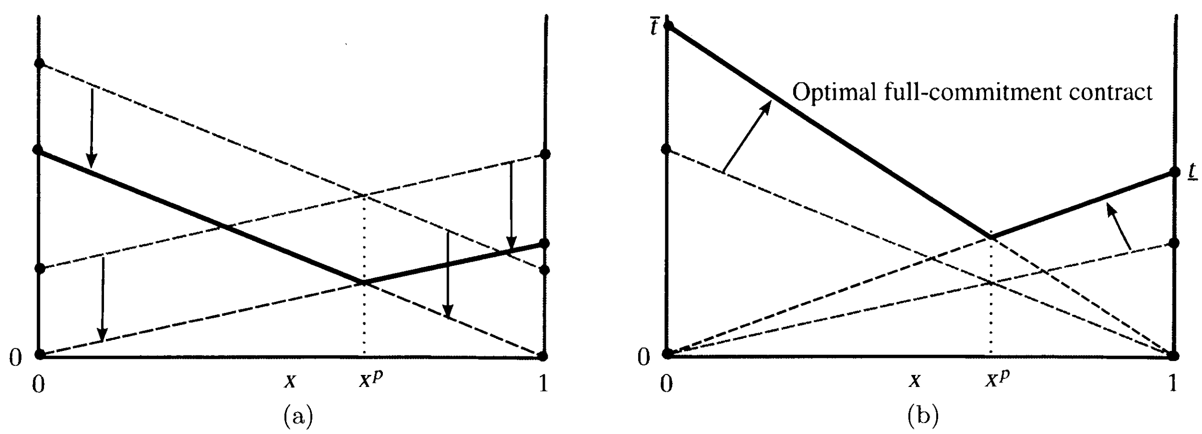


Figure 2.5.2: Optimal full-commitment contracts

search to contracts that induce two lotteries that make the expert indifferent at the prior, and that have the same low payment. In fact, in the contract that maximizes effort, the low payment of both lotteries must be zero, and the high payments are pinned down by U_0 and by the fact that the expert must be indifferent between the lotteries at the prior. To see this, suppose that we start with two lotteries with the same but positive low payment, and that give the expert an expected net utility of U_0 (given his optimal choice of effort). If we reduce all payments by a constant so that the lotteries' low payments become zero, we can implement the same level of effort, but reduce the expert's expected net utility (see Figure 2.5.2a). Then, as illustrated by Figure 2.5.2b, we can increase the lotteries' high payments to the point where the expert's expected net utility gets back to U_0. Since we are restricting attention to pairs of lotteries that make the expert indifferent at the prior, these high payments are uniquely pinned down by U_0. The resulting lotteries implement a higher level of effort.[34] Thus, the optimal full-commitment contract in this example is $\{(\bar{t} \textit{ if the decision succeeds}, d_1), (\bar{t} \textit{ if the decision fails}, d_2), (\underline{t} \textit{ if the decision succeeds}, d_2)\}$. The expert will choose the first option after observing posteriors in $[0, \frac{1}{2}]$, the second option after observing posteriors in $[\frac{1}{2}, x^p]$, and the third option after observing posteriors in $[x^p, 1]$. This contract implements the efficient decision rule (d_1 if $x \leq \frac{1}{2}$ and d_2 if $x > \frac{1}{2}$) and induces the two lotteries that maximize effort given U_0.

Under Expert-authority, Pareto-optimal outcomes implementable under full-commitment cannot be implemented. To see this note that, if a pair $(t(z), d_1)$ induces a lottery with payments $\bar{t}$ if the state is θ_1 and $\underline{t}$ if the state is θ_2, then the pair $(t(z), d_2)$ induces the lottery with payments $\bar{t}$ if the state is θ_2 and $\underline{t}$ if the state is θ_1. This implies that, under Expert-authority, the upper envelope of the lotteries induced by any contract is symmetric around $x = \frac{1}{2}$. Thus, even though with Expert-authority we can still implement the efficient decision rule (this is the case as long as all transfer schemes pay more after success than after failure), we are not able to induce as much effort as with full-commitment (given U_0).

Under Principal-authority, however, we can implement with a rubber-stamping scheme any Pareto-optimal outcome implementable under full-commitment. Indeed to implement the outcome implemented by the contract in Figure 2.5.2b we could let $M = \{d_1, d_2\}$ and offer the expert a contract with three payment options:

1. $\bar{t}$ if the decision succeeds and zero otherwise.

2. $\bar{t}$ if the decision fails and zero otherwise.

[34] From footnote 33 we can see that that if $T(x)$ is convex and changes its slope only once and at the prior, then the experts optimal choice of effort solves $(\Delta(1) - \Delta(0)) \int_0^{x^p} G(x)dx = c'(e)$, where $\Delta(x^0)$ is a subderivative of $T(x)$ at x^0.

3. $\underline{t}$ if the decision succeeds and zero otherwise.

This contract implements the outcome with a rubber-stamping scheme. Suppose the principal's beliefs only depend on the payment scheme selected by the expert (and not on his recommendation). Specifically, suppose that, if the expert selects the first option from the contract, the principal believes that he observed a posterior in $[0, \frac{1}{2}]$ (so the principal's posterior lies in $[0, \frac{1}{2}]$); if the expert selects the second option from the contract, the principal believes that he observed a posterior in $[\frac{1}{2}, x^p]$ (so the principal's posterior lies in $[\frac{1}{2}, x^p]$); if the expert selects the third option from the contract, the principal believes that he observed a posterior in $[x^p, 1]$ (so the principal's posterior lies in $[x^p, 1]$). The principal's best response given these beliefs is to choose d_1 after the expert selects the first option and d_2 after the expert selects the second or third options from the contract.[35] Finally, given the principal's behavior, it is a best response for the expert to select the first option from the contract and recommend d_1 after observing a posterior in $[0, \frac{1}{2}]$, the second option from the contract and recommend d_2 after observing a posterior in $[\frac{1}{2}, x^p]$, and the third option from the contract and recommend d_2 after observing a posterior in $[x^p, 1]$. Thus, the principal's beliefs were indeed consistent.

We have seen that, in environments where Assumption 2.3.1 is not satisfied, the allocation of authority in general matters for the outcomes that can be implemented. In the example that we studied, Principal-authority dominates Expert-authority. Whether this result is more general is an open question.

[35] Indeed, if the principal chooses d_2 instead of d_1 after the expert selects the first option, she would get $x^*(\bar{y} - \bar{t}) + (1 - x^*)\underline{y}$ instead of $x^*\underline{y} + (1 - x^*)(\bar{y} - \bar{t})$, where $x^* \in [0, \frac{1}{2}]$ is the principal's posterior. For this to be a profitable deviation, we would require $\bar{y} - \underline{y} - \bar{t} < 0$, but this would imply that the principal would prefer d_2 to d_1 for every posterior in $[0, \frac{1}{2}]$. Then the Pareto-optimal outcome that we were implementing under full-commitment must violate individual rationality, for the principal would have been better off always taking d_2 without hiring the expert.

After the expert selects the second option, the principal prefers d_2 to d_1 because d_2 maximizes expected output and minimized the expected payment to the expert given $x^* \in [\frac{1}{2}, x^p]$.

Finally, after the expert selects the third option, the principal prefers d_2 to d_1 because $\bar{y} - \underline{y} - \underline{t} \geq \bar{y} - \underline{y} - \bar{t} \geq 0$.

2.6 Concluding remarks

This paper has examined a general principal-expert model. We have identified conditions under which the allocation of authority is irrelevant in terms of the outcomes that can be implemented; under these conditions, any Pareto-optimal outcome implementable under full-commitment can be implemented regardless of who has the right to ultimately take the decision. For this result to hold, it must be possible to penalize the party in control after decisions that differ from the plan established by the full-commitment benchmark are taken, while, at the same time, keeping the expert's incentives to exert effort unaltered. In particular, this is possible if: 1) the decision taken can be identified ex-post; or 2) the realization of the state of nature can be observed ex-post. When the conditions provided are not met, the set of implementable outcomes in general depends on the allocation of authority (and is a subset of the set of outcomes implementable under full-commitment).

These results emphasize the value of *accountability*, understood as the possibility of limiting the authority of the party in control by penalizing her after taking decisions that deviate from the optimal plan under full-commitment. Importantly, it is valuable to penalize such deviations even if they lead to high output.[36] In environments where accountability is possible, compensation schemes can respond to changes in the allocation of authority, muting the effects found in setups where the allocation of authority is the only incentive instrument.

Moreover, the paper provided a general characterization of implementable outcomes under full-commitment that uncovers a central tradeoff in this problem: in environments where the decision taken affects the amount of information about the state that is revealed ex-post, distorting decision-making in favor of decisions that reveal more information can help provide better incentives for effort. This is the only reason to distort decision-making away from what would be efficient given the information available. Indeed, if all decisions reveal the same amount of information ex-post, optimal contracts induce efficient decision-making conditional on the information available. If decisions can be ranked in terms of the amount

[36] Thus, accountability is not attainable in setups where transfer schemes are restricted to be linear in output such as Athey and Roberts (2001) or Rantakari (2011).

of information that they reveal ex-post, optimal contracts distort decision-making in favor of decisions that reveal more information. If decisions cannot be ranked, however, the direction of optimal distortions is ambiguous in general, since it depends on the features of the information acquisition technology. These results illustrate how optimal organizational design can lead to endogenous conflicts of interest between the parties regarding the decisions that are to be taken. Indeed, when a contract induces inefficient decision-making, the principal and the expert often disagree on the decision that should be taken given the information available.

The results in the paper depend crucially on the fact that contracts can specify menus of transfer schemes from which the expert can choose. In Zermeño (2011*a*), I compare explicitly the outcomes that can be attained under full-commitment depending on whether contracts specify a menu of transfer schemes or a single transfer scheme. I show that, with menus, the fact that the same decision can be associated with more than one transfer scheme enables us to separate the expert's incentives to exert effort from the way decisions are taken. Thus, optimal distortions to decision-making are purely driven by differences in the amount of information about the state that different decisions reveal ex-post. Without menus, however, the inability to separate the expert's incentives from the principal's decision introduces an additional reason to distort decision-making. Thus, the intuition derived in this paper can be overturned; without menus, optimal distortions to decision-making can favor decisions that reveal *less* information about the state ex-post. The results on implementability under imperfect commitment also rely on the fact that contracts can specify menus of transfer schemes. The result that, under assumption 2.3.1, any Pareto-optimal outcome implementable under full-commitment can also be implemented under expert-authority can easily be extended to the case where contracts only specify a single transfer scheme. However, whether the same is true under principal-authority is still an open question.

2.7 Appendix

2.7.1 Completion of the type space

The notion of implementability established in definition 2.2.3 requires the expert to have a best response at the interim stage after observing *any* posterior $x^0 \in X$. Nevertheless, given the effort specified by the outcome to be implemented (e_0), the probability measure, $P_{e_0}(\cdot)$, need not have full support. Thus the requirement that a best response must exists for *every* posterior might be too strong, and more outcomes could potentially be implemented if a best response were required to exist only after observing posteriors in the support of $P_{e_0}(\cdot)$. Here we will see that this is actually not the case.

For any $e_0 \in E$, let $S(e_0) \subseteq X$ be the support of $P_{e_0}(\cdot)$. Let $d_{e_0} : S(e_0) \rightarrow D$ be a decision rule with support over $S(e_0)$. For consistency, $d(x)$ will still denote decision rules defined over the whole simplex. The following definition formalizes the weaker notion of implementability described above:

Definition 2.7.1. *The IR outcome $(e_0, d_{e_0}(x), \mathbb{T})$ is implementable if there exists a contract $\{(t_r(z), d_r)\}_{r\in R}$ and a report rule $r^* : S(e_0) \rightarrow R$ such that:*

1. *For all $x^0 \in S(e_0)$, $r^*(x^0) \in arg\max_{r\in R} \mathbb{E}_{x^0}[t_r(z(d_r, \theta))]$.*

2. $\mathbb{E}_{P_{e_0}}\left[\mathbb{E}_x\left[t_{r^*(x)}(z(d_{r^*(x)}, \theta))\right]\right] = \mathbb{T}$.

3. $e_0 \in arg\max_{e\in E} \quad \{\mathbb{E}_{P_e}\left[\sup_{r\in R}\{\mathbb{E}_x\left[t_r(z(d_r, \theta))\right]\}\right] - c(e)\}$.

