Online Appendix to Optimal Technology Design: Risk-averse Agent

Daniel Garrett, George Georgiadis, Alex Smolin, and Balázs Szentes

January 2023

Abstract

This appendix solves the agent's project design problem when the agent is risk averse, and where the agent's risk preferences are known to the principal. We build on several results provided in the main document. We show that the agent still finds it optimal to choose a binary project, and we characterize the optimal binary project.

A Risk-averse agent

We now study the natural possibility that the agent is risk averse, with a concave utility over payments. In particular, while the principal is still risk neutral and has the same preferences as in the paper, the agent has a payoff $v(w) - c(F)$ where w is the payment, v is a concave utility function, c represents the agent's technology (i.e., his cost function), and F is the agent's choice of output distribution.

Establishing the optimality of a binary project for the agent is more challenging in this environment for the following reason. The idea of the proof of Proposition 1 for a risk-neutral agent was that garbling output to determine binary output distributions with the same mean makes the agent more difficult to incentivize, so the agent receives a (weakly) higher expected payment to generate the same expected output. Using this idea, we showed that, starting with an arbitrary project c^*, it is possible to find a binary project in which the principal implements the same expected output and the agent is better off. With a risk-averse agent, garbling output still increases the expected payments that must be made to incentivize the agent. However, because the associated payments may be more risky, risk aversion could leave the agent worse off than without the garbling. The argument required to establish the optimality of binary technologies is therefore more complicated. It involves first studying optimal binary projects, and then combining the insights from this analysis with arguments that are similar in spirit to the proof of Proposition 1.

Given the structure of our arguments, we begin our analysis in Section A.2 below by characterizing optimal binary pojects, and determining the relationship between maximum agent payoffs and the profits that are attainable by the principal. Section A.3 then makes use of this analysis in establishing that binary projects are optimal.

We recap at the end of Section A.3 how this online appendix demonstrates the conclusions reported in Section 4 of the main text, under the heading "Risk Aversion". First, the arguments in Section A.3 will show that, for each project c^* and any equilibrium (w^*, F^*) in c^*, there is a binary project and equilibrium in that project in which (a) the agent generates output one with certainty, and (b) the outcome Pareto dominates the outcome in the original equilibrium. Moreover, if F^* does not put all its mass on output one, we will see that the Pareto improvement is strict, a conclusion which in turn implies that binary projects are optimal. In Section A.2, we will see that, in an optimal binary project, where the principal has some equilibrium payoff π^*, the cost function is defined by a cost of zero for generating output one with probability π^* and a marginal cost $v\left(1-\pi^*/\mu\right)$ for probabilities μ of ouput one with $\mu > \pi^*$.

A.1 Preliminaries

The setting is the same as in the main document, except we generalize to allow that the agent's payoff is given by $v\left(w\right) - c\left(F\right)$, where $w \geq 0$ is any payment to the agent, F is any distribution on $[0,1]$, and $v : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing, weakly concave and continuously differentiable function. For any $w \geq 0$, we refer to $v\left(w\right)$ as the agent's "felicity of consumption" from consumption w. We adopt the normalization that $v\left(0\right) = 0$. The inverse of v is given by the function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$. A central objective will be to establish the optimality of binary projects for a risk-averse agent, as follows immediately from Proposition A1 below. We characterize optimal binary projects in the process.

Due to the agent's risk aversion, lotteries over payments are no longer payoff equivalent to deterministic payments with the same mean. In fact, the principal always finds it optimal to choose payments that are deterministic conditional on output. This is because the agent's incentives are determined by the expected utility conditional on output, and a given level of expected utility is most cheaply provided (for the risk-neutral principal) by a deterministic payment. It is then easy to see that considering only equilibria in which the principal offers deterministic payments comes at no loss in the agent's project design problem.

In spite of the previous observation, the proof below makes use of payment schemes that are random conditional on output. We find this convenient in order to mimic the logic of Lemma 1 in the main document. Lemma 1 considered payments offered in a possibly non-binary original project c^* that were linear in output, thus rendering the agent's payoff linear in output. Given the agent's risk aversion, such linearity of the expected payoff in output can be obtained by considering instead payments that are randomized over a binary support, with the probability of the higher payment linear in output.

