A/B Contracts: Online Appendix

George Georgiadis and Michael Powell[29]

A Additional Results

Here, we present additional results and illustrations from our empirical exploration using the data of DellaVigna and Pope (2018).

A Additional Results on the Optimal Adjustments

Figure 6 presents disaggregated data for the optimal adjustments from every homogeneous A/B test. To be specific, the top panel illustrates, for each homogeneous A/B test and each status quo treatment $C \in \{2, \ldots, 7\}$, the maximum available gains, as well as the realized gains. The bottom panel illustrates the difference, in percentage terms, between agent's expected utility under the test optimal contract, and that under the status quo contract w^C.[30] The ratio $u^{AB}(w^C)/u(w^C)$ ranges between 0.97 and 1.07, and it is greater than one—as desired—in 31 out of the 42 cases. Two instance deserve further discussion: When the A/B test comprises treatments $(6, 7)$, and the status quo contract $C = 4$ and $C = 5$, the realized gains are greater than the maximum available gains. This occurs because in those instances, the test-optimal contract gives the agent approximately 3% fewer utils in expectation than the benchmark-optimal contract, as illustrated in the lower panel of the figure.

Figure 7 illustrates the benchmark-optimal contract and the test-optimal contracts using each of the seven homogeneous A/B tests for $C = 3$ (in the left panel) and $C = 7$ (in the right panel). Observe that these contracts pay the minimum wage up to a cutoff, and above that, prescribe similar pay increases. We make two remarks: First, although the test-optimal contracts sometimes prescribe very large payments for output realizations above say $x = 3000$, the probability of these output realizations is very low, and payments can be capped with virtually no loss in profits. Second, these contracts can be well-approximated

[29] Kellogg School of Management, Northwestern University, Evanston, IL 60208, U.S.A., g-georgiadis@kellogg.northwestern.edu and mike-powell@kellogg.northwestern.edu .

[30] Recall that our objective is to find, given an A/B test and a status quo contract w^C, a contract that maximizes the principal's profit and gives the agent at least as much utility as w^C. This property is not guaranteed because the principal does not know the production environment when she chooses the test-optimal contract. To assess the performance on this dimension, we compare the agent's expected utility under the test-optimal contract, $u^{AB}(w^C) = \int \tilde{v}\left(w^{AB}\left(w^C\right)(x)\right)\tilde{f}\left(x|a^{AB}\left(w^C\right)\right)dx - \tilde{c}(a^{AB}\left(w^C\right))$, to that under the status quo contract w^C, $u(w^C) = \int \tilde{v}(w^C(x))\tilde{f}(x|a(w^C))dx - \tilde{c}(a(w^C))$.

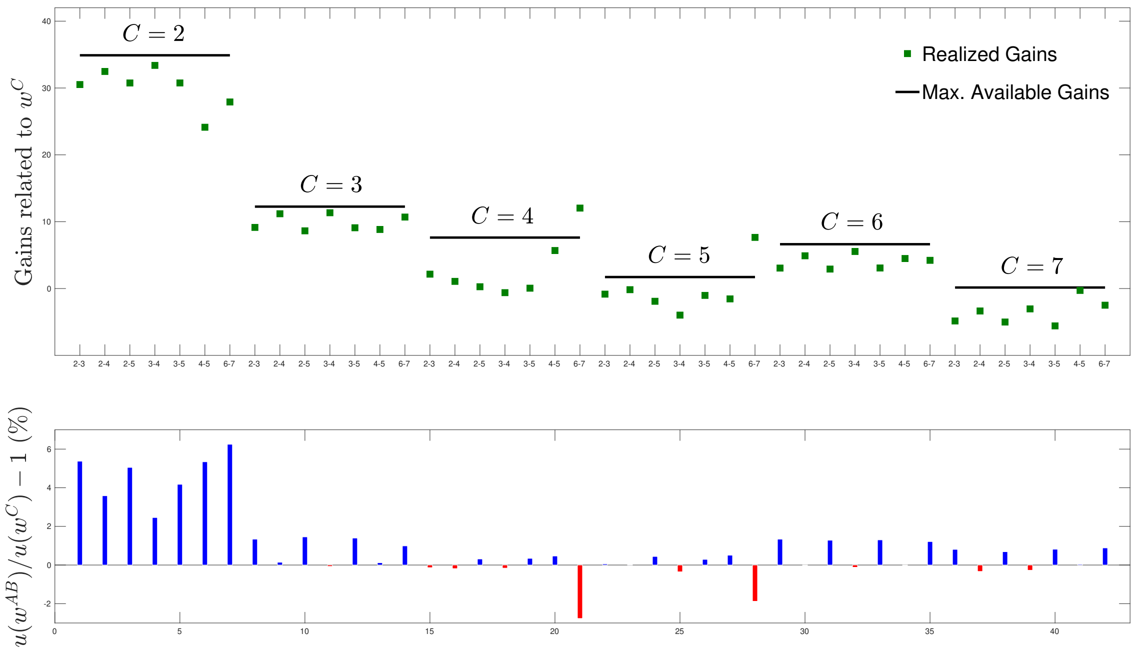

Figure 6: The top panel illustrates, for each homogeneous A/B test and each status quo treatment $C \in \{2, \dots, 7\}$, the maximum available gains, as well as the realized gains. The bottom panel illustrates the difference, in percentage terms, between agent's expected utility under the test optimal contract, and that under the status quo contract.

by simple, parametric contracts that comprise a simple piece-rate with a floor and a cap on wages, or a base wage plus a bonus paid when output exceeds a threshold.

Figure 8 plots, for each of the piece-rate treatments $C \in \{2, \dots, 5\}$, the empirical CDF $F_C(x)$ and the predicted CDF using every homogeneous A/B test, $\widehat{F}_C^{AB}(x)$. In brackets, we report the p-values for the Kolmogorov-Smirnov test, which tests the null hypothesis that the predicted distribution is identical to the empirical one. A/B tests comprising piece-rate treatments predict the output distribution quite accurately. Indeed, we cannot reject the null hypothesis at the 0.05 confidence level in 8 out of 12 cases. In contrast, A/B tests comprising bonus treatments perform less well, especially when predicting $F_2(x)$, which is the farthest out-of-sample contract in terms of the marginal incentives it induces.

Figure 9 plots for each of the bonus treatments $C \in \{6, 7\}$, the empirical CDF $F_C(x)$ and the predicted CDF using every homogeneous A/B test, $\widehat{F}_C^{AB}(x)$. While the predicted CDFs are closely clustered, they do not predict the output distribution very well, especially the lack of probability mass around $x = 2000$. Indeed, using the Kolmogorov-Smirnov test, we can reject the null hypothesis in all cases.

