Work for Image and Work for Pay

Roberta Dessí*and Aldo Rustichini‡

September 10, 2011

Abstract

Standard economic models with complete information predict a positive, monotonic relationship between pay and performance. This prediction does not always hold in experimental tests: offering a small payment may result in lower performance than not offering any payment.

We test experimentally two main explanations that have been put forward for this result: the "incomplete contract" hypothesis views the payment rule as a signal given to subjects on purpose of the activity. The "informed principal" hypothesis views it as a signal concerning the characteristics of the agent or of the task. The *incomplete contract* view appears to offer the best overall explanation for our results. We also find that high-powered monetary incentives do not "crowd out" intrinsic motivation, but may elicit "too much" effort when intrinsic motivation is very high.

1 Introduction

Standard economic models tend to predict a positive, monotonic relationship between monetary incentives and performance. This prediction seems to hold in a number of real and experimental contexts - but not always. For example Gneezy and Rustichini (2000) found a nonmonotonic relationship between the piece rate payment and performance on an IQ test. Performance when no piece rate was offered was *higher* than with a small piece rate. Performance did increase for sufficiently high piece rates.

*Toulouse School of Economics (IDEI and GREMAQ), and CEPR.

†University of Minnesota.

‡We would like to thank Samuele Centorrino and Olivier Darmouni for excellent research assistance.

Several possible explanations have been put forward for this result. Gneezy and Rustichini favored an "*incomplete contract*" interpretation. The payment rule (no payment, small payment or large payment) is perceived by the subject as a signal that the experimenter sends on the purpose of the activity. Participants received a fixed fee for participating in the experiment. If no payment for performance in the task is offered, the signal is implicitly stating that subjects have already been paid for the task, and to keep their side of the bargain they are expected to perform as well as they can. If instead some additional payment for performance is offered, this fact signals that the purpose of the activity is earning money. In this case the amount of payment is relevant to decide effort, in comparison perhaps with similar activities. Different signals complete differently the initial contract.

Alternatively, the result has also been interpreted as evidence in support of the Bénabou and Tirole (2003) analysis of "crowding-out" by monetary incentives. In their principal-agent model, the principal has better information than the agent about the agent's characteristics (his ability, for example), or about the task's characteristics: for example its difficulty, or how enjoyable it would be. This fact is known to both principal and agent. The principal's offer of a monetary reward for performance can then become a "bad" signal, increasing the likelihood of the negative characteristics: respectively implying that the agent has low ability or that the task is very hard, or boring.

These two different explanations for Gneezy and Rustichini's result seem plausible *a priori*, but they have quite different implications. We therefore designed an experiment to shed light on the relative importance of the two different explanations: the incomplete contract hypothesis, and the informed principal hypothesis.

In the experiment, subjects take an IQ test consisting of 27 questions from Raven's Advanced Matrices (RAM) test. In the RAM test a subject is presented with figures sharing a logical pattern, with one slot missing; the task is to detect an abstract rule in the figures and choose the missing figure in a set of feasible options. As in Gneezy and Rustichini, the experiment has a first treatment in which there is no mention of a payment per correct answer, a second treatment in which subjects are offered a very low payment (one cent of a euro) per correct answer, and a third treatment in which they are offered a high payment (one euro) per correct answer. In the first two treatments, subjects are given a fairly generous participation fee (15 euros), while no participation fee is given for the last treatment where the piece rate is high. These treatments are meant to replicate the previous findings. We expected performance to be higher in the first treatment and the third treatment than in the second, and highest in the third treatment. To test the

different explanations, we include a fourth treatment: subjects receive the participation fee (15 euros) and they are asked to *choose* whether they wish to receive, in addition, one cent per correct answer. Since they are choosing, there is no "bad" signal. A fifth treatment is added to test whether the flat payment, that is the 15 euro participation fee, is necessary to induce effort. In this treatment subjects do not receive any participation fee, and are offered only one cent per correct answer.

Let us compare the predictions of the two theories. According to the *informed principal hypothesis*, subjects who choose (rather than being given) the low piece rate in the choice treatment should perform better than those who are simply given the low piece rate by the experimenter, because they have not received any outside signal about their skill in the task, the difficulty of the task, or the effort cost. On the other hand they have a higher payment, hence standard incentive theory predicts higher effort and performance than without. According to the *incomplete contract hypothesis* there should be no difference whether the low piece rate is chosen by the experimenter or the subject. In both cases, the purpose of the task is earning money, and effort on the task should be consistent with the piece rate. When participants are asked to choose whether to receive a piece rate of one cent, and they choose not to, *they still may have inferred that the purpose of the activity is earning money*, and thus feel no sense of obligation to perform well. Thus the incomplete contract hypothesis predicts that performance should be better for subjects who are simply given the test without any mention of a payment per correct answer than for subjects who decide not to accept the piece rate.

Our first main finding is that *there is no significant difference in performance (i.e. test score) across treatments*. On average, subjects answer correctly 18 questions out of 27 (19 in the treatment with the one euro piece rate and the treatment where subjects choose the one cent piece rate, but the difference is not statistically significant). Also in the fifth treatment, with no participation fee, performance did not substantially fall: average performance for this treatment was, yet again, just over 18 correct answers. These results are striking: our participants seem to perform roughly as well when given one euro for each correct answer as when they are given one cent, or nothing. At first sight, this seems at odds with standard economic models and also with the previous findings by Gneezy and Rustichini. It implies that for a principal hiring our participants, the same performance costs 18 euros when obtained through high-powered incentives, 15 euros when obtained through a flat wage, and 18 cents when obtained through low-powered incentives.