4. *For all $x^* \in S(e_0)$, $d_{r^*(x^0)} = d_{e_0}(x^0)$.*

5. *All transfers are non-negative.*

The function $r^*(x)$ represents the expert's best response after observing posteriors $x^0 \in S(e_0)$. This definition requires best responses to exist only after observing posteriors that are in the support of $P_{e_0}(\cdot)$. If the expert does exert effort e_0, $r^*(x)$ must be such that the expert's expected transfer is in fact $\mathbb{T}$ (condition (2)), and the decision rule $d_{e_0}(x)$ is actually implemented (condition (4)). Condition (3) guarantees that the expert is in fact willing to exert effort e_0, and (5) is limited liability.

The following proposition shows that it is always possible to complete a contract (making sure that the same outcome is still implemented) to guarantee that the expert will have a best response after observing *any* possible posterior.

Proposition 2.7.1. *Let the IR outcome, $(e_0, d_{e_0}(x), \mathbb{T})$, contract, $\{(t_r(z), d_r)\}_{r\in R}$, and report rule, $r^*(x)$, satisfy the conditions in definition 2.7.1. Then, there exists a decision rule $d(x)$, with $d(x^0) = d_{e_0}(x^0)$ for all $x^0 \in S(e_0)$, such that the outcome $(e_0, d(x), \mathbb{T})$ is implementable according to definition 2.2.3. That is, there exists a contract, $\{(\hat{t}_r(z), \hat{d}_r)\}_{r\in\hat{R}}$ (and its induced conditional expected payment function, $\hat{T}(x)$), such that:*

1. $\mathbb{E}_{P_{e_0}}[\hat{T}(x)] = \mathbb{T}$.

2. $e_0 \in arg\max_{e\in E} \ \{\mathbb{E}_{P_e}[\hat{T}(x)] - c(e)\}$.

3. *For all $x^0 \in X$, there exists $r_0 \in arg\max_{r\in\hat{R}} \ \mathbb{E}_{x^0}[\hat{t}_r(z(\hat{d}_r, \theta))]$ such that $\hat{d}_{r_0} = d(x^0)$.*

4. *All transfers are non-negative.*

Proof. Let $R^* = \{r^*(x) \in R \mid x \in S(e_0)\}$, and $\hat{R} = R^* \bigcup (X/S(e_0))$. Let $\hat{t}_r(z) \equiv t_r(z)$ and $\hat{d}_r = d_r$ for all $r \in R^*$. Condition (1) of definition 2.7.1 implies that

$$r^*(x^0) \in argmax_{r\in R^*}\mathbb{E}_{x^0}[\hat{t}_r(z(\hat{d}_r, \theta))] \ \ \forall \ \ x^0 \in S(e_0). \tag{2.7.1}$$

Moreover, condition (3) of definition 2.7.1 implies that

$$e_0 \in arg\max_{e\in E}\left\{\mathbb{E}_{P_e}\left[\sup_{r\in R^*}\left\{\mathbb{E}_x\left[\hat{t}_r(z(\hat{d}_r, \theta))\right]\right\}\right] - c(e)\right\}. \tag{2.7.2}$$

So far we have only specified $\hat{t}_r(z)$ and $\hat{d}_r$ for $r \in R^*$. We need to construct these values for $r \in X/S(e_0)$ in a way such that:

1. $\hat{t}_{\tilde{x}}(z) \geq 0$ for all $\tilde{x} \in X/S(e_0)$.

2. $\mathbb{E}_{\tilde{x}}[\hat{t}_{\tilde{x}}(z(\hat{d}_{\tilde{x}},\theta))] = \sup_{r\in R^*}\mathbb{E}_{\tilde{x}}[\hat{t}_r(z(d_r,\theta))] \geq \mathbb{E}_{\tilde{x}}[\hat{t}_{x'}(z(\hat{d}_{x'},\theta))]$ for all $\tilde{x}, x' \in X/S(e_0)$.

3. $\mathbb{E}_{x^0}[\hat{t}_{r^*(x^0)}(z(\hat{d}_{r^*(x^0)},\theta))] \geq \mathbb{E}_{x^0}[\hat{t}_{\tilde{x}}(z(\hat{d}_{\tilde{x}},\theta))]$ for all $x^0 \in S(e_0)$ and $\tilde{x} \in X/S(e_0)$.

These three conditions and conditions 2.7.1 and 2.7.2 imply that conditions (1)-(4) in the proposition will be satisfied.

We start by showing that the vectors $\{\hat{t}_r(z)\}_{r\in R^*}$ must be uniformly bounded. WLOG, we can assume that, if $z_0 \neq z(d_r,\theta_i)$ for any $i \in \{1,...,N\}$, then $\hat{t}_r(z_0) = 0$. If these vectors were not bounded, since Θ is finite, there would exist $i_0 \in \{1,...,N\}$ such that $\hat{t}_r(z(\hat{d}_r,\theta_{i_0}))$ is not bounded. Moreover, since $x^p(e_0) \in int(X)$, there must exist $x^0 \in S(e_0)$ such that $x^0_{i_0} > 0$. Thus, condition 2.7.1 would not have been satisfied.

Take any $\tilde{x} \in X/S(e)$. Since $\{\hat{t}_r(z)\}_{r\in R^*}$ is uniformly bounded, $\sup_{r\in R^*}\mathbb{E}_{\tilde{x}}[\hat{t}_r(z(\hat{d}_r,\theta))]$ is finite. Let $(r_n)_{n\in\mathbb{N}}$ be a sequence, contained in R^*, such that

$$\lim_{n\to\infty}\mathbb{E}_{\tilde{x}}[\hat{t}_{r_n}(z(\hat{d}_{r_n},\theta))] = \sup_{r\in R^*}\mathbb{E}_{\tilde{x}}[\hat{t}_r(z(\hat{d}_r,\theta))].$$

Note that D is compact relative to the discrete metric, and the set of vectors, $\{\hat{t}_r(z)\}_{r\in R^*}$, is bounded (relative to the Euclidean metric). Thus, there must exist subsequences, $\hat{d}_{r_{n_k}}$ and $\hat{t}_{r_{n_k}}(z)$, that converge (relative to the discrete metric and Euclidean metric respectively). Moreover, it is easy to verify that the function $f(t(\cdot),d) = \mathbb{E}_{\tilde{x}}[t(z(d,\theta))]$ is continuous relative to those metrics. Define $\hat{d}_{\tilde{x}} = \lim_{k\to\infty}\hat{d}_{r_{n_k}}$ and $\hat{t}_{\tilde{x}}(z) = \lim_{k\to\infty}\hat{t}_{r_{n_k}}(z)$ (relative to their respective metrics). Note that $\hat{t}_{\tilde{x}}(z) \geq 0$. Furthermore, by continuity,

$$\mathbb{E}_{\tilde{x}}[\hat{t}_{\tilde{x}}(z(\hat{d}_{\tilde{x}},\theta))] = \lim_{k\to\infty}\mathbb{E}_{\tilde{x}}[\hat{t}_{r_{n_k}}(z(\hat{d}_{r_{n_k}},\theta))] = \sup_{r\in R^*}\mathbb{E}_{\tilde{x}}[\hat{t}_r(z(\hat{d}_r,\theta))].$$

Using this procedure, we can construct $\hat{t}_{\tilde{x}}(z)$ and $\hat{d}_{\tilde{x}}$ for all $\tilde{x} \in X/S(e_0)$. We just need to check that conditions (2) and (3) above are satisfied. Take $\tilde{x} \in X/S(e_0)$. By definition,

$$\mathbb{E}_{\tilde{x}}[\hat{t}_{\tilde{x}}(z(\hat{d}_{\tilde{x}},\theta))] = \sup_{r\in R^*}\mathbb{E}_{\tilde{x}}[\hat{t}_r(z(\hat{d}_r,\theta))].$$

Take $x' \in X/S(e)$. Then $\hat{t}_{x'}(z) = \lim_{n\to\infty}\hat{t}_{r_n}(z)$ and $\hat{d}_{x'} = \lim_{n\to\infty}\hat{d}_{r_n}$ for some sequence

$(r_n)_{n\in\mathbb{N}} \subset R^*$. Again, using continuity we obtain:

$$\mathbb{E}_{\tilde{x}}[\hat{t}_{x'}(z(d_{x'}, \theta))] = \lim_{n\to\infty} \mathbb{E}_{\tilde{x}}[\hat{t}_{r_n}(z(\hat{d}_{r_n}, \theta))] \leq \mathbb{E}_{\tilde{x}}[\hat{t}_{\tilde{x}}(z(\hat{d}_{\tilde{x}}, \theta))],$$

so condition (2) is satisfied. Showing that condition (3) is satisfied is analogous.

□

The argument used in the proof can be generalized into other mechanism design environments. In setups where the space of decisions is a compact metric space (so that every sequence of decisions has a converging subsequence within the set), and where the agents' utility is continuous in decisions, a similar argument would make any assumptions about the type space superfluous. For example, it should often be possible to "convexify" a non-convex type space without affecting the set of implementable outcomes.

2.7.2 Stochastic decision rules

This section illustrates two points. First, if full-commitment contracts could specify stochastic decision rules, the parties in general could do better. Thus, restricting attention to deterministic decision rules is with loss of generality. Second, outcomes with stochastic decision rules are difficult to implement under imperfect commitment. Specifically, we will see that any outcome with a stochastic decision rule implementable under expert-authority is Pareto-dominated by an outcome with a deterministic decision rule and also implementable under expert-authority.

Let $\sigma(x)$ denote a stochastic decision rule ($\sigma_d(x)$ is the probability of taking decision $d \in D$ after observing posterior $x \in X$). The following proposition shows that, if there exists a decision that can be identified ex-post and that reveals the realization of the state, then any outcome that would be implementable if all decisions were revealing can be arbitrarily approximated by an implementable outcome with a stochastic decision rule. Formally,

Proposition 2.7.2. *Suppose that there exists $d_0 \in D$ such that $z(d_0, \theta_0) \neq z(d', \theta')$ for all $(d_0, \theta_0) \neq (d', \theta')$. Let $\sigma(x)$ be such that $\sigma_{d_0}(x^0) > 0$ for all $x^0 \in X$, and such that the* *outcome* $(e_0, \sigma(x), \mathbb{T})$ *is IR. Then* $(e_0, \sigma(x), \mathbb{T})$ *is implementable if and only if it would be implementable in an environment with* $\hat{z}(d,\theta) \equiv (d,\theta)$.

Proof. Only one side of the proof is non-trivial. Let $\{(t_r(d,\theta), \sigma^r)\}_{r\in R}$ be the contract that implements $(e_0, \sigma(x), \mathbb{T})$ when $\hat{z}(d,\theta) \equiv (d,\theta)$. For each $r_0 \in R$, let $\hat{t}_{r_0}(z_0) = 0$ if $z_0 \notin \{z(d_0,\theta) \mid \theta \in \Theta\}$, and $\sigma^{r_0}_{d_0}(x)\hat{t}_{r_0}(z_0(d_0,\theta_i)) = \mathbb{E}_{\sigma^{r_0}}[t_{r_0}(d,\theta_i)]$, for all $\theta_i \in \Theta$. The function $\hat{t}_{r_0}(z)$ is well defined because, if $\theta_0 \neq \theta'$, $z(d_0,\theta_0) \neq z(d_0,\theta')$. Furthermore, $\hat{t}_{r_0}(z)$ is non-negative. Now, for any $\theta_0, \theta' \in \Theta$, if $d' \neq d_0$, then $z(d',\theta') \neq z(d_0,\theta_0)$, so $\hat{t}_{r_0}(z(d',\theta')) = 0$. Therefore, for any $\theta_i \in \Theta$ and $r_0 \in R$, $\mathbb{E}_{\sigma^{r_0}}[\hat{t}_{r_0}(z(d,\theta_i))] = \mathbb{E}_{\sigma^{r_0}}[t_{r_0}(d,\theta_i)]$, and the contract, $\{(\hat{t}_r(z), \sigma^r)\}_{r\in R}$, implements the same outcome. □

The idea is that, when there exists a decision that is identifiable, and that reveals the state ex-post, as long as it is always taken with positive probability, it is possible to place all the weight of a contract upon this decision. By doing this we can effectively make the state of nature contractible. However, for this to work it is crucial that the expert is risk-neutral, and that the parties are able to commit to stochastic decision rules.[37] Note that with a contract as the one constructed in the proof, without full-commitment neither one of the parties would actually be willing to carry out the stochastic decision rule. If the expert had the right to decide, he would always choose d_0, which is the only decision that allows him to get positive transfers. If the principal had the decision right, she would not want to choose d_0 with positive probability whenever her information was such that some other decision was ex-post efficient. In fact, the following proposition shows that stochastic decision rules would never be implemented under expert-authority.