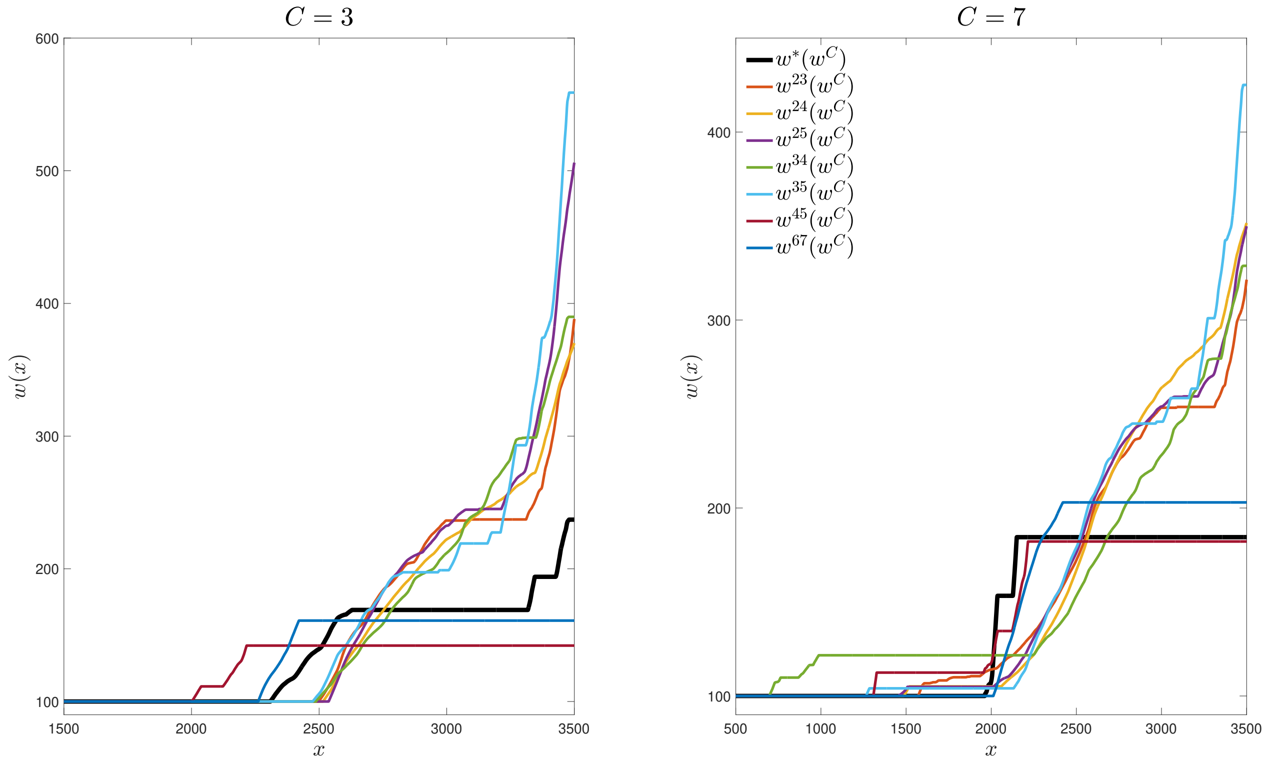

Figure 7: This figure illustrates the benchmark-optimal contract, as well as the test-optimal contracts using each of the seven homogeneous A/B tests. In the left panel, the status quo contract is treatment $C = 3$, whereas in the right panel it is $C = 7$.

Table 5 reports the summary statistics for the performance of optimal adjustments using hybrid A/B tests. In particular, it reports the average and maximum gains, the gains ratio, the average effort deviation, and the average overpayment for different values of the coefficient of RRA we used in the benchmark model, $\widetilde{\rho}$, the coefficient of RRA that the principal assumed to solve for the test-optimal contract given an A/B test, $\widehat{\rho}$, and the principal's profit margin, $\widetilde{m}$. In line with our findings in Section A, hybrid tests perform slightly worse than homogeneous tests, but the gains ratio varies little across the various parametric assumptions.

Finally, a natural question is whether the performance of optimal adjustments would continue to be insensitive to the assumptions about the agent's coefficient of RRA if stakes were higher. To examine this concern, we scaled the contracts as well as the profit margin, $\widetilde{m}$, by 100 times, and reconstructed Table 4, which presents the performance of optimal adjustments and a sensitivity analysis. The new table (with scaled payoffs) is reported below.

The pattern is very similar to the one reported in Table 4 *except for* Column VI. This

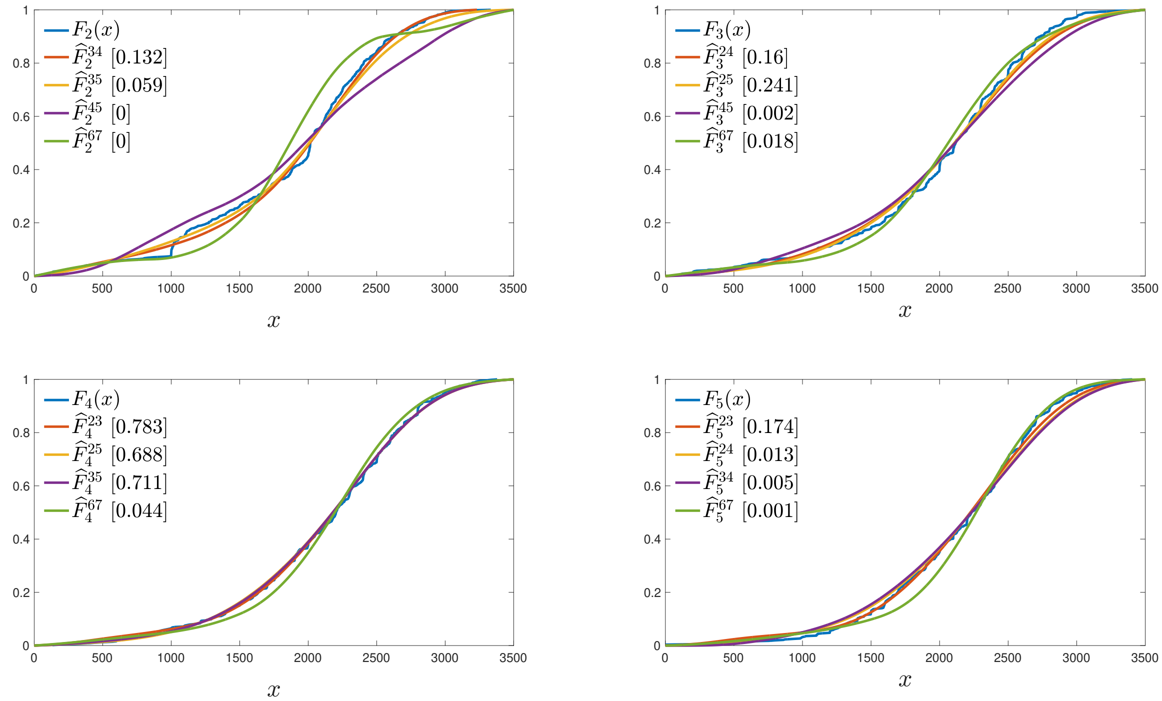

Figure 8: This figure illustrates, for each of the piece-rate treatments, the predicted CDF of output using every homogeneous A/B test, and compares it to the observed one. In brackets, we report the p-values for the Kolmogorov-Smirnov test, which tests the hypothesis that the predicted distribution is identical to the empirical one.

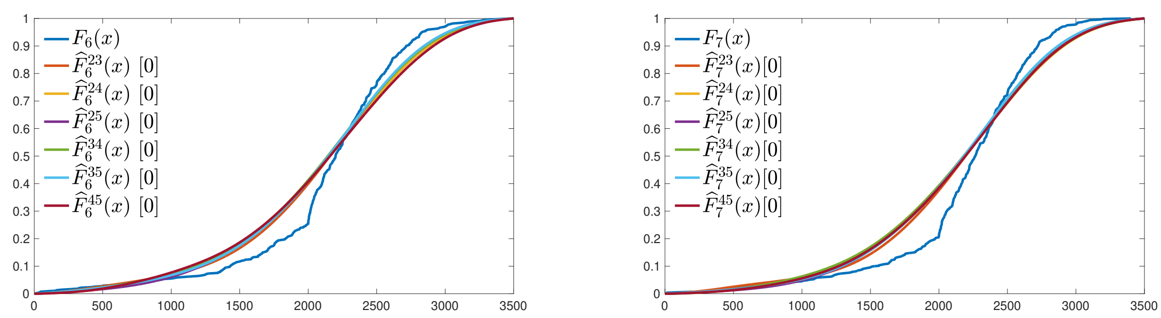

Figure 9: This figure illustrates, for each of the bonus treatments, the predicted CDF of output using every homogeneous A/B test, and compares it to the observed one. In brackets, we report the p-values for the Kolmogorov-Smirnov test, which tests the hypothesis that the predicted distribution is identical to the empirical one.

column corresponds to the case in which the principal acts as if the agent is close to risk neutral, whereas in reality he is not. To understand why the gains ratio is abnormally high in this case, in Figure 10 below, we report disaggregated data for the performance of each homogeneous A/B test under that column. In the majority of cases, the test-optimal contract

Table 5: Performance of Optimal Adjustments and Sensitivity Analysis for Hybrid A/B Tests

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Model coeff. of RRA ($\widetilde{\rho}$)	0.3	0.3	0.3	0.1	0.5	0.3	0.3
Test coeff. of RRA ($\widehat{\rho}$)	0.3	0.3	0.3	0.3	0.3	0.1	0.5
Profit margin ($\widetilde{m}$)	0.2	0.15	0.25	0.2	0.2	0.2	0.2
Average Gains ($)	6.39	4.48	8.61	6.36	6.52	6.62	5.67
Maximum Gains ($)	10.55	7.62	13.52	10.74	10.59	10.55	10.55
Gains Ratio (%)	60.62	58.80	63.70	59.23	61.58	62.75	53.80
Average Effort Deviation	-29.12	-32.55	-26.21	-30.76	-27.95	-26.86	-32.19
Average Overpayment ($)	4.36	4.36	4.25	4.19	4.57	1.86	4.69

