Robust Incentives for Information Acquisition

Gabriel Carroll, Microsoft Research and Stanford University

gabriel.d.carroll@gmail.com

October 2, 2012

Abstract

A principal needs to make a decision, and wishes to incentivize an expert to acquire relevant information, which requires costly effort. The expert can be rewarded based on his reported information, and on the state of nature, which is revealed ex post. Both parties are financially risk-neutral, and payments are constrained by limited liability. The principal is uncertain about the expert's information acquisition technology: she knows of some experiments the expert can perform, but there may also be other experiments available. In the face of this uncertainty, the principal evaluates incentive contracts using a maxmin expected utility criterion. Under quite general conditions, the optimal contract is a *restricted investment contract*, in which the expert chooses from a subset of the decisions available to the principal, and is paid proportionally to the value of his chosen decision in the realized state.

This is a preliminary draft. Please do not circulate without permission. A public version will be available by late October 2012 on my website at `http://research.microsoft.com/en-us/people/gabrcarr/` .

Thanks to (in random order) Sylvain Chassang, Bengt Holmström, Iván Werning, Alex Wolitzky, Ben Golub, Yusuke Narita, Abhijit Banerjee, Daron Acemoglu, and Luis Zermeño for helpful discussions and advice.

1 Introduction

We consider an agency problem in which a principal needs to make a decision, and seeks advice from an expert, who can privately obtain information relevant to the decision by exerting effort. The expert has no intrinsic preferences over the decision being made, but the principal can incentivize him to exert effort by making his payment depend on how well the information he reports corresponds to information revealed after the decision is made. More specifically, we assume that the state of nature (which determines the payoff from each possible decision) becomes publicly known ex post, and payments can depend on the reported information and the realized state. We assume both parties are financially risk-neutral, and payments to the expert are constrained by limited liability.

We build on the work of Zermeño [10, 11], who gave a very general formulation of the principal-expert problem. However, we depart from that work, and from most of the existing agency literature, by assuming that the expert's technology for acquiring information is not common knowledge. Instead, as in this author's previous work [2], we take a robust-contracting approach: The principal knows some actions (*experiments*) that the expert can perform to acquire information, but there may be other, unknown experiments available. The principal does not have a probabilistic belief about which experiments are and are not available. Rather, she evaluates incentive contracts based on their worst-case performance over all possible technologies the expert may have access to.

In simpler settings, this kind of robustness concern can lead to linear contracts, in which the agent is paid a constant share of the principal's payoff (see [2]). In the present problem, the optimal contract is a variation on a linear contract, which we call a *restricted investment contract*. Instead of paying the expert a constant share of the principal's own realized payoff, she pays him proportionally to the payoff that *would* have accrued (in the realized state) if she had been restricted to a certain subset of her possible decisions. By excluding risky decisions, the principal can make the limited liability constraint less severe and thus pay the expert less while maintaining incentives. In addition to this qualitative description, we also show that the optimal contract — a potentially complicated object — can actually be described in terms of a small number of parameters (one per state), so that the computational problem of optimizing in any given application is greatly simplified.

This paper aims to make two main points. One is literal: we give a recipe that can be used to write contracts in a particular agency setting, and the resulting contracts are optimally robust in a precise sense. Moreover, the form of the contracts that come out of the analysis is novel. The other point is methodological: we show how using a maxmin objective leads to a tidy and tractable model. By contrast, a Bayesian approach, with the principal maximizing expected utility with respect to some belief over the space of possible information acquisition technologies, would be unmanageable in general. (Even with common knowledge of the technology, the problem is difficult; [10] gives an explicit solution only for very special cases.)

Our assumption that the state is fully revealed ex post essentially means that the principal finds out what the payoff from any other decision would have been, not just the decision that was actually taken. This assumption is not realistic in all applications, but it is reasonable in some. For example, the principal may be an investor, seeking the expert's advice as to which assets to invest in, and ex post she can see the returns on the assets she didn't invest in, as well as those that she did. For another application, the principal may be a firm deciding how much to invest in developing each of several products. The returns to each product depend on exogenous future shocks to market demand, and the expert's job is to try to predict these shocks; ex post, the realized shocks can be inferred from market data.

By assuming that the state is fully revealed, we are able to separate the problem of choosing the decision from that of providing incentives to the expert. In a more general model, there would be a reason to distort decisions in favor of those that reveal more information about how hard the expert worked [10]. Here, all decisions are equally revealing, so the principal simply takes whichever decision is optimal condition on the reported information.

In the next section, we present the formal model. We then give a more detailed discussion of the intuition behind restricted investment contracts, before proceeding to the formal analysis showing that such contracts are optimal. The main proof is essentially an application of the same linear separation techniques used in [2]. In Section 3, we briefly discuss a number of extensions and variations, including the question of when an *un*restricted investment contract is or is not optimal.

Besides the work of Zermeño [10, 11] cited above, several other authors previously considered agency problems involving incentives to acquire information, in particular Demski and Sappington [5] (who introduced the term "expert") and Malcolmson [7]. However these models assumed that the decision is delegated to the expert, and that only the realized payoff is observed, not the entire state. Osband [8] considered a model more similar to the present paper in these respects, but focused on optimal screening with a prior belief over the expert's technology, in a very specific environment. In addition to this literature, the present paper also contributes to the literature on mechanism design with maxmin objectives, e.g. [1, 4, 6]; see the author's earlier paper [2] for more discussion.

2 Model and results

2.1 The setup

As already indicated, the model is quite similar to that of [10], with the main difference being the principal's non-quantifiable uncertainty about the expert's information acquisition technology.

The principal needs to choose a decision d from some compact space D. The payoff to each decision depends on a state of nature, ω, to be realized in the future. We assume the set Ω of possible states is finite. Payoffs are represented by a continuous function $u : D \times \Omega \to \mathbb{R}$. The principal has a prior belief about the state, $p_0 \in \Delta(\Omega)$, with $p_0(\omega) > 0$ for each $\omega \in \Omega$.

Before the principal makes her decision, she can hire the expert to obtain information about the state. The expert initially shares the prior p_0, but can obtain more information by performing an *experiment*. In principle, an experiment should be thought of as producing a joint distribution on states ω and observed signals (outcomes of the experiment), such that the marginal distribution over states is p_0. However, the signal will matter only through the expert's resulting posterior belief about the state, so we take the notational shortcut of representing experiments directly in terms of posteriors. Thus, we define an *experiment* to be a pair $(F, c) \in \Delta(\Delta(\Omega)) \times \mathbb{R}^+$ (here $\mathbb{R}^+$ is the set of nonnegative reals), such that F has mean p_0. The interpretation is that the expert can, at a cost of c, perform the experiment, and obtain a posterior (an element of $\Delta(\Omega)$) drawn from distribution F. The requirement that F should have mean p_0 is simply the law of iterated expectations — the posterior should, in expectation, be equal to the prior. We will typically use the variable p for a posterior.

We give $\Delta(\Delta(\Omega)) \times \mathbb{R}^+$ the natural product topology, and define an *information acquisition technology* (IAT) to be a compact subset of $\Delta(\Delta(\Omega)) \times \mathbb{R}^+$. An IAT then describes the set of experiments available to the agent. There is some exogenously given IAT, $\mathcal{I}_0$, consisting of the experiments which are mutually known to be available. From the principal's point of view, the true set of experiments (known only to the agent) may potentially be any IAT $\mathcal{I}$ such that $\mathcal{I} \supseteq \mathcal{I}_0$. We will sometimes consider $\mathcal{I}_0$ satisfying the following *full-support condition*: for every $(F, c) \in \mathcal{I}_0$, the support of F is all of $\Delta(\Omega)$, or else $F = \delta_{p_0}$ (a point mass on the prior).

After the contract is chosen, the expert can acquire information, and make a report to the principal, who then chooses decision $d \in D$. After the decision is made, the true state is revealed, and payments can be made contingent on all relevant observable information (the report and the realized state). Limited liability means that payments can never be negative. Thus, a *contract* is a pair (M, w), consisting of a *message space* M, some nonempty compact space; and a *payment function* $w : M \times \Omega \to \mathbb{R}^+$, which must be continuous. (The topological assumptions ensure that the expert's behavior is well-defined.)

Because we are interested specifically in *optimal* contracts for the principal, we may assume the expert reports the posterior as part of his message. Indeed, for any contract (M, w), we may consider a contract (M', w') with $M' = M \times \Delta(\Omega)$ and $w'(m, p, \omega) = w(m, \omega)$, so that the principal ignores the reported posterior when calculating payments, and uses it only to choose the decision d. Then the expert's incentives for information acquisition, and for the m component of his report, are exactly the same as under the original contract; and he is indifferent about the p component of his report, so he is willing to report the true posterior, which clearly allows the principal to choose at least as good a d as she could have chosen with only the message m. The full version of this argument depends on the description of behavior and on a more precise definition of contracts, but to avoid technical details we leave it to Appendix A, and for now simply assume henceforth that the expert reports p in addition to the message m. (In fact, we could apply the revelation principle more fully and assume that the expert reports *only* his posterior; this approach is used in [10, 11]. However, we will be interested in contracts that arguably are more naturally formulated using other message spaces.)

We can now summarize the timing of the game:

1. the principal offers a contract (M, w);
2. the expert, knowing $\mathcal{I}$, chooses experiment $(F, c) \in \mathcal{I}$;
3. a posterior $p \sim F$ is realized;
4. the expert chooses a message $m \in M$ to send, along with his posterior p;
5. the principal chooses decision d;
6. the state ω is revealed;
7. payoffs are received: $u(d, \omega) - w(m, \omega)$ for the principal; $w(m, \omega) - c$ for the expert.

It remains only to describe behavior. We give a brief summary here, and will introduce formal notation shortly. The expert knows $\mathcal{I}$, and he chooses (F, c) and then m to maximize his expected payoff. If the expert is indifferent between several choices, then we assume he acts so as to maximize the principal's expected payoff. The above behavior by the expert gives rise to an expected payoff for the principal for each possible IAT $\mathcal{I}$; we will notate this expected utility by $V_P(M, w, r|\mathcal{I})$. The principal evaluates each contract (M, w, r) by the worst possible value of V_P over all possible IAT's $\mathcal{I}$, knowing only that $\mathcal{I} \supseteq \mathcal{I}_0$. Finally, the question is how to design a contract to maximize this worst-case value.

Before developing the formal notation, we introduce some other useful objects. Suppose that the principal learns the expert's posterior is $p \in \Delta(\Omega)$. Then, clearly, she will choose d to maximize $E_p[u(d, \omega)]$. We denote this decision and the expected payoff by

$$d(p) = \arg\max_{d \in D} E_p[u(d, \omega)]; \qquad U(p) = \max_{d \in D} E_p[u(d, \omega)].$$

Note that U is convex in p, since it is the maximum of the affine functions $p \mapsto E_p[u(d, \omega)]$. Similarly, when the expert has posterior p, he will choose message m so as to maximize $E_p[w(m, \omega)]$. We denote

$$W(p) = \max_{m \in M} E_p[w(m, \omega)],$$

and call this function $W : \Delta(\Omega) \to \mathbb{R}^+$ the *reduced form* of the given contract. W is likewise convex in p.