Proposition 2.7.3. *If the IR outcome* $(e_0, \sigma, \mathbb{T})$ *is implementable under expert-authority, it is (at least weakly) Pareto-dominated by another outcome with a deterministic decision rule and implementable under expert-authority.*

Proof. Suppose that $(e_0, \sigma, \mathbb{T})$ is implementable under expert-authority. For the expert to be willing to randomize after he observed posterior x^0, he must be indifferent between the

[37] If the expert was risk-averse, the stochastic contracts considered here would destroy value.

decisions that he chooses with positive probability. Thus, using the same compensation scheme, it is possible to implement an outcome $(e_0, d(x), \mathbb{T})$, where

$$d(x^0) \in arg \max_{d_0 \in \{d \in D | \sigma_d(x^0) > 0\}} \mathbb{E}_{x^0}[y(d_0, \theta)].$$

The new outcome leaves the expert indifferent (the conditional expected payment function, $T(x)$ does not change), and makes the principal better off. $\square$

2.7.3 Proof of Proposition 2.3.1 (part (a))

Let the outcome $(e_0, d(x), \mathbb{T})$ be implementable under full-commitment with a contract $\{(t_r(z), d_r)\}_{r \in R}$. For each $x^0 \in X$, let $r^*(x^0) \in arg\max_{r \in R} \mathbb{E}_{x^0}[t_r(z(d_r, \theta))]$, and $d_{r^*(x^0)} = d(x^0)$. Let $R^* = \{r^*(x^0) \in R \mid x^0 \in X\}$. For each $r_0 \in R^*$, define $\hat{t}_{r_0}(z_0) = t_{r_0}(z_0)$ if $z_0 \in \{z(d_{r_0}, \theta) \mid \theta \in \Theta\}$, and $\hat{t}_{r_0}(z_0) = 0$ otherwise. Consider the expert's choice of $d \in D$ after having chosen a function $\hat{t}_{r_0}(z)$ with $r_0 \in R^*$. Suppose he picks some $d' \neq d_{r_0}$. Take any $\theta_i \in \Theta$. If $z(d', \theta_i) = z(d_{r_0}, \theta_i)$, then $\hat{t}_{r_0}(z(d', \theta_i)) = \hat{t}_{r_0}(z(d_{r_0}, \theta_i))$. If $z(d', \theta_i) \neq z(d_{r_0}, \theta_i)$, then, by assumption 2.3.1, $z(d', \theta_i) \notin \{z(d_{r_0}, \theta) \mid \theta \in \Theta\}$, and $\hat{t}_{r_0}(z(d', \theta_i)) = 0 \leq \hat{t}_{r_0}(z(d_{r_0}, \theta_i))$. Thus, the expert prefers d_{r_0} over d' when the transfer scheme is $\hat{t}_{r_0}(z)$. Then, it is clear that the contract, $\{\hat{t}_r(x)\}_{r \in R^*}$, implements $(e_0, d(x), \mathbb{T})$ under expert-authority.

2.7.4 Proof of Proposition 2.3.1 (part (b))

Let the Pareto-optimal outcome $(e_0, d(x), \mathbb{T})$ be implementable under full-commitment with a contract $\{(t_r(z), d_r)\}_{r \in R}$. For each $x^0 \in X$, let $r^*(x^0) \in arg\max_{r \in R} \mathbb{E}_{x^0}[t_r(z(d_r, \theta))]$, and $d_{r^*(x^0)} = d(x^0)$. Let $R^* = \{r^*(x^0) \in R \mid x^0 \in X\}$. The functions, $t_{r^*(x)}(z)$ and $d_{r^*(x)}$, generate a partition over the simplex given by $\{X^j\}_{j \in J}$, where $x_0 \in X^j$ if and only if $t_{r^*(x^0)}(z) \equiv t^j(z)$ and $d_{r^*(x^0)} = d_j$. The idea of the proof is to construct a contract, $\{\hat{t}_r(z)\}_{r \in R^*}$, satisfying:

1. The functions, $\hat{t}_{r^*(x)}(z)$ and $d_{r^*(x)}$, partition the simplex in the same way as above.

2. Given the choice of a transfer scheme $\hat{t}_{r_0}(z)$ and recommendation d_{r_0} (with $r_0 \in R^*$), the principal prefers d_{r_0} to any other decision.

3. For all $\theta_i \in \Theta$ and $r_0 \in R^*$, $\hat{t}_{r_0}(z(d_{r_0}, \theta_i)) = t_{r_0}(z(d_{r_0}, \theta_i))$. Moreover, $\hat{t}_{r_0}(z) \geq 0$.

These conditions and the fact that the original contract implemented $(e_0, d(x), \mathbb{T})$, imply that the contract, $\{\hat{t}_r(z)\}_{r \in R^*}$, implements this outcome with a rubber-stamping scheme under principal-authority. To see this note that, if the expert never deviates, and chooses $\hat{t}_{r^*(x^0)}(z)$ and recommends $d_{r^*(x^0)}$, with $x^0 \in X^j$, the principal's posterior must be an element of the convex hull of X^j, which is denoted by $x^j \in X$. The transfer scheme, $\hat{t}_{r^*(x^0)}(z) \equiv \hat{t}^j(z)$, must guarantee that the principal's preferred decision, given the posterior $x^j \in X$, is indeed $d_{r^*(x_0)} = d_j$. If the expert chooses a pair $(\hat{t}_{r_0}(z), d_0)$, that is never observed along the equilibrium path, we take the principal's off-equilibrium beliefs to be x^j, where $\hat{t}_{r_0}(z) \equiv \hat{t}^j(z)$. When this happens, the principal overrules d_0 and takes d_j. Thus, the expert's choice effectively reduces to picking an option from the menu $\{(\hat{t}_r(z), d_r)\}_{r \in R^*}$, which will not be overruled, and, by condition (3) above, the same outcome is implemented.

All we need to do is to construct these transfers. WLOG, assume that, for all $r_0 \in R^*$ and $z_0 \notin \{z(d_{r_0}, \theta) \mid \theta \in \Theta\}$, $t_{r_0}(z_0) = 0$. For all $x^0 \in X$, let $\hat{t}_{r^*(x^0)}(z_0) = t^j(z_0)$ if $x^0 \in X^j$ and $z_0 \in \{z(d_j, \theta) \mid \theta \in \Theta\}$, and $\hat{t}_{r^*(x^0)}(z_0) = \bar{t}_j$ otherwise, where $\bar{t}_j \in \mathbb{R}_+$ will be defined shortly. Clearly, these transfer schemes satisfy conditions (1) and (3) above.

Here we show that, for payments $\bar{t}_j$ appropriately defined, condition (2) is also satsfied. Consider the principal's decision problem after she observes the expert's choice, $(\hat{t}_{r^*(x^0)}(z)$, $d_{r^*(x^0)})$ with $x^0 \in X^j$. Suppose the principal is considering picking an alternative decision $d' \neq d_{r^*(x^0)} = d_j$. Let $\Theta^{d'} = \{\theta \in \Theta \mid z(d', \theta) \neq z(d_j, \theta)\}$. Then

$$\mathbb{E}_{x^j}\left[y(d_j, \theta) - \hat{t}^j(z(d_j, \theta)) - y(d', \theta) + \hat{t}^j(z(d', \theta))\right] =$$

$$\mathbb{E}_{x^j}\left[y(d_j, \theta) - y(d', \theta)\right] + Pr_{x^j}\left[\theta \in \Theta^{d'}\right] \mathbb{E}_{x^j}\left[\bar{t}_j - t^j(z(d_j, \theta)) \mid \theta \in \Theta^{d'}\right]$$

where the equality follows because, by assumption 2.3.1, $z(d',\theta_i) \neq z(d_{r^*(x^0)},\theta_i)$ implies $z(d',\theta_i) \notin \{z(d_j,\theta) \mid \theta \in \Theta\}$. Let $\underline{x}^j = min\{x_i^j \mid i = 1,...,N,\ x_i^j > 0\}$, and $B \in \mathbb{R}$ such that $|y(d_0,\theta_i)| \leq B$ for all $d_0 \in D$, $\theta_i \in \Theta$, and $t^j(z(d_j,\theta_i)) \leq B$ for all $\theta_i \in \Theta$. Define $\bar{t}_j = \frac{3B}{\underline{x}^j}$. Then, if $Pr_{x^j}\left[\theta \in \Theta^{d'}\right] > 0$, $\bar{t}_j$ ensures that the principal would not prefer d' to d_j. If $Pr_{x^j}\left[\theta \in \Theta^{d'}\right] = 0$, then for all $i \in \{1,...,N\}$ such that $x_i^j > 0$, $z(d',\theta_i) = z(d_j,\theta_i)$. Suppose that $\mathbb{E}_{x^j}[y(d_j,\theta) - y(d',\theta)] < 0$, so that the principal would deviate. We will show that this must imply that this outcome is Pareto-dominated, which is a contradiction. Since x^j is an element of the convex hull of X^j (it need not be an element of X^j), there must exists $x^0 \in X^j$ such that $\mathbb{E}_{x^0}[y(d_j,\theta) - y(d',\theta)] < 0$, and such that $x_i^j = 0$ implies $x_i^0 = 0$. Thus, for all $i \in \{1,...,N\}$, if $x_i^0 > 0$, then $x_i^j > 0$ and $z(d',\theta_i) = z(d_j,\theta_i)$. Therefore, with the original contract we could have implemented an outcome with one change: $d(x^0) = d'$ instead of d_j. This outcome Pareto-dominates the original outcome.

2.7.5 Proof of Theorem 2.4.1

We begin with necessity. Suppose that an outcome $(e_0, d(x), \mathbb{T})$ is implementable with a contract $\{(t_r(z), d_r)\}_{r\in R}$. Note that, for all $i \in \{1,...,N\}$, $t_r(z(d_r,\theta_i))$ must be bounded above (it is bounded below by limited liability).[38] This implies that the function, $T : \bar{X} \to \mathbb{R}$, defined as

$$T(x^0) \equiv \sup_{r\in R} \sum_{i=1}^{N} x_i^0 t_r(z(d_r,\theta_i)),$$

is well defined. For any $x^0 \in X$, let $r^*(x^0) \in \arg\max_{r\in R} \mathbb{E}_{x^0}[t_r(z(d_r,\theta))]$ (when $x^0 \in X$, the supremum is attained). We will see that $T(x)$ satisfies conditions (1)-(5) of Theorem 2.4.1. Conditions (1)-(2) must be satisfied by definition 2.2.3. Condition (3) is satisfied because $T(x)$ is the upper envelope of linear functions of x (so it is convex). To see that condition (4) is satisfied, take any $x^{-i} \in X^{-i}$, and consider the line segment that joins x^{-i} and x^i. For each $\alpha \in [0,1]$, let $\hat{x}(\alpha) \equiv \alpha x^i + (1-\alpha)x^{-i}$. By the envelope theorem (see Milgrom and

[38] Otherwise, for any $x^0 \in int(X)$, we would have $\arg\max_{r\in R} \mathbb{E}_{x^0}[t_r(z(d_r,\theta))] = \varnothing$.