Table 5: This table reports for different values of the parameters $\tilde{\rho}$, $\hat{\rho}$, and $\widetilde{m}$, the average and maximum gains, the gains ratio, the average effort deviation, and the average overpayment, averaged across $C \in \{2, \ldots, 7\}$. Column (I) represents our baseline parameters. In columns (II) and (III) we vary the profit margin, $\widetilde{m}$. In columns (IV) and (V) we vary the coefficient of RRA used in the benchmark model, $\widetilde{\rho}$. Finally, in columns (VI) and (VII) we vary the coefficient of RRA that the principal assumed to solve for the test-optimal contract given an A/B test, $\widehat{\rho}$.

gives the agent less utility than the status quo contract w^C, which in turn leads to large average realized gains (and hence a large gains ratio). This is because the principal, acting as if the agent is close to risk neutral, designs contracts that concentrate large payments over a small range of outputs. As the agent is in reality more risk averse, he does not value those payments as much as the principal assumes, and his expected utility ends up being smaller than what the principal predicts. This result suggests that if the principal is uncertain about the agent's coefficient of RRA, it is safer to assume he is at least moderately risk-averse.[31]

[31] This issue does not arise in column V, where the agent is again more risk-averse than the principal thinks, but $\widehat{\rho} = 0.3$ (instead of 0.1). This issue also does not arise if we scale payoffs by 0.01.

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Model coeff. of RRA ($\widetilde{\rho}$)	0.3	0.3	0.3	0.1	0.5	0.3	0.3
Test coeff. of RRA ($\widehat{\rho}$)	0.3	0.3	0.3	0.3	0.3	0.1	0.5
Profit margin ($\widetilde{m}$)	20	15	25	20	20	20	20
Panel A: Homogeneous A/B Tests							
Average Gains (\$)	717.20	521.31	925.99	724.80	715.08	813.09	731.02
Maximum Gains (\$)	1045.02	757.14	1337.04	1058.02	1049.52	1045.02	1045.02
Gains Ratio (%)	68.63	68.85	69.26	68.51	68.13	77.81	69.95
Average Effort Deviation	-5.24	-7.29	-4.98	-7.26	-4.88	-9.36	-5.07
Average Overpayment (\$)	191.28	146.93	225.64	168.42	188.22	53.06	152.74
Panel B: Hybrid A/B Tests							
Average Gains (\$)	639.22	447.88	861.10	636.23	652.09	898.66	575.95
Maximum Gains (\$)	1045.02	757.14	1337.04	1058.02	1049.52	1045.02	1045.02
Gains Ratio (%)	61.17	59.15	64.40	60.13	62.13	86.00	55.11
Average Effort Deviation	-27.62	-32.38	-24.88	-30.26	-26.29	-22.07	-30.19
Average Overpayment (\$)	439.61	438.71	426.75	426.23	463.70	13.79	475.84

Table 6: This table reports for different values of the parameters $\tilde{\rho}$, $\hat{\rho}$, and $\widetilde{m}$, the average and maximum gains, the gains ratio, the average effort deviation, and the average overpayment, averaged across $C \in \{2, \ldots, 7\}$ when payoffs are scaled by 100.

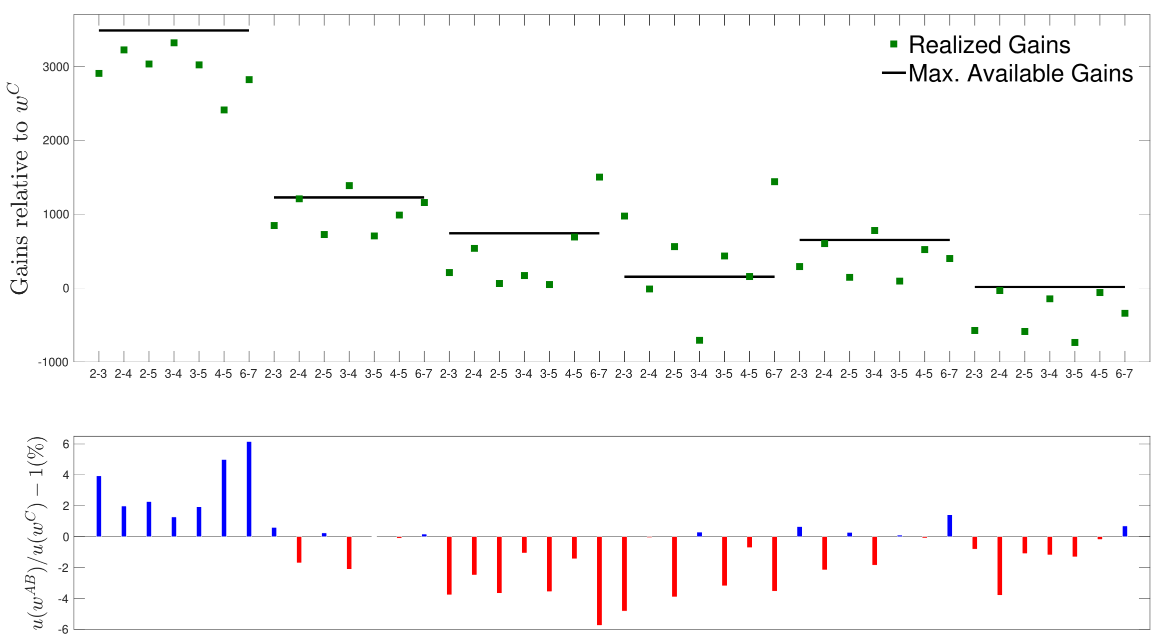

Figure 10: This figure illustrates disaggregated data for the performance of each homogeneous test A/B under column (VI) in Table 6; *i.e.*, when payoffs are scaled by 100 the true coefficient of relative risk aversion $\widetilde{\rho} = 0.3$ but test-optimal contracts assume that $\widehat{\rho} = 0.1$ instead.

B External Incentives

A limitation of our model is that both contracts in the A/B test must generate strictly positive marginal incentives, for otherwise it is impossible to recover the cost parameters β and ϵ. Moreover, it predicts zero effort for any contract that generates zero marginal incentives, which is inconsistent with the fact that in the experiment of DellaVigna and Pope (2018), subjects exert strictly positive effort even when they receive a fixed, performance-independent payment. This suggests that subjects may be motivated by factors beyond explicit performance pay. Such external incentives can be incorporated into our model by modifying Condition 2 so that the agent's cost function satisfies $c'(a) = e^{-\beta/\epsilon} a^{1/\epsilon} - I_0$ for some parameters β, ϵ, and I_0. The parameter I_0 captures the magnitude of the agent's external incentives. To recover these parameters, output data for three contracts is needed. To be specific, if the principal observes the output distribution for three contracts, (w^A, w^{B_1}, w^{B_2}), she can use the corresponding output distributions f^A, f^{B_1}, and f^{B_2} to determine the functions $g(\cdot)$ and $h(\cdot)$, and compute the marginal incentives for any contract. Then she can recover the cost parameters β, ϵ, and I_0 by solving the nonlinear system

$$\ln a(w) = \beta + \epsilon \ln [I(w) + I_0]$$

for $w \in \{w^A, w^{B_1}, w^{B_2}\}$.

Table 6 reports summary statistics for predicted performance when the principal has one versus two test contracts. Column I is identical to column IV of Table 2, and it corresponds to the case in which the principal has a single test contract and predicts out-of-sample effort under Conditions 1 and 2. For Column II, the principal has two test contracts, one of which is the no-incentives treatment w_1, and predicts effort using the modified cost function described above. In both cases, we assume that the agent's utility exhibits constant RRA with coefficient $\rho = 0.3$.