Let us then introduce the notation for random payments up front, while the details of how they are used in the argument appear in Section A.3 below. For any output $x \in [0, 1]$, we let the random payment conditional on x be determined by a cdf $G_x : \mathbb{R}_+ \to [0, 1]$, where $G_x(w)$ is the probability of a payment no greater than w conditional on output realization x. A collection of distributions is denoted $\mathcal{G} = (G_x)_{x\in[0,1]}$, which is the representation of the payment schedule when random payments conditional on output are permitted.

The notation introduced in Section 2 of the main document updates straightforwardly to random payments and a risk-averse agent. The agent's expected payoff in project c, when choosing output distribution F, given payments $\mathcal{G}$, is

$$U(c, \mathcal{G}, F) = \int_0^1 \int_0^\infty v(\widetilde{w})\, dG_x(\widetilde{w})\, dF(x) - c(F).$$

The principal's expected payoff from output distribution F and payments $\mathcal{G}$ is

$$\Pi(\mathcal{G}, F) = \int_0^1 \left[x - \int_0^\infty \widetilde{w} dG_x(\widetilde{w})\right] dF(x).$$

The value of an agent in project c for payments $\mathcal{G}$ is

$$u\left(c,\mathcal{G}\right)=\sup_{F\in\mathcal{F}}U\left(c,\mathcal{G},F\right).$$

We can then define $\mathbf{F}^{c,\mathcal{G}}$ by $\left(F_n\right)\in\mathbf{F}^{c,\mathcal{G}}$ if and only if $\lim_{n\to\infty}U\left(c,\mathcal{G},F_n\right)=u\left(c,\mathcal{G}\right)$. The principal's value in project c given payment policy $\mathcal{G}$ is then

$$\pi\left(c,\mathcal{G}\right)\equiv\sup\left\{\limsup_{n\to\infty}\Pi\left(\mathcal{G},F_n\right)\quad:\quad\left(F_n\right)\in\mathbf{F}^{c,\mathcal{G}}\right\}.$$

If $w:[0,1]\to\mathbb{R}_+$ is a deterministic payment schedule, we continue to use the notation of the main document and write w in place of $\mathcal{G}$. So, for instance, the players' values in project c, given deterministic payment schedule w, are $u\left(c,w\right)$ and $\pi\left(c,w\right)$.

A.2 Analysis of binary projects

Similar to Section 3.2 of the main document, we aim at a characterization of the players' payoffs in equilibria of binary projects. We begin by determining the highest payoff the agent can obtain in an outcome in which the principal earns payoff π. That is, we consider all binary projects in which there is an equilibrium of the subgame following project selection where the principal earns payoff π, and ask what is the highest payoff that can be obtained by the agent in such an equilibrium?

Lemma A1. *Conditional on any payoff $\pi\in[0,1]$ for the principal, the agent can obtain a maximal payoff, in any binary project, of*

$$\int_0^{v(1-\pi)}\frac{\pi}{1-\gamma\left(z\right)}dz=v\left(1-\pi\right)-\int_\pi^1 v\left(1-\frac{\pi}{z}\right)dz.$$

Proof. Consider binary projects in which there is an equilibrium involving the principal offering bonus $\widehat{b}\in[0,1]$ and the agent choosing probability of output one equal to $\widehat{\mu}\in[0,1]$. Let us now determine an upper bound on the payoff the agent can obtain in such an equilibrium.

Let $C:[0,1]\to\mathbb{R}_+$ be a binary project specifying the cost of obtaining each probability of output one. Write the agent's value when the bonus delivers felicity of consumption ω as

$$u(\omega) = \sup_{\mu \in [0,1]} \{\omega\mu - C(\mu)\}. \tag{1}$$

Note that the function u is identical to that for the risk-neutral case (where ω is equal to the bonus).

Let $\Gamma(\omega)$ denote the values μ such that there is a sequence (μ_n) with $\mu_n \to \mu$ and $\omega\mu_n - C(\mu_n) \to u(\omega)$. Take $\bar{\mu}(\omega) = \max \Gamma(\omega)$. From the same arguments as in the proof of Proposition 2 in the main document we have, for any $\omega \geq 0$,

$$u(\omega) = u(0) + \int_0^{\omega} \bar{\mu}(z)\, dz.$$

As in the main document, note that $u(0) \leq 0$; i.e., the agent cannot obtain a strictly positive payoff if the bonus is zero. Also, $u(\cdot)$ is non-decreasing and convex.

