Randomization is Optimal in the Robust Principal-Agent Problem*

Ashwin Kambhampati†

July 7, 2022

Abstract

A principal contracts with an agent, who takes a hidden action. The principal does not know all of the actions the agent can take and evaluates her payoff from any contract according to its worst-case performance. Carroll (2015) showed that there exists a linear contract that is optimal within the class of deterministic contracts. This paper shows that, whenever there is an optimal linear contract with non-zero slope, the principal can strictly increase her payoff by randomizing over deterministic contracts.

1 Introduction

A principal writes an incentive contract for an agent. The agent takes a productive, but hidden, action. Unfortunately for the principal, she does not know all available actions. Which contract yields her the highest worst-case payoff?

In path-breaking work, Carroll (2015) sets forth a new paradigm to answer this question. He proves, very generally, that the optimal deterministic contract is linear—

*I thank the editor, Tilman Börgers, and two anonymous referees for helpful comments and suggestions. I thank Max Rosenthal, Naomi Utgoff, and, especially, Juuso Toikka for encouraging and illuminating discussions. I gratefully acknowledge funding from the Volgenau Faculty Fellowship and the Naval Academy Research Council Seed Grant. The views expressed here are solely my own and do not in any way represent the views of the U.S. Naval Academy, U.S. Navy, or the Department of Defense.

†Department of Economics, United States Naval Academy; kambhamp@usna.edu.

the agent receives a constant fraction of the output she produces. This simple contract contrasts with the more complicated, detail-sensitive contracts predicted by the standard, Bayesian principal-agent model.

At the same level of generality as Carroll (2015), this paper shows that the principal can strictly increase her worst-case payoff by randomizing over deterministic contracts, as long as there is an optimal (deterministic) linear contract with non-zero slope (Theorem 1). Hence, if the principal believes that randomization can alleviate her ambiguity aversion, then restricting attention to the study of deterministic contracts is *with* loss of generality.

Section 2 presents the model, Section 3 states and proves the result, and Section 4 discusses a decision-theoretic justification for restricting attention to the study of deterministic contracts.

2 Model

In what follows, any Euclidean space is equipped with the Euclidean topology and any product of topological spaces is equipped with the product topology. The set of Borel distributions on any topological space $\mathcal{X}$ is denoted by $\Delta(\mathcal{X})$ and is always equipped with the topology of weak convergence.

2.1 Environment

There is a single principal and a single agent. The agent takes a costly, hidden action to produce stochastic, but observable, output. All parties are risk-neutral.

Let $Y \subset \mathbb{R}$ denote the set of possible output levels. It is assumed to be compact with $\min(Y) = 0$. To produce output, the agent chooses an action, a, which consists of a probability distribution, $F(a) \in \Delta(Y)$, and a cost of effort, $c(a) \in \mathbb{R}_+$. It is assumed that the set of actions available to the agent $A \subset \Delta(Y) \times \mathbb{R}_+$ is compact.

The principal can commit to a deterministic contract — a continuous function $w : Y \to \mathbb{R}_+$ — or a randomization over deterministic contracts. Non-negativity of wages reflects agent limited liability. Given a (deterministic) contract w, if the agent takes an action a and produces output y, then the agent's ex-post payoff is $w(y) - c(a)$ and the principal's ex-post payoff is $y - w(y)$.

The (non-empty) set of optimal actions for the agent under the action set A given a contract w is

$$\mathcal{A}(w, A) := \arg\max_{a \in A} E_{F(a)}[w(y)] - c(a).$$

It is assumed that if the agent is indifferent among several actions, she chooses the principal's most preferred action. Hence, the principal's expected payoff given w and A is

$$V(w, A) := \max_{a \in \mathcal{A}(w,A)} E_{F(a)}[y - w(y)].$$

2.2 Max-Min Problems

The principal knows only a compact subset of available actions to the agent $A_0 \subset \Delta(Y) \times \mathbb{R}_+$ when she writes a contract. She thus chooses one with the highest possible payoff guarantee across all compact supersets of her knowledge $A \supseteq A_0$.

Let the set of all deterministic contracts be denoted by $\mathcal{W}$ (equip it with the sup-norm topology). Let $\mathcal{S}$ denote the set of all compact supersets of $A_0 \subset \Delta(Y) \times \mathbb{R}_+$ (equip it with the topology induced by the Hausdorff metric). Carroll (2015) solved the principal's deterministic max-min optimization problem:

$$V_D^* := \sup_{w \in \mathcal{W}} \inf_{A \in \mathcal{S}} V(w, A). \tag{1}$$

In particular, he showed that there exists a linear contract

$$w^*(y) := \alpha^* y \quad \text{for some } \alpha^* \in [0, 1]$$

that obtains

$$V_D^* = \max_{\alpha \in [0,1], a_0 \in A_0} (1 - \alpha) \left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha} \right), \tag{2}$$

where if $\alpha = 0$ and $c(a_0) = 0$, $c(a_0)/\alpha$ is interpreted as 0, and if $\alpha = 0$ and $c(a_0) > 0$, $c(a_0)/\alpha$ is interpreted as $+\infty$. Given an optimal share $\alpha^* \in [0, 1]$, the bracketed expression corresponds to a tight lower bound on the agent's worst-case expected productivity.