C Multitasking with Distribution Manipulation

One of the premises of our model is that the agent chooses a one-dimensional action that determines the distribution of his output, and we normalized this action to equal mean output. Although this model predicts out-of-sample effort reasonably accurately, it appears that this assumption is sometimes violated. As an example, consider treatments 5 and 7: mean output is similar, but the output distributions have distinctly different patterns as illustrated in the left panel of Figure 11. In particular, for treatment 5, which is a piece-rate treatment, performance is roughly symmetrically distributed around the average. For

Table 67: Out-of-Sample Effort Predictions with one vs. two test contracts

# of Test Contracts	(I) One (A/B test)	(II) Two ($A/B^1/B^2$ test)
Panel A: Homogeneous A/B Tests		
Corr $(\hat{a}_C, a_C)$	0.94	0.95
Mean APE (%)	1.59	1.57
Within-class	0.66	0.68
Across-class: piece-rate predictions	0.99	n/a
Across-class: bonus predictions	2.71	2.46
Worst-case APE (%)	3.34	2.95
Within-class	2.35	2.35
Across-class: piece-rate predictions	2.39	n/a
Across-class: bonus predictions	3.34	2.95
Panel B: Hybrid A/B Tests		
Corr $(\hat{a}_C, a_C)$	0.84	0.81
Mean APE (%)	2.16	2.94
Worst-case APE (%)	10.70	10.44

Table 7: This table reports summary statistics for predicted performance when the principal has one versus two test contracts, and the agent's utility exhibits constant RRA with coefficient $\rho = 0.3$. To make the comparison meaningful, column (II) considers tests in which one of the contracts is the no incentives treatment w_1.

the bonus treatment 7, however, performance spikes just over $x = 2000$, the threshold for receiving the bonus. A similar pattern emerges when we consider the pairs $(4, 6)$ and $(5, 6)$.

We now show how one can enrich the model by allowing the agent to choose his output distribution in a richer manner if the principal has output data for an additional test contract. In doing so, we demonstrate how to apply some of the ideas from Section B. To be specific, suppose the agent chooses an effort $a \in \mathbb{R}$ at cost $c(a)$, which generates a "natural" output distribution $f(x|a)$. The agent can then engage in manipulation so that his output is drawn according to $\widetilde{f}(x)$ instead, by incurring additional cost $\kappa \left\| \widetilde{f}(x) - f(x|a) \right\|_2$ for some parameter $\kappa \geq 0$, which is unknown to the principal. The agent can choose any probability density function $\widetilde{f}(x)$ subject to the constraint that its mean is no larger than that of $f(x|a)$.[32] We assume that $f(x|a)$ and $c(a)$ satisfies Conditions 1 and 2, respectively.

[32]This model is inspired by Barron, Georgiadis, and Swinkels (2020). As long as $w(\cdot)$ is non-decreasing, the agent will optimally choose an $\widetilde{f}(x)$ that has the same mean as $f(x|a)$, which equals a by assumption.

Faced with a contract $w(x)$, an agent chooses a and $\widetilde{f}(x)$ by solving the following program:

$$\begin{aligned} \max_{a,\,\widetilde{f}(x)\geq 0}\ & \int v(w(x))\widetilde{f}(x)dx - c(a) - \kappa \int \left[\widetilde{f}(x) - g(x) - ah(x)\right]^2 dx \\ \text{s.t. } & \int \widetilde{f}(x)dx = 1 \\ & \int x\widetilde{f}(x)dx \leq a. \end{aligned} \tag{5}$$

Because manipulation is costly, the agent will optimally choose $\widetilde{f}(x) \equiv f(x|a)$ if $v(w(x))$ is affine in x, and he will choose an $\widetilde{f}(x)$ that is more (less) dispersed than $f(x|a)$ if $v(w(x))$ is strictly convex (concave). We can exploit this observation in the following way: Suppose the principal observes output data for two piece-rate treatments and one bonus treatment, denoted (f^A, f^{B_1}, f^{B_2}). Under the assumption that the agent is risk-neutral, using the data from the two piece rate treatments, the principal can infer the functions g and h, as well as the cost parameters β and ϵ as described in Section A. We now explain how to use the bonus treatment, w^{B_2}, to recover the parameter κ. First, the principal can infer that the corresponding effort, a^{B_2}, is equal to the mean output. For each κ, let $\widetilde{f}_\kappa$ denote the pdf which solves (5) when the agent faces the bonus contract w^{B_2} and $a = a^{B_2}$. Then the principal may infer that κ^* is the minimizer of the L^2-norm distance between $\widetilde{f}_\kappa$ and f^{B_2}, that is, κ^* minimizes $\int[\widetilde{f}_\kappa(x) - f^{B_2}(x)]^2dx$.

Having determined all parameters of the model, the principal can predict the effort and the output distribution for any contract by solving (5). Let $\widetilde{f}_C^{AB_1B_2}(\cdot;\kappa)$ denote the predicted output distribution for treatment C given output data for treatments (A, B_1, B_2) as a function of the cost parameter κ. The right panel of Figure 11 illustrates the principal's prediction for the bonus treatment $C = 7$ given output data for treatments $(4, 5, 6)$, and compares it to the observed distribution and to the predicted distribution using the method described in Section A. Notice that $\widetilde{f}_C^{AB_1B_2}(x;\kappa^*)$ is a more accurate prediction of the observed output distribution $\widehat{f}^7$ than $\widetilde{f}_C^{AB_1B_2}(x;\infty)$, reflecting the fact that the agent moves mass just above 2,000 points when facing a bonus contract. However, this approach improves the prediction accuracy for mean output only by about 0.2%: the mean absolute percentage error decreases from 2.93% (fifth line in Column I of Table 2) to 2.72%, while the worst-case percentage error decreases from 3.65% to 3.47%.

This extension highlights two points. First, it is important for the principal to take a stance on the nature and the dimensions of effort a priori. For example, a model where the agent chooses the mean and the variance of output would not be useful in this setting, because subjects appear to influence the distribution of output in a richer manner. Second,

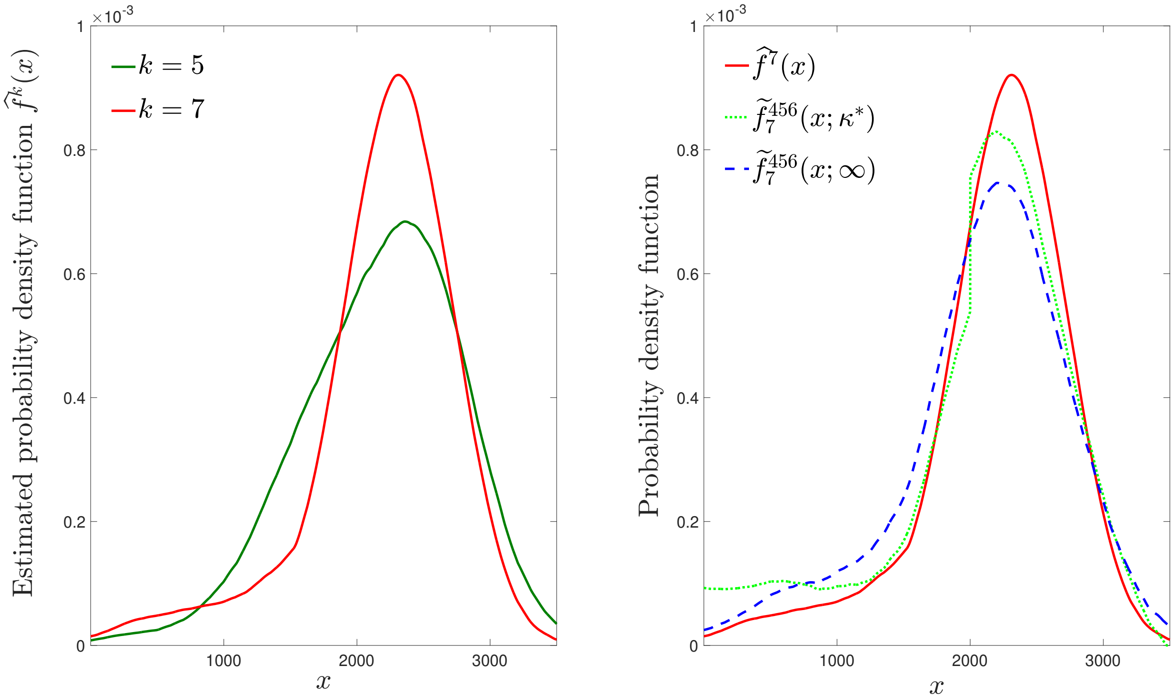

Figure 11: The left panel of this figure illustrates the estimated probability density functions for treatments 5 and 7. The right panel illustrates the principal's predicted output distribution for the bonus treatment $C = 7$ given output data for treatments $(4, 5, 6)$ using the estimated cost parameter κ^*, and compares it to the predicted distribution when she ignores the possibility of manipulation (*i.e.*, assumes $\kappa = \infty$) and to the observed output distribution.

it highlights that the right kind of contract variation may be needed to learn about different dimensions of effort: Recall that we use the data from two treatments that are not contaminated by manipulation to infer the functions (g, h) and the cost parameters (β, ϵ), and then we use the data from the bonus treatment to infer the manipulation parameter κ.