Denote $\widehat{\pi} = \widehat{\mu}\left(1 - \widehat{b}\right)$. Incentive compatibility of the principal offering bonus $\widehat{b}$ requires that, for all $z \geq 0$,

$$\begin{aligned} \widehat{\pi} &\geq \bar{\mu}(z)(1 - \gamma(z)) \\ &= u'_{+}(z)(1 - \gamma(z)). \end{aligned} \tag{2}$$

(Note that equality of $\bar{\mu}$ and u'_{+} is established in Lemma 4 of the main document.)

Now let us determine an upper bound on the agent's payoff, across binary projects C, that can occur for an equilibrium in which the principal offers bonus $\widehat{b}$ and the agent achieves output one with probability $\widehat{\mu}$. Letting $\widehat{\omega} = v(\widehat{b})$, we can write the agent's equilibrium payoff as

$$u(\widehat{\omega}) = u(0) + \int_0^{\widehat{\omega}} u'_{+}(z)\, dz.$$

Consider maximizing this value by choice of convex function $u : \mathbb{R}_{+} \to \mathbb{R}$ subject to the constraints (i) $u(0) \leq 0$, and (ii) $\widehat{\pi} \geq u'_{+}(z)(1 - \gamma(z))$ for all z. The first requirement reflects the above observation that the agent cannot obtain a positive payoff if the bonus is zero. The second condition is a re-statement of Condition (2). Any solution to this problem involves $u(0) = 0$ and

$$u'_{+}(z) = \frac{\widehat{\pi}}{1 - \gamma(z)}$$

for all $z \in [0, \widehat{\omega})$. In other words, Constraint (ii) above holds with equality over $z \in [0, \widehat{\omega})$. The optimal choice of u in the above problem therefore satisfies

$$u(\omega) = \int_0^{\omega} \frac{\widehat{\pi}}{1 - \gamma(z)} dz$$

on $[0, \widehat{\omega}]$.[1]

If $\widehat{\mu} = 0$, then both principal and agent must earn payoff zero. So suppose that $\widehat{\mu} > 0$. We have $\widehat{b} = 1 - \widehat{\pi}/\widehat{\mu}$ and so $\widehat{\omega} = v(1 - \widehat{\pi}/\widehat{\mu})$. An upper bound on the agent's payoff can then be written as

$$\int_0^{v\left(1 - \frac{\widehat{\pi}}{\widehat{\mu}}\right)} \frac{\widehat{\pi}}{1 - \gamma(z)} dz.$$

This is maximized by taking $\widehat{\mu} = 1$. It evaluates to zero if $\widehat{\pi} \in \{0, 1\}$ (i.e., the agent must obtain a payoff zero for these values of the principal's payoff), so suppose that $\widehat{\pi} \in (0, 1)$.

Let us now demonstrate that the agent can obtain the above payoff. Consider the binary project

$$C(\mu; \widehat{\pi}) = \begin{cases} \int_{\widehat{\pi}}^{\mu} v\left(1 - \frac{\widehat{\pi}}{\widetilde{\mu}}\right) d\widetilde{\mu} & \text{if } \mu \in [\widehat{\pi}, 1] \\ +\infty & \text{otherwise.} \end{cases}$$

The function $C(\cdot; \widehat{\pi})$ is strictly convex on $[\widehat{\pi}, 1]$. If the principal offers a bonus $b \in [0, 1 - \widehat{\pi}]$, generating felicity $\omega = v(b)$, the agent solves $\max_{\mu \in [0,1]} \{\mu\omega - C(\mu)\}$. The solution, $\mu^*(\omega)$ is unique and characterized by the first-order condition

$$\omega = C'(\mu^*(\omega); \widehat{\pi}) = v\left(1 - \frac{\widehat{\pi}}{\mu^*(\omega)}\right).$$

Hence,

$$\mu^*(\omega) = \frac{\widehat{\pi}}{1 - \gamma(\omega)}.$$

The principal obtains payoff $\widehat{\pi}$ from offering every bonus in $[0, 1 - \widehat{\pi}]$. So there is indeed an

[1] Note that the integrand in the expression for $u(\omega)$ remains bounded, so $u(\omega)$ is necessarily well-defined and finite. To see this, we only need to consider the case where $\widehat{\omega}$ takes its highest value, i.e. $\widehat{\omega} = v(1)$. In this case, $\widehat{\pi} = 0$ and so we immediately conclude $u(\omega) = 0$ on $[0, \widehat{\omega}]$.