As in Carroll (2015), it is assumed that there exists a known action generating strictly positive surplus, i.e., an action $a_0 \in A_0$ for which $E_{F(a_0)}[y] - c(a_0) > 0$. This ensures that $V_D^* > 0$; the optimal share α^* is strictly less than one and worst-case expected productivity is strictly positive,

$$\max_{a_0 \in A_0} \left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha^*} \right) > 0. \tag{3}$$

In addition, it is assumed that there exists a solution to (2) in which $\alpha^* > 0$, i.e., that there is an optimal linear contract with non-zero slope. A simple sufficient condition to ensure that this is the case is that any action, $a_0 \in A_0$, which generates strictly positive expected output, $E_{F(a_0)}[y] > 0$, has non-zero cost, $c(a_0) > 0$.

Under these (weak) hypotheses, I consider the more general problem in which the principal can commit to a randomization over deterministic contracts:

$$V_R^* := \sup_{\tilde{w} \in \Delta(\mathcal{W})} \inf_{A \in \mathcal{S}} V(\tilde{w}, A), \tag{4}$$

where $V(\tilde{w}, A) := E_{\tilde{w}}[V(w, A)]$. I note that, by standard arguments, it is without loss of generality to restrict Nature to choose a pure strategy, $A \in \mathcal{S}$, instead of allowing her to choose a mixed strategy, $\tilde{A} \in \Delta(\mathcal{S})$. That is, permitting Nature to randomize has no effect on the value of V_D^* or V_R^*.

3 Analysis

The main result of the paper follows below.

Theorem 1
Randomization strictly increases the principal's worst-case payoff:

$$V_R^* > V_D^*.$$

The proof is constructive; I exhibit a random contract that yields the principal a strictly higher worst-case payoff than any deterministic contract. (Appendix A.4 illustrates the difficulties that arise when trying to prove the result using a minimax theorem.) For this purpose, fix a linear contract $w^*(y) = \alpha^* y$ with $\alpha^* > 0$ that is optimal within the class of deterministic contracts. Then, define another deterministic, linear contract that yields the agent a smaller share of output than optimal:

$$w_\epsilon^*(y) := (\alpha^* - \epsilon)\, y,$$

where $\alpha^* > \epsilon > 0$. I show that, for ϵ sufficiently small, a contract that uniformly randomizes over the optimal deterministic contract and this alternative, sub-optimal contract, i.e.,

$$\tilde{w}_\epsilon := \frac{1}{2} \circ w^* + \frac{1}{2} \circ w_\epsilon^* \in \Delta(\mathcal{W}),$$

yields the principal a strictly higher worst-case payoff than w^*. Intuitively, the agent's worst-case productivity under $\tilde{w}_\epsilon$ is the same as under the optimal deterministic contract w^*, but the principal extracts more rent by, sometimes, paying the agent a smaller share of the output she produces.[1]

I first establish a lower bound on the principal's worst-case payoff from $\tilde{w}_\epsilon$.

Lemma 1

The worst-case payoff from the random contract $\tilde{w}_\epsilon$ is bounded below by the value function of a screening problem:

$$\inf_{A \in \mathcal{S}} V(\tilde{w}_\epsilon, A) \geq \underline{V}(\tilde{w}_\epsilon),$$

where

$$\begin{aligned} \underline{V}(\tilde{w}_\epsilon) := \quad & \min_{a^*, a_\epsilon^* \in \Delta(Y) \times \mathbb{R}_+} \quad \frac{1}{2}(1-\alpha^*)E_{F(a^*)}[y] + \frac{1}{2}(1-(\alpha^*-\epsilon))E_{F(a_\epsilon^*)}[y] \\ & \text{subject to} \\ & [IC_{w^*}] \quad \alpha^* E_{F(a^*)}[y] - c(a^*) \geq \max_{a_0 \in A_0} \alpha^* E_{F(a_0)}[y] - c(a_0) \\ & [IC_{w_\epsilon^* \to w^*}] \quad (\alpha^*-\epsilon)E_{F(a_\epsilon^*)}[y] - c(a_\epsilon^*) \geq (\alpha^*-\epsilon)E_{F(a^*)}[y] - c(a^*). \end{aligned} \tag{5}$$

Proof. See Appendix A.1. ☐

The value of (5) provides a lower bound on the principal's payoff under the random contract $\tilde{w}_\epsilon$. IC_{w^*} ensures that, relative to any known action, the agent prefers to take action a^* when the contract realization is w^*. $IC_{w_\epsilon^* \to w^*}$ ensures that the agent prefers to take action a_ϵ^* instead of a^* when the contract realization is w_ϵ^*.

I next establish properties that hold in any solution to (5).

Lemma 2

If $\epsilon > 0$ is sufficiently small, then there exists a solution to (5) and the following properties hold in any solution:

[1] A related intuition is explored in a team-production setting in Kambhampati (2022).

1. $IC_{w_\epsilon^* \to w^*}$ *and* IC_{w^*} *bind.*

2. $c(a^*) = c(a_\epsilon^*) = 0$ *and* $E_{F(a^*)}[y] = E_{F(a_\epsilon^*)}[y]$.