The expert's incentives to acquire information depend only on the reduced form of the contract. Zermeño [11] characterized the possible reduced forms of contracts, as Lipschitz convex functions satisfying certain boundary conditions. However, for our purposes, it will be more helpful simply to think of W as the upper envelope of the affine functions $\Delta(\Omega) \to \mathbb{R}^+$ given by $p \mapsto E_p[w(m, \omega)]$, for each $m \in M$.

Now we can describe behavior formally. The expert, given posterior p, chooses message m as above, leading to expected payment $W(p)$. At the earlier, experiment-choosing stage, he knows the IAT $\mathcal{I}$ and so chooses $(F, c) \in \mathcal{I}$ to maximize expected payoff $E_F[W(p)] - c$. We will write the value of the contract (M, w) to the expert as

$$V_E(M, w|\mathcal{I}) = \max_{(F,c) \in \mathcal{I}} (E_F[W(p)] - c)$$

and the expert's choice set as

$$I^*(M, w|\mathcal{I}) = \arg\max_{(F,c)\in\mathcal{I}} \left(E_F[W(p)] - c\right).$$

When the principal learns the posterior p, whe will make decision $d(p)$, gaining expected gross payoff $U(p)$. Thus, her expected net payoff from the contract under IAT $\mathcal{I}$ is

$$V_P(M, w|\mathcal{I}) = \max_{(F,c)\in I^*(M,w|\mathcal{I})} \left(E_F[U(p)] - E_F[W(p)]\right)$$

(with the assumption that the expert breaks indifference in favor of the principal). The principal evaluates each possible contract ex ante by its worst-case expected payoff over all IAT's $\mathcal{I}$:

$$V_P(M, w) = \inf_{\mathcal{I}\supseteq\mathcal{I}_0} V_P(M, w|\mathcal{I}).$$

The principal's problem, which we analyze, is then how to choose the contract (M, w) to maximize V_P.

From here on, we will maintain the *non-triviality* assumption that there exists some contract (M, w) with $V_P(M, w) > U(p_0)$. That is, the principal benefits from hiring the expert. We shall shortly see conditions on primitives that ensure that this assumption is satisfied.

2.2 Intuitions

How can the principal write a contract to guarantee herself a reasonably high expected payoff? A natural first try would be to use a linear contract: pay the expert some fixed fraction $\alpha \in (0, 1]$ of the principal's gross payoff, adjusted by some constant β chosen so that the limited-liability constraint binds. With contracts defined as above, this could be implemented by setting $M = D$ and

$$w(m, \omega) = \alpha(u(m, \omega) + \beta). \tag{2.1}$$

(Explicitly, the relevant value of β is $\beta = -\min_{d,\omega} u(d, \omega)$.) For any given posterior, the expert would then send a message equal to the principal's optimal decision, $m = d(p)$.

We can compute a payoff guarantee from such a contract just as in [2] (see also [3]). If the expert chooses experiment (F, c), his expected payoff will be $E_F[W(p)] - c =$

$\alpha(E_F[U(p)] + \beta) - c$. Hence we have

$$\alpha(E_F[U(p)] + \beta) \geq \alpha(E_F[U(p)] + \beta) - c = V_E(M, w|\mathcal{I}) \geq V_E(M, w|\mathcal{I}_0).$$

Whenever the expert earns some payoff $\alpha(y+\beta)$, the principal earns $(1-\alpha)y - \alpha\beta$. Hence, the principal's expected payoff will be

$$\begin{aligned}(1-\alpha)E_F[U(p)] - \alpha\beta &\geq \frac{1-\alpha}{\alpha} V_E(M, w|\mathcal{I}_0) - \beta \\ &= \frac{1-\alpha}{\alpha}\left[\max_{(F,c)\in\mathcal{I}_0} \alpha(E_F[U(p)] + \beta) - c\right] - \beta \\ &= \max_{(F,c)\in\mathcal{I}_0}\left((1-\alpha)E_F[U(p)] - \frac{1-\alpha}{\alpha}c\right) - \alpha\beta. \end{aligned} \tag{2.2}$$

Thus the guarantee from the linear contract, $V_P(M, w)$, is at least the right-hand side of (2.2).

In particular, let F be any distribution over posteriors such that $E_F[U(p)] > U(p_0)$ — that is, any distribution that potentially provides useful information for the decision. One can check that, as long as

$$c < \left(\sqrt{E_F[U(p)] + \beta} - \sqrt{U(p_0) + \beta}\right)^2, \tag{2.3}$$

there exists α such that the right side of (2.2) is greater than $U(p_0)$. (This can be seen by choosing α to maximize the expression in (2.2), then solving the resulting inequality for c.) Hence, if $\mathcal{I}_0$ contains some such (F, c), the non-triviality condition is satisfied. This is our sufficient condition on primitives, as promised.

In the simpler, pure moral hazard setting of [2], linear contracts were optimal: the only way to turn a guarantee on the agent's expected payoff (provided by the known technology) into a guarantee on the principal's payoff was to have a linear relationship between the two. Here, however, there is scope for improving on linear contracts in two ways.

First, suppose we change a payment function by adding a quantity that depends only on the realized state: given payment function $w(m, \omega)$, replace it by $w(m, \omega) + \beta(\omega)$. This provides exactly the same incentives as the original payment function, since the expert has no control over the extra term. In symbols, the expected payoff from any experiment

(F, c) is

$$\begin{aligned} E_F[\max_m(E_p[w(m,\omega)+\beta(\omega)])]-c &= E_F[W(p)+E_p[\beta(\omega)]]-c \\ &= E_F[W(p)]+E_{p_0}[\beta(\omega)]-c \end{aligned}$$

since F has mean p_0; this differs from $E_F[W(p)]-c$ by a constant, so the expert's incentives as to which experiment to perform are unchanged. Hence, instead of (2.1), the principal may as well adjust the linear contract with state-by-state constants so that limited liability binds in each state:

$$w(m,\omega)=\alpha(u(m,\omega)+\beta(\omega)), \qquad \beta(\omega)=-\min_{d\in D} u(d,\omega).$$

The second reason why linear contracts may not be optimal is subtler: By cutting out risky decisions, the principal can relax the limited liability constraint, and thus pay a lower β. To see this, consider a situation with four decisions and two states, and the following payoffs:

	ω_1	ω_2
d_1	10	0
d_2	8	6
d_3	6	8
d_4	0	10

Each decision is optimal for some range of posteriors. This is shown in Figure 1, where beliefs are represented as numbers in $[0, 1]$ (representing the probability of state ω_2); each decision is shown as an affine function, mapping the belief p to the corresponding expected payoff, and $U(p)$ is the upper envelope of these four lines. Now, suppose the principal chooses a linear contract, with, say, share $\alpha = 1/2$. The corresponding choice of β is zero (in each state), so if the posterior is such that, say, d_2 is optimal, she will have to pay 4 in state ω_1 and 3 in state ω_2. However, if she cuts out decisions d_1 and d_2 from the message space, leaving only $M = \{d_3, d_4\}$, then she can adjust payments by $\beta = -3$ without violating limited liability, and so never pay more than 1. Now, by cutting out the extreme decisions, the principal may weaken the expert's incentives for information acquisition. However, if the known experiments rarely yield posteriors in the ranges where d_1 or d_4 are optimal, then the effect on incentives is small, and the change in β is more important. Note that under the new contract, the expert's *decisions* would not be restricted — she could still choose d_1 or d_4 if the expert happened to report an extreme posterior, but the expert would not be paid according to these decisions.

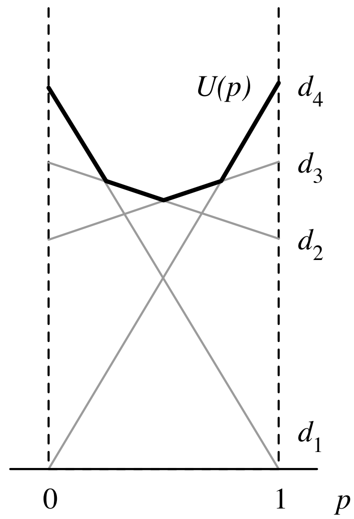

Figure 1: An example decision problem with risky decisions

To formalize this idea, we define a *restricted investment contract* as follows: the message space is some compact subset of decisions, $M \subseteq D$, and the payment function is given by

$$w(m, \omega) = \alpha(u(m, \omega) + \beta(\omega)) \tag{2.4}$$

for some $\alpha \in [0, 1]$ and some $\beta : \Omega \to \mathbb{R}$, such that the resulting payments are always nonnegative. The name emphasizes that the expert is allowed to "invest" in a restricted subset of decisions, and once the state is revealed, he is compensated according to the payoff that his chosen decision would have produced in that state.

What makes this setting different from the pure moral hazard setting, in which linear contracts are optimal? Here, there is a separation between the incentive instrument and the limited liability constraint: the expert's incentives are determined by the reduced form $W(p)$, whereas limited liability applies to the state-by-state payments $w(m, \omega)$. This separation means it is possible to relax limited liability appreciably while having only a small effect on the expert's incentives. As the arguments in the next subsection will make clear, relaxing limited liability is the only reason to depart from pure linearity.

2.3 Full analysis

So far, the discussion has emphasized restricted investment contracts. However, for the formal analysis, it will be useful to make a conceptual distinction between these and a related class, which we call *transform-bounded contracts*. Such a contract is parameterized by a value of $\alpha \in [0,1]$ and $\beta : \Omega \to \mathbb{R}$, and is defined as follows: the message space M is the set of all functions $m : \Omega \to \mathbb{R}^+$ such that

$$E_p[m(\omega)] \leq \alpha(U(p) + E_p[\beta(\omega)]) \qquad \text{for all } p \in \Delta(\Omega),$$

and the payment function is simply

$$w(m, \omega) = m(\omega).$$

As long as the values of β are large enough so that $U(p) + E_p[\beta(\omega)] \geq 0$ for all p, the resulting M is nonempty, and so we do obtain a contract.

We will say that two contracts are *equivalent* if they have the same reduced form. Notice that the values of $V_E(M, w|\mathcal{I})$, $I^*(M, w|\mathcal{I})$, $V_P(M, w|\mathcal{I})$, and therefore also $V_P(M, w)$ all depend on the contract (M, w) only through its reduced form; hence, any two equivalent contracts give the same payoff both to the expert and to the principal. Our analysis will show that transform-bounded contracts are optimal, and then that, under appropriate circumstances, transform-bounded contracts are equivalent to restricted investment contracts.

Thus, we state our first main result as:

Theorem 2.1. *There is an optimal contract that is transform-bounded. Moreover, if the known IAT $\mathcal{I}_0$ satisfies the full-support assumption, then* every *optimal contract is equivalent to a transform-bounded contract.*

For the second portion of the analysis, we will need one more definition. We will say that the decision problem (D, u) is *regular* if, for all $d, d' \in D$ and all $\lambda \in [0,1]$, there exists $\widehat{d} \in D$ such that

$$u(\widehat{d}, \omega) \geq \lambda u(d, \omega) + (1 - \lambda) u(d', \omega)$$

for all ω.