What could explain the apparent insensitivity to monetary rewards? A signal (either sent by an informed principal, or by an experimenter completing a contract) is effective only if the subject does not have a strong personal

signal on his own attitude to the task. If subjects perceive the task as enjoyable, interesting, or challenging, then they will be willing to provide the highest possible level of effort even without monetary incentives. This may explain the main result, and raises the next question: is effort responding to different treatments? Although we cannot measure cognitive effort, we do observe the time that each subject took to answer each question and complete the test. Using test completion time as a rough proxy for effort, we can investigate whether this was also, like performance, essentially the same across treatments. We find instead that there are two highly significant differences:

(i) *effort is significantly higher in the treatment with the one euro piece rate* (relative to one cent or no piece rate, and 15 euro participation fee);

(ii) *effort is significantly higher in the treatment with no participation fee and one cent piece rate* (relative to one cent or no piece rate, and 15 euro participation fee).

The first result shows that monetary incentives do have a positive impact on effort. However, this does not generate a corresponding increase in performance. Thus effort may be "too high" in the presence of monetary incentives, as subjects take more time to think about their answers without any appreciable gain in accuracy.

The second result is puzzling. A possible explanation, in the spirit of the incomplete contract view discussed earlier, is that the monetary rewards offered were so low (no participation fee, just one cent per correct answer) that subjects believed this part of the "contract" to be incomplete; i.e. they expected additional monetary rewards to be offered in the course of the experiment. In support of this interpretation, we note that our subjects volunteered to participate in the experiment knowing that experimental earnings on average were roughly ten euros but could vary a great deal depending on the experiment, the treatment, own behavior during the experiment and other participants' behavior. In this particular experiment, after taking the test, participants were asked to guess how many correct answers they had given, and the average number of correct answers in their session. In the treatment with the one cent piece rate and no participation fee, they were offered much higher payments for guessing correctly than in other treatments (10 euros for each correct guess). Thus it is quite plausible that subjects in this treatment, when taking the test, expected additional monetary rewards to be offered later in the experiment, as was indeed the case, without knowing what they might be. If this is the correct interpretation, it suggests that contractual incompleteness can be at least as effective as high-powered incentives in eliciting effort.

When it comes to performance (score) on the test, the main lesson from these findings appears to be not "pay enough or don't pay at all", as in

Gneezy and Rustichini (2000), but rather "don't pay too much". We conjecture that the difference is due to the nature of the task and the setting in which it was undertaken: in the experiment run by Gneezy and Rustichini the questions came from a test that would have been quite familiar to their student participants (a psychometric test used to scan applicants to the university, similar to GMAT). In contrast, our experiment used a real IQ test that was unfamiliar to our subjects. Moreover, participants in each session took the test under identical conditions, and were aware of this. The experiment therefore represented a unique opportunity for them to test their IQ relative to their peers. This seems to have provided sufficient intrinsic motivation for most of them, with little added benefit from high-powered monetary incentives.

If this conjecture is correct, we would expect that if the *same* subjects participated in a follow-up experiment, where they were asked to complete another IQ test, and monetary rewards were made very salient to all participants, then intrinsic motivation would be lower, and monetary incentives might start to matter for performance. This is what we find in our second experiment. Specifically, our results show that the *piece rate has a significant positive effect on performance in the new test*, and that this overall effect is *due primarily to the positive impact* of monetary incentives *on the more talented participants*. Moreover, our subjects behave as if they expected this: when asked to behave as "Principals" and set piece rates for their "Agents", they offer significantly *higher piece rates to the more talented individuals*.

Our combined findings from the two experiments suggest that for tasks requiring talent "pay enough or don't pay at all" applies in the following way: "don't pay at all" when intrinsic motivation is sufficiently high, because monetary incentives in this case elicit too much effort and not enough performance, but "pay enough" when intrinsic motivation is insufficient to motivate talented individuals.

The remainder of the paper is organized as follows. Section 2 describes in detail the experimental design and procedures. Our results are presented in section 3, and our conclusions in section 4.