equilibrium in $C(\cdot;\widehat{\pi})$ with the principal offering bonus $1-\widehat{\pi}$ and the agent choosing output one with certainty. The agent's value in this project, using the definition in (1), satisfies $u(0)=0$ and $u'_{+}(\omega)=\mu^{*}(\omega)$ for all $\omega\in[0,v(1-\widehat{\pi})]$. Hence, the agent indeed obtains a payoff

$$\int_{0}^{v(1-\widehat{\pi})}\frac{\widehat{\pi}}{1-\gamma(z)}dz.$$

This is also equal to the felicity of consumption $1-\widehat{\pi}$ less the agent's cost; i.e., $v(1-\widehat{\pi})-C(1;\widehat{\pi})$. This yields the expressions in the lemma. *QED*

Next consider the cost the agent incurs in an equilibrium of a binary project that yields the highest possible agent payoff, given principal expected payoffs π (this value of the agent's payoff is obtained in the previous result). We show that this cost is strictly convex in the principal's profits, a fact that will be important below.

Lemma A2. *The function $h(\pi)\equiv\int_{\pi}^{1}v(1-\pi/z)\,dz$ is strictly convex over $\pi\in[0,1]$.*

Proof. Let $\pi\in(0,1)$ and consider the change of variables

$$x=1-\frac{\pi}{z}.$$

Note that

$$\frac{dx}{dz}=\frac{\pi}{z^{2}}=\frac{(1-x)^{2}}{\pi}.$$

Therefore, we have

$$h(\pi)=\pi\int_{0}^{1-\pi}\frac{v(x)}{(1-x)^{2}}dx.$$

Differentiating with respect to π yields

$$h'(\pi)=\int_{0}^{1-\pi}\frac{v(x)}{(1-x)^{2}}dx-\frac{v(1-\pi)}{\pi}$$

and

$$h''(\pi)=\frac{v'(1-\pi)}{\pi}>0.$$

The result follows. *QED*

A.3 Optimality of binary projects

We now state our main result, which implies the optimality of binary projects. For this purpose, recall that B_x denotes a binary distribution with probability mass x on output 1.

Proposition A1. *In any optimal outcome for the agent, (c^*, w^*, F^*), we have $F^* = B_1$.*

The rest of this section proves Proposition A1. Consider an agent-optimal outcome (c^*, w^*, F^*) and suppose for a contradiction that $F^* \neq B_1$. Note that, because the agent can secure a strictly positive payoff (as demonstrated in the previous section), we have $\mu_{F^*} > 0$ as well as $\mathbb{E}_{F^*}[w^*] > 0$.

We reach a contradiction by determining a project (and equilibrium in that project) in which the agent has a strictly higher payoff than for the outcome (c^*, w^*, F^*). Our first aim is to construct a binary project $\check{c}$ in which the principal offers bonus

$$b^* = \frac{\mathbb{E}_{F^*}[w^*]}{\mu_{F^*}}$$

to implement B_1 and earns a profit $1 - b^* = (\mu_{F^*} - \mu_{F^*} b^*)/\mu_{F^*} = (\mu_{F^*} - \mathbb{E}_{F^*}[w^*])/\mu_{F^*}$ (we complete this task by Lemma A5 below). That is, the principal's profit is $1/\mu_{F^*}$ times the profit in the original outcome, and the agent's payment is also $1/\mu_{F^*}$ times the expected payment in the original. Note however that, because possibly $\check{c}(B_1) > c^*(B_1)$, we are unable to guarantee that the agent be better off in the binary project $\check{c}$. So the project $\check{c}$ will need to be further modified.

As with the proof of Proposition 1 in the main document, our first step is to define a binary project

$$\widehat{c}(B_\mu) = \begin{cases} \inf\{c^*(F) : \mu_F = \mu\} & \text{if } \mu < \mu_{F^*}. \\ \infty & \text{otherwise.} \end{cases}$$

As explained in the main document, in binary projects, the players view equivalently a payment schedule that pays a bonus b only for output one and the payment schedule w_b (with $w_b(x) = bx$ for $x \in [0,1]$, as defined in the main document). Different to the main document, we introduce a particular random payment policy, denoted $\mathcal{G}^b$. The policy $\mathcal{G}^b$ specifies, for each output realization $x \in [0,1]$, the payment distribution $G^b_x(w) = 1 - x + x\mathbf{1}_{w \geq b}(w)$

where $\mathbf{1}_{w \geq b}$ is the indicator function that takes value 1 when $w \geq b$ and zero otherwise. That is, G_x^b is the distribution that puts mass $1-x$ on payment zero and mass x on payment b. Note that the expected payment to the agent given policy $\mathcal{G}^b$ depends only on the mean of output: the expected payment when mean output is μ is equal to $b\mu$.