This can be thought of as a kind of convexity condition on the efficient frontier of decisions. For example, if we identify each decision with the vector of payoffs it generates (one for each state ω) so as to think of the decision space as a subset of $\mathbb{R}^\Omega$, then the decision space is regular if it is convex, e.g. if we allow for explicitly randomized decisions. Regularity is also satisfied in the application where decisions involve investment of wealth in several assets whose payoffs are state-dependent, if any asset allocation is possible, and payoffs are concave in investment within each state. (This could happen because of literally decreasing marginal returns, or constant returns but risk-aversion. To reconcile the latter interpretation with our assumption of risk-neutrality, we should think of the principal as risk-averse on the large scale of investment, but effectively risk-neutral on the scale of potential payments to the expert.)

The second part of our analysis is then encapsulated in the following, purely technical result:

Proposition 2.2. *Suppose the decision problem is regular. Then every transform-bounded contract is equivalent to a restricted investment contract. In fact, the transform-bounded contract with parameters α, β is equivalent to the restricted investment contract with the same parameters, and message space*

$$D_R(\beta) = \{d \in D \mid u(d, \omega) + \beta(\omega) \geq 0 \text{ for all } \omega\}, \tag{2.5}$$

unless $\alpha = 0$ in which case we can just take $M = D$.

Combining these two results immediately gives:

Corollary 2.3. *Suppose the decision problem is regular. Then there is some restricted investment contract that is optimal. If $\mathcal{I}_0$ satisfies the full-support assumption, then every optimal contract is equivalent to a restricted investment contract.*

Now that the results are stated, we begin the proofs. The proof of Theorem 2.1 closely parallels the analysis in [2]. We first characterize the worst-case payoff from any given contract. The description is straightforward, but because of the assumption that the expert's indifference is broken in favor of the principal, there are a few technical details to deal with.

We need to deal separately with *zero contracts*: those whose reduced form W satisfies $W(p) = 0$ for all p (equivalently, those satisfying $w(m, \omega) = 0$ for all m, ω). We denote the guarantee of a zero contract by $V_P(0)$. If there exists some experiment $(F, c) \in \mathcal{I}_0$ with $c = 0$, then for any $\mathcal{I}$, the agent will choose whichever such experiment maximizes the principal's expected payoff $E_F[U(p)]$; hence, $V_P(0)$ is simply $\max_{(F,0) \in \mathcal{I}_0} E_F[U(p)]$. If there is no such experiment in $\mathcal{I}_0$, then when the IAT is $\mathcal{I} = \mathcal{I}_0 \cup \{(\delta_{p_0}, 0)\}$, the agent will choose the latter experiment, and we see that $V_P(0) = U(p_0)$. (The principal will not do worse than this under any other IAT, by convexity of U.)

For nonzero contracts, we have the following.

Lemma 2.4. *Let (M, w) be any nonzero contract such that $V_P(M, w) \geq V_P(0)$. Let W be its reduced form. Then,*

$$V_P(M, w) = \min E_F[U(p) - W(p)] \qquad \text{over } F \in \Delta(\Delta(\Omega)) \text{ such that} \qquad (2.6)$$
$$E_F[p] = p_0 \text{ and } E_F[W(p)] \geq V_E[M, w|\mathcal{I}_0].$$

Moreover, if $V_P(M, w) > U(p_0)$, then for any F attaining the minimum, the expert's payoff constraint holds with equality: $E_F[W(p)] = V_E[M, w|\mathcal{I}_0]$.

Proof: One direction is trivial: For any $\mathcal{I} \supseteq \mathcal{I}_0$, and any experiment (F, c) chosen by the expert, $E_F[W(p)] \geq E_F[W(p)] - c \geq V_E[M, w|\mathcal{I}_0]$, so F satisfies the constraints in (2.6), and hence $V_P(M, w|\mathcal{I})$ is at least the indicated minimum. This holds for any $\mathcal{I}$, so $V_P(M, w)$ is at least the minimum in (2.6). (Note that this minimum is well-defined.)

To prove the reverse inequality, we begin by defining the affine function $Z : \Delta(\Omega) \to \mathbb{R}^+$ by $Z(p) = \sum_\omega p(\omega) W(\delta_\omega)$. By convexity, $W(p) \leq Z(p)$ for every p. Now let F be a distribution attaining the minimum in (2.6). Suppose F places positive probability on posteriors p such that $W(p) < Z(p)$. Then we have

$$E_F[W(p)] < E_F[Z(p)] = Z(p_0)$$

where the equality holds because Z is affine. For small $\epsilon > 0$, define a distribution F' as follows: with probability $1 - \epsilon$, F' chooses a posterior from F; with the remaining probability ϵ, F' picks a state $\omega \sim p_0$ and gives posterior δ_ω. Evidently, $E_{F'}[p] = p_0$, and

$$E_{F'}[W(p)] = (1 - \epsilon)E_F[W(p)] + \epsilon Z(p_0) > E_F[W(p)] \geq V_E[M, w|\mathcal{I}_0].$$

So if the IAT is $\mathcal{I} = \mathcal{I}_0 \cup \{(F', 0)\}$, the expert's unique optimal choice of experiment is $(F', 0)$. The principal's expected payoff $V_P(M, w|\mathcal{I})$ is then

$$E_{F'}[U(p) - W(p)] = (1 - \epsilon)E_F[U(p) - W(p)] + \epsilon E_{p_0}[U(\delta_\omega) - W(\delta_\omega)].$$

By taking $\epsilon \to 0$, we see that the principal cannot be guaranteed a payoff higher than $E_F[U(p) - W(p)]$, which is exactly the amount in (2.6).

If, on the other hand, $W(p) = Z(p)$ throughout the support of F, but $E_F[W(p)] > V_E[M, w|\mathcal{I}_0]$ strictly, then we can give a similar argument by taking $\mathcal{I} = \mathcal{I}_0 \cup \{(F, 0)\}$. This leaves us with the case where

$$V_E[M, w|\mathcal{I}_0] = E_F[W(p)] = E_F[Z(p)] = Z(p_0).$$

For this to happen, it must be that whatever experiment (F_0, c_0) is chosen under $\mathcal{I}_0$ satisfies $W(p) = Z(p)$ throughout the support of F_0, and $c_0 = 0$. However, in this case, we have

$$V_P(0) = E_{F_0}[U(p)] > E_{F_0}[U(p) - W(p)] = V_P(M, w|\mathcal{I}_0) \geq V_P(M, w).$$

(The strict inequality follows from the assumptions that W is nonzero, and F_0 has mean p_0 which has full support.) This contradicts our assumption that $V_P(M, w) \geq V_P(0)$, so this case cannot happen.

Finally, suppose $V_P(M, w) > U(p_0)$, and let F be a distribution attaining the minimum in (2.6). Then certainly

$$E_F[U(p) - W(p)] = V_P(M, w) > U(p_0) \geq U(p_0) - W(p_0).$$

If $E_F[W(p)] > V_E(M, w|\mathcal{I}_0)$ strictly, then define another distribution F' by drawing a posterior from F with probability $1 - \epsilon$, and placing the remaining probability mass ϵ on p_0. For small ϵ, F' still satisfies the constraints of (2.6), and

$$E_{F'}[U(p) - W(p)] = (1 - \epsilon)E_F[U(p) - W(p)] + \epsilon[U(p_0) - W(p_0)] < E_F[U(p) - W(p)],$$

contradicting minimality for F. □

We will also need the following result:

Lemma 2.5. *There exists an optimal contract.*

The proof is topological — we show that we can restrict attention to a compact set of contracts, and that V_P is upper semi-continuous on this set. (It is not continuous in general.) The details are in Appendix B.

Accepting that detail, we can commence the proof of Theorem 2.1. We first use a linear separation argument that shows that that for any given contract, there is some linear transform of the principal's reduced-form payoff — some function of the form $\alpha(U(p) +$

$E_p[\beta(\omega)]$ — that "looks like" the reduced form $W(p)$ from the point of view of the worst-case distribution F identified in Lemma 2.4. We can replace the given contract by the transform-bounded contract with parameters α, β, and check that the new contract is better for the expert. Because of the linear relationship between the transform-bounded contract and the reduced-form payoff U, this can be mapped back to show that the principal is better off as well.

Proof of Theorem 2.1: By Lemma 2.5, there exists an optimal contract, call it (M, w). In particular, we have $V_P(M, w) \geq V_P(0)$ and, by non-triviality, $V_P(M, w) > U(p_0)$.

Consider the following two subsets of $\Delta(\Omega) \times \mathbb{R} \times \mathbb{R}$: S is the convex hull of points $(p, W(p), U(p) - W(p))$; and T is the set of all points (p_0, y, z) such that $y \geq V_E[M, w|\mathcal{I}_0]$ and $z < V_P(M, w)$. These two sets are disjoint — otherwise there would be some distribution F such that $E_F[p] = p_0$, $E_F[W(p)] \geq V_E[M, w|\mathcal{I}_0]$, and $E_F[U(p) - W(p)] < V_P(M, w)$, contradicting Lemma 2.4. Applying a proper separating hyperplane theorem (e.g. [9, Theorem 11.3]) to these sets gives us $\lambda \in \mathbb{R}^\Omega$, $\mu, \nu, \xi \in \mathbb{R}$ such that

$$\sum_\omega \lambda_\omega p(\omega) + \mu W(p) + \nu(U(p) - W(p)) \leq \xi \qquad \text{for all } p \in \Delta(\Omega), \tag{2.7}$$

$$\sum_\omega \lambda_\omega p_0(\omega) + \mu V_E[M, w|\mathcal{I}_0] + \nu V_P(M, w) \geq \xi; \tag{2.8}$$

$\mu \geq 0$, $\nu \leq 0$; and we do not simultaneously have $\mu = \nu = 0$ and all the λ_ω equal to each other.

Moreover, we can let F^* be the distribution attaining the minimum in (2.6), and take the expectation over $p \sim F^*$ in (2.7). Since $E_{F^*}[W(p)] = V_E[M, w|\mathcal{I}_0]$ and $E_{F^*}[U(p) - W(p)] = V_P(M, w)$, we must have equality in (2.7) for all p in the the support of F^*, and also have equality in (2.8).

At least one of the inequalities $\mu \geq 0$, $\nu \leq 0$ must hold strictly: otherwise $\sum \lambda_\omega p(\omega) \leq \xi \leq \sum \lambda_\omega p_0(\omega)$ for all $p \in \Delta(\Omega)$, which would only be possible if all λ_ω were equal (because p_0 has full support), but proper separation means this cannot happen. Let's show that in fact both inequalities are strict: $\mu > 0$ and $\nu < 0$.

First, suppose $\nu = 0$. Then $\mu > 0$, and (2.7) implies $W(p)$ is bounded above by the affine function $Z(p) = (\xi - \sum_\omega \lambda_\omega p(\omega))/\mu$, with equality throughout the support of F^*. The equality in (2.8) then means that $V_E[M, w|\mathcal{I}_0] = E_{F^*}[W(p)] = E_{F^*}[Z(p)] = Z(p_0)$.