2 Experimental Design and Procedures

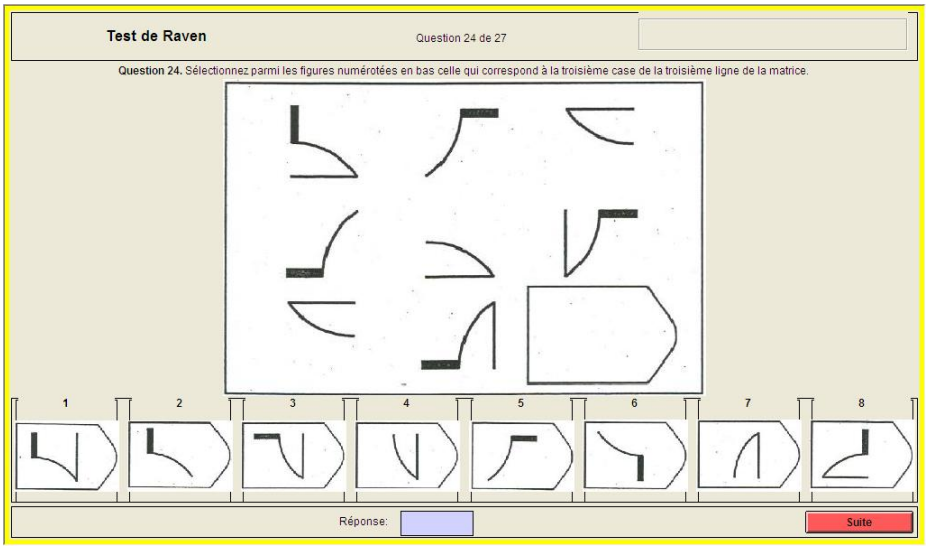
Our subjects were students at the University of Toulouse, and all the experimental sessions were carried out in the experimental laboratory of the Toulouse School of Economics, using the software z-Tree (Fischbacher (2007)).

2.1 Tests

We employed two tests during the experiments. In the first experiment, subjects were given 27 questions from Raven's Advanced Progressive Matrices (Series II). This is an IQ test: a subject is presented with figures sharing a logical pattern, with one slot missing; the task is to detect an abstract rule in the figures and choose the missing figure in a set of feasible options. An example question is shown below. We refer to this test as *Raven27* in what follows. Participants were told, before starting the test, that they would be given 27 questions from a Raven test, "which is often used as a test of intelligence quotient (IQ)".

In the second experiment, subjects were given 12 questions from Raven's Advanced Progressive Matrices (the ones from series II that had not been used previously, and questions from series I). We refer to this test as *Raven12*. Participants were simply told they would be given 12 questions from another Raven test, similar to the first one, but not the same.

Figure 1: **An example of the questions used in the RAM task.**



2.2 First Experiment: Design

Subjects were all given the same test (Raven27). For each question, they could choose one of 8 possible answers, or blank. They had 66 seconds to

reply (the time remaining was shown on their screens), then the question disappeared and the following one appeared on their screen.

Participants were randomly assigned to five different conditions (treatments), and were not aware of any condition except their own. Subjects in any given experimental session all faced the same condition, and this was made clear by reading out some of the instructions that also appeared on each participant's screen. The treatments were:

Treatment 1: in this condition, subjects were told they would be paid 15 euros for their participation in the experiment. There was no mention of any additional payment based on their performance in the test.

Treatment 2: this condition differed from Treatment 1 in one respect: subjects were told they would be paid, in addition to the 15 euro participation fee, 1 cent for each correct answer in the test.

Treatment 3: subjects in this condition were told they would be paid one euro for each correct answer on the test. There was no mention of a participation fee.

Treatment 4: in this condition subjects were asked whether they wished to receive, in addition to the 15 euro participation fee, 1 cent for each correct answer.

Treatment 5: subjects in this condition were told they would be paid one cent for each correct answer on the test. There was no mention of any participation fee.

Subjects in treatments with a participation fee and/or a piece rate were told the relevant amounts before taking the test. In treatments without a participation fee or a piece rate, these were simply not mentioned. This was all the information concerning payments given to participants before they took the test. After the test, all subjects were asked to guess their own score, and the average score in their session (number of correct answers). In all but one treatment they received one euro per correct guess, or three euros for making both guesses correctly. The exception was Treatment 5: here, since subjects were paid no participation fee and only a tiny piece rate on the test (one cent), the payment for each correct guess was much higher (10 euros). This treatment therefore generated some very low earnings. On the other hand, one third of subjects made at least one correct guess, and one participant guessed both own and average score correctly. Subjects had been told in advance (when deciding whether to volunteer for participation) that remuneration would depend on their answers and on the answers given by other participants in the experiment.

2.3 Second Experiment: Design

Subjects in treatments 1, 2, 4 and 5 of the first experiment (i.e. the treatments with zero or one cent piece rates, which nevertheless generated roughly the same average performance as the treatment with the one euro piece rate) were invited to participate in the second experiment. In each session of this experiment, participants were randomly and anonymously assigned to groups of three. Within a given group, each subject was the "Principal" with respect to one of the other two subjects (his "Agent"), and the "Agent" with respect to the other (his "Principal"). Roles were defined neutrally with letters A, B and C.

The Principal-Agent game to be played was explained to all subjects as follows. Each Agent would be given 12 questions from another Raven test (similar to the one used in the first experiment, but not the same). Each Principal would gain 1 euro per correct answer given by his Agent. Each Principal would have to tell his Agent in advance (before the test) his piece rate; that is, the amount that the Principal would pay his Agent for each correct answer.

Principals were asked to choose the piece rates conditional on different hypotheses about their Agent (i.e. using the strategy method). The applicable piece rate was then announced to the Agent before the test. The hypotheses are listed in the Appendix: essentially they distinguish between subjects that had participated in different treatments in the first experiment. This made monetary rewards salient by making participants aware of the different treatments that had been used in the first experiment (except for the one with the one euro piece rate). Importantly, for all but one treatment[1] the hypotheses also distinguish between subjects who had scored above or below 18, the average score. This enabled us to check whether subjects offered different piece rates to agents depending on their performance in the first experiment.