Proof. See Appendix A.2. □

The binding screening constraint $IC_{w_\epsilon^* \to w^*}$ prevents Nature from minimizing the principal's payoff contract-by-contract, i.e., Nature's worst-case response to the contract w^* constrains her worst-case response to w_ϵ^*. In addition, when $\epsilon > 0$ is sufficiently small, there is pooling in any solution to (5): $c(a^*) = c(a_\epsilon^*) = 0$ and $E_{F(a^*)}[y] = E_{F(a_\epsilon^*)}[y]$. That is, in worst-case scenarios, an agent receiving contract w^* takes an action with payoff-identical properties as when she receives w_ϵ^*.

Pooling means that the agent's worst-case expected productivity under w_ϵ^* is no lower than her worst-case expected productivity under w^*. And by the binding constraint IC_{w^*},

$$E_{F(a^*)}[y] = E_{F(a_\epsilon^*)}[y] = \max_{a_0 \in A_0} \left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha^*} \right) > 0,$$

where the strict inequality is from (3). These observations immediately yield that any solution to (5) results in a payoff for the principal strictly larger than V_D^*:

$$\underline{V}(\tilde{w}_\epsilon) = \frac{1}{2}\left[(1-\alpha^*)E_{F(a^*)}[y]\right] + \frac{1}{2}\left[(1-(\alpha^*-\epsilon))E_{F(a^*)}[y]\right] > (1-\alpha^*)E_{F(a^*)}[y] = V_D^*,$$

where the inequality follows because the principal, sometimes, pays the agent a share of output $\alpha^* - \epsilon$ instead of α^*.

Putting everything together, I have shown that, if ϵ is sufficiently small, then

$$V_R^* \geq \inf_{A \in \mathcal{S}} V(\tilde{w}_\epsilon, A) \geq \underline{V}(\tilde{w}_\epsilon) > V_D^*,$$

where the first inequality is by definition, the second is proved in Lemma 1, and the third is a corollary of Lemma 2. Theorem 1 has thus been proven; the principal strictly benefits from randomization.

I conclude the analysis with two remarks. First, Theorem 1 holds under more general utility functions for the agent, e.g., those reflecting risk aversion. Specifically, suppose that the agent has any increasing, bijective utility function over wages, $u_A : \mathbb{R} \to \mathbb{R}$ with $u_A(0) = 0$, and that the agent's overall utility is additively separable in utility from wages and disutility of effort. Under these conditions, a working paper version[2] of Carroll (2015) showed that there exists an optimal deterministic contract w^* for which there exist constants $\alpha^* \geq 0$ and $\beta^* \geq 0$ such that[3]

$$u_A(w^*(y)) = \alpha^*(y - w^*(y)) + \beta^*.$$

In Appendix A.3, I show that if there is an optimal deterministic contract with $\alpha^* > 0$, so that $w^*(y)$ is strictly increasing, then the principal strictly benefits from randomization. The proof mirrors that of Theorem 1, but, instead, considers the payoff guarantee of a contract that uniformly randomizes over w^* and another contract, w^*_ϵ, satisfying[4]

$$u_A(w^*_\epsilon(y)) = (\alpha^* - \epsilon)\,(y - w^*_\epsilon(y)) + \beta^*.$$

It is shown that this randomization yields the principal a strictly higher payoff than w^* if $\epsilon > 0$ is sufficiently small. Hence, the result that randomization strictly benefits the principal does not rely crucially on the agent's risk-neutrality or on the linearity of the optimal deterministic contract in output.

Second, the limited liability constraint is indispensable. If the limited liability constraint is omitted from the model and replaced with the requirement that the agent receives non-negative expected utility, then the principal can extract full surplus from the agent under A_0 with a deterministic contract $w(y) = y - s_0$, where $s_0 := \max_{a_0 \in A_0} E_{F(a_0)}[y] - c(a_0)$. Because it is impossible to obtain a higher worst-case expected payoff than the total surplus under A_0, the principal can do no better by randomizing.

4 Conclusion

In the spirit of Raiffa (1961)'s critique and in the tradition of the theory of zero-sum games, I have explored the possibility that randomization might be used to increase the principal's worst-case payoff in the robust principal-agent problem of Carroll

[2] See http://www.bu.edu/econ/files/2013/03/May-4-Caroll.pdf, dated December 21, 2012.

[3] The intercept β^* must be positive because, if not, then the limited liability constraint on w^* would be violated when $y = 0$. Notice also that though w^* is affine in the principal's utility, it need *not* be affine in output.

[4] The working paper version of Carroll (2015) contains a detailed proof that, for every $\epsilon > 0$, there exists a unique continuous function w^*_ϵ satisfying this equation.

(2015). I proved that the principal does, in fact, achieve a strictly higher worst-case payoff by randomizing.

How should the optimal deterministic contract be interpreted? Building upon Ellsberg (1961), Saito (2015) argues that it might be reasonable for a decision maker to believe that randomization will not resolve her ambiguity aversion (see also Ke and Zhang (2020)). In particular, the principal might believe that Nature moves only after a deterministic contract is realized. Hence, under such beliefs, the optimal deterministic contract cannot be improved upon. The validity of restricting attention to the study of deterministic contracts thus depends crucially upon the principal's beliefs about the timing of the resolution of uncertainty.