On the other hand, if (F, c) is the experiment the expert chooses under IAT $\mathcal{I}_0$, then

$$V_E[M, w|\mathcal{I}_0] = E_F[W(p)] - c \leq E_F[Z(p)] - c = Z(p_0) - c,$$

so it must be that $c = 0$. Thus, there is an experiment available for free in $\mathcal{I}_0$ that gives the principal at least $V_P(M, w)$. But then the principal could have gotten a strictly higher payoff by using a zero contract, since she would get at least $E_F[U(p)] > E_F[U(p) - W(p)] \geq V_P(M, w)$. This contradicts the assumption that $V_P(M, w) \geq V_P(0)$. This shows that $\nu < 0$ after all. Next, suppose $\mu = 0$. Then (2.7) for p_0 and the inequality in (2.8) imply

$$U(p_0) - W(p_0) \geq \frac{\sum_\omega \lambda_\omega p(\omega) - \xi}{-\nu} = V_P(M, w),$$

contradicting the assumption $V_P(M, w) > U(p_0)$. Thus $\mu > 0$ strictly.

Now we can return to the main argument. We can rearrange (2.7) to obtain

$$W(p) \leq \frac{-\sum_\omega \lambda_\omega p(\omega) - \nu U(p) + \xi}{\mu - \nu} \tag{2.9}$$

for all p. Put

$$\alpha = \frac{-\nu}{\mu - \nu}, \qquad \beta(\omega) = \frac{-\lambda_\omega + \xi}{-\nu} \text{ for each } \omega.$$

Let (M', w') be the transform-bounded contract with these parameters (note indeed $\alpha \in [0, 1]$), and let W' be its reduced form.

Let m be any message in the original contract. Then for all p, we have

$$E_p[w(m, \omega)] \leq W(p) \leq \alpha(U(p) + E_p[\beta(\omega)])$$

where the first inequality is by definition of $W(p)$ and the second is from (2.9). Thus, the map $\omega \mapsto w(m, \omega)$ is in M'. In particular, for all p, $W'(p) \geq E_p[w(m, \omega)]$. Taking the maximum over m gives $W'(p) \geq W(p)$: the reduced form of the new contract dominates the reduced form of the original contract (for the expert). On the other hand, we still have

$$W'(p) \leq \alpha(U(p) + E_p[\beta(\omega)]) = \frac{-\sum_\omega \lambda_\omega p(\omega) - \nu U(p) + \xi}{\mu - \nu} \tag{2.10}$$

by definition of the transform-bounded contract. This can be rearranged to

$$-\nu U(p) \geq \sum_\omega \lambda_\omega p(\omega) + (\mu - \nu) W'(p) - \xi,$$

or

$$-\nu(U(p) - W'(p)) \geq \sum_{\omega} \lambda_\omega p(\omega) + \mu W'(p) - \xi. \tag{2.11}$$

The relation $W'(p) \geq W(p)$ for all p implies immediately that $V_E(M', w'|\mathcal{I}_0) \geq V_E(M, w|\mathcal{I}_0)$. Under the new contract, for any IAT $\mathcal{I}$, the expert will choose an experiment (F, c) such that $E_F[W'(p)] \geq E_F[W'(p)] - c = V_E(M', w'|\mathcal{I}_0)$ and so

$$\begin{aligned}
E_F[U(p) - W'(p)] &\geq E_F\left[\frac{\sum_\omega \lambda_\omega p(\omega) + \mu W'(p) - \xi}{-\nu}\right] \qquad \text{(by (2.10))} \\
&= \frac{\sum_\omega \lambda_\omega p_0(\omega) + \mu V_E(M', w'|\mathcal{I}_0) - \xi}{-\nu} \\
&\geq \frac{\sum_\omega \lambda_\omega p_0(\omega) + \mu V_E(M, w|\mathcal{I}_0) - \xi}{-\nu} \\
&= V_P(M, w). \qquad \text{(from equality in (2.8))}
\end{aligned}$$

This shows that $V_P(M', w'|\mathcal{I}) \geq V_P(M, w)$ for all $\mathcal{I}$, and so $V_P(M', w') \geq V_P(M, w)$. Since (M, w) was assumed to be an optimal contract, this must be an equality, and the transform-bounded contract (M', w') is again optimal.

Moreover, suppose that the full-support assumption holds, and suppose that W' is not identically equal to W. Then, we have $V_E(M', w'|\mathcal{I}_0) > V_E(M, w|\mathcal{I}_0)$, since the experiment chosen by the expert under (M, w) and $\mathcal{I}_0$ has full support (it cannot be distribution δ_{p_0} because of our assumption $V_P(M, w) > U(p_0)$) and so gives the expert strictly higher expected payoff under (M', w'). Together with $\mu > 0$, this implies that the third line of the chain of inequalities above is a strict inequality; hence, $V_P(M', w') > V_P(M, w)$. This contradicts the optimality of $V_P(M, w)$. Thus, W' must be equal to W after all — that is, (M, w) is equivalent to the transform-bounded contract (M', w'). □

It remains only to make the leap from transform-bounded contracts to restricted investment contracts. In effect, we need to show that, for any given transform-bounded contract, the payoff represented by its reduced form at any given posterior p is actually attained by some decision available in the corresponding restricted decision space.

Proof of Proposition 2.2: Suppose that the decision problem is regular. Let (M, w) be the transform-bounded contract with parameters α, β, and (M', w') the corresponding restricted investment contract identified in the proposition statement. Write W, W' for the corresponding reduced forms. We wish to show that W and W' are identical. The $\alpha = 0$ case is trivial (both W and W' are zero), so assume $\alpha > 0$.

First the easy direction: For any decision $d \in M'$, the message taking each state ω to

$w'(d,\omega) = \alpha(u(d,\omega) + \beta(\omega))$ is in M. Thus for every posterior p, $E_p[w'(d,\omega)] \leq W(p)$. Therefore $W'(p) = \max_d E_p[w'(d,\omega)] \leq W(p)$.

Now for the reverse inequality. Fix a posterior p. Consider the following subset S of $\mathbb{R}^\Omega$: S is the set of all functions $m : \Omega \to \mathbb{R}$ such that there exists some $d \in D$ with $\alpha(u(d,\omega) + \beta(\omega)) \geq m(\omega)$ for all ω. S is closed, and the regularity assumption implies it is also convex. Now, by definition of the reduced form W, there is some $m^* \in M$ such that $E_p[m^*(\omega)] = W(p)$.

We will show that $m^* \in S$. Suppose not. Then we can apply a strict separating hyperplane theorem to conclude the existence of $\lambda \in \mathbb{R}^\Omega$ and ξ such that

$$\sum_\omega \lambda_\omega m(\omega) < \xi \qquad \text{for all } m \in S, \tag{2.12}$$

$$\sum_\omega \lambda_\omega m^*(\omega) > \xi. \tag{2.13}$$

(2.12) implies that $\lambda_\omega \geq 0$ for all ω. Then, we can normalize to assume that $\sum_\omega \lambda_\omega = 1$, so that λ equals some probability distribution $q \in \Delta(\Omega)$. Consider the decision $d = d(q)$, and $m(\omega) = \alpha(u(d,\omega) + \beta(\omega))$; then (2.12) and (2.13) give us

$$E_q[\alpha(u(d,\omega) + \beta(\omega))] < \xi < E_q[m^*(\omega)].$$

But the definition of the transform-bounded contract requires that

$$E_q[m^*(\omega)] \leq \alpha(U(q) + E_q[\beta(\omega)]) = E_q[\alpha(u(d,\omega) + \beta(\omega))],$$

and so we have a contradiction.

Thus, $m^* \in S$, which means that there is some $d \in D$ satisfying

$$\alpha(u(d,\omega) + \beta(\omega)) \geq m^*(\omega) \geq 0$$

for all ω. Dividing through by $\alpha > 0$, we see that this decision d lies in the restricted decision space M' as well. Then the value of the restricted investment contract at posterior p satisfies

$$W'(p) \geq E_p[\alpha(u(d,\omega) + \beta(\omega))] \geq E_p[m^*(\omega)] = W(p).$$

Now we have $W'(p) \leq W(p)$ and $W'(p) \geq W(p)$ for every p, so we are finished. $\square$

Before wrapping up this section, we should comment a bit further on how to identify the parameters of the optimal (transform-bounded) contract. Although it is not possible to express all the parameters in closed form, we can give an implicit characterization that will be useful for further investigation in Subsection 3.6.

Given $\beta : \omega \to \mathbb{R}$, put

$$M(\beta) = \{m : \Omega \to \mathbb{R}^+ \mid E_p[m(\omega)] \leq U(p) + E_p[\beta(\omega)] \text{ for all } p \in \Delta(\Omega)\}$$

and then put

$$U_R(p; \beta) = \max_{m \in M(\beta)} \left(E_p[m(\omega)] - E_p[\beta(\omega)] \right).$$

Clearly $U_R(p; \beta) \leq U(p)$. In the regular case, we have an especially straightforward interpretation for U_R: Proposition 2.2 implies that $U_R(p; \beta) = \max_{d \in M} E_p[u(d, \omega)]$ where M is just the restricted decision space coming from a restricted investment contract with parameter β.

We now have the following:

Proposition 2.6. *Let (M, w) be the transform-bounded contract with parameters α, β. Then, its payoff guarantee satisfies*

$$V_P(M, w) \geq \max_{(F,c) \in \mathcal{I}_0} \left((1 - \alpha) E_F[U_R(p; \beta)] - \frac{1 - \alpha}{\alpha} c \right) - \alpha \cdot E_{p_0}[\beta(\omega)] \tag{2.14}$$

where, if $\alpha = 0$, we interpret the $-((1 - \alpha)/\alpha)c$ term as 0 for $c = 0$ and $-\infty$ otherwise.

Moreover, there are some α, β such that the corresponding transform-bounded contract is optimal and (2.14) is an equality.

The seemingly odd wording of the last sentence appears because we do not rule out the possibility that different choices of α, β might give rise either to equivalent contracts, or to distinct optimal contracts.

Proof: If $\alpha = 0$, then if there does not any $(F, 0) \in \mathcal{I}_0$, the right side of (2.14) is $-\infty$ and the contract cannot be optimal; and if there is such an experiment, the right side is $\max_{(F,0) \in \mathcal{I}_0} E_F[U_R(p; \beta)] \leq \max_{(F,0) \in \mathcal{I}_0} E_F[U(p)]$ which is exactly the payoff guarantee from that contract. Equality holds if β is chosen large enough so that $U_R(p; \beta) = U(p)$.

Now suppose $\alpha > 0$. The inequality is proven in the same way as the guarantee from a linear contract (2.2). The reduced form of the transform-bounded contract is precisely

$$W(p) = \alpha(U_R(p; \beta) + E_p[\beta(\omega)]). \tag{2.15}$$

Once the expert attains posterior p, his expected payoff is $W(p) \leq \alpha(U(p) + E_p[\beta(\omega)])$, while the principal's is

$$U(p) - W(p) \geq (1-\alpha)U(p;\beta) - \alpha E_p[\beta(\omega)] \\ \geq \frac{1-\alpha}{\alpha}W(p) - E_p[\beta(\omega)].$$

Therefore, whatever experiment (F, c) the expert performs, the principal's expected payoff $V_P(M, w|\mathcal{I})$ is

$$\begin{aligned} E_F[U(p) - W(p)] &\geq E_F\left[\frac{1-\alpha}{\alpha}W(p) - E_p[\beta(\omega)]\right] \\ &= \frac{1-\alpha}{\alpha}E_F[W(p)] - E_{p_0}[\beta(\omega)]. \end{aligned}$$

Since

$$E_F[W(p)] \geq E_F[W(p)] - c \geq V_E(M, w|\mathcal{I}_0) = \max_{(F,c)\in\mathcal{I}_0}(E_F[W(p)] - c),$$

we can plug in to obtain

$$V_P(M, w|\mathcal{I}) \geq \frac{1-\alpha}{\alpha} \cdot \max_{(F,c)\in\mathcal{I}_0}(E_F[W(p)] - c) - E_{p_0}[\beta(\omega)].$$

Now plugging in from (2.15) gives exactly the right side of (2.14). Since this applies for all $\mathcal{I}$, (2.14) follows.