2.4 Experimental Procedures

Subjects were recruited by visiting the first or last 5 minutes of lectures given to undergraduates and Master's students in Economics, Business and Finance, and Law, at the University of Toulouse 1. We informed students that they could, if they wished, volunteer to participate in experiments on decision-making in the Experimental Economics Laboratory of TSE (Toulouse School of Economics), by registering on the Laboratory's recruitment web-

[1]The exception was Treatment 4, where we distinguished instead between subjects who chose to be paid the one cent piece rate and those who chose not to.

site. We told them that sessions could take up to 90 minutes, inclusive of individual confidential payments at the end of the experiment. Payments would depend on their decisions and those of other participants.

In total, 278 subjects participated in the first experiment. We invited those who had attended sessions for treatments 1, 2, 4 and 5 to participate in the second experiment. We organized a smaller number of sessions for the second experiment, yielding a total of 81 subjects.

Table 1 describes participation in the different treatments of the first experiment.

3 Results

3.1 First Experiment

All subjects in this experiment took an IQ test consisting of 27 questions from Raven's Advanced Matrices ("Raven27"). They all took the test under the same time conditions (66 seconds per question). What differed across treatments was the participation fee and the piece rate per correct answer, as described in detail in section 2. Table 2 summarizes the resulting IQ test scores (number of correct answers).

Our first main finding is that there are *no significant differences in performance across treatments*, with pairwise Mann-Whitney tests failing to reject the hypothesis that the data comes from the same population[2]. This finding is confirmed by the results of a Tobit regression[3] for the test score (number of correct answers), reported in Table 3. None of the treatment dummies are statistically significant. In fact, gender and age are the only statistically significant effects. The negative coefficient for age may well reflect the presence in the sample of some "repeat" students among the undergraduates; that is, undergraduates who failed their exams and had to retake the year[4].

These results are quite striking: monetary rewards seem to have very little impact on performance, with an average of 18 correct answers out of

[2]In treatment 4, only 7 out of 50 subjects chose not to receive the one cent piece rate. The difference in performance between subjects who opted to receive the piece rate and those who did not is insignificant ($p = 0.102$).

[3]We just allow for the possibility of censoring at zero, since there are no observations with a full score (i.e. 27 correct answers).

[4]The French educational system allows students who fail to retake the year more than once. Thus a small minority of undergraduates are significantly older than the average undergraduate or Master's student.

Table 1: First Experiment: participation

Treatment	Participants	mean age	% female	% graduates
1	51	21	51	24
2	36	21	61	19
3	105	20	49	37
4	50	21	56	16
5	36	20	53	0

Table 2: First experiment: number of correct answers on Raven27

Test score	Mean	Observations
treatment1	18.04	51
treatment2	18.19	36
treatment 3	19.35	105
treatment4 one cent	18.77	43
treatment4 no cent	16.71	7
treatment 5	18.19	36
Pairwise MW tests		p-value
treatment 1 versus 2		0.779
treatment 1 versus 3		0.133
treatment 2 versus 4 one cent: p=0.560		
treatment 1 versus 5: p=0.976		

Table 3: Tobit regression for IQ test score (Raven27)

Variable	Coefficient	p-value
treatment2	0.129	0.884
treatment3	0.948	0.181
treatment4 one cent	0.667	0.431
treatment4 no cent	-1.097	0.507
treatment 5	0.049	0.956
female	-0.966*	0.051
graduate	0.735	0.221
age	-0.189**	0.050

Observations: 278

27 when subjects are given no piece rate or a one cent piece rate, and 19 correct answers when they choose the one cent piece rate themselves or they are given a one euro piece rate. Moreover, these small differences across treatments are not statistically significant.

From the perspective of a "Principal" hiring our participants as agents, it would seem that the cost of obtaining a performance of 18 correct answers can be as high as 18 euros if he provides high-powered incentives (one euro piece rate, no participation fee) and as low as 18 cents if he provides low-powered incentives (one cent piece rate, no participation fee).

What could explain the apparent insensitivity to monetary rewards? A first conjecture might be that subjects' effort on the task did not respond to monetary incentives: for example, because the task was so enjoyable and interesting that participants were willing to provide the highest possible level of effort even without monetary incentives. Although we cannot measure cognitive effort, we do observe the time that each subject took to answer each question and complete the test. Using test completion time as a rough proxy for effort, we can investigate whether this was also, like performance, essentially the same across treatments.

To calculate completion times, we added the times taken to answer each question. Some subjects occasionally reached the timeout of 66 seconds without recording an answer (neither one of the suggested eight possible answers, nor blank). For these questions, it seems reasonable to assume on average that the subject was not paying attention (providing effort); we therefore do not include the 66 seconds for that particular question in our measure of completion time (effort). In some cases, it is possible that the timeout was reached without doing anything because the subject was thinking hard about the answer and could not decide until the end. We therefore repeated the analysis excluding all the subjects who had reached the timeout on at least one occasion during the test (24 in total, leaving us with 254 participants); this yielded very similar results, reported in the Appendix.

