Randomization is Optimal in the Robust Principal-Agent Problem*

Ashwin Kambhampati†

March 1, 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) identified the principal's optimal deterministic contract. I prove that the principal can strictly increase her payoff by randomizing over deterministic contracts.

*I thank Max Rosenthal, Naomi Utgoff, and, especially, Juuso Toikka for encouraging and illuminating discussions.

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

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—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.

I prove, at the same level of generality as Carroll (2015), that the principal can strictly increase her worst-case payoff by randomizing over deterministic contracts (Theorem 1). Hence, restricting attention to the study of deterministic contracts is *with* loss of generality, provided that the principal believes that randomization can alleviate her ambiguity aversion. (A more nuanced discussion of these preferences is contained at the end of the paper.)

Section 2 outlines the model, Section 3 states and proves the result, and Section 4 concludes.

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 agents $A \subseteq \Delta(Y) \times \mathbb{R}_+$ 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. Let $\mathcal{A}(w, A)$ denote the (non-empty) set of optimal actions for the agent under the action set A given a deterministic contract w. It is assumed that if the agent is indifferent among several actions, she chooses the principal's most preferred action. Hence, the principal's 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 subset of available actions to the agent $A_0 \subseteq \Delta(Y) \times \mathbb{R}_+$ when she writes a contract. She thus chooses one with the highest possible payoff guarantee across all supersets of her knowledge $A \supseteq A_0$. To avoid trivial solutions to the principal's contracting problem, it is assumed, as in Carroll (2015), that there exists an action, $a_0 \in A_0$, generating strictly positive surplus, i.e. $E_{F(a_0)}[y] - c(a_0) > 0$.

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 $A^0 \subseteq \Delta(Y) \times \mathbb{R}_+$ (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

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

where the bracketed expression corresponds to a tight lower bound on the agent's worst-case expected productivity. I consider the more general problem in which the principal can commit to a randomization over deterministic contracts:

$$V_R^* := \sup_{\tilde{w} \in \Delta(\mathcal{W})} \inf_{A \in \mathcal{S}} V(\tilde{w}, A), \tag{3}$$

where $V(\tilde{w}, A) := E_{\tilde{w}}[V(w, A)]$.[1]

3 The Result

The main result of the paper follows below.[2]

Theorem 1

If there exists an optimal deterministic linear contract different from the zero contract, i.e. one with a share parameter $\alpha^* > 0$, then randomization strictly increases the principal's worst-case payoff:

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

The proof is constructive; I exhibit a random contract that strictly outperforms the optimal linear contract. It is instructive, however, to first see the difficulties that arise when trying to establish the result using the minimax theorem.

3.1 Unsuccessful, Minimax Proof Approaches

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

$$V_D^* = \sup_{w \in \mathcal{W}} \inf_{A \in \mathcal{S}} V(w, A) < \inf_{A \in \mathcal{S}} \sup_{w \in \mathcal{W}} V(w, A) := \overline{V}_D. \tag{4}$$

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 Nuemann'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:

$$V_R^* = \sup_{\tilde{w} \in \Delta(\mathcal{W})} \inf_{A \in \mathcal{S}} V(\tilde{w}, A) = \inf_{A \in \mathcal{S}} \sup_{\tilde{w} \in \Delta(\mathcal{W})} V(\tilde{w}, A) := \overline{V}_R, \tag{5}$$

[1] Of course, permitting Nature to randomize has no effect on the value of V_D^* or V_R^*.

[2] As shown by Carroll (2015), a necessary and sufficient condition for existence of an optimal linear contract different from the zero contract is the existence of a pair

$$(\alpha^*, a_0^*) \in \operatorname*{argmax}_{\alpha \in [0,1], a_0 \in A_0} (1 - \alpha) \left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha} \right) \quad \text{for which } \alpha^* > 0,$$

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 $+\infty$. There exists such an $\alpha^* > 0$ if, for instance, all known actions that generate positive surplus are costly (the optimal share parameter is $\alpha^* = \sqrt{c(a_0^*)}/\sqrt{E_F(a_0^*)[y]}$).

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 (4).