Segal (2002) or Krishna and Maenner (2001)),

$$T(\hat{x}(\alpha_0)) \equiv T(x^{-i}) + \int_0^{\alpha_0} \Delta(\alpha) d\alpha,$$

where $\Delta(\alpha) \equiv t_{r^*(\hat{x}(\alpha))}(z(d_{r^*(\hat{x}(\alpha))}, \theta_i)) - \mathbb{E}_{x^{-i}}[t_{r^*(\hat{x}(\alpha))}(z(d_{r^*(\hat{x}(\alpha))}, \theta))]$.[39] By the convexity of $T(x)$, $\Delta(\alpha)$ must be non-decreasing. Then we have:

$$T(x^{-i}) + T'(x^{-i} \mid x^i - x^{-i}) = \mathbb{E}_{x^{-i}}[t_{r^*(\hat{x}(0))}(z(d_{r^*(\hat{x}(0))}, \theta))] + \Delta(0^+)$$

$$\geq \mathbb{E}_{x^{-i}}[t_{r^*(\hat{x}(0))}(z(d_{r^*(\hat{x}(0))}, \theta))] + \Delta(0)$$

$$= t_{r^*(\hat{x}(0))}(z(d_{r^*(\hat{x}(0))}, \theta_i)) \geq 0,$$

where $\Delta(0^+) = \lim_{\alpha \to 0^+} \Delta(\alpha)$. The first equality follows because

$$T'(x^{-i} \mid x^i - x^{-i}) = \lim_{h \to 0^+} \frac{T(x^{-i} + h(x^i - x^{-i})) - T(x^{-i})}{h} = \lim_{h \to 0^+} \frac{T(\hat{x}(h)) - T(\hat{x}(0))}{h} = \Delta(0^+).$$

Finally, to see that condition (5) is satisfied, take any $x^0 \in X$ and $v \in span(\{x^i - x^j \mid z(d(x^0), \theta_i) = z(d(x^0), \theta_j)\})$. Then, there exists vectors, $\{x^{i_k} - x^{j_k}\}_{k=1}^K$, and scalars, $\{\lambda_k\}_{k=1}^K$, such that $v = \sum_{k=1}^K \lambda_k (x^{i_k} - x^{j_k})$ and $z(d(x^0), \theta_{i_k}) = z(d(x^0), \theta_{j_k})$, for all $k \in \{1, ..., K\}$. By definition,

$$T(x^0 + v) \geq \sum_{i=1}^N (x_i^0 + v_i) t_{r^*(x^0)}(z(d_{r^*(x^0)}, \theta_i)) = T(x^0) + \sum_{i=1}^N t_{r^*(x^0)}(z(d(x^0), \theta_i)) \sum_{k=1}^K \lambda_k (x_i^{i_k} - x_i^{j_k})$$
$$= T(x^0) + \sum_{k=1}^K \lambda_k \sum_{i=1}^N (x_i^{i_k} - x_i^{j_k}) t_{r^*(x^0)}(z(d(x^0), \theta_i)) = T(x^0),$$

[39] This follows because

$$\mathbb{E}_{\hat{x}(\alpha)}[t_r(z(d_r, \theta))] \equiv \mathbb{E}_{x^{-i}}[t_r(z(d_r, \theta))] + \alpha \left[t_r(z(d_r, \theta_i)) - \mathbb{E}_{x^{-i}}[t_r(z(d_r, \theta))] \right].$$

where the last line follows because $\sum_{i=1}^{N}(x_i^{i_k} - x_i^{j_k})t_{r^*(x^0)}(z(d(x^0),\theta_i)) = 0$, for all $k \in \{1,...,K\}$.

Now we prove sufficiency. Suppose that the function $T(x)$ satisfies conditions (1)-(5) in Theorem 2.4.1 given the outcome $(e_0, d(x), \mathbb{T})$. For each $x^0 \in X$, let $v(x^0) \in \mathbb{R}^N$ be a subgradient of $T(x)$ at x^0 with the property that, if $z(d(x^0),\theta_i) = z(d(x^0),\theta_j)$, then $v(x^0) \cdot (x^i - x^j) = 0$.[40] Such subgradients can be constructed because $T(x)$ satisfies condition (5). For each $x^0 \in X$ and $i \in \{1,...,N\}$, define:

$$t_{x^0}(z(d(x^0),\theta_i)) = T(x^0) + v(x^0) \cdot (x^i - x^0).$$

Note that:

1. These transfers are well defined. If $z(d(x^0),\theta_i) = z(d(x^0),\theta_j)$, $t_{x^0}(z(d(x^0),\theta_i)) - t_{x^0}(z(d(x^0),\theta_j)) = v(x^0) \cdot (x^i - x^j) = 0$.

2. For all $x^0 \in X$, $\mathbb{E}_{x^0}[t_{x^0}(z(d(x^0),\theta))] = T(x^0)$.

3. For all $x^0, x' \in X$, $T(x^0) \geq \mathbb{E}_{x^0}[t_{x'}(z(d(x'),\theta))]$. To see this, take any $x^0, x' \in X$, and note that, by construction,

$$T(x^0) \geq T(x') + v(x') \cdot (x^0 - x') = \mathbb{E}_{x^0}[t_{x'}(z(d(x'),\theta))].$$[41]

4. For any $i \in \{1,...,N\}$ and $x^0 \in X/X^{-i}$, $t_{x^0}(z(d(x_0),\theta_i)) \geq 0$. To see this, take any $x^0 \in X/X^{-i}$ and consider the line segment that joins x^i with the set X^{-i} going through x^0. Let $x^{-i} = x^{-i}(x^0)$ be the point in X^{-i} that lies in this line segment. Let $\hat{x}(\alpha) = (1-\alpha)x^{-i} + \alpha x^i$ describe the line segment ($\alpha \in [0,1]$). Note that, since $v(\hat{x}(\alpha_0))$ is a subgradient of $T(x)$ at $\hat{x}(\alpha_0)$, $\Delta(\alpha_0) \equiv v(\hat{x}(\alpha_0)) \cdot (x^i - x^{-i})$ is a subderivative of $T(\hat{x}(\alpha))$ at α_0. Moreover, as $T(\hat{x}(\alpha))$ is convex and continuous in $[0,1]$, it can be

[40] A vector, $v(x^0) \in \mathbb{R}^N$, is a *subgradient* of the convex function, $T(x)$, at $x^0 \in X$ if $T(x) \geq T(x^0) + v(x^0) \cdot (x - x^0)$, for all $x \in \bar{X}$.

[41] The last equality follows because $\sum_{i=1}^{N} x_i^0 t_{x'}(z(d(x'),\theta_i)) = \sum_{i=1}^{N} x_i^0 \left(T(x') + v(x') \cdot (x^i - x')\right) = T(x') + v(x') \cdot (x^0 - x')$.

expressed as

$$T(\hat{x}(\alpha_0)) \equiv T(\hat{x}(0)) + \int_0^{\alpha_0} \Delta(\alpha) d\alpha,$$

where $\Delta(\alpha)$ is non-decreasing in α. Moreover,

$$t_{\hat{x}(\alpha)}(z(d(\hat{x}(\alpha)), \theta_i)) \equiv T(\hat{x}(\alpha)) + v(\hat{x}(\alpha)) \cdot (x^i - \hat{x}(\alpha)) \equiv T(\hat{x}(\alpha)) + (1-\alpha)\Delta(\alpha)$$

is non-decreasing in α.[42] Thus, for any $\alpha > 0$,

$$t_{\hat{x}(\alpha)}(z(d(\hat{x}(\alpha)), \theta_i)) \geq T(\hat{x}(0^+)) + \Delta(0^+) = T(x^{-i}) + T^{'}(x^{-i} \mid x^i - x^{-i}) \geq 0.$$

In particular, this inequality holds when $\hat{x}(\alpha) = x^0$.

Finally we will modify the transfers that we constructed to guarantee that limited liability is satisfied for *all* $x^0 \in X$, and z. We must do so in such a way that the conditions (1)-(3) that we just verified are still met. For each $x^0 \in X$, let $\hat{t}_{x^0}(z) = t_{x^0}(z)$ if, for some $i \in \{1, .., N\}$, $z = z(d(x^0), \theta_i)$ and $x_i^0 > 0$. Let $\hat{t}_{x^0}(z) = 0$ otherwise. Note that $\hat{t}_x(z)$ is well defined, and satisfies limited liability. Moreover, $\hat{t}_{x^0}(z(d(x^0), \theta_i)) = t_{x^0}(z(d(x^0), \theta_i))$, for all $i \in \{1, ..., N\}$ such that $x_i^0 > 0$. Thus, for all $x^0 \in X$, $\mathbb{E}_{x^0}[\hat{t}_{x^0}(z(d(x^0), \theta))] = T(x^0)$.

In order to prove that the contract $\{(\hat{t}_x(z), d(x))\}_{x \in X}$ implements the outcome $(e_0, d(x), \mathbb{T})$, we must only show that $T(x^0) \geq \mathbb{E}_{x^0}[\hat{t}_{x'}(z(d(x'), \theta))]$, for all $x^0, x' \in X$. Suppose that there existed $x', x^0 \in X$ such that $\mathbb{E}_{x^0}[\hat{t}_{x'}(z(d(x'), \theta))] > T(x^0)$. Define

$$I = \{i \in \{1, ..., N\} \mid \hat{t}_{x'}(z(d(x'), \theta_i)) \neq t_{x'}(z(d(x'), \theta_i)) \text{ or } x_i^0 = 0\}.$$

[42] Take any $\alpha > \alpha'$. Then,

$$\begin{aligned} t_{\hat{x}(\alpha)}(z(d(\hat{x}(\alpha)), \theta_i)) - t_{\hat{x}(\alpha')}(z(d(\hat{x}(\alpha')), \theta_i)) &= \int_{\alpha'}^{\alpha} \Delta(s) ds + \Delta(\alpha)(1-\alpha) - \Delta(\alpha')(1-\alpha') \\ &\geq (\alpha - \alpha')\Delta(\alpha') + \Delta(\alpha)(1-\alpha) - \Delta(\alpha')(1-\alpha') \\ &= (1-\alpha)(\Delta(\alpha) - \Delta(\alpha')) \geq 0. \end{aligned}$$

Let $K = \#(I)$. Note that $K < N$. Otherwise, we would have

$$\mathbb{E}_{x^0}[\hat{t}_{x'}(z(d(x'),\theta))] = 0 \leq \mathbb{E}_{x^0}[\hat{t}_{x^0}(z(d(x^0),\theta))] = T(x^0).$$

This follows because $x_i^0 > 0$ would imply $\hat{t}_{x'}(z(d(x'),\theta_i)) \neq t_{x'}(z(d(x'),\theta_i))$, which means that $\hat{t}_{x'}(z(d(x'),\theta_i)) = 0$. Furthermore, there must exists $i \in I$ such that $x_i^0 > 0$. Otherwise, we would have that

$$\mathbb{E}_{x^0}[\hat{t}_{x'}(z(d(x'),\theta))] = \mathbb{E}_{x^0}[t_{x'}(z(d(x'),\theta))] \leq T(x^0),$$

since $\hat{t}_{x'}(z(d(x'),\theta_i)) \neq t_{x'}(z(d(x'),\theta_i))$ would imply $x_i^0 = 0$.

For $\varepsilon \in (0,1)$, define $x^\varepsilon \in int(X)$ as $x_i^\varepsilon = \frac{\varepsilon}{K}$ if $i \in I$, and $x_i^\varepsilon = (1-\varepsilon)\frac{x_i^0}{\sum_{j \notin I} x_j^0}$ if $i \notin I$. We will show that, for ε small enough, $\mathbb{E}_{x^0}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta))] > T(x^0)$, which is a contradiction. Let $\Theta^I = \{\theta_i \mid i \in I\}$. Note that

$$\mathbb{E}_{x^0}\left[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I\right] \geq 0 = \mathbb{E}_{x^0}\left[\hat{t}_{x'}(z(d(x'),\theta)) \mid \theta \in \Theta^I\right],$$

where the equality follows because, if $\theta_i \in \Theta^I$ and $x_i^0 > 0$, $\hat{t}_{x'}(z(d(x'),\theta_i)) = 0$. The conditional expectations are well defined because there exists $i \in I$ with $x_i^0 > 0$. Moreover,

$$\begin{aligned}
\mathbb{E}_{x^0}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \notin \Theta^I] =& \mathbb{E}_{x^\varepsilon}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \notin \Theta^I] \\
=& \frac{T(x^\varepsilon) - \varepsilon \mathbb{E}_{x^\varepsilon}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I]}{1-\varepsilon} \\
\geq& \frac{\mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta))] - \varepsilon \mathbb{E}_{x^\varepsilon}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I]}{1-\varepsilon} \\
=& \mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta)) \mid \theta \notin \Theta^I] \\
&+ \frac{\varepsilon}{1-\varepsilon} \mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta)) - t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I] \\
=& \mathbb{E}_{x^0}[\hat{t}_{x'}(z(d(x'),\theta)) \mid \theta \notin \Theta^I] \\
&+ \frac{\varepsilon}{1-\varepsilon} \mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta)) - t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I],
\end{aligned}$$

where each line can be obtained by using the definitions of $\hat{t}_x(z)$, Θ^I and x^ε. Then we have:

$$\begin{aligned}
\mathbb{E}_{x^0}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta))] &\geq \mathbb{E}_{x^0}[\hat{t}_{x'}(z(d(x'),\theta))] + \varepsilon \mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta)) - t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I] \\
&> T(x^0) + \varepsilon \mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta)) - t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I].
\end{aligned}$$

Note that, for all $i \in \{1, ..., N\}$, $t_{x^\varepsilon}(z(d(x^\varepsilon),\theta_i))$ must be bounded above (it is bounded below because, since $x^\varepsilon \in int(X)$ for all $\varepsilon \in (0,1)$, it must be non-negative).[43] Thus,

$$\lim_{\varepsilon \to 0} \varepsilon \mathbb{E}_{x^\varepsilon}[t_{x'}(z(d(x'),\theta)) - t_{x^\varepsilon}(z(d(x^\varepsilon),\theta)) \mid \theta \in \Theta^I] = 0,$$

and, for ε small enough, $\mathbb{E}_{x^0}[t_{x^\varepsilon}(z(d(x^\varepsilon),\theta))] > T(x^0)$, a contradiction.