Now consider an optimal contract (M, w). Lemma 2.4 identifies its payoff guarantee to the principal; let F^* be the worst-case distribution given by that lemma. In the proof of Theorem 2.1, we obtained parameters (which were called α, β in that proof, but which we here refer to as α', β' for distinctness) such that the corresponding transform-bounded contract (M', w'), with reduced form W', satisfies:

- $W'(p) \geq W(p)$ for all p, hence $V_E(M', w'|\mathcal{I}_0) \geq V_E(M, w|\mathcal{I}_0)$;
- $V_P(M', w') \geq V_P(M, w)$ (from which (M', w') is again an optimal contract); and
- $W'(p) = W(p) = \alpha'(U(p) + E_p[\beta'(\omega)])$ for all p in the support of F^*.

The proof of Theorem 2.1 also showed that we cannot have $V_E(M', w'|\mathcal{I}_0) > V_E(M, w|\mathcal{I}_0)$ strictly, because that would imply $V_P(M', w') > V_P(M, w)$, contradicting optimality of the original contract. Hence $V_E(M', w'|\mathcal{I}_0) = V_E(M, w|\mathcal{I}_0)$, and so F^* satisfies the constraints of (2.6). Since $E_{F^*}[U(p) - W'(p)] = E_{F^*}[U(p) - W(p)]$, F^* must actually attain the minimum in (2.6) (for the new contract) — otherwise the minimum would be strictly lower, and we would have $V_P(M', w') < V_P(M, w)$, a contradiction.

Now we show that (2.14) is satisfied for the new contract (M', w'). We may suppose that $\alpha' > 0$ (otherwise we have the zero contract, and we have already covered this case). From (2.6) we have

$$\begin{aligned} V_P(M', w') &= E_{F^*}[U(p) - W'(p)] \\ &= E_{F^*}[(1-\alpha')U(p) - \alpha' E_p[\beta'(\omega)]] \\ &= \frac{1-\alpha'}{\alpha'} E_{F^*}[W'(p)] - E_{p_0}[\beta'(\omega)] \\ &= \frac{1-\alpha'}{\alpha'} V_E[M', w' | \mathcal{I}_0] - E_{p_0}[\beta'(\omega)]. \end{aligned}$$

And finally plugging in again $W'(p) = \alpha'(U_R(p; \beta') + E_p[\beta'(\omega)])$ gives us exactly (2.14) — with equality — for the new contract with its parameters (α', β'). ☐

Proposition 2.6 immediately implies:

Corollary 2.7. *The payoff guarantee from the optimal contract is equal to*

$$\max_{\substack{\alpha,\beta \\ (F,c)\in\mathcal{I}_0}} \left((1-\alpha) E_F[U_R(p;\beta)] - \frac{1-\alpha}{\alpha} c \right) - \alpha \cdot E_{p_0}[\beta(\omega)]. \tag{2.16}$$

Moreover, this guarantee is attained by a transform-bounded contract with the corresponding values of α, β.

This is a natural time to emphasize that transform-bounded contracts are not only a technical stepping stone, even though restricted investment contracts may be more immediately interpretable. Transform-bounded contracts are also computationally convenient. Indeed, while arbitrary contracts are very high-dimensional objects, and restricted investment contracts as we have defined them are still quite complicated (because the restricted decision space $M \subseteq D$ can be anything), transform-bounded contracts are parameterized by just $|\Omega| + 1$ numbers. Hence, Theorem 2.1 makes the search for the optimal contract potentially computationally tractable. Moreover, Corollary 2.7 makes a further step in this direction. In particular, if $\mathcal{I}_0$ consists of only a small number of experiments, one can try out a grid of values of α and β, and calculate the value of (2.16) in each case by considering each $(F, c) \in \mathcal{I}_0$ and computing $E_F[U_R(p; \beta)]$ by simulation. And we can further shave down one parameter by easily solving for the optimal α, given β. But at this point, we defer further discussion to Subsection 3.6.

3 Extensions and variations

We now consider a number of possible extensions to the model (independently of each other). Several of these extensions are analogous to extensions of the basic robust moral hazard model from [2], and the arguments are identical to arguments in that paper. In these cases, we only describe the ideas briefly.

3.1 Participation constraint

Suppose that the principal needs to guarantee the expert some expected payoff $\overline{U}_E > 0$ in order to hire him. Thus, the principal is restricted to contracts that satisfy $E_F[W(p)] - c \geq \overline{U}_E$ for some $(F, c) \in \mathcal{I}_0$. We maintain non-triviality: assume there exists some such contract (M, w) with $V_P(M, w) > U(p_0)$.

As in [2], adding a participation constraint changes nothing. The compactness argument that ensures existence of an optimal contract still holds when we add the participation constraint; and since the main step of the proof of Theorem 2.1 replaced the given contract (M, w) by a transform-bounded contract that is weakly better for both the expert and the principal, we see that if the original contract satisfies the participation constraint, so does the new one. Then the results of Subsection 2.3 still hold.

3.2 Smaller sets of experiments

We have written the model so that the principal evaluates contracts by their worst case over all possible IAT's $\mathcal{I} \supseteq \mathcal{I}_0$. This is perhaps an unrealistically large class of IAT's. In fact, we do not need such drastic uncertainty: As in [2], the results hold as long as every IAT of the form $\mathcal{I}_0 \cup \{(F, c)\}$, for some (F, c), is considered possible. Moreover, we can restrict to distributions F whose support consists of at most $2|\Omega|+2$ posteriors. This holds because the minimum in (2.6) is attained by some distribution whose support has size at most $|\Omega| + 1$. Indeed, if F^* is any distribution attaining the minimum, Carathéodory's Theorem (e.g. [9, Theorem 17.1]) implies that there exists F' with support contained inside that of F^*, such that $E_{F'}[p] = E_{F^*}[p] = p_0$, $E_{F'}[W(p)] = E_{F^*}[W(p)] = V_E(M, w|\mathcal{I}_0)$, and $E_{F'}[U(p) - W(p)] = E_{F^*}[U(p) - W(p)] = V_P(M, w)$; and the support of F' has size at most $|\Omega| + 2$. (In fact, with a little more work we can trim this down to $|\Omega| + 1$.) The extra $|\Omega|$ support points are needed to construct the auxiliary distributions used in the proof of Lemma 2.4.

3.3 Screening on technology

In general, the principal's minmax payoff typically is strictly greater than her maxmin payoff: that is, there is some payoff $\overline{V}_P$ strictly greater than the guarantee of the maxmin-optimal contract such that, if she knew the IAT $\mathcal{I}$ with certainty when choosing the contract, she could achieve an expected payoff at least $\overline{V}_P$, no matter what $\mathcal{I} \supseteq \mathcal{I}_0$ she faced. This can be shown by the same arguments as in [2]. This suggests the possibility of screening experts according to their IATs, by offering a *menu* of contracts (M, w), which the expert chooses from before performing any experiment, so that experts with different IATs may choose different contracts.

In fact, the principal cannot guarantee herself a better payoff with screening than she can without screening. The argument is exactly the same as that given in [2]. If the principal can guarantee herself some payoff V_P^* using a menu of contracts, we show she can also do so using just the contract (M_0, w_0) that the expert with IAT $\mathcal{I}_0$ would choose from this menu. If this were not the case, then there would be some experiment (F', c') such that, under $\mathcal{I}_1 = \mathcal{I}_0 \cup \{(F', c')\}$, the expert strictly prefers to perform experiment (F', c'), and the principal's resulting expected payoff is less than V_P^*. Then, under the IAT $\mathcal{I}_1$, the expert might choose a different contract from the menu, but she would again perform experiment (F', c'), which means the principal's payoff can only be even worse than under (F', c') and (M_0, w_0) (since the principal gets the same distribution over posteriors but now has to pay the expert more). This means that the menu of contracts does not guarantee V_P^*, a contradiction.

3.4 Observable posteriors

Our model has assumed that information acquisition is private: when the expert performs an experiment, only he observes the outcome. What if instead we assume that the principal also observes the outcome, and so updates to the same posterior p as the expert? What contract provides the optimal worst-case guarantee in this setting?

In this case, there is no need to provide incentives for truthful reporting, only for exertion of effort in acquiring information. So instead of having payments depend on the realized state, we can have them depend solely on the posterior: a contract is now any continuous function $W : \Delta(\Omega) \to \mathbb{R}^+$. Since the principal's gross payoff from posterior p is $U(p)$, this looks like the simple robust moral hazard model of [2]. This suggests that the optimal contract should simply be affine in $U(p)$ — that is, of the form $W(p) = \alpha(U(p) + \beta)$.

This is almost correct, except that the actions potentially available to the expert do not produce all possible distributions over posteriors, only those distributions with mean p_0. Hence, the model actually fits into the extension with a lower bound on costs in [2]. Here all of p is observable, and the lower bound on the cost of an experiment is 0 if it has mean p_0 and ∞ otherwise. The results from that model then show that the optimal contract is of the form $W(p) = \alpha(U(p) + E_p[\beta(\omega)])$ for some function $\beta : \Omega \to \mathbb{R}$. That is, it is simply an *un*restricted investment contract, with state-by-state adjustments just as for our restricted investment contracts.

Since the choice of β can have no effect on incentives (this term contributes $\alpha \cdot E_{p_0}[\beta(\omega)]$ to expected payoff regardless of the experiment performed), its only role can be to relax the limited liability constraint. Therefore the optimal choice of β is such that

$$\min_p (U(p) + E_p[\beta(\omega)]) = U(p_0) + E_{p_0}[\beta(\omega)] = 0$$

— that is, $\beta(\omega)$ follows a subdifferential of U at the point p_0.

3.5 Influencing states

What if we adopt the main model, except that we allow the expert's actions to influence the distribution of states, in addition to providing information? That is, we drop the requirement $E_F[p] = p_0$ from the definition of an experiment, so an experiment is now any element of $\Delta(\Delta(\Omega)) \times \mathbb{R}^+$, and the rest of the model is left unchanged. (Zermeño [11] allows for this as well.)

We can repeat the analysis, and find that Lemma 2.4 is the same except that the constraint $E_F[p] = p_0$ is dropped from (2.6). Then, in the separation argument used to prove Theorem 2.1, inequalities (2.7) and (2.8) hold with all coefficients λ_ω equal to zero. Hence, the same proof now shows that there is an optimal contract that is transform-bounded with the adjustment term $\beta(\omega)$ constant across all states ω. In the regular case, this means likewise that a restricted investment contract with β constant across all states is optimal.

3.6 Restricted versus unrestricted investment

In this subsection we briefly consider the question of when the optimal contract is an *un*restricted investment contract, with $M = D$. Such contracts have a natural interpretation, as delegating the decision directly to the expert, who is then paid a fraction α of the payoff plus the state-by-state adjustment β. If it also happens that the worst possible payoff in each state is zero ($\min_d u(d,\omega) = 0$ for each ω), then these contracts can be interpreted even more simply: the expert chooses the decision and is paid a fixed fraction of the realized payoff. This would bridge the gap between our model and some of the prior principal-expert literature [5, 7] which assumed that the decision must be delegated to the expert and compensation could depend only on the realized payoff.