Unfortunately, to my knowledge, existing extensions of von Nuemann's minimax theorem to infinite-dimensional spaces cannot be used to establish (5). For instance, Sion's minimax theorem (Theorem 3.4 of Sion (1958)) requires, among other conditions, that $V(w, \cdot)$ is lower semicontinuous. But $V(w, \cdot)$ is not lower semicontinuous, as demonstrated by the following example.

Example 1. Suppose $A_0 = \{a_0\}$, with $E_{F(a_0)}[y] = \frac{1}{2}$ and $c(a_0) = \frac{1}{8}$, and $Y = \{0, 1\}$. Fix $w(y) = \frac{1}{2}y$. To show that $V(w, \cdot)$ is not lower semicontinuous, take a sequence of action sets (A_n) where $A_n := \{a_0, a_n\}$, with $c(a_n) = 0$ and $E_{F(a_n)}[y] = \frac{1}{4} + \frac{1}{n+1}$. For A_n, $\mathcal{A}(w, A_n) = \{a_n\}$ yielding the principal an expected payoff of $V(w, A_n) = \frac{1}{2}(\frac{1}{4} + \frac{1}{n+1})$. Hence, $\lim_{n\to\infty} V(w, A_n) = \frac{1}{8}$. However, the limit of (A_n) in the Hausdorff metric is $A^* := \{a_0, a^*\}$, with $c(a_n) = 0$ and $E_{F(a_n)}[y] = \frac{1}{4}$. Under this action set, $\mathcal{A}(w, A_n) := \{a_0, a^*\}$. Principal-preferred action selection then leads the principal to choose a_0, yielding her a payoff of $V(w, A^*) = \frac{1}{4} > \frac{1}{8}$. Hence, $V(w, \cdot)$ is not lower semicontinuous at A^*.[3]

To avoid lower semicontinuity issues, one might instead invoke the Kneser-Fan minimax theorem (e.g. Theorem 4.2 of Sion (1958)). This theorem requires, among other conditions, that $V(\tilde{w}, \cdot)$ is convexlike. But $V(\tilde{w}, \cdot)$ is not convexlike, as demonstrated by the following example.

Example 2. Suppose again that $A_0 = \{a_0\}$, with $E_{F(a_0)}[y] = \frac{1}{2}$ and $c(a_0) = \frac{1}{8}$, and $Y = \{0, 1\}$. V is *convexlike* in $\mathcal{S}$ if, for every $A, A' \in \mathcal{S}$ and every $0 \leq t \leq 1$, there is an $A'' \in \mathcal{S}$ such that

$$tV(\tilde{w}, A) + (1 - t)V(\tilde{w}, A') \geq V(\tilde{w}, A'') \quad \text{for all } \tilde{w} \in \Delta(\mathcal{W}).$$

To see that $V(\tilde{w}, \cdot)$ is not convexlike, consider $A := \{a_0, a\}$ and $A' := \{a_0, a'\}$, where $c(a) = c(a') = 0$ and $E_{F(a)}[y] = \frac{1}{4} < \frac{1}{3} = E_{F(a')}[y]$. Set $t = \frac{1}{2}$ and consider the class of

[3]To fix this issue, one might assume principal least-preferred action selection. But then $V(\cdot, A)$ would not be upper semicontinuous, another necessary condition for Sion's minimax theorem. Relatedly, an optimal deterministic contract need not exist.

linear deterministic contracts $w(y) = \alpha y$, where $0 \leq \alpha \leq \frac{3}{4}$. Then,

$$\frac{1}{2}V(w,A) + \frac{1}{2}V(w,A') = \begin{cases} (1-\alpha)\frac{7}{24} & \text{for } 0 \leq \alpha < \frac{1}{2} \\ (1-\alpha)\frac{10}{24} & \text{for } \frac{1}{2} \leq \alpha \leq \frac{3}{4}. \end{cases}$$

Now, towards contradiction, suppose that there existed an $A'' \in \mathcal{S}$ such that

$$V(w,A'') \leq \frac{1}{2}V(w,A) + \frac{1}{2}V(w,A')$$

for all $\alpha > 0$. Then, for every $\alpha = \frac{1}{2+\epsilon}$, where $\epsilon > 0$, there must exist an action $a''_\epsilon \in \mathcal{A}(w,A'')$ for which

$$E_{F(a''_\epsilon)}[y] \leq \frac{7}{24}. \tag{6}$$

Similarly, for $\alpha = \frac{3}{4}$, there must exist an action $a'' \in \mathcal{A}(w,A'')$, for which

$$E_{F(a'')}[y] \leq \frac{10}{24}, \tag{7}$$

and which satisfies

$$\frac{3}{4}E_{F(a'')}[y] - c(a'') > \frac{3}{4}E_{F(a_0)}[y] - c(a_0) \iff E_{F(a'')}[y] > \frac{8}{24} + \frac{4}{3}c(a''). \tag{8}$$