A Proofs

A.1 Proof of Lemma 1

Fix $\tilde{w}_\epsilon$ and take any strategy of Nature $A \in \mathcal{S}$. If w^* is realized, then the agent chooses an action $a^* \in \mathcal{A}(w^*, A)$, which necessarily satisfies

$$\alpha^* E_{F(a^*)}[y] - c(a^*) \geq \max_{a \in A} \alpha^* E_{F(a)}[y] - c(a).$$

Similarly, if w^*_ϵ is realized, then the agent chooses an action $a^*_\epsilon \in \mathcal{A}(w^*_\epsilon, A)$, which necessarily satisfies

$$(\alpha^* - \epsilon) E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \geq \max_{a \in A} (\alpha^* - \epsilon) E_{F(a)}[y] - c(a).$$

Because w^* and w^*_ϵ are realized with equal probability, if these two actions are taken, then the principal obtains an expected payoff of

$$\frac{1}{2} E_{F(a^*)}[y] + \frac{1}{2}(1 - (\alpha^* - \epsilon)) E_{F(a^*_\epsilon)}[y].$$

It follows that

$$\inf_{A \in \mathcal{S}} V(\tilde{w}_\epsilon, A) \geq \hat{V}(\tilde{w}_\epsilon),$$

where

$$\begin{aligned}
\hat{V}(\tilde{w}_\epsilon) := \quad & \min_{A \in \mathcal{S}} \quad \frac{1}{2}(1-\alpha^*)E_{F(a^*)}[y] + \frac{1}{2}(1-(\alpha^*-\epsilon))E_{F(a^*_\epsilon)}[y] \\
& \text{subject to} \\
& [IC_{w^*}] \quad \alpha^* E_{F(a^*)}[y] - c(a^*) \geq \max_{a \in A} \alpha^* E_{F(a)}[y] - c(a) \\
& [IC_{w^*_\epsilon}] \quad (\alpha^*-\epsilon)E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \geq \max_{a \in A}(\alpha^*-\epsilon)E_{F(a)}[y] - c(a).
\end{aligned}$$

I next claim that $\hat{V}(\tilde{w}_\epsilon) \geq \underline{V}(\tilde{w}_\epsilon)$, where $\underline{V}(\tilde{w}_\epsilon)$ is defined in the statement of the Lemma. First, observe that any $A \in \mathcal{S}$ containing more than two "unknown" actions for the agent can be replaced with $A_0 \cup \{a^*, a^*_\epsilon\} \subset A$, where a^* is the agent's action under w^* and a^*_ϵ is their action under w^*_ϵ. The resulting program,

$$\begin{aligned}
& \min_{a^*, a^*_\epsilon \in \Delta(Y) \times \mathbb{R}_+} \quad \frac{1}{2}(1-\alpha^*)E_{F(a^*)}[y] + \frac{1}{2}(1-(\alpha^*-\epsilon))E_{F(a^*_\epsilon)}[y] \\
& \text{subject to} \\
& [IC_{w^*}] \quad \alpha^* E_{F(a^*)}[y] - c(a^*) \geq \max_{a_0 \in A_0} \alpha^* E_{F(a_0)}[y] - c(a_0) \\
& [IC_{w^*_\epsilon}] \quad (\alpha^*-\epsilon)E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \geq \max_{a \in A_0}(\alpha^*-\epsilon)E_{F(a_0)}[y] - c(a_0) \\
& [IC_{w^* \to w^*_\epsilon}] \quad \alpha^* E_{F(a^*)}[y] - c(a^*) \geq \alpha^* E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \\
& [IC_{w^*_\epsilon \to w^*}] \quad (\alpha^*-\epsilon)E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \geq (\alpha^*-\epsilon)E_{F(a^*)}[y] - c(a^*),
\end{aligned}$$

has fewer inequality constraints. Therefore, its solution results in a (weakly) smaller payoff for the principal. Similarly, removing the constraints $IC_{w^*_\epsilon}$ and $IC_{w^* \to w^*_\epsilon}$ results in an even smaller payoff for the principal. It follows that

$$\inf_{A \in \mathcal{S}} V(\tilde{w}_\epsilon, A) \geq \hat{V}(\tilde{w}_\epsilon) \geq \underline{V}(\tilde{w}_\epsilon).$$

A.2 Proof of Lemma 2

I first prove that if $\epsilon \in (0, \alpha^*)$ is sufficiently small, then it must be that $c(a^*_\epsilon) = 0$ in any solution to (5). Observe that in any solution it must be that

$$\alpha^* E_{F(a^*)}[y] - c(a^*) > 0.$$

This follows from IC_{w^*} and $\max_{a_0 \in A_0} \alpha^* E_{F(a_0)}[y] - c(a_0) > 0$, which is implied by (3). Therefore, there exists a value $\epsilon_1 \in (0, \alpha^*)$ such that, for all $\epsilon < \epsilon_1$,

$$(\alpha^* - \epsilon) E_{F(a^*)}[y] - c(a^*) > 0.$$

Now, let $\epsilon < \epsilon_1$ and, towards contradiction, suppose that $c(a^*_\epsilon) > 0$. By $IC_{w^*_\epsilon \to w^*}$, it must be that $E_{F(a^*_\epsilon)}[y] > 0$. So, a^*_ϵ can be replaced with an alternative action $\hat{a}$ that has cost $c(\hat{a}) = 0$ and a mixture distribution $F(\hat{a})$ which places probability $\delta > 0$ on the Dirac measure with unit mass on $y = 0$ and probability $(1-\delta)$ on $F(a^*_\epsilon)$, where δ is chosen so that the agent's utility is unchanged. Then, $E_{F(\hat{a})}[y] < E_{F(a^*_\epsilon)}[y]$ and all incentive constraints are satisfied. Since the objective function is strictly increasing in the agent's productivity, it follows that the principal's payoff could not have been minimized with $c(a^*_\epsilon) > 0$, the desired contradiction.