Table 4 presents the main results for the full sample of 278 subjects. Unlike performance (test scores), effort measured by test completion time does vary significantly across treatments. Specifically, pairwise Mann-Whitney tests reveal two significant differences relative to the treatment with participation fee and no piece rate. First, *effort is significantly higher in the treatment with the high* (one euro) *piece rate*. Second, *effort is significantly higher in the treatment with no participation fee and low piece rate* (one cent).

We also estimated a tobit regression, including controls for age, gender and educational status. The estimates are shown in table 5 and lend further support to our finding: the coefficients on the dummy variables for the same two treatments are large and positive, and highly significant.

Table 4: First experiment: time to complete IQ test (Raven27)

Completion time	Mean	Observations
treatment1	990.14	51
treatment2	972.47	36
treatment 3	1089.87	105
treatment4 one cent	1012.33	43
treatment4 no cent	1002.86	7
treatment 5	1103.06	36
Pairwise MW tests		p-value
treatment 1 versus 3		0.001***
treatment 1 versus 5		0.002***

treatment 2 versus 4 one cent: p=0.342
treatment 1 versus 2: p=0.856

Table 5: Tobit regression for IQ test completion time (Raven27)

Variable	Coefficient	p-value
treatment2	-7.163	0.885
treatment3	123.908***	0.002
treatment4 one cent	34.018	0.472
treatment4 no cent	-22.936	0.803
treatment 5	124.892**	0.014
female	-6.770	0.806
graduate	-44.724	0.182
age	14.309***	0.008

Observations: 278

These results shed some light on the insensitivity of performance to monetary rewards found earlier. The first conjecture, that subjects' effort was largely unaffected by monetary incentives, does not seem consistent with the data on test completion times. Offering a high piece rate (one euro) does have a significant positive impact on effort. Surprisingly, so does the offer of a tiny piece rate with no participation fee! A possible explanation for this second finding is that the monetary rewards offered were so low in this case that subjects believed this part of the "contract" to be incomplete; i.e. they expected additional monetary rewards to be offered in the course of the experiment (as was the case), and these expectations were sufficiently optimistic to elicit considerable effort.

However, significantly higher effort did not generate a corresponding increase in performance. Thus high-powered monetary incentives (or optimistic expectations) may elicit "too much" effort, as subjects take more time to think about their answers without any appreciable gain in accuracy. Why? We conjecture that the explanation lies in the nature of the task and the setting in which it was undertaken: our experiment used a real IQ test that was unfamiliar to our subjects. Moreover, participants in each session took the test under identical conditions, and were aware of this. The experiment therefore represented a unique opportunity for them to test their IQ relative to their peers. This appears to have provided sufficient intrinsic motivation for most of them, with little added benefit from high-powered monetary incentives.

If this conjecture is correct, we would expect that if the *same* subjects participated in a follow-up experiment, where they were asked to complete another IQ test, and monetary rewards were made very salient to all participants, then intrinsic motivation would be lower, and monetary incentives might start to matter for performance. We explored this possibility in our second experiment.

3.2 Second Experiment

All subjects in this experiment had previously participated in the first experiment. In the second experiment they all took a new IQ test, consisting of 12 new questions from Raven's Advanced Matrices (Raven12).

The new experimental setting made monetary rewards particularly salient for all participants. Subjects were randomly and anonymously assigned to groups of three. Within each group, every subject played the role of "Principal" relative to one member of the group, and the role of "Agent" relative to the other group member. In his role as Principal, each participant was told he would receive one euro per correct answer to the test given by his

Agent; before the test, he was asked to specify the piece rate he would pay his Agent. In his role as Agent, each participant was told before starting the test the piece rate his Principal would pay him.

3.2.1 Performance on the IQ test

Our data from the test in the first experiment provides a rough proxy for an individual's talent (ability): the ratio of the number of correct answers to the time taken to complete the test. This should be higher for more talented individuals. The variable "talent" is equal to this ratio multiplied by one hundred. The variable "effort" represents the time taken to complete the new test (Raven12). The variable "pay" is equal to the piece rate for the individual taking the new test (known to him when he takes the test).

Our conjecture is that monetary incentives will have a significant impact on performance in this setting, in contrast with the first experiment. We therefore estimate tobit regressions for the test score (number of correct answers to Raven12). The results are presented in table 6.

Column 1 shows that *the piece rate has a highly significant positive effect on performance, controlling for talent*, which is consistent with our conjecture. Column 2 reveals that this effect is due to the positive impact of the piece rate on the performance of *talented* individuals. We also examine the impact of effort on performance: as expected, column 3 shows that it is positive and highly significant, while column 4 makes clear that *what matters is the effort of talented individuals.*

3.2.2 Principals' offers

Our results on performance suggest that a Principal is likely to be better off offering a higher piece rate to talented individuals, thereby achieving a significant increase in performance, than to less talented individuals, which would simply increase the Principal's costs with little gain in performance. Do our experimental subjects offer higher piece rates to their more talented agents? We can easily check this, because piece rates were elicited using the strategy method, giving us piece rate offers by each participant, contingent on the agent's previous performance. Table 7 summarizes average piece rate offers contingent on the treatment the Agent had participated in and his performance in the first experiment (IQ test score above or below average).