Moreover, since (8) and non-negativity of $c(a'')$ implies that $E_{F(a'')}[y] > \frac{7}{24}$, it must be that the agent strictly prefers a''_1 to a''_ϵ for each $\alpha = \frac{1}{2+\epsilon}$:

$$\frac{1}{2+\epsilon}E_{F(a''_\epsilon)}[y] - c(a''_\epsilon) > \frac{1}{2+\epsilon}E_{F(a'')}[y] - c(a'') \iff$$

$$E_{F(a'')}[y] - E_{F(a''_\epsilon)}[y] < (2+\epsilon)(c(a'') - c(a''_\epsilon)). \tag{9}$$

But, there do not exist values $(E_{F(a''_\epsilon)}[y], c(a''_\epsilon))_\epsilon$, each with $E_{F(a''_\epsilon)}[y] \geq 0$ and $c(a''_\epsilon) \geq 0$, and values $E_{F(a'')}[y] \geq 0$ and $c(a'') \geq 0$, that simultaneously satisfy (6), (7), (8), and (9) for all $\epsilon > 0$, the desired contradiction.[4]

To avoid both continuity and convexity issues, one might permit Nature to ran-

[4]To see why, notice that (9) is most relaxed subject to $c(a''_\epsilon) \geq 0$ and (6) when $c(a''_\epsilon) = 0$ and $E_{F(a''_\epsilon)} = \frac{7}{24}$ for all $\epsilon > 0$. But then it must both be that $E_{F(a'')}[y] < \frac{7}{24} + (2+\epsilon)c(a'')$ for all $\epsilon > 0$, i.e. $E_{F(a'')}[y] \leq \frac{7}{24} + 2c(a'')$, and $E_{F(a'')}[y] > \frac{8}{24} + \frac{4}{3}c(a'')$. To satisfy both inequalities, it must be that $c(a'') > \frac{1}{16}$, in which case $E_{F(a'')}[y] > \frac{10}{24}$, contradicting (7).

domize, i.e. extend her strategy space from $\mathcal{S}$ to $\Delta(\mathcal{S})$. Then, the minimax equality becomes

$$V_R^* = \sup_{\tilde{w}\in\Delta(\mathcal{W})} \inf_{\tilde{A}\in\Delta(\mathcal{S})} V(\tilde{w},\tilde{A}) = \inf_{\tilde{A}\in\Delta(\mathcal{S})} \sup_{\tilde{w}\in\Delta(\mathcal{W})} V(\tilde{w},\tilde{A}) := \overline{V}_{RR}, \tag{10}$$

where $V(\tilde{w},\tilde{A}) := E_{\tilde{w},\tilde{A}}[V(w,A)]$. However, establishing (10) is insufficient to establish $V_R^* > V_D^*$ because it need not be the case that $\overline{V}_{RR} \geq \overline{V}_D$. Put differently, (10) can hold even when $V_R^* = V_D^*$.

3.2 The Actual, Constructive Proof

To prove Theorem 1, I instead construct a random contract that strictly outperforms the best linear contract $w^*(y) = \alpha^* y$, $\alpha^* > 0$. For this purpose, consider a 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.[5]

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:

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

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

where

$$\begin{aligned}
\underline{V}(\tilde{w}_\epsilon) := \quad & \min_{a^*, a^*_\epsilon \in \Delta(Y)\times\mathbb{R}_+} \quad \frac{1}{2}(1-\alpha^*)E_{F(a^*)}[y] + \frac{1}{2}(1-(\alpha^*-\epsilon))E_{F(a^*_\epsilon)}[y] \\
& \text{subject to} \\
& [IC_{w^*}] \quad \alpha^* E_{F(a^*)}[y] - c(a^*) \geq \max_{a_0 \in A_0} \alpha^* E_{F(a_0)}[y] - c(a_0) \\
& [IC_{w^*_\epsilon}] \quad (\alpha^*-\epsilon)E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \geq \max_{a \in A_0} (\alpha^*-\epsilon)E_{F(a_0)}[y] - c(a_0) \\
& [IC_{w^* \to w^*_\epsilon}] \quad \alpha^* E_{F(a^*)}[y] - c(a^*) \geq \alpha^* E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \\
& [IC_{w^*_\epsilon \to w^*}] \quad (\alpha^*-\epsilon)E_{F(a^*_\epsilon)}[y] - c(a^*_\epsilon) \geq (\alpha^*-\epsilon)E_{F(a^*)}[y] - c(a^*).
\end{aligned} \tag{11}$$

Proof. See Appendix A.1. □

The solution to (11) provides a lower bound on Nature's worst-case response to the contract $\tilde{w}_\epsilon$. IC_{w^*} ensures that, relative to any known action, the agent prefers to take action a^* when receiving contract w^*. $IC_{w^* \to w^*_\epsilon}$ ensures that, relative to a^*_ϵ , the agent prefers to take action a^* when receiving contract w^*. Analogous statements hold for $IC_{w^*_\epsilon}$ and $IC_{w^*_\epsilon \to w^*}$.

I next establish properties that hold in any solution to (11).

Lemma 2

If $\epsilon > 0$ is sufficiently small, then the following properties hold in any solution to (11):

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 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 particular, at any solution to (11), $IC_{w^*_\epsilon \to w^*}$ binds. In addition, when $\epsilon > 0$ is sufficiently small, there is pooling: $c(a^*) = c(a^*_\epsilon) = 0$ and $E_{F(a^*)}[y] = E_{F(a^*_\epsilon)}[y]$. That is, 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^*. Moreover, the binding

constraint IC_{w^*}, pins down

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

where the bracketed expression corresponds to the tight lower bound on worst-case productivity under the optimal deterministic contract (see (2)). These observations immediately yield that the solution to (11) 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 now, 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.

4 Discussion

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 minimax payoff in the robust principal-agent problem of Carroll (2015). I proved that the principal does, in fact, achieve a strictly higher worst-case payoff by randomizing. Hence, restricting attention to the study of deterministic contracts is *with* loss of generality.

How should the optimal deterministic (linear) 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. 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 deterministic contracts thus depends crucially upon the principal's beliefs about the timing of the resolution of uncertainty.

In ongoing research, I study the optimality and structure of random contracts under varying beliefs that randomization can eliminate ambiguity aversion (preferences reflecting these attitudes have been axiomatized by Saito (2015)). It is my hope that this paper spurs related research at the intersection of the frontiers of decision theory and contract theory.

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

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

Because $\hat{w}^*$ and $\hat{w^*}_\epsilon$ 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

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

where

$$\hat{V}(w_\epsilon^*) := \frac{1}{2} \quad \min_{A \in \mathcal{S}} \quad (1-\alpha^*)E_{F(a^*)}[y] + (1-\alpha^*+\epsilon)E_{F(a_\epsilon^*)}[y]$$

$$\text{subject to}$$

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

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

To see why it suffices to replace $\hat{V}(w_\epsilon^*)$ with $\underline{V}(\tilde{w}_\epsilon)$, notice 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 has fewer inequality constraints. Therefore, its solution results in a (weakly) smaller payoff for the principal. It follows that $\underline{V}(\tilde{w}_\epsilon) = \hat{V}(w_\epsilon^*)$.

A.2 Proof of Lemma 2

I consider the properties of the solution to (11). I first prove that, in any solution, it must be that $c(a_\epsilon^*) = 0$. Towards contradiction, suppose that $c(a_\epsilon^*) > 0$ in some solution. I claim that the objective function can be strictly reduced if Nature replaces this action with the alternative action $\hat{a}$ with cost $c(\hat{a}) = 0$ and with distribution satisfying

$$E_{F(\hat{a})}[y] = E_{F(a_\epsilon^*)}[y] - \frac{c(a_\epsilon^*)}{\alpha^* - \epsilon} < E_{F(a_\epsilon^*)}[y].$$