We now provide a result that is analogous to Lemma 1 in the main document.

Lemma A3. *For all* $b \in [0,1]$, $\pi(\widehat{c}, w_b) \leq \pi\left(c^*, \mathcal{G}^b\right)$.

Proof. By the definition of $u(\widehat{c}, w_b)$ and $\pi(\widehat{c}, w_b)$, there is a sequence (μ_n) such that

$$\mu_n v(b) - \widehat{c}(B_{\mu_n}) \to u(\widehat{c}, w_b)$$

and

$$\Pi(w_b, B_{\mu_n}) = \mu_n(1-b) \to \pi(\widehat{c}, w_b).$$

For each $k \in \mathbb{N}$, there exists n_k such that

$$\mu_{n_k} v(b) - \widehat{c}\left(B_{\mu_{n_k}}\right) + \frac{1}{k} \geq \mu v(b) - \widehat{c}(B_\mu)$$

for all $\mu \in [0,1]$. Therefore, for all k, and all $\mu \in [0, \mu_{F^*})$,

$$\mu_{n_k} v(b) - \inf\{c^*(F) : \mu_F = \mu_{n_k}\} + \frac{1}{k} \geq \mu v(b) - \inf\{c^*(F) : \mu_F = \mu\}.$$

Hence, there is a sequence (F_{n_k}) with means μ_{n_k} (for each k) such that, for every F with mean in $[0, \mu_{F^*})$,

$$\mu_{n_k} v(b) - c^*(F_{n_k}) + \frac{2}{k} \geq \mu_F v(b) - c^*(F). \tag{3}$$

There are then two cases. The first is where the inequality (3) holds for all k and *all* F. In this case,

$$\lim_{k\to\infty} U\left(c^*, \mathcal{G}^b, F_{n_k}\right) = u\left(c^*, \mathcal{G}^b\right)$$

and hence

$$\pi\left(c^*, \mathcal{G}^b\right) \geq \lim_{k\to\infty} \Pi\left(\mathcal{G}^b, F_{n_k}\right) = \lim_{k\to\infty} \Pi\left(w_b, B_{\mu_{n_k}}\right) = \pi(\widehat{c}, w_b)$$

as desired. In the second case, the inequality (3) does not hold for some k and F, which implies

$$u\left(c^*,\mathcal{G}^b\right)=\sup\left\{\mu_F v\left(b\right)-c^*\left(F\right):F\in\mathcal{F}\right\}>\sup\left\{\mu_F v\left(b\right)-c^*\left(F\right):F\in\mathcal{F},\mu_F<\mu_{F^*}\right\}.$$

This means that there is a sequence of distributions along which the agent's payoff converges to $u\left(c^*,\mathcal{G}^b\right)$ and for which every distribution has mean at least μ_{F^*}. We therefore have

$$\pi^*\left(c^*,\mathcal{G}^b\right)\geq\mu_{F^*}\left(1-b\right)\geq\pi\left(\widehat{c},w_b\right),$$

where the second inequality follows because any distribution with mean at least μ_{F^*} is assigned an infinite cost in project $\widehat{c}$. *QED*

Our next goal is to define a binary project $\widetilde{c}$ and an equilibrium in $\widetilde{c}$ in which the principal offers the bonus payment b^*, and the agent chooses distribution $B_{\mu_{F^*}}$. Note that the principal's profit is then the same as in the original equilibrium: $\Pi\left(w_{b^*},B_{\mu_{F^*}}\right)=\Pi\left(w^*,F^*\right)=\mu_{F^*}\left(1-b^*\right)$.

Let then $\bar{c}$ be determined by

$$\mu_{F^*}v\left(b^*\right)-\bar{c}=\sup_{\mu\in[0,1]}\left\{\mu v\left(b^*\right)-\widehat{c}\left(B_\mu\right)\right\}.$$

Define the binary project $\widetilde{c}$ by

$$\widetilde{c}\left(F\right)=\begin{cases}\bar{c} & \text{if } F=B_{\mu_{F^*}}\\ \widehat{c}\left(F\right) & \text{if } F\neq B_{\mu_{F^*}}.\end{cases}$$

Note that the agent has a best response in project $\widetilde{c}$ to bonus b^* equal to the distribution $B_{\mu_{F^*}}$.