Characterizing an optimal adjustment in this extension is challenging, because the principal's problem is not a convex program. Our results allow one, however, to predict the profits under counterfactual contracts. Because contracts in practice typically consist of a finite set of parameters (*e.g.*, a base wage, a piece-rate that kicks in for performance above some cutoff, and a cap), the principal's problem can be solved by brute force as long as the number of parameters is not too large.

B A Second Empirical Exploration[33]

In this section, we apply our model to the third version of the Amazon MTurk experiment conducted by DellaVigna and Pope (2021). That experiment was designed to evaluate the robustness of experimental findings to a change in the task being performed. Instead of pressing the 'a' and 'b' keys in alternating order, subjects were assigned to code the occupation field of World War II enrollment cards. In this experiment, subjects were paid a fixed \$1 fee to complete 40 WWII cards. After they were done, they saw the following message: "If you are willing, there are 20 additional cards to be coded. Doing this additional work is not required for your HIT to be approved or for you to receive the \$1 promised payment." In the treatments we focus on, which are summarized in Table 7, subjects were paid different piece-rates for each additional card they completed. For example in the first treatment, subjects were told that "The number of additional cards you complete will not affect your payment in any way." Subjects in the second piece-rate treatment, w^3, were informed "as a bonus, you will be paid an extra 1 cent for every 2 additional cards you complete."

Table 8: Experimental Treatments in Extra-Work Task in DellaVigna and Pope (2019)

	Contract	Avg. #Additional Cards	Std. Deviation	#Subjects
No-Incentives	$w^1(x) = 100$	8.63	9.37	158
Piece-Rate	$w^2(x) = 100 + 0.05x$	9.94	9.67	155
	$w^3(x) = 100 + 0.5x$	12.63	9.24	136
	$w^4(x) = 100 + 2x$	15.21	8.08	136
	$w^5(x) = 100 + 5x$	17.39	6.16	154

Table 8: This table describes five experimental treatments from the data-entry task in DellaVigna and Pope (2019) that differed in the monetary incentives offered to the subjects. The second column describes the implied incentive contract, denominated in cents. The remaining columns describe, for each treatment, the average number of additional cards coded, the standard deviation, and the number of subjects.

We now report the findings from the first exercise described in Section A. That is, we take each pair of piece-rate treatments in Table 7 to constitute an A/B test, and use our model to predict the average number of additional cards completed in each of the remaining piece-rate treatments. We will assume that at the outset of the experiment, each subject observes the contract he or she is offered and chooses "effort" a. Then the number of additional cards completed, $x \in \{0, \ldots, 20\}$, is drawn from some probability distribution with mean a. We therefore interpret effort to be the average number of additional cards completed in a particular treatment. We will assume that Conditions 1 and 2 hold, that is, the output

[33] We are very grateful to Stefano DellaVigna and Devin Pope for sharing the data used in this section.

distribution is affine in effort and $c'(a) = e^{-\beta/\epsilon} a^{1/\epsilon}$ for some parameters β and ϵ. We will also assume that each subject has constant-relative-risk-aversion preferences over money, so that $v'(\omega) = \omega^{-\rho}$, and we will vary the parameter $\rho \in [0, 1]$.

This setting presents a new challenge in that between 85% and 90% of the subjects in each treatment completed either none or all of the 20 additional cards. Due to those mass points, the kernel density estimation approach used in Section V is not applicable, and we must therefore modify our methodology. We do so as follows: Imagine that the principal has outcome data from two treatments, say A and B. For these treatments, she can compute the empirical distribution functions $\widehat{F}^A(x)$ and $\widehat{F}^B(x)$. Let $d\widehat{F}^A(0) = \widehat{F}^A(0)$ and $d\widehat{F}^A(x) = \widehat{F}^A(x) - \widehat{F}^A(x-1)$ for every integer $x \geq 1$, and $d\widehat{F}^B(x)$ be similarly defined. The counterpart of Condition 1 in this discrete probability space is that $dF(x|a) = g(x) + ah(x)$ for some functions $g(x)$ and $h(x)$ such that $\sum_{x=0}^{20} g(x) = 1$ and $\sum_{x=0}^{20} h(x) = 0$. Letting $\widehat{a}^A$ and $\widehat{a}^B$ denote the mean output in treatment A and B, respectively, we can construct the functions

$$\widehat{h}^{AB}(x) = \frac{d\widehat{F}^A(x) - d\widehat{F}^B(x)}{\widehat{a}^A - \widehat{a}^B} \quad \text{and} \quad \widehat{g}^{AB}(x) = d\widehat{F}^A(x) - \widehat{a}^A \widehat{h}^{AB}(x).$$

For each triple (A, B, C), we then compute the predicted marginal incentives under contract C according to

$$\widehat{I}_C^{AB} = \sum_{x=0}^{20} v(w^C(x)) \widehat{h}^{AB}(x).$$

Using the estimates of the agent's marginal incentives under contracts A and B, we can estimate the parameters of the agent's cost function $\widehat{\beta}^{AB}$ and $\widehat{\epsilon}^{AB}$. Finally, our prediction for the average number of additional cards completed in treatment C is $\ln \widehat{a}_C^{AB} = \widehat{\beta}^{AB} + \widehat{\epsilon}^{AB} \ln \widehat{I}_C^{AB}$.

Figure 12 plots our predictions against the actual performance for each treatment. The red stars represent predictions using A/B tests and the procedure described above. The blue triangles represent predictions using two test contracts, one of which is the no-incentives treatment w^1, and the procedure described in Appendix B adapted to this setting. To be specific, we assume that each subject's cost function satisfies $c'(a) = e^{-\beta/\epsilon} a^{1/\epsilon} - I_0$ for some parameters β, ϵ, and I_0, which we estimate using outcome data from three (rather than two treatments). Recall that I_0 represents the agent's external incentives, which may, for example, be due to intrinsic motivation.

Table 8 reports summary statistics of the effort predictions for different values of the coefficient of RRA ρ. In particular, it reports the correlation between actual and predicted effort, the mean and the worst-case absolute percentage error. Columns (I)-(IV) focus on

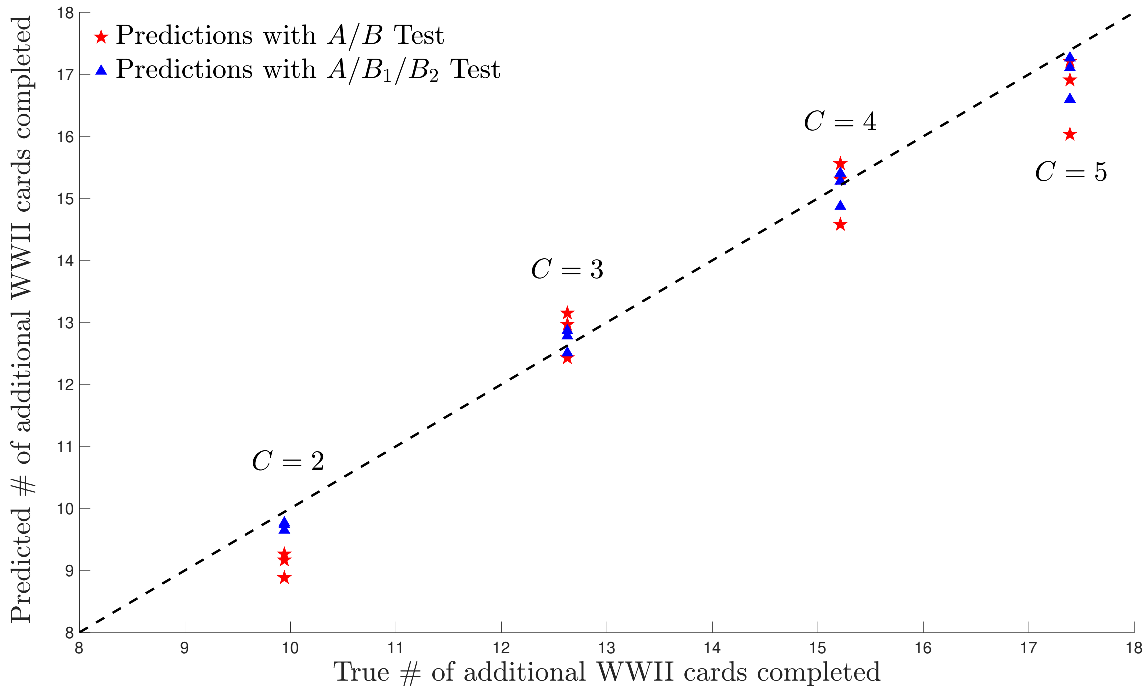

Figure 12: This figure plots our predictions against the actual performance for each treatment. The horizontal axis, depicts the actual number of additional WWII cards completed, while the vertical axis plots predicted performance. The red stars represent predictions using A/B tests and the procedure described above. The blue triangles represent predictions using two test contracts, one of which is the no-incentives treatment w^1, and the procedure described in Appendix B.