[43] Suppose not, and take any $\varepsilon_0 \in (0,1)$. Then there must exists $\varepsilon' \in (0,1)$ such that $\mathbb{E}_{x^{\varepsilon_0}}[t_{x^{\varepsilon'}}(z(d(x^{\varepsilon'}),\theta))] > T(x^{\varepsilon_0})$, a contradiction.

Chapter 3

A Reputational Model of Expertise

3.1 Introduction

After taking advice from an expert (a doctor, for example), we often make inference about the quality of his advice based on the outcome of the decision that we took. When our assessment about the expert's quality is positive, we are more likely to return for advice or to recommend him to other people. Thus, experts have an interest in fostering their reputation amongst their clients.[1] This paper addresses the question of how the expert's concern for his reputation can influence the amount of information that he can transmit to his clients, and the way decisions are ultimately taken.

I examine a model with two parties, a principal or decision-maker (she), and an expert (he). The principal has to choose between two decisions. Her ex-post utility depends on the decision taken and on the state of nature, which is assumed to be binary. The expert can be *informed* or a *charlatan*. If the expert is a charlatan, he does not have any private information about the state, and his prior belief about the likelihood of each state coincides with that of the principal. If the expert is informed, he receives a private and non-verifiable informative signal about the state of nature. The expert is the only one that knows his type. The parties play the following game: first, the expert observes his type and, if he

[1] Ehrbeck and Waldmann (1996), Chevalier and Ellison (1997) and Chevalier and Ellison (1999) provide evidence on the influence of reputational concerns over the behavior of experts.

is informed, the realization of the informative signal about the state. Second, the expert makes a cheap-talk report to the principal (no restrictions are imposed on the set of feasible reports). Third, the principal observes the report and takes a decision. Fourth, the principal observes the realization of the state of nature and updates her belief that the expert was informed. The expert has no direct interest in the principal's decision, he is only concerned about his reputation. More precisely, the expert's ex-post utility is a continuous and strictly increasing function of the principal's posterior belief that he is informed. Note that, absent the expert's reputational concerns, there would be no conflict of interest in this model. Thus, in line with Holmström (1999)'s seminal paper, in this model career concerns are the source rather than the cure of the incentive problem.

The main result of the paper is that, no matter how small is the principal's prior belief that the expert is a charlatan, the expert's concern for his reputation severely reduces the value of his advice. The game played by the parties has a unique equilibrium outcome in which some information is transmitted by the expert. This outcome can be implemented with an equilibrium in which the expert only sends two reports, so the information that the expert can transmit is endogenously coarse. Moreover, the expert must be indifferent between the two reports precisely when his posterior about the state of nature is equal to the prior. Thus, the expert sends one report or the other depending on whether his posterior about the state is above or below the prior (with two states of nature, posteriors are one-dimensional). This fact implies that the model yields a precise prediction about the way the principal's decision will be distorted. Indeed, if in equilibrium the expert's advice influences the principal's decision, decision-making is biased away from the status quo, understood as the decision that the principal would take without the expert's advice.[2] Remarkably, this distortion does not disappear as the principal's prior belief that the expert is informed approaches one.

The paper most related to this one is Levy (2004). She studies a model very similar to the

[2] The charlatan randomizes between the two reports, so he sometimes induces the principal to inefficiently deviate from the status quo. The informed expert also distorts the principal's decision. When his posterior about the state lies between the prior and the efficient threshold to change the decision, he will induce the principal to inefficiently deviate from the status quo.

one analyzed here. The main difference is in the models' information structures. In Levy's model, the expert observes the realization of a *binary* informative signal about the state (which is also binary). The expert has a continuum of types which determine the precision of the signal. The main result of the paper is that there is unique equilibrium where two reports are sent (only equilibria where at most two reports are sent are considered). Under the specific utility function for the principal considered, this equilibrium leads to decision-making being distorted away from the status quo (what Levy calls *anti-herding*), as happens here. Importantly, in contrast to what happens in the two-report equilibrium of my model, in Levy's equilibrium the expert is *not* indifferent between the two reports when his posterior is equal to the prior. This means that there are utility functions for the principal under which Levy's equilibrium would actually lead to distortions to decision-making in *favor* of the status quo. This does not happen in my model because the expert is indifferent between the two reports exactly at the prior. Intuitively, in my model, if the expert was truthful, deviations from the status quo would only be recommended by an informed expert. Thus, in equilibrium, recommending deviations from the status quo carries a reputational premium for the expert, who ends up inducing such deviations too frequently.[3] This argument breaks down in Levy's model because a *relatively* uninformed expert may also recommend deviations from the status quo under truth-telling.

This paper is also closely related to Ottaviani and Sørensen (2006*b*) and Ottaviani and Sørensen (2006*c*), who analyze single-expert games similar to the one studied here, but with the important distinction that the expert does not know his type. They also find that in the presence of reputational concerns, the information that the expert can credibly transmit becomes endogenously coarse. Another strand of the literature, which began with Scharfstein and Stein (1990)'s influential paper, has emphasized how reputational concerns can create incentives for herding in multiple-expert models where the beliefs about each expert's quality are formed based on the reports made by all the experts.[4]

[3] This result has a similar flavor to that in Prendergast and Stole (1996), where young experts react too much to their information to signal their ability.

[4] See, for example, Trueman (1994), Zwiebel (1995), Avery and Chevalier (1999), Effinger and Polborn (2001) and Ottaviani and Sørensen (2006*a*).

The rest of the paper is structured as follows. Section 3.2 presents the model. Section 3.3 shows that the game played by the parties has a unique equilibrium outcome in which some information is transmitted by the expert, and that this outcome is implementable by an equilibrium in which only two reports are sent. Section 3.4 discusses how the potential presence of charlatans reduces the value of the expert's advice. In particular, it illustrates how the expert's concern for his reputation distorts the principal's decision-making away from the status quo. Finally, Section 3.5 concludes.

3.2 The model

Consider an environment with two parties, a principal (she) and an expert (he). The principal's utility, $u_P(d, \theta)$, depends on a decision, $d \in \{d_0, d_1\} = D$, and on the binary state of nature, $\theta \in \{\theta_0, \theta_1\}$. Both parties start with the same prior belief, $x_p \in (0, 1)$, that the state is θ_1. The decision d_0 represents the status quo in the sense that it is assumed to maximize the principal's expected utility under the prior.

The expert can have two types, $t \in \{Ch, I\}$; he can be a *charlatan* (Ch) or *informed* (I). The expert is the only one that knows his type; the principal starts with a prior belief, $z_p \in (0, 1)$, that the expert is informed. A *charlatan* expert has no additional information about the state. That is, he believes that the state is θ_1 with probability x_p. An *informed* expert, however, privately observes the realization of a signal that contains information about the state. The realization of the signal cannot be verified by the principal. Without loss of generality, the signal can be described by a CDF over posteriors denoted by $F(x)$, where $x \in [0, 1]$ is the informed expert's posterior probability that $\theta = \theta_1$. By the law of iterated expectations, we must have $\mathbb{E}_F[x] = x_p$. We only impose the restriction that $F(x)$ must be continuos at x_p.[5] The expert has no direct interest in the principal's decision; he is only concerned about his reputation. We assume that the expert's utility is an strictly increasing and continuous function, $u_E(z)$, where z is the principal's posterior belief that the expert is

[5] This assumption is not crucial for the main results but simplifies the analysis.

informed. Note that $u_E(\cdot)$ does not depend on the expert's type.[6]

The parties play the following game. First, the expert observes his type. If he is a charlatan his posterior belief that the state is θ_1 is x_p. If he is informed, he observes a posterior, $x \in [0,1]$, drawn from the CDF, $F(x)$. Based on these observations, the expert chooses a lottery over *reports*, $\sigma_x^t \in \Delta R$.[7] The principal observes the expert's report, $r \in R$, and, given her conjecture about the expert's behavior ($\hat{\sigma}_x^t$), updates her information about the state. Let $x^{\hat{\sigma}_x^t}(r)$ be the principal's posterior belief that the state is θ_1 after observing the expert's report, $r \in R$, given her conjecture about the expert's behavior, $\hat{\sigma}_x^t$. Based on this information, the principal takes a decision. Finally, the principal observes the realization of the state, and updates her beliefs about the expert's type.[8] Denote the principal's posterior belief that the expert is informed (given her conjecture $\hat{\sigma}_x^t$, the expert's report and the realization of the state) by $z^{\hat{\sigma}_x^t}(r,\theta)$. Contingent transfers between the parties are not feasible.[9] The following timeline summarizes the sequence of events:

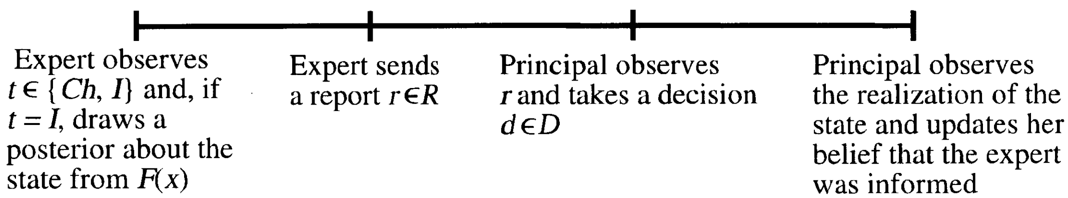


Figure 3.2.1: Timeline

The analysis will focus on identifying the properties of the Perfect Bayesian Nash equi-

[6] An specification like this one can be micro-founded with a two-period model such as those in Holmström and I Costa (1986) or Holmström (1999). Reduced-form models of reputation like the one presented here have previously been studied by Scharfstein and Stein (1990), Levy (2004), Ottaviani and Sørensen (2006*b*) and Ottaviani and Sørensen (2006*a*), for example.

[7] As is typical in cheap-talk games, the set R is specified by the equilibrium (see Crawford and Sobel (1982)).

[8] If $u(d,\theta_0) \neq u(d,\theta_1)$ for all $d \in D$, the assumption that the principal observes the realization of the state is equivalent to the principal observing her utility ex-post.

[9] This assumption is made to enable the model to isolate the effect of the expert's concern for his reputation on the amount of information transmitted and on the way the decision is made. Holmström and I Costa (1986), Gibbons and Murphy (1992) and Caruana and Celentani (2001), for example, study how contingent payment schemes should be structured in models where the agent is concerned about his reputation.

libria of the game. Formally, we have:

Definition 3.2.1. *The expert's report rule, $\sigma_x^t : \{Ch, I\} \times [0,1] \to \Delta R$, and principal's decision rule, $d : R \to D$, are a Perfect Bayesian Nash equilibrium (equilibrium hereafter) of the game if*

1. *The principal's beliefs, $z^{\sigma_x^t}(r,\theta)$ and $x^{\sigma_x^t}(r)$, are generated through Bayes rule whenever r is sent given the principal's correct conjecture, σ_x^t. If r is never sent given σ_x^t, then $z^{\sigma_x^t}(r,\theta) = 0$ and $x^{\sigma_x^t}(r) = x_p$.*

2. *For all $x \in [0,1]$ and $t \in \{Ch, I\}$, $r_0 \in supp(\sigma_x^t)$ implies $r_0 \in \arg\max_{r \in R} \mathbb{E}_x[u_E(z^{\sigma_x^t}(r,\theta))]$.*

3. *For all $r \in R$, $d(r) \in \arg\max_{d \in D} \mathbb{E}_{x^{\sigma_x^t}(r)}[u_P(d,\theta)]$.*

The first condition states that the principal's beliefs must be consistent with the expert's behavior. The second condition guarantees that σ_x^t is actually a best response for the expert given his information and the principal's (correct) conjecture. The third condition states that the principal must take the decision that maximizes her expected utility given her information. Finally, we have

Definition 3.2.2. *An outcome of the game is a pair, $(U_E(x), \delta(x))$, where $U_E : [0,1] \to \mathbb{R}$ specifies the expert's expected utility conditional on the posterior that he observed, and $\delta : [0,1] \to \Delta D$ specifies the decision (or lottery over decisions) implemented as a function of the posterior observed by the expert.*[10]

The outcome of the game determines the parties' expected payoffs as a function of the information *about the state* available at the interim stage. That is, if the equilibrium report rule, σ_x^t, and decision rule, $d(r)$, implement the outcome, $(U_E(x), \delta(x))$, we must have $U_E(x) \equiv max_{r \in R} E_x[u_E(z^{\sigma_x^t}(r,\theta))]$, and $\delta(x)(d_0) = Pr[Ch \mid x] Pr[d(r) = d_0 \mid \sigma_x^{Ch}] + Pr[I \mid x] Pr[d(r) = d_0 \mid \sigma_x^I]$ for all $x \in [0,1]$ and $d_0 \in D$.