We find indeed that piece rate offers are *higher* for Agents who obtained a score above average in the IQ test given to subjects in the first experiment (Raven27). This is true across treatments, and the differences are statistically significant (using the matched pairs Mann-Whitney-Wilcoxson test). No

Table 6: Second experiment: Tobit regressions for IQ test score (Raven12)

	1	2	3	4
Variable				
pay	0.080***	-0.062		
(p-value)	(0.002)	(0.407)		
talent	-0.488	-0.941**	0.133	-1.865**
(p-value)	(0.215)	(0.046)	(0.762)	(0.026)
female	0.608	0.509	0.861	0.546
(p-value)	(0.433)	(0.502)	(0.262)	(0.437)
age	-0.248	-0.228	-0.313*	-0.189
(p-value)	(0.170)	0.198	(0.086)	(0.262)
talent*pay		0.074**		
(p-value)		(0.047)		
effort			0.011***	-0.005
(p-value)			(0.003)	(0.363)
talent*effort				0.010***
(p-value)				(0.000)

Dependent variable: Test score on Raven12 .
Score: number of correct answers.
Talent: (Score for Raven27/completion time)*100
Effort: completion time for Raven12
Pay: piece rate for Raven12
Observations: 81

other differences are significant.

4 Conclusions

We can now provide an overall interpretation of our results. In the first experiment a very strong and clear signal given by the explicit statement that the test "is normally used as a test of IQ (intelligence quotient)" leaves no ambiguity. Subjects acquire a strong belief on the purpose of the activity - it is about testing intelligence, and this has considerable motivational force, particularly in a setting where subjects are competing directly and visibly with their peers, under identical conditions. The additional signal contained in the payment rule then has only a secondary, less important impact. This effect is still strong enough to affect the level of effort.

The situation is reversed in the second experiment, where the payment

rule is made much more salient and becomes the strongest signal concerning the purpose of the activity. The interpretation in terms of signals concerning the purpose of the activity seems to offer a way of reconciling the results of both experiments, which is appealing.

It can account for the fact that in the first experiment, although effort is higher in the presence of high-powered monetary incentives, there is very little difference in performance (test score): the strong explicit signal concerning the purpose of the activity (testing intelligence) already provides sufficient motivation, with little added benefit from monetary rewards.

The same logic helps to shed light on the other significant finding from the first experiment: the higher effort by subjects in the treatment with no participation fee and only a tiny (one cent) piece rate. One way of interpreting this is that the strong signal concerning the purpose of the activity (testing intelligence) leads subjects to be optimistic about additional rewards that may be offered during the course of the experiment, since proving one's intelligence is normally expected to be correlated with tangible rewards.

In the same light, we can also account for the greater pay-performance sensitivity in the second experiment, where the setting sends a strong signal that the purpose of the activity is now earning money.

5 References

[1] Bénabou, R. and J. Tirole (2003) "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*, 70, 489-520.

[2] Fischbacher, U. (2007) "z-tree: Zurich toolbox for ready-made economic experiments", *Experimental Economics*, 10(2), 171-178.

[3] Gneezy, U. and A. Rustichini (2000), "Pay Enough or Don't Pay at All", *Quarterly Journal of Economics*, 115(3), 791-810.

6 Appendix

6.1 Instructions for the first experiment

Welcome; you are about to participate in an economics experiment. Your answers and decisions during this experiment will have no consequence for your grades or your degree. This experiment studies decision-making. During the experiment, you will be asked first of all to answer some questions from a Raven test. This is a test that is often used to test IQ (intelligence quotient).

The experiment will last approximately 40 minutes. {*In treatments with a participation fee, the following sentence was included here: "Each participant will receive 15 euros for his participation." In treatment 3, the following sentence was included instead: "Each participant will be remunerated for each correct answer he or she gives. The amount will be specified in the instructions."*}

We are going to explain the general rules for the experiment in a moment. We now ask you to please switch off your mobile phones. We also ask you not to talk to each other during the experiment. If you have a question, raise your hand and we will come to answer. Are there any questions?

If there are no questions, we can start. You will see some instructions on your screen. Read them carefully before clicking on "next" to proceed to the following screen. If you see the sentence "Waiting for other players" on your screen, it means that everyone must click on "next" before the next screen appears. If you have a question during the experiment, raise your hand.

General instructions:

We are now going to explain the general rules for this experiment. They will be followed by specific instructions.

Your answers during this experiment will have no consequence for other participants, and their answers will have no consequence for you. It is important for you to know that your answers will remain completely anonymous. If you have any questions, raise your hand. If there are no questions, we can give you the specific instructions.

Specific instructions:

You will now be given 27 questions from a Raven test. You will have roughly 30 minutes to answer the 27 questions (more precisely, 66 seconds for each question). Each time, you will see a 3*3 matrix of abstract figures. The third cell in the third line of the matrix will be empty, and you will be asked to select the corresponding figure from a set. We will show you an example before starting the test. At the end, you will see a screen saying "End of the 27 questions", before proceeding to the following screen.

{*In treatments with a piece rate, the following sentence was included here:*

"IMPORTANT : you will receive one cent (in treatment 3, one euro) for each correct answer". In treatment 4, the following sentence was included instead: "Before showing you the example and starting the test, we ask you to choose one of the following options:

OPTION 1: I wish to receive one cent for each correct answer (in addition to the 15 euros for participating in the experiment)

OPTION 2: I do not wish to receive one cent for each correct answer (in addition to the 15 euros for participating in the experiment)."}

Here is an example. You should choose from the set of eight figures below the one that corresponds to the third cell in the third line of the matrix.