Imposing the necessary condition for optimality that $c(a^*_\epsilon) = 0$, the following problem remains:

$$\begin{aligned}
&\min_{E_{F(a^*)}[y], E_{F(a^*_\epsilon)}[y] \in [0, \max(Y)], c(a^*) \geq 0} \quad \frac{1}{2}(1-\alpha^*) E_{F(a^*)}[y] + \frac{1}{2}(1 - (\alpha^* - \epsilon)) E_{F(a^*_\epsilon)}[y] \\
&\text{subject to} \\
&[IC_{w^*}] \quad E_{F(a^*)}[y] \geq U_0 + \frac{c(a^*)}{\alpha^*} \\
&[IC_{w^*_\epsilon \to w^*}] \quad E_{F(a^*_\epsilon)}[y] \geq E_{F(a^*)}[y] - \frac{c(a^*)}{(\alpha^* - \epsilon)},
\end{aligned} \tag{6}$$

where

$$U_0 := \max_{a_0 \in A_0} \left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha^*} \right) > 0$$

and Nature's choices of $F(a^*)$ and $F(a^*_\epsilon)$ have been replaced with their expected values, $E_{F(a^*)}[y]$ and $E_{F(a^*_\epsilon)}[y]$.[5]

Now, consider the relaxed problem in which the constraints $E_{F(a^*)}[y], E_{F(a^*_\epsilon)}[y] \in [0, \max(Y)]$ are omitted. In any solution to this problem, IC_{w^*} must bind. If IC_{w^*} does not bind, then Nature can reduce $E_{F(a^*)}[y]$ by a small amount while satisfying all incentive constraints and strictly reduce the objective function. In addition,

[5] Replacement with expectations is without loss of generality because only these properties of the distributions affect the objective function and constraints. In addition, any expected value between 0 and $\max(Y)$ can be obtained through an appropriately defined mixture distribution.

$IC_{w^*_\epsilon \to w^*}$ must bind. If not, then Nature could reduce $E_{F(a^*_\epsilon)}[y]$ by a small amount and strictly reduce the objective function. Eliminating $E_{F(a^*)}[y]$ and $E_{F(a^*_\epsilon)}[y]$ from Nature's problem using the binding constraints yields

$$\min_{c(a^*)\geq 0} \quad \frac{1}{2}(1-\alpha^*)\left(U_0 + \frac{c(a^*)}{\alpha^*}\right) + \frac{1}{2}(1-(\alpha^*-\epsilon))\left(U_0 + \frac{c(a^*)}{\alpha^*} - \frac{c(a^*)}{(\alpha^*-\epsilon)}\right).$$

The objective function is strictly increasing in $c(a^*)$ if and only if

$$\frac{(1-\alpha^*)(\alpha^*-\epsilon)}{(1-\alpha^*+\epsilon)} > \epsilon.$$

Since the left-hand side converges to $\alpha^* > 0$ as ϵ converges to zero and the right-hand side converges to zero as ϵ converges to zero, there exists an $\epsilon_2 > 0$ such that if $\epsilon < \min\{\epsilon_1, \epsilon_2\}$, then the inequality is satisfied. Hence, for $\epsilon < \min\{\epsilon_1, \epsilon_2\}$, it is optimal to make $c(a^*)$ as small as possible, i.e., any solution to the relaxed problem has $c(a^*) = 0$. When this is the case, IC_{w^*} and $IC_{w^*_\epsilon \to w^*}$ binding yields $E_{F(a^*_\epsilon)} = U_0 = E_{F(a^*_\epsilon)}[y] \in [0, \max(Y)]$. So, any solution to the relaxed problem is feasible in the original problem and the solution sets of the two problems coincide.

In summary, if $\epsilon > 0$ is sufficiently small, then $c(a^*) = c(a^*_\epsilon) = 0$ in any solution to Nature's screening problem. Removing $c(a^*)$ from the choice variables in (6), I observe that existence of a solution to Nature's screening problem follows from compactness of the constraint set and continuity of the principal's objective function in the remaining choice variables. Finally, IC_{w^*} and $IC_{w^*_\epsilon \to w^*}$ binding yields $E_{F(a^*)}[y] = U_0 = E_{F(a^*_\epsilon)}[y]$ in any solution. Hence, all properties stated in the Lemma have been proven.