[10] In this definition we are implicitly assuming that $supp(F(x)) \bigcup \{x_p\} = [0,1]$. Alternatively, we could define the functions in the definition only for posteriors that can be observed by the expert.

3.3 The equilibrium outcomes

This section derives the properties of the game's equilibrium outcomes. The main result is that there is a unique informative equilibrium outcome in which the information transmitted by the expert is severely limited. Indeed, this outcome can be implemented by an equilibrium where an informed expert only sends two reports depending on whether his posterior belief that the state is θ_1 is above or below x_p, and the charlatan randomizes between the two reports.

As is standard in cheap-talk games, in this game there is always an uninformative or *babbling* equilibrium in which the expert always sends the same report and no information is transmitted. Here we will focus on the informative equilibria of the game. The first result, which is a consequence of the following two propositions, is that the information transmitted in any informative equilibrium, can be transmitted in an equilibrium where only two reports are sent. Let $\{z^{\hat{\sigma}_x^t}(r_0, \theta_i)\}_{i=0}^{1}$ be the *lottery over reputations* induced by the report $r_0 \in R$, given the principal's conjecture about the expert's behavior, $\hat{\sigma}_x^t$. Then we have:

Proposition 3.3.1. *In any equilibrium, $(\sigma_x^t, d(r))$, reports sent along the equilibrium path can be divided into three groups: R_0, R_1 and R_p. Reports belonging to the same group induce the same lottery over reputations. If a report is sent after the expert observes a posterior smaller than the prior, it belongs to R_0; if a report is sent after the expert observes a posterior greater than the prior, it belongs to R_1; and if a report is sent only after the expert observes the prior, it belongs to R_p. It must be the case that $Pr[r \in R_p \mid \sigma_x^t] = 0$.*

Proof. Take any equilibrium, $(\sigma_x^t, d(r))$. Any report sent along the equilibrium path must be sent sometimes by the charlatan. Otherwise, such a report would lead the principal to believe that the expert is informed with probability one, and the charlatan would actually always want to send it. Thus, the expert must be indifferent between all the reports that are sent along the equilibrium path when his posterior about the state is equal to the prior.

The next part of the argument is illustrated in Figure 3.3.1. The horizontal axis represents the posterior about the state observed by the expert. There is a line associated to each

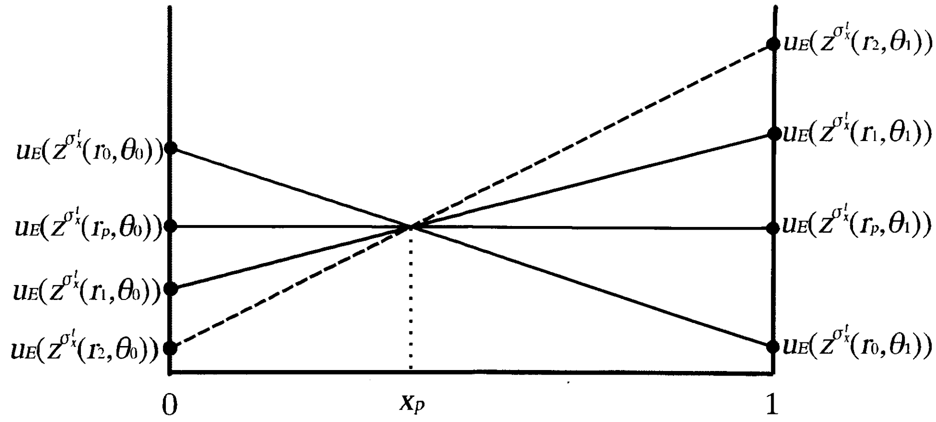


Figure 3.3.1: The expert's decision problem

different report. The height of the line associated to a report, r, at a posterior, x, gives the expert's expected utility (given x) derived from reporting r. From our previous discussion, the lines associated to any report sent along the equilibrium path must intersect at the prior, as is illustrated in the figure. Suppose two reports are sent in equilibrium after the expert observes posteriors greater than the prior. Since $u_E(\cdot)$ is strictly increasing, both reports must induce the same lottery over reputations. Otherwise, as shown in the figure, the expert would strictly prefer the report that induces the highest reputation after θ_1 is realized. Similarly, if two reports are sent in equilibrium after the expert observes posteriors smaller than the prior, they must induce the same lottery over reputations.

Finally, if a report, r_p, is only sent after the expert observes the prior, it must be the case that $z^{\sigma_x^t}(r_p, \theta_0) = z^{\sigma_x^t}(r_p, \theta_1) = \bar{z} > 0$, as illustrated in the figure. Indeed,

$$z^{\sigma_x^t}(r_p, \theta) = \frac{z_p Pr[r_p \mid I, \sigma_x^t] Pr[\theta \mid r_p, I, \sigma_x^t]}{z_p Pr[r_p \mid I, \sigma_x^t] Pr[\theta \mid r_p, I, \sigma_x^t] + (1 - z_p) Pr[r_p \mid Ch, \sigma_x^t] Pr[\theta \mid r_p, Ch, \sigma_x^t]}$$

$$= \frac{z_p Pr[r_p \mid I, \sigma_x^t]}{z_p Pr[r_p \mid I, \sigma_x^t] + (1 - z_p) Pr[r_p \mid Ch, \sigma_x^t]},$$

which does not depend on θ. The second equality follows because $Pr[\theta \mid r_p, I, \sigma_x^t] = Pr[\theta \mid$

$r_p, Ch, \sigma_x^t]$. Moreover, it must be the case that $Pr[r \in R_p \mid \sigma_x^t] = 0$. Otherwise we would have,

$$0 = \frac{Pr[I, r \in R_p \mid \sigma_x^t]}{Pr[r \in R_p \mid \sigma_x^t]} = Pr[I \mid r \in R_p, \sigma_x^t] = \mathbb{E}[Pr[I \mid r, \sigma_x^t] \mid r \in R_p, \sigma_x^t] = \bar{z} > 0,$$

where the first equality uses the assumption that $F(x)$ is continuous at x_p.

□

Proposition 3.3.1 implies that, in any equilibrium, there can be at most two lotteries over reputations induced with positive probability. However, reports that contain the same information about the expert's type (i.e. that induce the same lottery over reputations) could potentially contain different information about the state of nature. The following proposition proves that this is actually not the case.

Proposition 3.3.2. *In any equilibrium, if $r, r' \in R$ are sent along the equilibrium path and induce the same lottery over reputations, then we must have $x^{\sigma_x^t}(r) = x^{\sigma_x^t}(r')$, where σ_x^t is the principal's (correct) equilibrium conjecture about the expert's behavior.*

Proof. See appendix 3.6.1. □

Therefore, in any informative equilibrium the principal can end up (with positive probability) with at most two different posteriors about the state of nature. As suggested by this observation, the next result, which is a corollary of Propositions 3.3.1 and 3.3.2, states that there is no loss of generality in considering only equilibria in which at most two reports are sent. Formally,

Corollary 3.3.1. *Any equilibrium outcome can be induced by an equilibrium in which at most two reports are sent.*

Proof. Take any equilibrium. From Propositions 3.3.1 we know that reports sent along the equilibrium path can be divided into three groups: R_0, R_1 and R_p. Suppose that the expert merges all the reports that belong to each group and sends a single report, r_i whenever he was gong to send a report in R_i, with $i = 0, 1, p$. By Propositions 3.3.1 and 3.3.2, for $i = 0, 1, p$,

r_i must indue the same lottery over reputations and the same posterior about the state of nature as each report in R_i. Finally, suppose that the expert sends r_0 whenever he was going to send r_p. Without loss of generality, assume that r_0 was being sent with positive probability by both the informed expert and the charlatan. Then, since, by Proposition 3.3.1, r_p was sent with probability zero by the charlatan and the informed expert, this change does not affect the lotteries over reputations or the posteriors about the state induced by r_0 and r_1. Note that, in the new equilibrium, the expert's expected utility conditional on his information is the same as in the original equilibrium (see Figure 3.3.1). Moreover, the principal is induced to take her decision conditional on the information available in the same way as she did originally (we assume that whenever she was indifferent between the decisions, she always chose the status quo). Thus, the outcome implemented by the equilibrium with two reports coincides with the original outcome. □

It remains to be seen that there actually exists an informative equilibrium where two reports are sent. The following proposition states that, indeed, such equilibrium always exists and it is unique.

Proposition 3.3.3. *There exists a unique informative equilibrium in which two reports are sent. The expert must be indifferent between the reports when his posterior about the state is equal to the prior.*

Proof. See appendix 3.6.2. □

Therefore, as we wanted to show, this model has a unique equilibrium outcome in which some information is transmitted by the expert.

3.4 The value of information in the presence of charlatans

The purpose of this section is to emphasize how, if the expert is concerned about his reputation, the mere possibility (no matter how small) that he may be a charlatan severely reduces the value of the information that he possesses. Indeed, for any prior belief that the expert is informed, $z_p \in (0,1)$, in the unique informative equilibrium outcome of the game decision-making is distorted away from the status quo. That is, there are posteriors under which d_0 would be the efficient decision, but where d_1 is taken instead. There is a discontinuity in the value of the expert's information in the sense that the induced distortion to decision-making does not disappear as z_p tends to one.

We start by considering what would happen in the model if the expert's type was known to the principal (or if the expert was not concerned about his reputation). In this case, it would always be incentive compatible for the expert to reveal his information, and the principal would always take the efficient decision conditional on the information available. Figure 3.4.1 illustrates the principal's decision problem. The horizontal axis represents the expert's

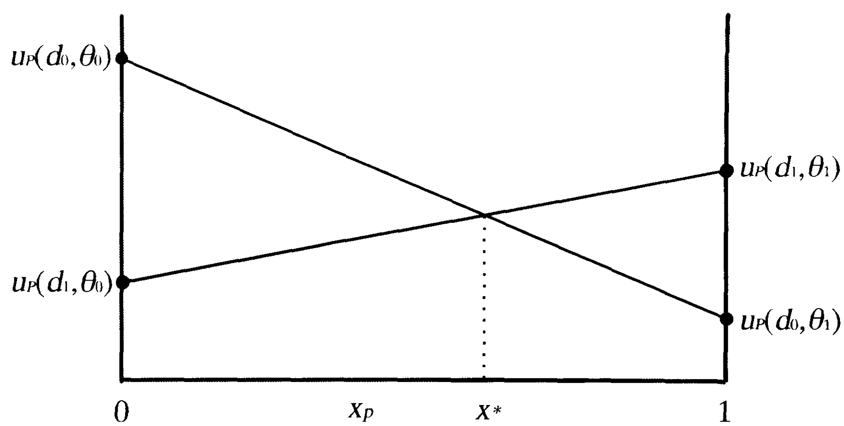


Figure 3.4.1: The principal's decision problem

posterior about the state, and each line gives the principal's expected utility conditional on each posterior after taking one of the decisions. As we can see, d_0 is the efficient decision if the posterior is to the left of x^*, and d_1 is the efficient decision otherwise. Note that, as long as $F(x^*) < 1$, under truthful revelation, the expert's information influences the principal's decision, so the expert's advice is valuable.

When the expert is concerned about his reputation, the outcome of the game drastically changes. As we saw in the previous section, in this case there is a unique informative equilibrium outcome implementable by an equilibrium in which the expert only sends two reports, say r_0 and r_1. The informed expert sends r_0 when his posterior belief that the state is θ_1 is smaller than x_p, and r_1 otherwise. The charlatan randomizes between the two reports. For each $z_p \in (0,1)$, let $\sigma_x^{t*}(z_p)$ denote the expert's equilibrium report rule, and $x^{\sigma_x^{t*}(z_p)}(r_i)$ be the principal's posterior belief that the state is θ_1 after observing the report r_i $(i = 0,1)$. Now, even if $F(x^*) < 1$, it is no longer true that the expert's advice will always influence the principal's decision. Indeed, this is only the case if $x^{\sigma_x^{t*}(z_p)}(r_1) > x^*$. If this inequality does not hold, the expert's advice becomes worthless. Moreover, even when the inequality holds, the expert's information loses value with respect to the truth-telling benchmark. This happens because, whenever the charlatan or the informed expert with a posterior in the interval (x_p, x^*) send the report r_1, the principal is induced to take decision d_1, while d_0 would have been efficient. Formally, the following is a corollary of Corollary 3.3.1 and Proposition 3.3.3.