Here the correct answer is "5".

If you have any question, raise your hand.

Click on "Next" to begin the 27 questions.

{*Test*}

End of the 27 questions.

Now everyone has completed the Raven test. You are asked to guess the number of correct answers you have given, and the average number of correct answers given in your group. {*Each treatment here specified the payment rule for correct guesses: one euro per correct guess in treatments 1, 2 and 4; three euros for two correct guesses in treatment 3; ten euros per correct guess in treatment 5*}

{*Results screen*}

The experiment has ended. We now ask you to answer the following questionnaire.

6.2 Instructions for the second experiment

Welcome; you are about to participate in an economics experiment. Your answers and decisions during this experiment will have no consequence for your grades or your degree. This experiment studies decision-making. There are no right or wrong decisions - you should simply decide according to your preferences.

The experiment will be remunerated. The amount will depend on your decisions and those of the other participants.

We are going to explain the rules for the experiment in a moment. We now ask you to please switch off your mobile phones. We also ask you not to talk to each other during the experiment. If you have a question, raise your hand and we will come to answer. Are there any questions?

If there are no questions, we can start. You will see some instructions on your screen. Read them carefully before clicking on "next" to proceed to the following screen. If you see the sentence "Waiting for other players" on your

screen, it means that everyone must click on "next" before the next screen appears. If you have a question during the experiment, raise your hand.

General instructions:

We are now going to explain the general rules for this experiment. They will be followed by specific instructions.

During this experiment you will sometimes be asked to take decisions that will have consequences for you and for other participants. It is important for you to know that your decisions will remain completely anonymous.

Each individual will be assigned to a group of three, using the ID code you selected at the beginning. You will never know who were the other members of your group, and they will never know that you were in their group. In each group of three, there will be a participant "A", a participant "B", and a participant "C". Each individual will learn his role (A, B or C) shortly. When referring to other members of your group, we will always use their letter (A, B or C) and never their ID code or other information that might allow you to identify them.

If you have any questions, raise your hand. If there are no questions, we can give you the specific instructions.

Specific instructions:

In this experiment, each participant will answer 12 questions from a Raven test (similar to the one used in the last experiment, but not the same).

In each group of three, "A" will receive 1 Euro per correct answer given by "B". "B" will receive 1 Euro per correct answer given by "C". "C" will receive 1 Euro per correct answer given by "A".

Before starting the test, "A" has to choose the amount he will pay "B" for each correct answer given by "B". Similarly "B" has to choose the amount he will pay "C" for each correct answer given by "C", and "C" has to choose the amount he will pay "A" for each correct answer given by "A". The chosen amounts will be communicated to the receivers before starting the test.

Example: "A" chooses 50 cents, this amount is communicated to "B"; "B" chooses 40 cents, this amount is communicated to "C"; "C" chooses 30 cents, this amount is communicated to "A". In this case, "A" knows that he will receive 50 cents per correct answer given by "B" (1 Euro minus the 50 cents he has chosen to pay "B"). He also knows he will receive 30 cents from "C" for each correct answer he will give himself.

{*Each subject then receives the instructions corresponding to his role. To save space we report those for "A"; the ones for "B" and "C" are identical except for the letters*}

You are the "A" member of the group. You will shortly be asked to choose the amount X that you will give to "B" for each correct answer. You will then receive 1 Euro minus X for each correct answer given by "B". You will

be asked to specify your choice under several possible hypotheses concerning "B"; we will use the choice corresponding to the correct hypothesis.

Reminder: this choice will be communicated to "B" before starting the test.

Important: all participants in today's experiment also participated in the last experiment. In that experiment, each participant answered the same 27 questions from a Raven test, under the same time constraint (maximum 66 seconds for each question). However, some aspects of the experiment were not the same in each session. We will give you more information about these shortly, in the form of ten hypotheses, before asking you each time to specify your choice. Take the time you need to read carefully each hypothesis.

{*The ten hypotheses are presented two at a time on the screen, with decisions being taken on each screen before proceeding to the following*}

{*Screen 1*}: In the last experiment, "B" participated in a session where the instructions specified: "During the experiment, you will be asked first of all to answer some questions from a Raven test. This is a test that is often used to test IQ (intelligence quotient). The experiment will last approximately 40 minutes. Each participant will receive 15 euros for his participation." Concerning the Raven test, the instructions specified: "You will have roughly 30 minutes to answer the 27 questions (more precisely, 66 seconds for each question). At the end of the experiment, you will see on your screen the number of correct answers you gave, and the average number of correct answers in your group".

Before starting the test, each participant had to decide whether he wished to receive 1 cent per correct answer on the Raven test or not.

Hypothesis 1: "B" decided not to receive 1 cent per correct answer.

Hypothesis 2: "B" decided to receive 1 cent per correct answer.