As in the case of a risk-neutral agent, we can show that $\bar{c}\leq c^*\left(F^*\right)$. Suppose for a contradiction that $\bar{c} > c^* (F^*)$. Then

$$\begin{aligned}
\mu_{F^*} v (b^*) - c^* (F^*) &> \mu_{F^*} v (b^*) - \bar{c} \\
&= \sup \{\mu v (b^*) - \widehat{c} (B_\mu) : \mu \in [0, 1]\} \\
&= \sup \{\mu_F v (b^*) - c^* (F) : F \in \mathcal{F}, \mu_F < \mu_{F^*}\} .
\end{aligned}$$

By continuity of v, there is then $b < b^*$ such that

$$\mu_{F^*} v (b) - c^* (F^*) > \sup \{\mu_F v (b) - c^* (F) : F \in \mathcal{F}, \mu_F < \mu_{F^*}\} .$$

This means that, if the principal offers payment schedule $\mathcal{G}^b$ in project c^*, the principal's value must be at least $\mu_{F^*} (1 - b) > \mu_{F^*} (1 - b^*) = \Pi (w^*, F^*)$. This contradicts the incentive compatibility of the payment schedule w^* in c^*.

We now want to show that $\left(w_{b^*}, B_{\mu_{F^*}}\right)$ is an equilibrium of project $\widetilde{c}$. To do so, we will rely on the following lemma.

Lemma A4. *For all $b \in [0, b^*)$, $\pi (\widetilde{c}, w_b) = \pi (\widehat{c}, w_b)$.*

Proof. The argument is identical to Lemma 2 in the main document, after noting that the agent has a felicity $v (b)$ (rather than b) when receiving bonus payment b. *QED*

Now let us show that $\left(w_{b^*}, B_{\mu_{F^*}}\right)$ is an equilibrium of project $\widetilde{c}$. We already saw that $B_{\mu_{F^*}}$ is incentive compatible for the agent in subgame $(\widetilde{c}, w_{b^*})$ (by choice of its cost $\bar{c}$). So we need to show that w_{b^*} is incentive compatible in $\widetilde{c}$. It is immediate that the principal does not want to deviate to a bonus $b > b^*$ (because the principal cannot attain expected output higher than μ_{F^*}; see also the argument in the main document). If $b < b^*$, then

$$\pi (\widetilde{c}, w_b) = \pi (\widehat{c}, w_b) \leq \pi \left(c^*, \mathcal{G}^b\right) \leq \Pi (w^*, F^*) = \Pi \left(w_{b^*}, B_{\mu_{F^*}}\right) .$$

The first equality follows by Lemma A4. The first inequality follows by Lemma A3. The second inequality follows because w^* is incentive compatible for the principal in c^*. The final equality follows by definition of b^*. Thus, the principal does not gain by deviating to $b < b^*$.

As we observed, the principal obtains the same payoff in outcome $\left(\widetilde{c}, w_{b^*}, B_{\mu_{F^*}}\right)$ as in the outcome (c^*, w^*, F^*). We have been unable, however, to determine whether the agent is better off in $\left(\widetilde{c}, w_{b^*}, B_{\mu_{F^*}}\right)$. Although the expected payment is the same in both outcomes, we have not ruled out that the agent could earn a lower payoff in outcome $\left(\widetilde{c}, w_{b^*}, B_{\mu_{F^*}}\right)$ due to the uncertainty in payments (i.e., due to worse insurance). Therefore, we make further modifications to the project.

First, we determine the project $\check{c}$ mentioned above, together with an equilibrium in which the agent chooses distribution B_1 and payoffs are those in the equilibrium $\left(w_{b^*}, B_{\mu_{F^*}}\right)$ of project $\widetilde{c}$, multiplied by $1/\mu_{F^*}$. This project is defined by $\check{c}\left(B_\mu\right) = \widetilde{c}\left(B_{\mu\mu_{F^*}}\right)/\mu_{F^*}$ for all $\mu \in [0,1]$. We show the following.