Corollary 3.4.1. *In any equilibrium in which the expert's report influences the principal's decision, decision-making is biased away from the status quo.*

This result is rather intuitive. Under truth-telling, d_1 would only be taken if the expert is informed. Thus, in the unique informative equilibrium outcome of the game, d_1 carries a reputational premium for the expert, who ends up inducing this decision too frequently.

Note that the size of the information loss caused by the presence of charlatans depends on the nature of the principal's decision problem. For instance, in the extreme case where $x^* = x_p$, information would never be wasted, for the principal's decision would always be taken efficiently given the information available. In this case, the expert's advice is always valuable. More generally, other things equal, the closer x^* gets to x_p (the more insecure the principal is about the decision she should take under the prior), the smaller the waste of information and, thus, the more likely it becomes that the expert's advice will influence the principal's decision.

Finally, the last result shows that, for any $x^* > x_p$, when the prior probability that the expert is informed becomes small enough, the expert's advice stops influencing the principal's decision, and so becomes worthless.

Proposition 3.4.1. *The function* $x^{\sigma_x^{t*}(z_p)}(r_0)$ *is strictly decreasing in* z_p*, the function* $x^{\sigma_x^{t*}(z_p)}(r_1)$ *is strictly increasing in* z_p*, and* $\lim_{z_p \to 0^+} x^{\sigma_x^{t*}(z_p)}(r_0) = \lim_{z_p \to 0^+} x^{\sigma_x^{t*}(z_p)}(r_1) = x_p$.

Proof. See appendix 3.6.3. □

Thus, even though the informed expert has valuable information for the principal's decision problem (as long as $F(x^*) < 1$), when the principal's prior belief that the expert is informed becomes sufficiently small, she will prefer to not consult the expert.

3.5 Concluding remarks

This paper has analyzed a model in which an expert who is concerned about appearing to be informed advices a principal before she takes a decision. We have shown how the mere possibility that the expert can be a charlatan drastically reduces the value of the expert's information. The model has a unique informative equilibrium outcome which can be implemented with an equilibrium in which the expert only sends two reports. Thus, the information that can be transmitted by the expert is endogenously coarse. Moreover, in this equilibrium the expert must be indifferent between the two reports when his posterior about the state is equal to the prior. This implies that, if the expert's advice influences the principal's decision, decision-making is biased away from the status quo. Remarkably, this distortion does not disappear as the principal's prior belief that the expert is informed approaches one.

Coarse information transmission seems to be a robust feature of communication games where the sender is concerned about appearing well informed.[11] However, the result that the expert can only send two reports containing different information seems more delicate. Indeed, the geometric argument provided in the proof of Proposition 3.3.1 relies on the fact

[11] For instance, Ottaviani and Sørensen (2006*b*) show that, under very general conditions, truth-telling cannot be an equilibrium regardless of whether the expert knows his type. In Ottaviani and Sørensen (2006*a*), they study a model where, as here, any equilibrium outcome can be induced by an equilibrium where only two reports are sent.

that in this model there are only two states of nature, and cannot be generalized to the case where there are more states. Similarly, even though the result that decision making is biased away from the status quo is intuitive, it should be taken with a grain of salt. As discussed in the introduction, if there were a continuum of types for the expert, distortions to decision-making could go in the opposite direction.

A potentially interesting direction for future research could be to look at how contingent payments could be designed to correct the distortions introduced by the reputational concerns of experts.[12] In particular, it would be useful to understand how such incentive schemes would compare to those designed to induce effort, and whether they would resemble the contracts that we actually observe.[13]

3.6 Appendix

3.6.1 Proof of Proposition 3.3.2

Suppose that $r, r' \in R$ lead to the same lottery over reputations given the principal's conjecture about the expert's behavior, σ_x^t. That is, we have

$$z^{\sigma_x^t}(r,\theta_1) = \frac{1}{1+\frac{1-z_p}{z_p}\frac{Pr[r|Ch,\sigma_x^t]}{Pr[r|I,\sigma_x^t]}\frac{Pr[\theta_1|r,Ch,\sigma_x^t]}{Pr[\theta_1|r,I,\sigma_x^t]}} = \frac{1}{1+\frac{1-z_p}{z_p}\frac{Pr[r'|Ch,\sigma_x^t]}{Pr[r'|I,\sigma_x^t]}\frac{Pr[\theta_1|r',Ch,\sigma_x^t]}{Pr[\theta_1|r',I,\sigma_x^t]}} = z^{\sigma_x^t}(r',\theta_1), \tag{3.6.1}$$

and

$$z^{\sigma_x^t}(r,\theta_0) = \frac{1}{1+\frac{1-z_p}{z_p}\frac{Pr[r|Ch,\sigma_x^t]}{Pr[r|I,\sigma_x^t]}\frac{Pr[\theta_0|r,Ch,\sigma_x^t]}{Pr[\theta_0|r,I,\sigma_x^t]}} = \frac{1}{1+\frac{1-z_p}{z_p}\frac{Pr[r'|Ch,\sigma_x^t]}{Pr[r'|I,\sigma_x^t]}\frac{Pr[\theta_0|r',Ch,\sigma_x^t]}{Pr[\theta_0|r',I,\sigma_x^t]}} = z^{\sigma_x^t}(r',\theta_0), \tag{3.6.2}$$

Note that $Pr[\theta_1 \mid r, Ch, \sigma_x^t] = Pr[\theta_1 \mid r', Ch, \sigma_x^t] = x_p$. Thus, if $Pr[\theta_1 \mid r, I, \sigma_x^t] > Pr[\theta_1 \mid$

[12] Holmström and I Costa (1986) and Caruana and Celentani (2001), for example, work along these lines.

[13] For models in which contingent payments are designed to induce experts to acquire costly information see, for example, Demski and Sappington (1987), Diamond (1998), Athey and Roberts (2001), Malcomson (2009), Zermeño (2011*b*) and Zermeño (2011*a*).

$r', I, \sigma_x^t]$, for equation 3.6.1 to hold we must have $\frac{Pr[r|Ch,\sigma_x^t]}{Pr[r|I,\sigma_x^t]} > \frac{Pr[r'|Ch,\sigma_x^t]}{Pr[r'|I,\sigma_x^t]}$. But we also have $Pr[\theta_0 \mid r, I, \sigma_x^t] < Pr[\theta_0 \mid r', I, \sigma_x^t]$, so equation 3.6.2 could not be satisfied. By a similar argument, it could not be the case that $Pr[\theta_1 \mid r, I, \sigma_x^t] < Pr[\theta_1 \mid r', I, \sigma_x^t]$, so we must have $Pr[\theta_1 \mid r, I, \sigma_x^t] = Pr[\theta_1 \mid r', I, \sigma_x^t]$.

An equivalent way of writing equations 3.6.1 and 3.6.2 is:

$$z^{\sigma_x^t}(r, \theta_1) = z_p \frac{Pr[r \mid I, \sigma_x^t]}{Pr[r]} \frac{Pr[\theta_1 \mid r, I, \sigma_x^t]}{Pr[\theta_1 \mid r, \sigma_x^t]} = z_p \frac{Pr[r' \mid I, \sigma_x^t]}{Pr[r']} \frac{Pr[\theta_1 \mid r', I, \sigma_x^t]}{Pr[\theta_1 \mid r', \sigma_x^t]} = z^{\sigma_x^t}(r', \theta_1),$$

and

$$z^{\sigma_x^t}(r, \theta_0) = z_p \frac{Pr[r \mid I, \sigma_x^t]}{Pr[r]} \frac{Pr[\theta_0 \mid r, I, \sigma_x^t]}{Pr[\theta_0 \mid r, \sigma_x^t]} = z_p \frac{Pr[r' \mid I, \sigma_x^t]}{Pr[r']} \frac{Pr[\theta_0 \mid r', I, \sigma_x^t]}{Pr[\theta_0 \mid r', \sigma_x^t]} = z^{\sigma_x^t}(r', \theta_0).$$

Since $Pr[\theta_1 \mid r, I, \sigma_x^t] = Pr[\theta_1 \mid r', I, \sigma_x^t]$, for both equation to hold we need $Pr[\theta_1 \mid r, \sigma_x^t] = Pr[\theta_1 \mid r', \sigma_x^t]$, so $x^{\sigma_x^t}(r) = x^{\sigma_x^t}(r')$ as we wanted to show.

3.6.2 Proof of Proposition 3.3.3

The proof is constructive. From Proposition 3.3.1, we know that for an equilibrium with two reports to be informative, each report must induce a different lottery over reputations and the expert must be indifferent between the reports when his posterior about the state is x_p. Thus, the expert must strictly prefer to send one report (say, r_1) after observing posteriors greater than the prior, and the other report (say, r_0) after observing posteriors smaller than the prior. Since $F(x)$ is continuous at x_p, what the report that the informed expert chooses at the prior is irrelevant for the outcome. Thus, the only variable we have to play with is the probability with which the charlatan sends r_0, $\sigma_{x_p}^{Ch}(r_0)$ (he sends r_1 with the complementary probability). The principal's consistent beliefs about the expert's type are:

$$z^{\sigma_x^t}(r_0, \theta_1) = \frac{1}{1 + \frac{1-z_p}{z_p} \frac{x_p \sigma_{x_p}^{Ch}(r_0)}{\int_{-\infty}^{x_p} x dF(x)}};$$

$$z^{\sigma_x^t}(r_0,\theta_0) = \frac{1}{1+\frac{1-z_p}{z_p}\frac{(1-x_p)\sigma_{x_p}^{Ch}(r_0)}{\int_{-\infty}^{x_p}(1-x)dF(x)}};$$

$$z^{\sigma_x^t}(r_1,\theta_1) = \frac{1}{1+\frac{1-z_p}{z_p}\frac{x_p(1-\sigma_{x_p}^{Ch}(r_0))}{\int_{x_p}^{\infty}xdF(x)}};$$

$$z^{\sigma_x^t}(r_1,\theta_0) = \frac{1}{1+\frac{1-z_p}{z_p}\frac{(1-x_p)(1-\sigma_{x_p}^{Ch}(r_0))}{\int_{x_p}^{\infty}(1-x)dF(x)}}.$$

Note that $z^{\sigma_x^t}(r_0,\theta_0) > z^{\sigma_x^t}(r_0,\theta_1)$ if and only if $x_p\int_{\infty}^{x_p}dF(x) > \int_{-\infty}^{x_p}xdF(x)$, which holds. Similarly, $z^{\sigma_x^t}(r_1,\theta_1) > z^{\sigma_x^t}(r_1,\theta_0)$. Thus, indeed, if the expert is indifferent between the reports at x_p, he strictly prefers to send r_0 after observing $x < x_p$, and r_1 after observing $x > x_p$.

The last part of the proof is to show that there exists a unique $\sigma_{x_p}^{Ch}(r_0) \in [0,1]$ that generates lotteries over reputations that make the expert indifferent between the reports at x_p. That is, we must show that there exists a unique $\sigma_{x_p}^{Ch}(r_0) \in [0,1]$ such that

$$x_p u_E(z^{\sigma_x^t}(r_0,\theta_1)) + (1-x_p)u_E(z^{\sigma_x^t}(r_0,\theta_0) = x_p u_E(z^{\sigma_x^t}(r_1,\theta_1)) + (1-x_p)u_E(z^{\sigma_x^t}(r_1,\theta_0).$$

Note that the LHS of the equation is continuous and strictly decreasing in $\sigma_{x_p}^{Ch}(r_0)$, and the RHS is continuous and strictly increasing. Moreover, if $\sigma_{x_p}^{Ch}(r_0) = 0$, then LHS>RHS, and, if $\sigma_{x_p}^{Ch}(r_0) = 1$, then LHS<RHS. Thus, there exists a unique $\sigma_{x_p}^{Ch}(r_0) \in (0,1)$ for which the equality holds.