A.3 General Agent Utility Function

Suppose that the agent has any increasing, bijective utility function over wages, $u_A : \mathbb{R} \to \mathbb{R}$ with $u_A(0) = 0$, and that the agent's overall utility is additively separable in utility from wages and disutility of effort. Let $w^*(y)$ be an optimal deterministic contract satisfying

$$u_A(w^*(y)) = \alpha^*(y - w^*(y)) + \beta^*,$$

where $\alpha^* > 0$ and $\beta^* \geq 0$. This contract yields the principal a worst-case payoff of[6]

$$V_D^* := \frac{1}{\alpha^*}U_0 - \frac{\beta^*}{\alpha^*} > 0, \tag{7}$$

where

$$U_0 := \max_{a_0 \in A_0} E_{F(a_0)}[u_A(w^*(y))] - c(a_0) > 0.$$

For each $0 < \epsilon < \alpha^*$, define w_ϵ^* to be the unique contract satisfying

$$u_A(w_\epsilon^*(y)) = (\alpha^* - \epsilon)(y - w_\epsilon^*(y)) + \beta^*.$$

Now, consider the random contract

$$\tilde{w}_\epsilon := \frac{1}{2} \circ w^*(y) + \frac{1}{2} \circ w_\epsilon^*(y).$$

I show that, for $\epsilon > 0$ sufficiently small, $\tilde{w}_\epsilon$ yields the principal a strictly higher worst-case payoff than w^*.

The following version of Lemma 1 holds in this model.

Lemma 3
The worst-case payoff from the random contract $\tilde{w}_\epsilon$ is bounded below by the value function of a screening problem:

$$\inf_{A \in \mathcal{S}} V(\tilde{w}_\epsilon, A) \geq \underline{V}(\tilde{w}_\epsilon) - \left(\frac{1}{2}\frac{\beta^*}{\alpha^*} + \frac{1}{2}\frac{\beta^*}{(\alpha^* - \epsilon)}\right),$$

where

$$\begin{aligned}
\underline{V}(\tilde{w}_\epsilon) := \quad & \min_{a^*, a_\epsilon^* \in \Delta(Y) \times \mathbb{R}_+} \quad \frac{1}{2\alpha^*} E_{F(a^*)}\left[u_A(w^*(y))\right] + \frac{1}{2(\alpha^* - \epsilon)} E_{F(a_\epsilon^*)}\left[u_A(w_\epsilon^*(y))\right] \\
& \text{subject to} \\
& [IC_{w^*}] \quad E_{F(a^*)}[u_A(w^*(y))] - c(a^*) \geq U_0 \\
& [IC_{w_\epsilon^* \to w^*}] \quad E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))] - c(a_\epsilon^*) \geq E_{F(a^*)}[u_A(w_\epsilon^*(y))] - c(a^*).
\end{aligned} \tag{8}$$

The proof of Lemma 3 is almost identical to that of Lemma 1 and is therefore omitted. It will be useful to make a few simplifications before proving an analog of Lemma 2. First, the argument that, if $\epsilon > 0$ is sufficiently small, then it must be

[6] This is an immediate implication of Lemma 3.1 in the working paper version of Carroll (2015).

that $c(a^*_\epsilon) = 0$ mirrors the argument in the proof of Lemma 2 and is omitted. The following problem remains:

$$\underline{V}(\tilde{w}_\epsilon) := \min_{F(a^*)\in\Delta(Y), F(a^*_\epsilon)\in\Delta(Y), c(a^*)\geq 0} \quad \frac{1}{2\alpha^*} E_{F(a^*)}\left[u_A(w^*(y))\right] + \frac{1}{2(\alpha^*-\epsilon)} E_{F(a^*_\epsilon)}\left[u_A(w^*_\epsilon(y))\right]$$

$$\text{subject to}$$

$$[IC_{w^*}] \quad E_{F(a^*)}[u_A(w^*(y))] - c(a^*) \geq U_0$$

$$[IC_{w^*_\epsilon \to w^*}] \quad E_{F(a^*_\epsilon)}[u_A(w^*_\epsilon(y))] \geq E_{F(a^*)}[u_A(w^*_\epsilon(y))] - c(a^*).$$

Observe that both IC_{w^*} and $IC_{w^*_\epsilon \to w^*}$ must bind in this problem because the objective function is strictly increasing in $E_{F(a^*)}\left[u_A(w^*(y))\right]$ and $E_{F(a^*_\epsilon)}\left[u_A(w^*_\epsilon(y))\right]$ and because decreasing $E_{F(a^*)}\left[u_A(w^*(y))\right]$ necessarily decreases $E_{F(a^*)}[u_A(w^*_\epsilon(y))]$. Hence, any strict relaxation of $IC_{w^*_\epsilon \to w^*}$ must result in a strict decrease in $\underline{V}(\tilde{w}_\epsilon)$.