Lemma A5. *For all $b \geq 0$, $\pi\left(\check{c}, w_b\right) = \pi\left(\widetilde{c}, w_b\right)/\mu_{F^*}$.*

Proof. We first show $\pi\left(\check{c}, w_b\right) \geq \pi\left(\widetilde{c}, w_b\right)/\mu_{F^*}$ for any $b \geq 0$. Fix any such b and suppose that (μ_n) is a sequence for which $U\left(\widetilde{c}, w_b, B_{\mu_n}\right) \to u\left(\widetilde{c}, w_b\right)$ and $\Pi\left(w_b, B_{\mu_n}\right) \to \pi\left(\widetilde{c}, w_b\right)$. Then there is a subsequence (μ_{n_k}) such that, for all k and all $\mu \in [0,1]$,

$$\mu_{n_k} v\left(b\right) - \widetilde{c}\left(B_{\mu_{n_k}}\right) + \frac{1}{k} \geq \mu v\left(b\right) - \widetilde{c}\left(B_\mu\right).$$

Then, for all $\mu \in [0, \mu_{F^*}]$,

$$\frac{\mu_{n_k}}{\mu_{F^*}} v\left(b\right) - \frac{1}{\mu_{F^*}} \widetilde{c}\left(B_{\mu_{n_k}}\right) + \frac{1}{\mu_{F^*} k} \geq \frac{\mu}{\mu_{F^*}} v\left(b\right) - \frac{1}{\mu_{F^*}} \widetilde{c}\left(B_\mu\right).$$

Hence, for all $\mu \in [0,1]$,

$$\frac{\mu_{n_k}}{\mu_{F^*}} v\left(b\right) - \check{c}\left(B_{\frac{\mu_{n_k}}{\mu_{F^*}}}\right) + \frac{1}{\mu_{F^*} k} \geq \mu v\left(b\right) - \check{c}\left(B_\mu\right).$$

This implies that $U\left(\check{c}, w_b, B_{\frac{\mu_{n_k}}{\mu_{F^*}}}\right) \to u\left(\check{c}, w_b\right)$ and $\Pi\left(w_b, B_{\mu_{n_k}/\mu_{F^*}}\right) \to \pi\left(\widetilde{c}, w_b\right)/\mu_{F^*}$. This establishes the claim.

We now show that $\pi\left(\check{c}, w_b\right) \leq \pi\left(\widetilde{c}, w_b\right)/\mu_{F^*}$. Suppose that (μ_n) is a sequence for which $U\left(\check{c}, w_b, B_{\mu_n}\right) \to u\left(\check{c}, w_b\right)$ and $\Pi\left(w_b, B_{\mu_n}\right) \to \pi\left(\check{c}, w_b\right)$. Then there is a subsequence (μ_{n_k})

such that, for all k and all $\mu \in [0,1]$,

$$\mu_{n_k} v\left(b\right) - \frac{\widetilde{c}\left(B_{\mu_{n_k}\mu_{F^*}}\right)}{\mu_{F^*}} + \frac{1}{k} \geq \mu v\left(b\right) - \frac{\widetilde{c}\left(B_{\mu\mu_{F^*}}\right)}{\mu_{F^*}}.$$

Then, for all k and all $\mu \in [0,1]$,

$$\mu_{F^*}\mu_{n_k} v\left(b\right) - \widetilde{c}\left(B_{\mu_{n_k}\mu_{F^*}}\right) + \frac{\mu_{F^*}}{k} \geq \mu_{F^*}\mu v\left(b\right) - \widetilde{c}\left(B_{\mu\mu_{F^*}}\right).$$

Letting, for each k, $\mu'_{n_k} = \mu_{F^*}\mu_{n_k}$, we have

$$\mu'_{n_k} v\left(b\right) - \widetilde{c}\left(B_{\mu'_{n_k}}\right) + \frac{\mu_{F^*}}{k} \geq \mu v\left(b\right) - \widetilde{c}\left(B_{\mu}\right)$$

for all $\mu \in [0, \mu_{F^*}]$, and hence all $\mu \in [0,1]$. Therefore, $U\left(\widetilde{c}, w_b, B_{\mu_{F^*}\mu_{n_k}}\right) \to u\left(\widetilde{c}, w_b\right)$, while $\Pi\left(w_b, B_{\mu_{F^*}\mu_{n_k}}\right) \to \mu_{F^*}\pi\left(\check{c}, w_b\right)$, which establishes the claim. *QED*