We will maintain several assumptions throughout this subsection. To avoid uninteresting cases, we will assume for this subsection that the zero contract is not optimal. (For example, this is the case if the known IAT $\mathcal{I}_0$ does not contain any experiment $(F, 0)$, except possibly $F = \delta_{p_0}$.) We will also assume that in every state, not all decisions give the same payoff. Moreover, we will assume that the decision problem is regular. Thus Proposition 2.2 applies, and we can describe the optimal contract either as a transform-bounded contract or as a restricted investment contract with the same parameters α, β. We immediately see that $\beta(\omega) \leq -\min_{d\in D} u(d,\omega)$ for all ω (otherwise $\beta(\omega)$ can be decreased without changing the restricted decision space M, thereby reducing the payments to the expert without changing incentives); and we have an unrestricted investment contract if and only if equality holds:

$$\beta(\omega) = -\min_{d\in D} u(d,\omega) \qquad \text{for each } \omega.$$

Denote this β by β_0. Finally, we assume that dominated decisions have been eliminated a priori: for any distinct $d, d' \in D$, there exists ω such that $u(d,\omega) > u(d',\omega)$. This eliminates uninteresting cases where there is some decision giving extremely low payoffs in every state, making an unrestricted investment contract extremely costly (via limited liability forcing $\beta(\omega)$ to be high) even though the expert would never invest in that decision.

We cannot give a complete characterization of when restriction is or is not optimal, but we can give partial results that express some relevant intuitions. Recall the function $U_R(p;\beta)$ defined toward the end of Section 2.3. Define, for each posterior p and each state ω, the one-sided partial derivative

$$\psi(p,\omega) = \left.\frac{\partial U_R^-}{\partial \beta(\omega)}\right|_{(p,\beta_0)}.$$

The main result in this subsection is the following:

Proposition 3.1. *Fix the decision problem (D,u), which is enough to define β_0 and ψ.*

The value $\psi(p,\omega)$ is well-defined and nonnegative.

If there is some state ω such that $\psi(p,\omega) = 0$ for all p, then unrestricted investment contracts cannot be optimal.

Conversely, if there is no such ω exists, then it is possible to choose $\mathcal{I}_0$ so that an unrestricted investment contract is optimal. We can do this while satisfying the non-triviality and full-support assumptions.

To help in understanding this proposition, consider Figure 2, which plots possible decision problems. Each panel shows a decision problem with state space $\Omega = \{\omega_1, \omega_2\}$; each decision is represented by the pair of payoffs $(u(d,\omega_1), u(d,\omega_2))$. The thick line depicts the set of available decisions D. Within each state, payoffs have been normalized so that the payoff of the worst decision is 0. For any posterior p, $U(p)$ corresponds to the maximal distance in direction $(p(\omega_1), p(\omega_2))$ represented by any available decision. Then, $\psi(p,\omega_i)$ represents how much this distance decreases when decisions are restricted by requiring the state-ω_i payoff to be incrementally larger than zero. Panel (a) shows a decision problem for which $\psi(p,\omega_i) = 0$ for each p and each ω_i: slightly trimming away the endpoints of the decision space has only a second-order effect on the expected payoff for any possible posterior. In panel (b), we have $\psi(p,\omega_1) = 0$ for all p, but $\psi(p,\omega_2) > 0$ if p puts sufficient weight on ω_2.

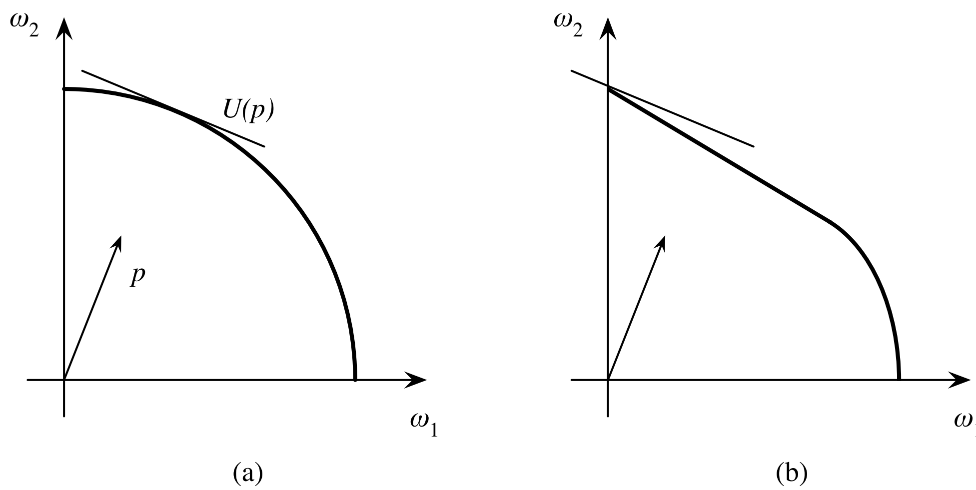

Figure 2: Example decision problems

We give a more general example in keeping with (b). Suppose that the decision space is the efficient frontier of the convex hull of a finite set of decisions. That is, suppose we begin with finite set of decisions D_0; we identify each decision d with the corresponding vector of payoffs in $\mathbb{R}^\Omega$, and form the convex hull D_1; and then D consists of all $d \in D_1$ such that there does not exist $d' \in D_1$ with $d'_\omega \geq d_\omega$ for all ω, strictly for some ω. By elimination of dominated decisions, we may as well assume $D_0 \subseteq D$. In this situation, we claim that the last condition in Proposition 3.1 is necessarily satisfied. We outline the argument here: Choose any state ω_0. We can pick some decision d for which $u(d, \omega_0)$ is minimal, and can take d to be a pure decision ($d \in D_0$). If there are several such decisions, choose an extreme one (not contained in the convex hull of the others). Then, using the assumption that d is not dominated, an appropriate application of the separating hyperplane theorem gives a posterior p under which d gives strictly higher expected payoff than any other decision. Consequently, if the decision space is restricted to (mixed) decisions with payoff at least $u(d, \omega_0) + \epsilon$ in state ω_0, any such mixture must place weight proportional to ϵ on decisions other than d, and so the maximum expected payoff under p is reduced by an amount proportional to ϵ. More concretely, we get

$$\psi(p, \omega_0) = \frac{E_p[u(d, \omega)] - \max_{d' \in D_0 \setminus \{d\}} E_p[u(d, \omega)]}{\max_{d' \in D_0} u(d', \omega_0) - u(d, \omega_0)} > 0.$$

The proof of Proposition 3.1, which is in Appendix C, is straightforward: We consider the payoff guarantee in (2.16), and solve for the optimal value of α given β; then we study conditions under which $\beta = \beta_0$ is or is not optimal.

Proposition 3.1 shows that for some decision problems an unrestricted investment contract cannot possibly be optimal. Can the reverse situation hold — can there be decision problems where restricting investment cannot be optimal, regardless of $\mathcal{I}_0$? The answer is essentially no, by the logic of Figure 1, which is quite general. If the known IAT only allows for posteriors in the neighborhood of the prior, or posteriors far away are possible but very unlikely, then the principal can relax the incentive constraints by trimming away extreme decisions at a low cost in terms of incentives.

We only need to rule out the knife-edge case in which extreme decisions are always optimal. (This case can arise, for example, if all decisions are coplanar — that is, there are positive weights λ_ω such that $\sum_\omega \lambda_\omega u(d, \omega)$ is constant over all $d \in D$.) Call a decision d *extreme* if $u(d, \omega) = \min_{d' \in D} u(d', \omega)$ for some ω.

Proposition 3.2. *Take as given any decision problem satisfying the assumptions of this subsection. Assume that there exists an open ball of beliefs $P \subseteq \Delta(\Omega)$ such that there is no one decision that is optimal for all $p \in P$, and such that extreme decisions cannot be* *optimal for any $p \in P$.*

Then there exists a prior $p_0 \in P$ and a known IAT $\mathcal{I}_0$ satisfying non-triviality and full support, such that a restricted investment contract, with $M \subset D$ strictly, is optimal.

Proof: Let F be any continuous distribution over posteriors with support P. The assumption that no decision is optimal throughout P ensures that $E_F[U(p)] > U(E_F[p])$. By continuity, this remains true when we replace P by a sufficiently large closed sub-ball $\overline{P}$ and F with the distribution conditioned on $\overline{P}$, call it $\overline{F}$. Let the prior be $p_0 = E_{\overline{F}}[p]$.

As noted at the beginning of Subsection 2.2, if $c > 0$ satisfies the bound (2.3), then $\mathcal{I}_0 = \{(\delta_{p_0}, 0), (\overline{F}, c)\}$ satisfies the non-triviality assumption. Let F' be any distribution with full support on $\Delta(\Omega)$ and with mean p_0. By continuity of the formula (2.16), we can take $F = (1-\epsilon)\overline{F} + \epsilon F'$ for sufficiently small $\epsilon > 0$, and then $\mathcal{I}_0 = \{(\delta_{p_0}, 0), (F, c)\}$ still satisfies non-triviality. It clearly satisfies full support as well.

We will show that for ϵ sufficiently small, under this $\mathcal{I}_0$, a restricted investment contract is optimal. Compactness implies that for all $p \in P$, optimal decisions are uniformly bounded away from extreme decisions; that is, one can choose $\beta_1 : \Omega \to \mathbb{R}$ with $\beta_1(\omega) < \beta_0(\omega)$, such that for all $p \in P$, the optimal decision $d(p)$ satisfies $u(d(p), \omega) + \beta_1(\omega) \geq 0$ for each ω.

In particular, $U_R(p; \beta_1) = U(p) = U_R(p; \beta_0)$ for each such p.

Now consider the formula (2.16), but with β fixed at β_0 but α allowed to assume the maximizing value. The discussion around (2.3) ensures that the optimum is attained with the experiment (F, c) (not $(\delta_0, 0)$), and α is bounded away from 0, say $\alpha \geq \underline{\alpha} > 0$, as ϵ varies near 0. We will show that if we now replace β_0 by β_1, the maximum in (2.16) strictly increases. Indeed, the $(1-\alpha)E_F[U_R(p, \beta)]$ term changes by at most $\epsilon \cdot (\max u(p, \omega) - \min u(p, \omega))$, since there is no change for any of the realizations $p \in P$ (which occur with probability at least $1 - \epsilon$). On the other hand, the final term changes by

$$-\alpha \cdot E_{p_0}[\beta_1(\omega)] + \alpha \cdot E_{p_0}[\beta_0(\omega)] \geq \underline{\alpha} \cdot E_{p_0}[\beta_1(\omega) - \beta_0(\omega)] > 0.$$

Thus, this term changes by at least a positive amount that is independent of ϵ. So for ϵ small enough, the change from β_0 to β_1 leads to a strict increase in the value of the maximand in (2.16).