3.6.3 Proof of Proposition 3.4.1

Note that

$$x^{\sigma_x^t}(r_0) = Pr[\theta_1 \mid r_0, \sigma_x^t] \equiv \frac{z_p\int_{-\infty}^{x_p}xdF(x) + (1-z_p)x_p\sigma_{x_p}^{Ch}(r_0)}{z_pF(x_p) + (1-z_p)\sigma_{x_p}^{Ch}(r_0)}.$$

Thus,

$$\frac{(1-z_p)}{z_p}\sigma_{x_p}^{Ch}(r_0) = \frac{x^{\sigma_x^t}(r_0)F(x_p) - \int_{-\infty}^{x_p}xdF(x)}{x_p - x^{\sigma_x^t}(r_0)}.$$

Substituting in the expression derived in the proof of Proposition 3.3.3, we obtain:

$$z^{\sigma_x^t}(r_0,\theta_1) = \frac{1}{1+\frac{x^{\sigma_x^t}(r_0)F(x_p)-\int_{-\infty}^{x_p} xdF(x)}{x_p-x^{\sigma_x^t}(r_0)}\frac{x_p}{\int_{-\infty}^{x_p} xdF(x)}};$$

$$z^{\sigma_x^t}(r_0,\theta_0) = \frac{1}{1+\frac{x^{\sigma_x^t}(r_0)F(x_p)-\int_{-\infty}^{x_p} xdF(x)}{x_p-x^{\sigma_x^t}(r_0)}\frac{(1-x_p)}{\int_{-\infty}^{x_p}(1-x)dF(x)}};$$

$$z^{\sigma_x^t}(r_1,\theta_1) = \frac{1}{1+\left(\frac{1-z_p}{z_p}-\frac{x^{\sigma_x^t}(r_0)F(x_p)-\int_{-\infty}^{x_p} xdF(x)}{x_p-x^{\sigma_x^t}(r_0)}\right)\frac{x_p}{\int_{x_p}^{\infty} xdF(x)}};$$

$$z^{\sigma_x^t}(r_1,\theta_0) = \frac{1}{1+\left(\frac{1-z_p}{z_p}-\frac{x^{\sigma_x^t}(r_0)F(x_p)-\int_{-\infty}^{x_p} xdF(x)}{x_p-x^{\sigma_x^t}(r_0)}\right)\frac{(1-x_p)}{\int_{x_p}^{\infty}(1-x)dF(x)}}.$$

From Proposition 3.3.3 we know that for each $z_p \in (0,1)$, there exists a unique $\sigma_x^{t*}(z_p)$ such that

$$x_p u_E(z^{\sigma_x^{t*}}(r_0,\theta_1)) + (1-x_p)u_E(z^{\sigma_x^{t*}}(r_0,\theta_0) = x_p u_E(z^{\sigma_x^{t*}}(r_1,\theta_1)) + (1-x_p)u_E(z^{\sigma_x^{t*}}(r_1,\theta_0).$$

Note that the LHS is strictly decreasing in $x^{\sigma_x^t}(r_0)$, and the RHS is strictly increasing. As z_p decreases, the RHS decreases, so to restore the equality $x^{\sigma_x^t}(r_0)$ must increase. Thus, $x^{\sigma_x^{t*}(z_p)}(r_0)$ is an strictly decreasing function of z_p. Moreover, as $z_p \to 0^+$, the RHS goes to infinity, so we must have $\lim_{z_p\to 0^+} x^{\sigma_x^{t*}(z_p)}(r_0) = x_p$.

With a similar argument, we can show that $x^{\sigma_x^{t*}(z_p)}(r_1)$ is an strictly increasing function of z_p, and that it must be the case that $\lim_{z_p\to 0^+} x^{\sigma_x^{t*}(z_p)}(r_1) = x_p$.

Bibliography

Aghion, P., and J. Tirole. 1997. "Formal and real authority in organizations." *Journal of Political Economy*, 105(1): 1–29.

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

Alonso, R., W. Dessein, and N. Matouschek. 2008. "When does coordination require centralization?" *The American economic review*, 98(1): 145–179.

Arrow, Kenneth J. 1974. *The limits of organization.* W.W. Norton & Company.

Athey, S., and J. Roberts. 2001. "Organizational design: Decision rights and incentive contracts." *American Economic Review*, 91(2): 200–205.

Avery, C.N., and J.A. Chevalier. 1999. "Herding over the career." *Economics Letters*, 63(3): 327–333.

Baker, G., R. Gibbons, and K.J. Murphy. 1999. "Informal authority in organizations." *Journal of Law, Economics, and Organization*, 15(1): 56.

Bergemann, D., and J. Välimäki. 2002. "Information acquisition and efficient mechanism design." *Econometrica*, 70(3): 1007–1033.

Blackwell, D. 1953. "Equivalent comparisons of experiments." *The Annals of Mathematical Statistics*, 24(2): 265–272.

Caruana, G., and M. Celentani. 2001. "Career concerns and contingent compensation." *UC3M Working Paper.*

Chade, H., and E. Schlee. 2002. "Another Look at the Radner-Stiglitz Nonconcavity in the Value of Information." *Journal of Economic Theory*, 107(2): 421–452.

Chade, Hector, and Natalia Kovrijnykh. 2011. "Information acquisition, moral hazard, and rewarding for bad news." Unpublished manuscript, Arizona State University.

Chevalier, J., and G. Ellison. 1997. "Risk Taking by Mutual Funds as a Response to Incentives." *The Journal of Political Economy*, 105(6): 1167–1200.

Chevalier, J., and G. Ellison. 1999. "Career concerns of mutual fund managers." *The Quarterly Journal of Economics*, 114(2): 389–432.

Crawford, V.P., and J. Sobel. 1982. "Strategic information transmission." *Econometrica*, 50(6): 1431–1451.

Crémer, J., F. Khalil, and J.C. Rochet. 1998*a*. "Contracts and Productive Information Gathering." *Games and Economic Behavior*, 25(2): 174–193.

Crémer, J., F. Khalil, and J.C. Rochet. 1998*b*. "Strategic information gathering before a contract is offered." *Journal of Economic Theory*, 81(1): 163–200.

Demski, J.S., and D.E.M. Sappington. 1987. "Delegated expertise." *Journal of Accounting Research*, 68–89.

Dessein, W. 2002. "Authority and communication in organizations." *Review of Economic Studies*, 69(4): 811–838.

Dewatripont, M., and J. Tirole. 1999. "Advocates." *Journal of Political Economy*, 107(1): 1–39.

Diamond, P. 1998. "Managerial incentives: on the near linearity of optimal compensation." *The Journal of Political Economy*, 106(5): 931–957.

Effinger, M.R., and M.K. Polborn. 2001. "Herding and anti-herding: A model of reputational differentiation." *European Economic Review*, 45(3): 385–403.

Ehrbeck, T., and R. Waldmann. 1996. "Why are professional forecasters biased? Agency versus behavioral explanations." *The Quarterly Journal of Economics*, 111(1): 21–40.

Fudenberg, D., and J. Tirole. 1991. *Game Theory.* MIT Press.

Gibbons, R., and K.J. Murphy. 1992. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *The Journal of Political Economy*, 100(3): 468–505.

Gromb, D., and D. Martimort. 2007. "Collusion and the organization of delegated expertise." *Journal of Economic Theory*, 137(1): 271–299.

Grossman, S.J., and O.D. Hart. 1983. "An analysis of the principal-agent problem." *Econometrica*, 7–45.

Harris, M., and A. Raviv. 1979. "Optimal incentive contracts with imperfect information." *Journal of Economic Theory*, 20(2): 231–259.

Holmström, B. 1977. "On incentives and control in organizations." PhD diss. Stanford University.

Holmström, B. 1979. "Moral hazard and observability." *The Bell Journal of Economics*, 10(1): 74–91.

Holmström, B. 1984. "On the theory of delegation." *Bayesian models in economic theory*, 115.

Holmström, B. 1999. "Managerial incentive problems: A dynamic perspective." *Review of Economic Studies*, 66(1): 169–182.

Holmström, B., and J.R. I Costa. 1986. "Managerial incentives and capital management." *The Quarterly Journal of Economics*, 101(4): 835–860.

Holmström, B., and P. Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, and Organization*, 7(special issue): 24.

Inderst, R., and M. Klein. 2007. "Innovation, endogenous overinvestment, and incentive pay." *The RAND Journal of Economics*, 38(4): 881–904.

Jehiel, P., and B. Moldovanu. 2001. "Efficient design with interdependent valuations." *Econometrica*, 69(5): 1237–1259.

Kim, S.K. 1995. "Efficiency of an information system in an agency model." *Econometrica*, 63(1): 89–102.

Krishna, V., and E. Maenner. 2001. "Convex potentials with an application to mechanism design." *Econometrica*, 69(4): 1113–1119.

Krishna, V., and J. Morgan. 2001. "A model of expertise." *The Quarterly Journal of Economics*, 116(2): 747.

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 J. Tirole. 1986. "Using cost observation to regulate firms." *The Journal of Political Economy*, 614–641.

Lambert, R.A. 1986. "Executive effort and selection of risky projects." *The Rand Journal of Economics*, 17(1): 77–88.

Levitt, S.D., and C.M. Snyder. 1997. "Is no news bad news? Information transmission and the role of "early warning" in the principal-agent model." *RAND Journal of Economics*, 28(4): 641–661.

Levy, G. 2004. "Anti-herding and strategic consultation." *European Economic Review*, 48(3): 503–525.

Lewis, T.R., and D.E.M. Sappington. 1997. "Information management in incentive problems." *Journal of Political Economy*, 105(4): 796–821.

Malcomson, J.M. 2009. "Principal and Expert Agent." *The BE Journal of Theoretical Economics*, 9(1).

Malcomson, J.M. 2011. "Do managers with limited liability take more risky decisions? An information acquisition model." *Journal of Economics & Management Strategy*, 20(1): 83–120.

Milgrom, P., and I. Segal. 2002. "Envelope theorems for arbitrary choice sets." *Econometrica*, 70(2): 583–601.

Mirrlees, J.A. 1976. "The optimal structure of incentives and authority within an organization." *The Bell Journal of Economics*, 105–131.

Müller, R., A. Perea, and S. Wolf. 2007. "Weak monotonicity and Bayes-Nash incentive compatibility." *Games and Economic Behavior*, 61(2): 344–358.

Osband, K. 1989. "Optimal forecasting incentives." *The Journal of Political Economy*, 97(5): 1091–1112.

Ottaviani, M., and P.N. Sørensen. 2006*a*. "Professional advice." *Journal of Economic Theory*, 126(1): 120–142.

Ottaviani, M., and P.N. Sørensen. 2006*b*. "Reputational cheap talk." *The Rand journal of economics*, 37(1): 155–175.

Ottaviani, M., and P.N. Sørensen. 2006*c*. "The strategy of professional forecasting." *Journal of Financial Economics*, 81(2): 441–466.

Persico, N. 2000. "Information acquisition in auctions." *Econometrica*, 68(1): 135–148.

Prendergast, C. 1993. "A Theory of Yes Men." *The American Economic Review*, 83(4): 757–770.

Prendergast, C. 2002. "The Tenuous Trade-off between Risk and Incentives." *Journal of Political Economy*, 110(5): 1071–1102.

Prendergast, C., and L. Stole. 1996. "Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning." *Journal of Political Economy*, 1105–1134.

Radner, R., and J. Stiglitz. 1984. "A Nonconcavity in the Value of Information." *Bayesian models in economic theory*, 5: 33–52.

Rantakari, H. 2008. "Governing Adaptation." *Review of Economic Studies*, 75(4): 1257–1285.

Rantakari, H. 2010. "Employee Initiative and Managerial Control." Unpublished manuscript, USC.

Rantakari, H. 2011. "Organizational design and environmental volatility." *Journal of Law, Economics, and Organization.* Forthcoming.

Rochet, J.C. 1987. "A necessary and sufficient condition for rationalizability in a quasilinear context." *Journal of Mathematical Economics*, 16(2): 191–200.

Rochet, J.C., and P. Chone. 1998. "Ironing, sweeping, and multidimensional screening." *Econometrica*, 783–826.

Rogerson, W.P. 1985. "The first-order approach to principal-agent problems." *Econometrica*, 1357–1367.

Rothschild, M., and J.E. Stiglitz. 1970. "Increasing risk: I. A definition." *Journal of Economic theory*, 2(3): 225–243.

Scharfstein, D.S., and J.C. Stein. 1990. "Herd behavior and investment." *The American Economic Review*, 465–479.

Shavell, S. 1979. "Risk sharing and incentives in the principal and agent relationship." *The Bell Journal of Economics*, 55–73.

Szalay, D. 2005. "The economics of clear advice and extreme options." *Review of Economic Studies*, 72(4): 1173–1198.

Szalay, D. 2009. "Contracts with endogenous information." *Games and Economic Behavior*, 65(2): 586–625.

Trueman, B. 1994. "Analyst forecasts and herding behavior." *Review of financial studies*, 7(1): 97–124.

Zermeño, Luis. 2011*a*. "A principal-expert model and the value of menus." Unpublished manuscript, MIT.

Zermeño, Luis. 2011*b*. "The role of authority in a general principal-expert model." Unpublished manuscript, MIT.

Zwiebel, J. 1995. "Corporate conservatism and relative compensation." *Journal of Political Economy*, 1–25.