Lemma A5 implies that the principal's incentives to offer different bonuses in project $\check{c}$ are the same as in $\widetilde{c}$. The optimal bonus for the agent that is incentive-compatible for the principal is b^*, with the agent best responding in subgame $(\check{c}, b^*)$ with the distribution B_1. To see this, recall that $(b^*, B_{\mu_{F^*}})$ is an equilibrium in project $\widetilde{c}$. Therefore, for all $\mu \in [0,1]$,

$$\mu_{F^*} v\left(b^*\right) - \widetilde{c}\left(B_{\mu_{F^*}}\right) \geq \mu v\left(b^*\right) - \widetilde{c}\left(B_{\mu}\right).$$

The claim follows because this is equivalent to the statement that, for all $\mu \in [0,1]$,

$$v\left(b^*\right) - \check{c}\left(B_1\right) \geq \mu v\left(b^*\right) - \check{c}\left(B_{\mu}\right).$$

We have established then that $\check{c}$ is a binary project in which the agent (in the agent-optimal equilibrium) receives a payment equal to $1/\mu_{F^*}$ times the expected payment $\mathbb{E}_{F^*}\left[w^*\right]$ of the original equilibrium and achieves output one with certainty. The principal's profits are $1 - b^*$. Consider now an *optimal* binary project for the agent in which the principal earns profits $1 - b^*$. Recalling the analysis in the previous section, there is an optimal binary project where the principal pays b^* and the agent achieves output one with probability one.

Recalling the definition in Lemma A2, the agent's cost is $h\left(1-b^*\right)$. The agent's payoff satisfies

$$v\left(b^*\right)-h\left(1-b^*\right)\geq v\left(b^*\right)-\check{c}\left(B_1\right)\geq v\left(b^*\right)-\frac{c^*\left(F^*\right)}{\mu_{F^*}}.$$

The first inequality follows because $\check{c}$ is a (not-necessarily optimal) binary project. The second inequality follows by construction of $\check{c}\left(B_1\right)$.

We can conclude that

$$h\left(1-b^*\right)\leq\frac{c^*\left(F^*\right)}{\mu_{F^*}}.$$

Because h is strictly convex, and because $h\left(1\right)=0$, we have

$$h\left(1-\mu_{F^*}b^*\right)<c^*\left(F^*\right).$$

Therefore,

$$v\left(\mu_{F^*}b^*\right)-h\left(1-\mu_{F^*}b^*\right)>v\left(\mu_{F^*}b^*\right)-c^*\left(F^*\right)\geq U\left(c^*,w^*,F^*\right),$$

where the second inequality follows by Jensen's inequality and concavity of v (as well as the observation that the expected payment determined by distribution F^* and payment schedule w^* is $\mu_{F^*}b^*$). The left-hand side represents the agent's payoff in an agent-optimal binary project in which the principal obtains payoff $1-\mu_{F^*}b^*$ (as determined in the previous section). The right-hand side is the agent's expected payoff in the original project. That the agent does better in the aforementioned binary project contradicts the optimality for the agent of the original, as desired. This completes the proof of Proposition A1.

Let us conclude by relating the claims in this online appendix to those made at the end of Section 4 (under the heading "Risk Aversion"). First note that Proposition A1 implies that binary projects are optimal. While the claim in the proposition only states that it is optimal for the agent to choose a project where the principal implements distribution B_1, recall from the discussion in Appendix B (under the heading "Uniqueness beyond binary projects") that any such project can be converted to a binary one. The proof of Proposition A1 showed in particular that, for an outcome $\left(c^*,w^*,F^*\right)$ where $\left(w^*,F^*\right)$ is an equilibrium of c^*, and where

$F^* \neq B_1$, there exists a binary project and equilibrium of that project which represents a strict Pareto improvement for both players. In particular, we constructed a binary outcome where the principal obtains a payoff $1 - \mu_{F^*} b^* = 1 - \mathbb{E}_{F^*}[w^*] > \mu_{F^*} - \mathbb{E}_{F^*}[w^*]$. For non-binary projects where B_1 is implemented in equilibrium, the previous observation that the project can be converted to a binary project applies. In this case, the equilibrium payoffs of the players are unaffected by the conversion. Finally, the claim that the marginal costs of probabilities μ above π^* are given by $v\left(1 - \frac{\pi^*}{\mu}\right)$ in an optimal project follows from the specification of $C(\mu; \pi^*)$ in the proof of Lemma A1.