Thus, in this case, the value of the maximum in (2.16) is attained for some $\beta \neq \beta_0$, which means that the corresponding optimal contract is a restricted investment contract, with $M \subset D$. $\square$

4 Conclusion

We wrap up by briefly recapitulating our results and putting them in context. We considered a principal-expert model, with risk-neutrality and limited liability, and ex-post revelation of the state of nature. We assumed unquantifiable uncertainty about the expert's information acquisition technology, represented by a maxmin objective. This led quite generally to a novel form of contracts being optimal: transform-bounded contracts, or equivalently (under a regularity assumption) restricted investment contracts, in which the expert chooses to invest in one of a restricted subset of less-risky decisions, and is paid proportionally to the payoff of the chosen decision in the realized state. The result reflects the usual intuition about the robustness of linear payment rules, with the added twist that by prohibiting investment in risky decisions, the principal can relax the limited liability constraints and so pay the expert less.

The direct interpretation of our results is that they show how one can go about optimally providing incentives to experts in uncertain environments. They offer both qualitative insights into the shape of an optimal contract and an approach to actually computing it.

More broadly, this work illustrates the value of the worst-case methodology. The standard Bayesian version of the principal-expert problem, even under risk-neutrality and limited liability as we have assumed here, seems to be intractable except in very special cases [10]. The corresponding maxmin version offers traction quite generally — without any functional-form assumptions on the known information acquisition technology — by extending the linear separation methods of [2]. As in the earlier paper, we also saw that the model can be extended in various directions without changing the underlying methodology. The worst-case environments are not extreme, and the qualitative intuitions that emerge seem to be valid even if one rejects the worst-case approach, or insists on a more nuanced version of it.

A natural task for future work is to find as large as possible a class of contracting problems (or other mechanism design problems) where a maxmin approach similarly offers new ways forward. At a glance, we can identify some of the features of the problem in this paper that make the tools successful. We identified the worst-case environment for a given contract as the solution to a linear program (Lemma 2.4), and then used the dual variables to construct from this a new contract in a small parameterized family that performs at least as well as the old one. Hence, some important features are that contracts are complex, high-dimensional objects (so that a result narrowing them down to a parameterized family is not trivial); and that the effective space of uncertainty (in this case, the space of possible experiments) is defined by a small number of linear constraints, so that the LP tools apply to identify the worst case.

A On direct reporting of posteriors

Here we elaborate formally the revelation-principle-style argument given intuitively in Subsection 2.1: when looking for optimal contracts, we may assume that the expert directly reports his posterior as one component of his message.

First we should clarify why this is not trivial. In order to define the principal's payoff guarantee form a contract, we need to describe behavior, and in particular we need to specify how the expert chooses when he is indifferent between several messages. Following the usual principal-agent literature, we would like to assume that the expert breaks ties in the way that is best for the principal, as indicated in Subsection 2.1. However, the principal's expected payoff from a given message depends on her endogenous interpretation of messages. That is, there is a signaling game in which the expert chooses which message to send, and the principal makes a decision based on the message; this game may have multiple equilibria.

One way to deal with this would be to require contracts to specify an equilibrium of the game. (This is the setup termed *principal-authority* in [11].) However, we then run into the thorny conceptual problem of how to define equilibrium in a Bayesian game when one of the players is not Bayesian. Instead of taking a stand on this, we will define contracts more broadly, and require them to specify strategies for both players, in which only the expert's strategy needs to be optimal (*full commitment* in [11]). We then show that any such contract is weakly outperformed by a contract in which the expert reports the posterior truthfully as part of his message, and the principal takes the corresponding optimal decision; and we argue that the latter contract would survive any reasonable equilibrium criterion.

Let $\mathcal{IAT}$ denote the set of all possible IAT's $\mathcal{I} \supseteq \mathcal{I}_0$. We properly define a *contract* to consist of five parts:

- a compact message space M;
- a continuous payment function $w : M \times \Omega \to \mathbb{R}^+$;
- an experiment strategy $\sigma : \mathcal{IAT} \to \Delta(\Delta(\Omega)) \times \mathbb{R}^+$, such that $\sigma(\mathcal{I}) \in \mathcal{I}$ for each $\mathcal{I}$;

- a reporting strategy $\rho : \mathcal{IAT} \times \Delta(\Omega) \to M$;
- a continuous decision rule $r : M \to D$.

Thus (σ, ρ) forms the expert's strategy, and r the principal's strategy. (We could allow for mixed strategies; this would change nothing.)

We want to restrict to contracts in which the expert's strategy is optimal; this requires defining payoffs. So for any posterior p, let

$$M^*(M, w, r|p) = \arg\max_{m \in M} E_p[w(m, \omega)],$$

the set of optimal messages, and define $W(p) = \max_{m \in M} E_p[w(m, \omega)]$ as before. Then, we define the optimal payoff for the principal, given that the expert chooses an optimal message:

$$\widehat{U}(M, w, r|p) = \max_{m \in M^*(M,w,r|p)} (E_p[u(r(m), \omega)] - W(p)).$$

We say that the reporting strategy ρ is *incentive-compatible* if

$$\rho(\mathcal{I}, p) \in \arg \max_{m \in M^*(M,w,r|p)} E_p[u(r(m), \omega)]$$

for all $\mathcal{I}$ and p. That is, the expert chooses a message to lexicographically maximize his own expected payoff and then the principal's. (The $W(p)$ can be omitted on the right side since it is independent of the choice of $m \in M^*(M, w, r|p)$.)

Now for the experimenting stage: For any IAT $\mathcal{I}$, we define

$$I^*(M, w, r|\mathcal{I}) = \arg \max_{(F,c) \in \mathcal{I}} (E_F[W(p)] - c), \qquad V_E(M, w, r|\mathcal{I}) = \max_{(F,c) \in \mathcal{I}} (E_F[W(p)] - c)$$

(note these are independent of ρ) and say that the experiment strategy σ is *incentive-compatible* if

$$\sigma(\mathcal{I}) \in \arg \max_{(F,c) \in I^*(M,w,r|\mathcal{I})} E_F[\widehat{U}(M, w, r|p)].$$

We now call the contract *acceptable* if both ρ and σ are incentive-compatible. As a side note, given any M, w, r, there exist incentive-compatible choices of ρ and σ, from the continuity and compactness assumptions.

Define the principal's payoff as before:

$$V_P(M, w, r|\mathcal{I}) = \max_{(F,c) \in I^*(M,w,r|\mathcal{I})} E_F[\widehat{U}(M, w, r|p)], \quad V_P(M, w, r) = \inf_{\mathcal{I} \supseteq \mathcal{I}_0} V_P(M, w, r|\mathcal{I}).$$

(These payoffs do not depend on the specification of σ, ρ.) At last, we can give the formal justification for our assumption that the expert directly reports his posterior.

Proposition A.1. *Let (M, w, σ, ρ, r) be any acceptable contract. Then define a new contract $(M', w', \sigma', \rho', r')$ by taking:*

- $M' = M \times \Delta(\omega)$*;*
- $w'(m, p, \omega) = w(m, \omega)$*;*
- $\rho'(\mathcal{I}, p) = (\rho(\mathcal{I}, p), p)$*;*
- $r'(p) = d(p)$ *(recall this was defined as* $\arg\max_{d \in D} E_p[u(d, \omega)]$*); and*
- $\sigma'(\mathcal{I})$ *is taken to be any* $(F, c) \in \mathcal{I}$ *that lexicographically maximizes* $E_F[W(p)] - c$ *and then* $E_F[U(p) - W(p)]$*, where* W *is the reduced form of the original contract.*

Then the new contract is also acceptable, and has at least as high a payoff guarantee V_P *as the old contract.*

Proof: At the reporting stage, the set of optimal messages is exactly those whose first component was optimal under the old contract: $M^*(M', w', r'|p) = M^*(M, w, r|p) \times \Delta(\Omega)$. And the reduced form of the new contract equals the old, $W'(p) = W(p)$. Given the product structure for $M^*(M', w', r'|p)$, the expert optimally chooses which posterior $\widehat{p}$ to report by maximizing $E_p[u(r'(\widehat{p}), \omega)]$. By construction, this is accomplished by reporting $\widehat{p} = p$. Hence, ρ' is incentive-compatible. Also, the fact that the principal now takes the optimal decision given the posterior implies that her decision-stage payoff under the new contract satisfies

$$\widehat{U}(M', w', r'|p) \geq \widehat{U}(M, w, r|p).$$

Incentive-compatibility of σ' is immediate from the definition, given that the new contract has the same reduced form as the original, and $\widehat{U}(M', w', r'|p) = U(p) - W(p)$. Notice also that the expert's choice set (before maximizing the principal's payoff) is the same under both contracts, $I^*(M', w', r'|\mathcal{I}) = I^*(M, w, r|\mathcal{I})$.

Now the new contract is acceptable, since we showed that both ρ' and σ' were incentive-compatible. And for any IAT $\mathcal{I}$,

$$\begin{aligned} V_P(M', w', r'|\mathcal{I}) &= \max_{(F,c) \in I^*(M', w', r'|\mathcal{I})} E_F[\widehat{U}(M', w', r'|p) \\ &\geq \max_{(F,c) \in I^*(M, w, r|\mathcal{I})} E_F[\widehat{U}(M, w, r|p)] \\ &= V_P(M, w, r|\mathcal{I}). \end{aligned}$$

Taking the inf over $\mathcal{I}$ gives $V_P(M', w', r) \geq V_P(M, w, r)$ as claimed. $\square$

As mentioned earlier, we would like to argue that the new contract given by Proposition A.1 is robust to any reasonable kind of Bayes-like refinement requiring the principal's decision rule r' to be a best reply to beliefs. Under the expert's strategy (σ', ρ'), any message (m, p) can only be sent if the true posterior really is p. Hence, for messages that can be sent along the equilibrium path, any reasonable specification of the principal's beliefs following (m, p) should put probability 1 on the expert's posterior really being p, in which case $r'(p) = d(p)$ is indeed the best decision for the principal to take. And for off-path messages (m, p), we can specify any beliefs, so in particular we can specify beliefs that still put probability 1 on the posterior being p.

One final comment on the more general class of indirect contracts we have defined in this appendix: Because, in Proposition A.1, the new contract is equivalent to the original contract (they have the same reduced form), we can see the equivalence results such as the last statement of Theorem 2.1 continue to hold even with the indirect contracts considered here.

B Existence of the optimum

Proof of Lemma 2.5:

Let $\overline{U} = \max_p U(p)$, and $\overline{W} = (\overline{U} - U(p_0))/(\min_\omega p_0(\omega))$. We may restrict attention to contracts whose reduced form satisfies $W(\delta_\omega) \leq \overline{W}$ for all ω. To see this, note that if there were some message guaranteeing the expert payoff higher than $(\overline{U} - U(p_0))/p_0(\omega)$ in some state ω, then whatever experiement the expert performs, he can always force the principal to pay more than $\overline{U} - U(p_0)$ in expectation (by just always sending this message), so the principal's expected payoff must be less than $U(p_0)$, and the principal is worse off than by not hiring the expert.

Since W is convex, this restriction implies $W(p) \leq \overline{W}$ for all p. Now say that a function $W : \Delta(\Omega) \to \mathbb{R}^+$ is *a reduced-form contract* if it is the reduced form of some contract. We make two claims:

- Claim 1: The set of reduced-form contracts $W : \Delta(\Omega) \to [0, \overline{W}]$ is compact in the sup-norm topology.