A/B tests, while columns (V)-(VIII) focus on A/B_1/B_2 tests in which one of the contracts is the no-incentives treatment. Notice that the prediction accuracy using A/B tests is somewhat worse than in the 'a-b' typing task reported in Table 2.[34] It improves substantially however, when we have outcome data for an additional test contract and use it to incorporate external incentives into the model. This finding echoes DellaVigna and Pope's discussion that the WWII coding task is "more motivating" than the 'a-b' typing task.

Next, we report the findings from the second exercise described in Section B. Table 9 reports summary statistics for the performance of optimal adjustments for different values of the coefficient of RRA we used in the benchmark model, $\widetilde{\rho}$, the coefficient of RRA that the principal assumed to solve for the test-optimal contract given an A/B test, $\widehat{\rho}$, and the

[34]Note also that performance varies substantially more across treatments in this task compared to the 'a-b' typing task. To be specific, the average absolute percentage performance difference across treatments is 31.93%, whereas in the 'a-b' typing task, it is 6.4%.

Table 9: Effort Predictions for the Extra-Work task in DellaVigna and Pope (2021)

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Coefficient of RRA ρ	0	0.3	0.5	1	0	0.3	0.5	1
Number of test contracts	One				Two			
Corr $(\hat{a}_C, a_C)$	0.98	0.98	0.97	0.96	0.995	0.99	0.99	0.98
Mean APE (%)	4.37	5.14	5.67	7.04	1.81	2.31	2.64	3.44
Worst-case APE (%)	10.68	13.64	15.73	21.37	4.56	5.52	6.14	7.64

Table 9: This table reports, for different coefficients of RRA, the correlation between actual and predicted performance, the mean and the worst-case absolute percentage error. Columns (I)-(IV) focus on A/B tests, while columns (V)-(VIII) focus on A/B_1/B_2 tests in which one of the contracts is the no-incentives treatment.

principal's profit margin, $\widetilde{m}$. The gains ratio, which is the ratio of the average profit increase the principal can achieve using just an A/B test to the profit increase she could achieve if she knew the production environment, varies between 70% and 80%. Moreover, the test-optimal contract implements an effort level very close to the optimal one; almost the entire gap between the average gains and the maximum gains is due to the principal implementing the optimal effort at too high a cost.

Table 10: Performance of Optimal Adjustments and Sensitivity Analysis for the Extra-Work task

	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
Model coeff. of RRA $(\widetilde{\rho})$	0.3	0.1	0.5	0.3	0.3	0.3	0.3
Test coeff. of RRA $(\widehat{\rho})$	0.3	0.3	0.3	0.1	0.5	0.3	0.3
Profit margin $(\widetilde{m})$	5	5	5	5	5	4	6
Average Gains ($)	17.44	17.06	17.77	17.03	17.73	13.29	22.07
Maximum Gains ($)	20.34	19.77	20.93	20.34	20.34	15.53	25.40
Gains Ratio (%)	75.28	76.30	74.15	70.33	79.60	78.47	76.48
Average Effort Deviation	-0.075	-0.104	-0.104	-0.092	-0.075	-0.021	-0.100
Average Overpayment ($)	2.84	2.66	3.09	3.26	2.55	2.22	3.26

Table 10: This table reports for different values of the parameters $\tilde{\rho}$, $\hat{\rho}$, and $\widetilde{m}$, the average and maximum gains, the gains ratio, the average effort deviation, and the average overpayment, averaged across $C \in \{2, \ldots, 5\}$. Column (I) represents our baseline parameters. In columns (II) and (III) we vary the coefficient of RRA used in the benchmark model, $\widetilde{\rho}$. In columns (IV) and (V) we vary the coefficient of RRA that the principal assumed to solve for the test-optimal contract given an A/B test, $\widehat{\rho}$. Finally, in columns (VI) and (VII) we vary the profit margin, $\widetilde{m}$.

Figure 13 presents disaggregated data for the optimal adjustments from every A/B test when the principal's profit margin $\widetilde{m} = 5¢$ per additional WWII card and the agent's coefficient of RRA $\widetilde{\rho} = \widehat{\rho} = 0.3$. The top panel illustrates, for each A/B test and each status quo treatment $C \in \{2, \ldots, 5\}$, the maximum available gains, as well as the realized gains. The bottom panel illustrates the difference, in percentage terms, between agent's expected utility under the test-optimal contract, and that under the status quo contract w^C. We remark that when the status quo contract $C = 2$ or $C = 3$, under both the benchmark-optimal and the test-optimal contracts, the agent's participation constraint is slack. Overall, the test-optimal contract delivers to the agent strictly more utils in expectation than the contract w^C in all but one case.

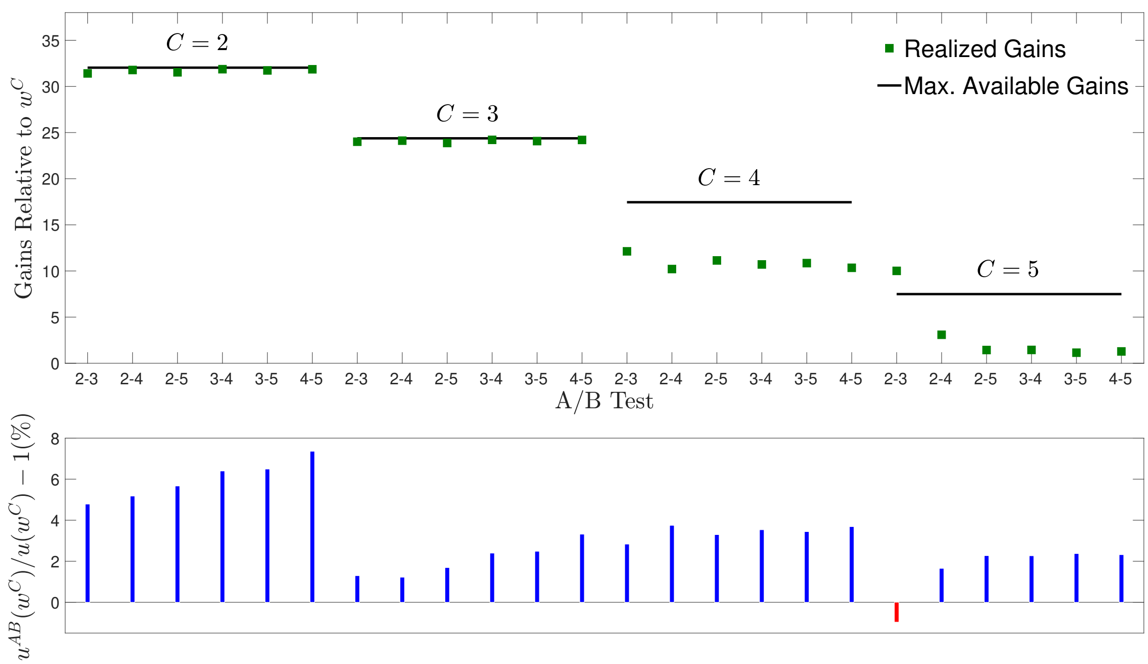

Figure 13: The top panel illustrates, for each A/B test and each status quo treatment $C \in \{2, \ldots, 5\}$, the maximum available gains, as well as the realized gains. The bottom panel illustrates the difference, in percentage terms, between agent's expected utility under the test optimal contract, and that under the status quo contract.