For the purpose of constructing such a relaxed problem, observe that for any $F(a^*) \in \Delta(Y)$ with $E_{F(a^*)}[y] > 0$, which must hold for any action a^* satisfying IC_{w^*}, it must be that

$$\begin{aligned} E_{F(a^*)}[u_A(w^*(y)) - u_A(w^*_\epsilon(y))] &= \epsilon E_{F(a^*)}[y] - \alpha^* E_{F(a^*)}[w^*(y)] + (\alpha^* - \epsilon) E_{F(a^*)}[w^*_\epsilon(y)] \\ &< \epsilon E_{F(a^*)}[y] - \alpha^* E_{F(a^*)}[w^*(y)] + (\alpha^* - \epsilon) E_{F(a^*)}[w^*(y)] \\ &= \epsilon E_{F(a^*)}[y - w^*(y)] \\ &= \frac{\epsilon}{\alpha^*}\left[E_{F(a^*)}[u_A(w^*(y))] - \beta^*\right], \end{aligned}$$

where the inequality follows because $w^*_\epsilon(y) < w^*(y)$ for all $y > 0$. Hence,

$$E_{F(a^*)}[u_A(w^*_\epsilon(y))] > \left(1 - \frac{\epsilon}{\alpha^*}\right) E_{F(a^*)}[u_A(w^*(y))] + \frac{\epsilon}{\alpha^*}\beta^*,$$

for any action a^* satisfying IC_{w^*}. Using this observation, consider the following relaxed screening problem:

$$\underline{V}^*(\tilde{w}_\epsilon) :=$$
$$\min_{E_{F(a^*)}[u_A(w^*(y))], E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))], c(a^*)\geq 0} \quad \frac{1}{2\alpha^*} E_{F(a^*)}\left[u_A(w^*(y))\right] + \frac{1}{2(\alpha^*-\epsilon)} E_{F(a_\epsilon^*)}\left[u_A(w_\epsilon^*(y))\right]$$

subject to

$$\begin{aligned}
&[IC_{w^*}] \quad E_{F(a^*)}[u_A(w^*(y))] - c(a^*) \geq U_0 \\
&[IC_{w_\epsilon^* \to w^*}] \quad E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))] \geq (1 - \frac{\epsilon}{\alpha^*}) E_{F(a^*)}[u_A(w^*(y))] + \frac{\epsilon}{\alpha^*}\beta^* - c(a^*) \\
&[N_1] \quad E_{F(a^*)}[u_A(w^*(y))] \in [0, u_A(w^*(\max(Y)))] \\
&[N_2] \quad E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))] \in [0, \max\{U_0 + \frac{\epsilon}{\alpha^*}\beta^*, u_A(w^*(\max(Y)))\}],
\end{aligned} \tag{9}$$

which necessarily satisfies $\underline{V}^*(\tilde{w}_\epsilon) < \underline{V}(\tilde{w}_\epsilon)$.

I now prove an analog of Lemma 2.

Lemma 4

If $\epsilon > 0$ is sufficiently small, then there exists a solution to (9) and any solution satisfies the following properties:

1. *$IC_{w_\epsilon^* \to w^*}$ and IC_{w^*} bind.*
2. *$c(a^*) = 0$, $E_{F(a^*)}[u_A(w^*(y))] = U_0$, and $E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))] = (1 - \frac{\epsilon}{\alpha^*})U_0 + \frac{\epsilon}{\alpha^*}\beta^*$.*

Proof. Consider (9) without N_1 and N_2. IC_{w^*} and $IC_{w_\epsilon^* \to w^*}$ must bind in this problem because the objective function is strictly increasing in both $E_{F(a^*)}[u_A(w^*(y))]$ and $E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))]$ and because decreasing $E_{F(a^*)}\left[u_A(w^*(y))\right]$ relaxes $IC_{w_\epsilon^* \to w^*}$. Using the binding constraints to eliminate $E_{F(a^*)}[u_A(w^*(y))]$ and $E_{F(a_\epsilon^*)}[u_A(w_\epsilon^*(y))]$ from Nature's problem yields

$$\min_{c(a^*)\geq 0} \quad \frac{1}{2\alpha^*} c(a^*) - \frac{1}{2(\alpha^*-\epsilon)} \frac{\epsilon}{\alpha^*} c(a^*),$$

where constants in the objective function have been removed. This expression is strictly increasing in $c(a^*)$ whenever $1 > \frac{\epsilon}{\alpha^*-\epsilon}$, which holds if $\epsilon > 0$ is sufficiently small. So, in these cases, it is optimal to set $c(a^*) = 0$. Existence of a solution to (9) then follows from compactness of the constraint set and continuity of the objective function in the choice variables. The binding constraint IC_{w^*} yields $E_{F(a^*)}[u_A(w^*(y))] = U_0 \in$

$[0, u_A(w^*(\max(Y))]$. The binding constraint $IC_{w^*_\epsilon \to w^*}$ yields $E_{F(a^*_\epsilon)}[u_A(w^*_\epsilon(y))] = (1 - \frac{\epsilon}{\alpha^*})U_0 + \frac{\epsilon}{\alpha^*}\beta^* \in [0, U_0 + \frac{\epsilon}{\alpha^*}\beta^*]$. So, N_1 and N_2 are satisfied and the properties of the solutions to the relaxed problem hold in any solution to the original problem. $\square$

For $\epsilon > 0$ sufficiently small, Lemma 4 yields

$$\underline{V}^*(\tilde{w}_\epsilon) = \frac{1}{2\alpha^*}U_0 + \frac{1}{2(\alpha^* - \epsilon)}\left((1 - \frac{\epsilon}{\alpha^*})U_0 + \frac{\epsilon}{\alpha^*}\beta^*\right).$$