- Claim 2: V_P is upper semi-continuous on the set of reduced-form contracts (with respect to the sup-norm topology).

Together, these claims imply that V_P attains a maximum over the reduced-form contracts whose values never exceed $\overline{W}$, which is then a global maximum, as needed.

To prove Claim 1, let $W_1, W_2, \ldots$ be a sequence of reduced-form contracts taking values in $[0, \overline{W}]$. Note that each W_k must be a Lipschitz function with constant $\overline{W}$ (relative to the L_1 norm on $\Delta(\Omega)$). By passing to a subsequence, we may assume that W_k converges pointwise at each *rational* point $p \in \Delta(\Omega)$. Let $W_\infty(p) = \lim_k W_k(p)$ for each such p. Define M as the set of all affine functions $m : \Delta(\Omega) \to [0, \overline{W}]$ such that $m(p) \le W_\infty(p)$ for each rational p. Notice that M is a compact subset of $\mathbb{R}^\Omega$. Define $w : M \times \Omega \to [0, \overline{W}]$ by $w(m, \omega) = m(\omega)$. Then (M, w) is a contract, with reduced form $W(p) = \max_{m \in M} m(p)$.

For each rational p, we have $W(p) = W_\infty(p)$. The direction $W(p) \le W_\infty(p)$ is immediate from the definition of W. For the reverse inequality, fix p. For each k, since W_k is a reduced-form contract, there is some affine $m_k : \Delta(\Omega) \to [0, \overline{W}]$ such that $m_k(p) = W_k(p)$ and $m_k(p') \le W_k(p')$ for all other rational p'. There is some subsequence along which the m_k converge to some m_∞. Then, $m_\infty(p) = W_\infty(p)$, and $m_\infty(p') \le W_\infty(p')$ for each other rational p', so that $m_\infty \in M$. Therefore $W(p) \ge W_\infty(p)$, and equality follows.

Now we claim that $W_k \to W$ in sup norm. If not, there exists some $\epsilon > 0$ and a subsequence of k's and points p_k along which $|W_k(p_k) - W(p_k)| > \epsilon$. Again by taking a subsequence, we may assume the p_k converge to some point p. Now, we can find a rational point q such that the L_1 distance between p and q is at most $\epsilon/4\overline{W}$. Then, for k high enough, $|W_k(q) - W(q)| \le \epsilon/4$ (because $W(q) = W_\infty(q)$ by the previous paragraph) and therefore

$$|W_k(p_k) - W(p_k)| \le |W_k(p_k) - W_k(q)| + |W_k(q) - W(q)| + |W(q) - W(p_k)| \le \frac{3\epsilon}{4},$$

a contradiction.

This shows that the set of reduced-form contracts taking values in $[0, \overline{W}]$ is sequentially compact, proving Claim 1.

For Claim 2, suppose we have a convergent sequence of reduced-form contracts $W_k \to W$. Consider any IAT $\mathcal{I}$, and let (F_k, c_k) be the expert's chosen experiment under contract W_k. Then, again by passing to a subsequence, we may assume that (F_k, c_k) converges to some limit (F, c). It follows that $(F, c) \in \mathcal{I}$, and since

$$E_F[W(p)] - E_{F_k}[W_k(p)] = (E_F[W(p)] - E_{F_k}[W(p)]) + (E_{F_k}[W(p)] - E_{F_k}[W_k(p)]) \to 0$$

(the first parenthesized expression goes to 0 by weak convergence of F_k, and the second by sup-norm convergence of W_k), we can conclude that (F, c) is an optimal experiment for the expert under W_k and $\mathcal{I}$: If it were outperformed by some (F', c'), then this experiment would also outperform (F_k, c_k) under W_k for high enough k, a contradiction.

This same double-convergence argument also implies that

$$E_F[U(p) - W(p)] - E_{F_k}[U(p) - W_k(p)] \to 0$$

from which

$$V_P(W|\mathcal{I}) \geq E_F[U(p) - W(p)] = \lim_k E_{F_k}[U(p) - W_k(p)] = \lim_k V_P(W_k|\mathcal{I}) \geq \limsup_k V_P(W_k).$$

Since this holds for all $\mathcal{I}$, we have

$$V_P(W) \geq \limsup_k V_P(W_k)$$

which proves Claim 2. □

C On the optimality of restricting investment

Here we prove Proposition 3.1. We begin by considering Corollary 2.7, which gives the formula (2.16) for the parameters of the optimal contract, and solving out for α: given β and (F, c), the α that attains the maximum is $\alpha = \sqrt{c/(E_F[U_R(p; \beta)] + E_{p_0}[\beta(\omega)])}$. Plugging in this value of α, (2.16) turns into

$$\max_{\substack{\beta:\Omega\to\mathbb{R} \\ (F,c)\in\mathcal{I}_0}} \left(E_F[U_R(p; \beta)] + c - 2\sqrt{c\,(E_F[U_R(p; \beta)] + E_{p_0}[\beta(\omega)])}. \right) \tag{C.1}$$

Moreover, the optimal contract uses the corresponding value of β. Thus if the maximum in (C.1) is not attained at $\beta = \beta_0$, then an unrestricted investment contract cannot be optimal; if it is uniquely attained at $\beta = \beta_0$, then an unrestricted investment contract must be optimal.

Proof of Proposition 3.1: First we show that U_R is concave in β. For values β and β', let $m \in M(\beta)$ and $m' \in M(\beta')$ attain the respective maxima in the definitions of $U_R(p; \beta), U_R(p; \beta')$. Suppose $\beta'' = \lambda\beta + (1 - \lambda)\beta'$ with $\lambda \in [0, 1]$. It is immediate that

$m'' = \lambda m + (1-\lambda)M' \in U_R(p; \lambda\beta + (1-\lambda)\beta')$, from which

$$U_R(p;\beta'') \geq E_p[m''(\omega)] - E_p[\beta''(\omega)] = \lambda U_R(p;\beta) + (1-\lambda)U_R(p;\beta').$$

We also can see that U_R is increasing in β, since if $\beta \leq \beta'$ componentwise, and m attains the maximum for β, then $m'(\omega) = m(\omega) + \beta'(\omega) - \beta(\omega)$ is in $M(\beta')$ and gives the same value for the maximand in defining $U_R(p;\beta')$; thus $U_R(p;\beta') \geq U_R(p;\beta)$.

Since U_R is concave and increasing on its domain, we see that the one-sided derivative $\psi(p,\omega)$ is indeed well-defined and nonnegative.

Next, suppose there is some ω with $\psi(p,\omega) = 0$ for all p. For $\epsilon \geq 0$, define $\beta_\epsilon : \Omega \to \mathbb{R}$ by $\beta_\epsilon(\omega) = \beta(\omega) - \epsilon$, and $\beta_\epsilon(\omega') = \beta(\omega')$ for other ω'. The concavity of U_R implies that $(U_R(p;\beta) - U_R(p;\beta_\epsilon))/\epsilon$ decreases as $\epsilon \to 0$, so we can invoke the monotone convergence theorem to conclude that

$$\frac{d}{d\epsilon^+} E_F[U_R(p;\beta_\epsilon)] = E_F\left[\frac{d}{d\epsilon^+} U_R(p;\beta_\epsilon)\right] = 0.$$

Consequently, no matter what $\mathcal{I}_0$ is given, if we consider the maximand in (C.1), plugging in $\beta = \beta_\epsilon$ and choosing the (F,c) that gives the maximal value for $\beta = \beta_0$, then the derivative of the maximand with respect to ϵ is equal to

$$p_0(\omega)\sqrt{\frac{c}{E_F[U_R(p;\beta_0)] + E_{p_0}[\beta_0(\omega)]}} > 0.$$

Thus, the maximum in (C.1) is not attained at β_0, and an unrestricted investment contract cannot be optimal.

Conversely, suppose that for each ω there is some $p(\omega)$ such that $\psi(p(\omega),\omega) \geq \eta > 0$. Let F be any distribution with full support, and mass at least ϵ on each $p(\omega)$. We will show that an unrestricted investment contract is optimal under $\mathcal{I}_0 = \{(\delta_0, 0), (F,c)\}$, for sufficiently small $c > 0$. Note that full support and non-triviality are satisfied.

We claim that for any β that is $\leq \beta_0$ componentwise,

$$E_F[U_R(p,\beta)] \leq E_F[U_R(p,\beta_0)] - \frac{\epsilon\eta}{|\Omega|}\sum_\omega(\beta_0(\omega) - \beta(\omega)). \tag{C.2}$$

Indeed, choose the state ω^* for which $\beta_0(\omega) - \beta(\omega)$ is largest. Put $\beta_1(\omega^*) = \beta(\omega^*)$ and

$\beta_1(\omega) = \beta(\omega)$ for other states ω. Then increasingness and concavity imply

$$U_R(p;\beta) \le U_R(p;\beta_1) \le U_R(p;\beta_0) - (\beta(\omega^*) - \beta_0(\omega^*))\psi(p,\omega^*).$$

Applying expectations under F, noting that $p = p(\omega^*)$ arises with probability at least ϵ, and $\beta(\omega^*) - \beta_0(\omega^*) \ge \sum_\omega (\beta(\omega) - \beta_0(\omega))/|\Omega|$ by choice of ω^*, leads to (C.2).

Thus, $E_F[U_R(p,\beta)]$ is bounded above by an affine function of β that is uniquely maximized at β_0, and this upper bound (C.2) holds with equality at $\beta = \beta_0$. Since the quantity $\sqrt{E_F[U_R(p;\beta)] + E_{p_0}[\beta(\omega)]}$ is locally Lipschitz in β near β_0, adding a small multiple of it will not change this fact. Hence, for c sufficiently small, the maximand in (C.1) is still uniquely maximized over β by taking $\beta = \beta_0$. (The optimum with respect to choice of experiment must be given by (F, c), not $(\delta_{p_0}, 0)$, by non-triviality.) It follows that only an unrestricted investment contract can be optimal. □

References

[1] Dirk Bergemann and Stephen Morris (2005), "Robust Mechanism Design," *Econometrica* 73 (6), 1771-1813.

[2] Gabriel Carroll (2012), "Robustness and Linear Contracts," working paper, Microsoft Research.

[3] Sylvain Chassang (2011), "Calibrated Incentive Contracts," working paper, Princeton University.

[4] Kim-Sau Chung and J. C. Ely (2007), "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies* 74 (2), 447-476.

[5] Joel S. Demski and David E. M. Sappington (1987), "Delegated Expertise," *Journal of Accounting Research* 25 (1), 68-89.

[6] Alexander Frankel (2011), "Aligned Delegation," working paper, University of Chicago Booth School of Business.

[7] James M. Malcolmson (2009), "Principal and Expert Agent," *B.E. Journal of Theoretical Economics*, 9 (1) (Contributions), Article 17.

[8] Kent Osband (1989), "Optimal Forecasting Incentives," *Journal of Political Economy* 97 (5), 1091-1112.

[9] R. Tyrrell Rockafellar (1970), *Convex Analysis*. Princeton: Princeton University Press.

[10] Luis Zermeño (2011), "A Principal-Expert Model and the Value of Menus," working paper, Massachusetts Institute of Technology.

[11] Luis Zermeño (2012), "The Role of Authority in a General Principal-Expert Model," working paper, Massachusetts Institute of Technology.