We conclude this section with a brief discussion of the shape of the test-optimal contracts. When the status quo contract $C \in \{2, 3\}$, each test-optimal contract pays the minimum wage, which is set to \$1, plus a lump-sum bonus if the agent completes all 20 additional WWII cards. When $C \in \{4, 5\}$, the test-optimal contracts feature trinary wages: they pay a base wage when $x = 0$, a slightly higher wage for any $x \in \{1, \ldots, 19\}$, and an even higher wage when $x = 20$.

C Proofs

Proof of Lemma 1. Fix arbitrary upper semi-continuous functions w and t (to ensure that the desired Gateaux derivatives are well-defined), and consider the contract $w + \theta t$ for some $\theta > 0$. The agent's effort satisfies the first-order condition

$$\int v(w(x) + \theta t(x)) f_a(x|a(w + \theta t))dx = c'(a(w + \theta t)).$$

Differentiating this equation with respect to θ and taking the limit as $\theta \to 0$ yields

$$\int t(x)v'(w(x))f_a(x|a(w)) = \left[c''(a(w)) - \int v(w(x))f_{aa}(x|a(w))\right] \mathcal{D}a(w,t)\,,$$

and using the definition $\mathcal{D}I(w,t) := \int t(x)v'(w(x))f_a(x|a(w))$, we obtain the desired expression for $\mathcal{D}a(w,t)$.

Next, consider the agent's expected utility. Faced with contract $w + \theta t$, the agent's expected utility is

$$u(w + \theta t) = \int v(w(x) + \theta t(x))f(x|a(w + \theta t))dx - c(a(w + \theta t))\,.$$

Differentiating with respect to θ and taking the limit as $\theta \to 0$ yields

$$\begin{aligned} \mathcal{D}u(w,t) &= \int t(x)v'(w(x))f(x|a(w))\,dx + \left[\int v(w(x))f_a(x|a(w)) - c'(a(w))\right] \mathcal{D}a(w,t) \\ &= \int t(x)v'(w(x))f(x|a(w))\,dx\,, \end{aligned}$$

where the second equality follows because the term in brackets is equal to 0 by the agent's first-order condition. □

Proof of Lemma 2. This lemma follows immediately from the fact that for any t, $\mathcal{D}\pi(w^A,t)$ and $\mathcal{D}u(w^A,t)$ depend only on $f^A(\cdot)$, $f^A_a(\cdot)$, and $\mathcal{D}a(w^A,t)$, and no other parameters of the production environment. □

Proof of Proposition 1. Note that

$$\begin{aligned} LAB\left(w^A, w^B \middle| P\right) &= \left(f^A, f^A_a, \mathcal{D}a\left(w^A, w^B \middle| P\right)\right) \\ LAB\left(w^A, w^B \middle| \tilde{P}\right) &= \left(\tilde{f}^A, \tilde{f}^A_a, \mathcal{D}a\left(w^A, w^B \middle| \tilde{P}\right)\right). \end{aligned}$$

If the first statement is true, then the second is obviously true. Next, suppose $LAB\left(w^A, w^B\middle| P\right) = LAB\left(w^A, w^B\middle| \tilde{P}\right)$. It is immediate that $f^A = \tilde{f}^A$ and $f_a^A = \tilde{f}_a^A$. Next, note that for all t,

$$\mathcal{D}I\left(w^A, t\middle| P\right) = \int tv'\left(w^A\right) f_a = \int tv'\left(w^A\right) \tilde{f}_a = \mathcal{D}I\left(w^A, t\middle| \tilde{P}\right),$$

and by Lemma 1,

$$\begin{aligned}
\mathcal{D}a\left(w^A, t|P\right) &= \frac{\mathcal{D}a\left(w^A, w^B|P\right)}{\mathcal{D}I\left(w^A, w^B|P\right)}\mathcal{D}I\left(w^A, t|P\right) \\
\mathcal{D}a\left(w^A, t|\tilde{P}\right) &= \frac{\mathcal{D}a\left(w^A, w^B|\tilde{P}\right)}{\mathcal{D}I\left(w^A, w^B|\tilde{P}\right)}\mathcal{D}I\left(w^A, t|\tilde{P}\right),
\end{aligned}$$

so $\mathcal{D}a\left(w^A, t\middle| P\right) = \mathcal{D}a\left(w^A, t\middle| \tilde{P}\right)$ for all t if and only if $\mathcal{D}a\left(w^A, w^B\middle| P\right) = \mathcal{D}a\left(w^A, w^B\middle| \tilde{P}\right)$, which is true by supposition. □

Proof of Proposition 2. The optimization program given in (Adj_{local}) can be rewritten as

$$\begin{aligned}
\max_t \quad & \mu^* \int tv'(w^A)f_a^A dx - \int tf^A dx \\
\text{s.t.} \quad & \int tv'(w^A)f^A dx \geq 0 \\
& \int t^2 dx \leq 1
\end{aligned}$$

where

$$\mu^* := \frac{\left(m - \int w_A f_a^A dx\right) \mathcal{D}a(w^A, w^B)}{\int w^B v'(w^A) f_a^A dx},$$

and we have used that for any t, $\mathcal{D}a(w^A, t) = \mathcal{D}a(w^A, w^B) \int tv'(w^A)f_a^A dx / \int w^B v'(w^A) f_a^A dx$ by Lemma 1. Letting $\lambda \geq 0$ and $\nu \geq 0$ denote the dual multipliers associated with the first and the second constraint, we have the Lagrangian

$$L(\lambda, \nu) = \max_t \left\{ \nu + \int \left[t\left(\lambda v'(w^A)f^A + \mu^* v'(w^A) f_a^A - f^A\right) - \nu t^2 \right] dx \right\}. \tag{6}$$

For any $\lambda \geq 0$ and $\nu > 0$, we can optimize the integrand with respect to t pointwise. Noting that the integrand is differentiable with respect to t, the corresponding first-order condition implies that

$$t_{\lambda,\nu} = \frac{\left(\lambda f^A + \mu^* f_a^A\right) v'(w^A) - f^A}{2\nu},$$

where t, f^A, f_a^A, and w^A are functions of x.[35]

Next, we pin down the optimal λ and ν, by solving the following dual problem:

$$\min_{\lambda \geq 0\,,\, \nu \geq 0} L(\lambda, \nu)\,.$$

This problem is convex, and using $t_{\lambda,\nu}$, the corresponding first-order conditions yield

$$\lambda^* = \max\left\{0, \frac{\int \left[f^A - \mu^* v'(w^A) f_a^A\right] v'(w^A) f^A dx}{\int \left[v'(w^A) f^A\right]^2 dx}\right\} \tag{7}$$

and

$$\nu^* = \frac{1}{2}\sqrt{\int \left[(\lambda^* f^A + \mu^* f_a^A)\, v'(w^A) - f^A\right]^2 dx}\,. \tag{8}$$

Thus, the optimal adjustment direction,

$$t^* = t_{\lambda^*,\nu^*} = \frac{\left[\lambda^* f^A(x) + \mu^* f_a^A(x)\right] v'(w^A(x)) - f^A(x)}{\sqrt{\int \left[(\lambda^* f^A(x) + \mu^* f_a^A(x))\, v'(w^A(x)) - f^A(x)\right]^2 dx}} \propto T(x, \lambda^*, \mu^*)\,.$$

Insofar, we have shown than t^* solves the dual problem. To show that t^* solves the primal problem given in (Adj_{local}), we will now establish that strong duality holds. Towards this goal, let Π^* denote the optimal value of the primal. Weak duality implies that $L(\lambda^*, \nu^*) \geq \Pi^*$. Moreover, it is straightforward to verify that $t\,(\lambda^*, \nu^*)$ is feasible for (Adj_{local}), and λ^* and ν^* is strictly positive if and only if the respective (primal) constraint binds; *i.e.*, the complementary slackness conditions are satisfied: $\lambda^* \int t^* v'(w^A) f^A dx = 0$ and $\nu^* (\int t^2 dx - 1) = 0$. This implies that the objective of (Adj_{local}) evaluated at $t\,(\lambda^*, \nu^*)$ is equal to $L(\lambda^*, \nu^*)$, and feasibility implies that $L(\lambda^*, \nu^*) \leq \Pi^*$. Therefore, we conclude that $L(\lambda^*, \nu^*) = \Pi^*$, which proves that strong duality holds, and $t\,(\lambda^*, \nu^*)$ solves (Adj_{local}). Finally, if w^A is locally optimal, then it must be the case that $t^*(x) = 0$ and hence $T\,(x, \lambda^*, \mu^*) = 0$ for all x. □

Proof of Lemma 3. We can write the agent's maximized utility given contract w as:

$$\begin{aligned}
u\,(w) &= \max_a \int v\,(w\,(x))\,[g\,(x) + ah\,(x)]\, dx - c\,(a) \\
&= \int v\,(w\,(x))\, g\,(x)\, dx + \max_a \left\{ a \int v\,(w\,(x))\, h\,(x)\, dx - c\,(a) \right\} \\
&= \int v\,(w\,(x))\, g\,(x)\, dx + \max_a \left\{ aI\,(w) - c\,(a) \right\}.
\end{aligned}$$

[35] If $\nu = 0$, then the integrand of (6) is linear in t, and so $L(\lambda, 0) = \infty$.