So,

$$\underline{V}^*(\tilde{w}_\epsilon) - \left(\frac{1}{2}\frac{\beta^*}{\alpha^*} + \frac{1}{2}\frac{\beta^*}{(\alpha^* - \epsilon)}\right) = \frac{U_0}{\alpha^*} - \frac{\beta^*}{\alpha^*} = V^*_D,$$

where the final equality follows from (7). Hence,

$$\inf_{A \in \mathcal{S}} V(\tilde{w}_\epsilon, A) \geq \underline{V}(\tilde{w}_\epsilon) - \left(\frac{1}{2}\frac{\beta^*}{\alpha^*} + \frac{1}{2}\frac{\beta^*}{(\alpha^* - \epsilon)}\right) > \underline{V}^*(\tilde{w}_\epsilon) - \left(\frac{1}{2}\frac{\beta^*}{\alpha^*} + \frac{1}{2}\frac{\beta^*}{(\alpha^* - \epsilon)}\right) = V^*_D,$$

where the first inequality is from Lemma 3, the second from $\underline{V}(\tilde{w}_\epsilon) > \underline{V}^*(\tilde{w}_\epsilon)$, and the third was just established as a corollary of Lemma 4.

A.4 Unsuccessful, Minimax Proof Approaches

Carroll (2015) observed that there is no saddle point in problem (1), i.e.,

$$V^*_D = \sup_{w \in \mathcal{W}} \inf_{A \in \mathcal{S}} V(w, A) < \inf_{A \in \mathcal{S}} \sup_{w \in \mathcal{W}} V(w, A) := \overline{V}_D. \tag{10}$$

He then remarked that this "suggests that [the principal] should be able to improve her worst-case guarantee by randomizing over contracts", an intuition coming from von Neumann's minimax theorem (von Neumann (1928)) and the existence of mixed-strategy saddle points in finite zero-sum games (von Neumann and Morgenstern (1944)). In particular, suppose that the following minimax equality holds:

$$V^*_R = \sup_{\tilde{w} \in \Delta(\mathcal{W})} \inf_{A \in \mathcal{S}} V(\tilde{w}, A) = \inf_{A \in \mathcal{S}} \sup_{\tilde{w} \in \Delta(\mathcal{W})} V(\tilde{w}, A) := \overline{V}_R, \tag{11}$$

where the principal's strategy space is extended from $\mathcal{W}$ to $\Delta(\mathcal{W})$. Then, the observation that $\overline{V}_R \geq \overline{V}_D$ yields $V^*_R > V^*_D$ by (10).

Unfortunately, it is not clear how to establish (11) using a minimax theorem.

First, though $\mathcal{S}$ is compact under the topological assumptions made in Section 2.2, it is not apparent that $\mathcal{S}$ can be made convex in a manner that makes $V(\tilde{w}, \cdot)$ quasi-concave or quasi-convex. Such a condition is necessary to apply Sion (1958)'s minimax theorem. Second, it is not clear that $V(\tilde{w}, \cdot)$ is convexlike, a necessary condition to use Fan (1953)'s minimax theorem. Finally, $V(\tilde{w}, \cdot)$ is not continuous under principal most-preferred action selection and $V(\cdot, \tilde{A})$ is not continuous if, instead, principal least-preferred action selection is assumed.

A natural approach to obviate all issues described and to make Nature's strategy space "well-behaved" is to extend it from $\mathcal{S}$ to $\Delta(\mathcal{S})$. Then, the minimax equality becomes

$$V_R^* = \sup_{\tilde{w} \in \Delta(\mathcal{W})} \inf_{\tilde{A} \in \Delta(\mathcal{S})} V(\tilde{w}, \tilde{A}) = \inf_{\tilde{A} \in \Delta(\mathcal{S})} \sup_{\tilde{w} \in \Delta(\mathcal{W})} V(\tilde{w}, \tilde{A}) := \overline{V}_{RR}, \tag{12}$$

where $V(\tilde{w}, \tilde{A}) := E_{\tilde{w}, \tilde{A}}[V(w, A)]$. However, (12) is insufficient to establish $V_R^* > V_D^*$ because it need not be the case that $\overline{V}_{RR} \geq \overline{V}_D$.

References

Carroll, G. "Robustness and Linear Contracts." *American Economic Review*, Vol. 105 (2015), pp. 536–563.

Ellsberg, D. "Risk, ambiguity, and the Savage axioms." *The Quarterly Journal of Economics*, (1961), pp. 643–669.

Fan, K. "Minimax Theorems." *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 39 (1953), pp. 42–47.

Kambhampati, A. "Robust Performance Evaluation." *Unpublished manuscript*, (2022).

Ke, S. and Zhang, Q. "Randomization and ambiguity aversion." *Econometrica*, Vol. 88 (2020), pp. 1159–1195.

Raiffa, H. "Risk, ambiguity, and the Savage axioms: comment." *The Quarterly Journal of Economics*, Vol. 75 (1961), pp. 690–694.

Saito, K. "Preferences for flexibility and randomization under uncertainty." *American Economic Review*, Vol. 105 (2015), pp. 1246–71.

Sion, M. "On general minimax theorems." *Pacific Journal of Mathematics*, Vol. 8 (1958), pp. 171–176.

von Neumann, J. “Zur theorie der gesellschaftsspiele.” *Mathematische Annalen*, Vol. 100 (1928), pp. 295–320.

von Neumann, J. and Morgenstern, O. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.