Since the objective function is strictly increasing in $E_{F(a_\epsilon^*)}[y]$, it suffices to show that $\hat{a}$ is incentive compatible. By construction, $IC_{w_\epsilon^*}$ and $IC_{w_\epsilon^* \to w^*}$ are satisfied. As no change to a^* has been made, IC_{w^*} remains satisfied. Finally, $IC_{w^* \to w_\epsilon^*}$ is satisfied because

$$E_{F(a^*)}[y] - \frac{c(a^*)}{\alpha^*} \geq E_{F(a_\epsilon^*)}[y] - \frac{c(a_\epsilon^*)}{\alpha^*} > E_{F(a_\epsilon^*)}[y] - \frac{c(a_\epsilon^*)}{\alpha^* - \epsilon} = E_{F(\hat{a})}[y].$$

Given that $c(a_\epsilon^*) = 0$, it suffices to consider the following minimization problem:

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

subject to

$$[IC_{w^*}] \quad E_{F(a^*)}[y] \geq \frac{c(a^*)}{\alpha^*} + \max_{a_0 \in A_0}\left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha^*}\right)$$

$$[IC_{w^*_\epsilon}] \quad E_{F(a^*_\epsilon)}[y] \geq \max_{a \in A_0}\left(E_{F(a_0)}[y] - \frac{c(a_0)}{(\alpha^*-\epsilon)}\right)$$

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

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

I next observe that $IC_{w^* \to w^*_\epsilon}$ is implied by $IC_{w^*_\epsilon}$ and IC_{w^*}. To see why, subtract $IC_{w^*_\epsilon}$ from IC_{w^*} to obtain

$$E_{F(a^*)}[y] - E_{F(a^*_\epsilon)}[y] \geq \frac{c(a^*)}{\alpha^*} + U_0 - U_0^\epsilon,$$

where

$$U_0 := \max_{a_0 \in A_0}\left(E_{F(a_0)}[y] - \frac{c(a_0)}{\alpha^*}\right) \quad \text{and} \quad U_0^\epsilon := \max_{a \in A_0}\left(E_{F(a_0)}[y] - \frac{c(a_0)}{(\alpha^*-\epsilon)}\right).$$

Since $U_0 - U_0^\epsilon \geq 0$ for any $\epsilon \geq 0$, it follows that $IC_{w^* \to w^*_\epsilon}$ must hold if $IC_{w^*_\epsilon}$ and IC_{w^*} hold.

Dropping $IC_{w^* \to w^*_\epsilon}$, I now inspect the solution to

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

subject to

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

$$[IC_{w^*_\epsilon}] \quad E_{F(a^*_\epsilon)}[y] \geq U_0^\epsilon$$

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

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.

Eliminating $E_{F(a^*)}[y]$ from Nature's problem using the binding constraint yields

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

Now, consider a relaxed problem without $IC_{w^*_\epsilon}$. I solve this program for $\epsilon > 0$ small and show that its solution satisfies $IC_{w^*_\epsilon}$. In any solution to this relaxed problem, the remaining constraint $IC_{w^*_\epsilon \to w^*}$ must bind. If not, then Nature could reduce $E_{F(a^*_\epsilon)}[y]$ by a small amount and strictly reduce the objective function.

Substituting the binding constraint $IC_{w^*_\epsilon \to w^*}$ into the objective function, I solve for the optimal value of $c(a^*)$:

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

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

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

If $\epsilon > 0$ is sufficiently small, the inequality is satisfied (the right-hand approaches zero and the left-hand side approaches a strictly positive number). Hence, it is optimal to make $c(a^*)$ as small as possible, i.e. set $c(a^*) = 0$.

In summary, when $\epsilon > 0$ is sufficiently small, $c(a^*) = c(a^*_\epsilon) = 0$ and both IC_{w^*} and $IC_{w^*_\epsilon \to w^*}$ bind in the solution to the relaxed problem. IC_{w^*} and $IC_{w^*_\epsilon \to w^*}$ binding yields $E_{F(a^*)}[y] = U_0 = E_{F(a^*_\epsilon)}[y]$. Hence, $IC_{w^*_\epsilon}$ is also satisfied:

$$E_{F(a^*_\epsilon)}[y] = U_0 \geq U_0^\epsilon.$$

It follows that the solution to the relaxed problem solves Nature's constrained problem and all properties stated in the Lemma have been established.

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.

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

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.