{*Screen 2*}: In the last experiment, "B" participated in a session where the instructions specified: "During the experiment, you will be asked first of all to answer some questions from a Raven test. This is a test that is often used to test IQ (intelligence quotient). The experiment will last approximately 40 minutes. Each participant will receive 15 euros for his participation." Concerning the Raven test, the instructions specified: "You will have roughly 30 minutes to answer the 27 questions (more precisely, 66 seconds for each question)." The instructions did not specify that at the end each participant would learn the number of correct answers he had given, and the average number of correct answers in his group.

Before starting the test, each participant had to decide whether he wished to receive 1 cent per correct answer on the Raven test or not.

Hypothesis 3: "B" decided not to receive 1 cent per correct answer.

Hypothesis 4: "B" decided to receive 1 cent per correct answer.

{*Screen 3*}: In the last experiment, "B" participated in a session where the instructions specified: "During the experiment, you will be asked first of all to answer some questions from a Raven test. This is a test that is often used to test IQ (intelligence quotient). The experiment will last approximately 40 minutes. Each participant will receive 15 euros for his participation." Concerning the Raven test, the instructions specified: "You will have roughly 30 minutes to answer the 27 questions (more precisely, 66 seconds for each question)." The instructions did not specify that at the end each participant would learn the number of correct answers he had given, and the average number of correct answers in his group.

Hypothesis 5: There was no mention of a payment per correct answer (beyond the 15 euros for participating in the experiment). The number of correct answers given by "B" was greater than or equal to 18.

Hypothesis 6: The instructions specified that 1 cent would be paid for each correct answer (beyond the 15 euros for participating in the experiment). The number of correct answers given by "B" was greater than or equal to 18.

{*Screen 4*}: In the last experiment, "B" participated in a session where the instructions specified: "During the experiment, you will be asked first of all to answer some questions from a Raven test. This is a test that is often used to test IQ (intelligence quotient). The experiment will last approximately 40 minutes. Each participant will receive 15 euros for his participation." Concerning the Raven test, the instructions specified: "You will have roughly 30 minutes to answer the 27 questions (more precisely, 66 seconds for each question)." The instructions did not specify that at the end each participant would learn the number of correct answers he had given, and the average number of correct answers in his group.

Hypothesis 7: There was no mention of a payment per correct answer (beyond the 15 euros for participating in the experiment). The number of correct answers given by "B" was less than 18.

Hypothesis 8: The instructions specified that 1 cent would be paid for each correct answer (beyond the 15 euros for participating in the experiment). The number of correct answers given by "B" was less than 18.

{*Screen 5*}: In the last experiment, "B" participated in a session where the instructions specified: "During the experiment, you will be asked first of all to answer some questions from a Raven test. This is a test that is often used to test IQ (intelligence quotient). The experiment will last approximately 40 minutes." Concerning the Raven test, the instructions specified: "You will have roughly 30 minutes to answer the 27 questions (more precisely, 66 seconds for each question)." The instructions did not specify that at the end each participant would learn the number of correct answers he

had given, and the average number of correct answers in his group. Concerning payment, the instructions specified that 1 cent would be paid for each correct answer. There was no mention of any payment for participating in the experiment.

Hypothesis 9: The number of correct answers given by "B" was greater than or equal to 18.

Hypothesis 10: The number of correct answers given by "B" was less than 18.

{*After making all ten choices, "A" saw the following instructions (B and C saw identical ones except for the letters)*}

"C" has decided to give you ??? cents for each correct answer you give to the 12 questions in the test. For each question you will have maximum 66 seconds to answer. Click on "Next" to start the test.

{*Test*}

You chose to pay ??? cents per correct answer given by "B". How many correct answers do you think "B" has given? You will receive an additional euro if your answer is correct.

{*Results screen*}

The experiment has ended. We now ask you to answer the following questionnaire.

6.3 Further results: robustness

As noted in section 3, we repeated the analysis of effort in the first experiment excluding all the subjects who had reached the 66 second timeout on at least one occasion during the test (24 in total, leaving us with 254 participants). The results, analogous to those presented in section 3 for the full sample of 278 subjects, are presented below.

Table 7: Second Experiment: Principals' piece rate offers

Treatment	Agent's score in first test	Mean offer (cents)
1	score>18	14.679
1	score<18	13.062
2	score>18	15.691
2	score<18	12.593
5	score>18	16.272
5	score<18	13.617

Agent chose one-cent piece rate: mean offer is 14.482
Agent rejected one-cent piece rate: mean offer is 16.605

Table 8: First experiment: time to complete IQ test (Raven27), excluding subjects who reached the timeout at least once during the test

Completion time	Mean	Observations
treatment1	968.32	44
treatment2	967.00	34
treatment 3	1089.41	99
treatment4 one cent	1020.24	38
treatment4 no cent	1002.86	7
treatment 5	1098.60	32
Pairwise MW tests		p-value
treatment 1 versus 3		0.0002***
treatment 1 versus 5		0.0006***

treatment 2 versus 4 one cent: p=0.313
treatment 1 versus 2: p=0.936

Table 9: Tobit regression for IQ test completion time (Raven27), excluding subjects who reached the timeout at least once during the test

Variable	Coefficient	p-value
treatment2	7.240	0.883
treatment3	145.623***	0.000
treatment4 one cent	-3.508	0.969
treatment4 no cent	63.001	0.191
treatment 5	135.415***	0.008
female	-22.997	0.404
graduate	-66.678**	0.048
age	13.779**	0.014

Observations: 254