Next, define the function

$$V(I) = \max_a aI - c(a).$$

Since $\tilde{a}(I)$ is everywhere continuous, by the envelope theorem, V is continuously differentiable, and we have

$$V'(I) = \tilde{a}(I).$$

By the fundamental theorem of calculus, for any I and $\tilde{I}$,

$$V(I) - V\left(\tilde{I}\right) = \int_{\tilde{I}}^{I} \tilde{a}(i)\, di.$$

We therefore have

$$\begin{aligned} u(w) - u(\tilde{w}) &= \int v(w(x))\, g(x)\, dx + V(I(w)) - \int v(\tilde{w}(x))\, g(x)\, dx - V(I(\tilde{w})) \\ &= \int [v(w(x)) - v(\tilde{w}(x))]\, g(x)\, dx + \int_{I(\tilde{w})}^{I(w)} \tilde{a}(i)\, di, \end{aligned}$$

which establishes the first claim. The second claim is immediate. □

Proof of Proposition 3. Recall that by definition

$$AB\left(w^A, w^B|P\right) = \left(f^A, f^B,\right) \quad , \quad AB\left(w^A, w^B|\tilde{P}\right) = \left(\tilde{f}^A, \tilde{f}^B\right) \text{ , and } \int x f(x|a) dx = a.$$

First, suppose that statement (i) is true. Noting that $I(w) = \int v(w(x))h(x)dx = \int v(w(x))\tilde{h}(x)dx = \tilde{I}(w)$, it follows from Condition 2 that $a(w^i) = \tilde{a}(w^i)$ for each $i \in \{A, B\}$, where we abuse notation and use (no) tildes to denote quantities under environment $\tilde{P}$ (P). Then, by Condition 1, for each $i \in \{A, B\}$, $f^i(x) = g(x) + a(w^i|P)h(x) = \tilde{g}(x) + a(w^i|\tilde{P})\tilde{h}(x) = \tilde{f}^i(x)$, implying statement (ii).

Next, suppose that (ii) holds; *i.e.*, $AB\left(w^A, w^B\middle| P\right) = AB\left(w^A, w^B\middle| \tilde{P}\right)$. Then it follows from Condition 2 that $g = \tilde{g}$ and $h = \tilde{h}$. Moreover, for each $i \in \{A, B\}$, we have $a(w^i) = \tilde{a}(w^i)$, and by Condition 1, $I(w^i) = \tilde{I}(w^i)$. Thus, by Condition 2, $\varepsilon = \tilde{\varepsilon}$ *and* $\beta = \tilde{\beta}$, and so statement (i) is true.

Finally, it is straightforward that the solution to (Adj) depends only on the parameters g, h, ϵ, and β, in addition to the agent's utility function, $v(\cdot)$. This completes the proof. □

Proof of Proposition 4. Given A/B test $AB\left(w^A, w^B\right)$, a type-ϕ agent's effort arc-elasticity satisfies

$$\varepsilon = \frac{\ln a^{\phi}\left(w^A\right)/a^{\phi}\left(w^B\right)}{\ln I^A/I^B}.$$

Since this holds for all ϕ, it must be the case that for all ϕ,

$$\frac{a^{\phi}\left(w^A\right)}{a^{\phi}\left(w^B\right)} = \frac{\overline{a}^A}{\overline{a}^B},$$

where this holds because for any real vectors $(x_1, \ldots, x_N)$, $(y_1, \ldots, y_N)$, and $(z_1, \ldots, z_N)$,

$$\frac{x_1}{y_1} = \cdots = \frac{x_N}{y_N} \Rightarrow \frac{x_i}{y_i} = \frac{\sum_i z_i x_i}{\sum_i z_i y_i}.$$

The principal therefore correctly estimates the arc elasticity parameter, since

$$\hat{\varepsilon} = \frac{\ln \bar{a}^A - \ln \bar{a}^B}{\ln I^A - \ln I^B} = \frac{\ln a^{\phi}\left(w^A\right) - \ln a^{\phi}\left(w^B\right)}{\ln I^A - \ln I^B} = \varepsilon.$$

Next, notice that a type-ϕ agent's effort-cost coefficient satisfies $\phi = \ln a^{\phi}\left(w^A\right) - \varepsilon \ln I^A$, and so his actual effort choice under contract $\tilde{w}$ is

$$a^{\phi}\left(\tilde{w}\right) = e^{\phi} I\left(\tilde{w}\right)^{\varepsilon} = a^{\phi}\left(w^A\right)\left(\frac{I\left(\tilde{w}\right)}{I^A}\right)^{\varepsilon}.$$

Mean effort under contract $\tilde{w}$ is therefore

$$\overline{a}\left(\tilde{w}\right) = \sum_{\phi} p^{\phi} a^{\phi}\left(\tilde{w}\right) = \sum_{\phi} p^{\phi} a^{\phi}\left(w^A\right)\left(\frac{I\left(\tilde{w}\right)}{I^A}\right)^{\varepsilon} = \overline{a}\left(w^A\right)\left(\frac{I\left(\tilde{w}\right)}{I^A}\right)^{\varepsilon}.$$

The principal's predicted mean effort is

$$\hat{a}\left(\tilde{w}\right) = e^{\hat{\phi}} I\left(\tilde{w}\right)^{\hat{\varepsilon}} = \overline{a}\left(w^A\right)\left(\frac{I\left(\tilde{w}\right)}{I^A}\right)^{\varepsilon},$$

which is equal to $\overline{a}\left(\tilde{w}\right)$.

We conclude the proof by showing that an aggregate A/B test suffices to solve the principal's problem. By Lemma 3, for each type ϕ, $u_{\phi}\left(\widetilde{w}\right) = u_{\phi}\left(w^A\right) + \int\left[v\left(\widetilde{w}\left(x\right)\right) - v\left(w^A\left(x\right)\right)\right] g\left(x\right) dx + \int_{I(w^A)}^{I(\widetilde{w})} \tilde{a}_{\phi}\left(i\right) di$, where $\widetilde{a}_{\phi}\left(I\right)$ is implicitly defined by the equation $c_{\phi}'\left(\widetilde{a}_{\phi}\left(I\right)\right) = I$. Taking averages across types, and rearranging, we see that

$$\sum_{\phi\in\Phi} p_\phi \left[u_\phi\left(\widetilde{w}\right) - u_\phi\left(w^A\right)\right] = \int \left[v\left(\widetilde{w}\left(x\right)\right) - v\left(w^A\left(x\right)\right)\right] g\left(x\right) dx + \int_{I(w^A)}^{I(\widetilde{w})} \widetilde{\mathbf{a}}\left(i\right) di,$$

where $\widetilde{\mathbf{a}}\left(I\right) = \sum_{\phi\in\Phi} p_\phi \widetilde{a}_\phi\left(I\right)$. We can therefore rewrite the principal's problem as

$$\max_{\widetilde{w}} \; m\bar{a}\left(\widetilde{w}\right) - \int \widetilde{w}\left(x\right)\left[g\left(x\right) + \bar{a}\left(\widetilde{w}\right) h\left(x\right)\right] dx$$

subject to

$$\int \left[v\left(\widetilde{w}\left(x\right)\right) - v\left(w^A\left(x\right)\right)\right] g\left(x\right) dx + \int_{I(w^A)}^{I(\widetilde{w})} \widetilde{\mathbf{a}}\left(i\right) di \geq 0.$$

Given the first part of this proposition, this problem and therefore its solution, depends only on the aggregate information $AB\left(w^A, w^B\right) = \left(\bar{f}^A, \bar{f}^B\right)$. $\square$