WORKING PAPERS

N° TSE - 567

April 2015

# "Strong intrinsic motivation"

Roberta Dessi and Aldo Ristichini

INSTITUT D'ECONOMIE INDUSTRIELLE

Toulouse School of Economics

# STRONG INTRINSIC MOTIVATION

ROBERTA DESSÍ AND ALDO RUSTICHINI

ABSTRACT. A large literature in psychology, and more recently in economics, has argued that monetary rewards can reduce intrinsic motivation. We investigate whether the negative impact persists when intrinsic motivation is strong, and test this hypothesis experimentally focusing on the motivation to undertake interesting and challenging tasks, informative about individual ability.

We find that this type of task can generate strong intrinsic motivation, that is impervious to the effect of monetary incentives, particularly when the individual is "racing against himself". In our experiments, monetary incentives have no significant impact on performance. In a second experiment using the same kind of task but a setting designed to weaken intrinsic motivation, monetary incentives do have a significant, positive, effect on performance. This result confirms that our experimental setup may, with appropriate conditions, replicate the known crowding out effects.

## 1. INTRODUCTION

A large literature in psychology, and more recently in economics, has argued that monetary rewards can reduce intrinsic motivation. We investigate this experimentally, focusing on the motivation to undertake interesting and challenging tasks, informative about individual ability. From solving puzzles to climbing mountains or writing music, from scientific research to art, technological innovation or design, these tasks are at the heart of human activities, at work and at play. Intrinsic motivation refers to individuals' motivation to perform well on such tasks independently of any external, notably monetary, rewards.

Experimental evidence by Gneezy and Rustichini (GR, [6]) suggests that offering performance-contingent monetary rewards can indeed have a detrimental effect on intrinsic motivation. They compare the performance of student subjects on a quiz consisting of 50 questions from a test normally used to scan applicants for university admission. The questions were chosen

*Date*: April 20, 2015.

We would like to thank Gary Charness, Uri Gneezy, Paul Seabright, Jean Tirole, seminar and workshop participants in HKUST, NUS and TSE, for helpful comments and suggestions, and Samuele Centorrino and Olivier Darmouni for excellent research assistance. Supported in part by grants from the National Science Foundation (SES1357877) and the ANR (ANR-09-FASHS-050).

with an emphasis on reasoning and computation rather than general knowledge. In the experiment, all student participants took the same test under the same conditions, except for the reward scheme which varied across treatments. In the baseline treatment, participants were simply offered a fixed participation fee (60 New Israeli Shekels, NIS), and then given the test. In the other treatments, participants were offered, in addition to the participation fee, a payment for each correct answer on the test. The results reveal a non-monotonic relationship between the piece rate (payment per correct answer) and performance on the test (number of correct answers). Specifically, average performance was higher in the baseline treatment, where there was no mention of any payment per correct answer, than in the treatment where participants were offered a low piece rate (0.1 NIS). With a higher piece rate (1 NIS or more), average performance was higher than in the baseline treatment.

A variety of possible explanations have been put forward for this result. For example, List and Rasul ([8]) argue that "This might either be because small monetary incentives crowd out intrinsic motivation ... or because the individual is reluctant to signal his willingness to accept low wages"; while Falk and Kosfeld ([4]) suggest that "incentives may undermine motivation... because they insult the agent". GR favored an "incomplete contract" interpretation: the experimental instructions can be viewed as an incomplete contract guiding the interaction between experimenter and subjects. When a participation fee is offered, and no mention is made of any payment per correct answer, subjects "complete" the contract by inferring that they have been paid for their participation, and they are now expected to do as well as they can on the test - this is their side of the bargain. In contrast, when a piece rate is offered, subjects infer that the effort they provide on the test should respond to the piece rate. Thus a small piece rate may elicit very little effort.

In this paper, we shed light on this debate, and on the broader question of the impact of monetary incentives on intrinsic motivation. We design and carry out an experiment similar to the one originally carried out by GR, but with some important differences. The task faced by participants in our experiment is to answer 27 questions taken 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. 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)". The experiment was computerized: each subject saw each question on his screen, and had 66 seconds to answer, before the image on the screen disappeared and the test moved on to the following question.

Subjects in the experiment were randomly assigned to one of five treatments. Three of these are analogous to treatments in GR: in the first, there

is no mention of a payment per correct answer; in the second, subjects are offered a very low payment (one cent of a euro) per correct answer; in the third, 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 third treatment where the piece rate is high. A fourth treatment asks subjects to *choose* whether they wish to receive, in addition to the 15-euro participation fee, one cent per correct answer. Thus subjects who would find the very low piece rate insulting can choose not to receive it; similarly, subjects who do not wish to signal a willingness to work for a low piece rate can refuse the piece rate (while keeping the participation fee). The last treatment simply offers one cent per correct answer (and no participation fee)[1].

We find that in the treatment where subjects are given the option to choose whether to receive one cent per correct answer or not, 86% choose to receive the one cent piece rate, suggesting that the majority of participants do not feel insulted by a low piece rate, and are not concerned about signaling their willingness to work for a low piece rate. More strikingly when compared to previous results, we also find that *small monetary incentives do not appear to crowd out intrinsic motivation*: subjects offered a very small piece rate (one cent) answered correctly an average of 18.19 questions, as compared with an average of 18.04 correct answers for subjects in the treatment with no mention of a payment per correct answer[2]. Why do we not replicate the earlier finding of a non-monotonic relationship between performance and the piece rate?

In fact, our results differ from those obtained by GR ([6]) more broadly, in the following way: we find that performance, measured by the number of correct answers on the IQ test, shows no significant difference across treatments. In other words, *the relationship between performance and the piece rate is essentially flat.* In particular, subjects answer correctly approximately 18 questions out of 27 irrespective of treatment (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). These results are surprising: 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, as well as with the previous findings by GR. It implies that for a principal hiring our participants as agents, 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? One possible explanation lies in the nature of the task: this represented an opportunity for participants to test their ability (IQ), in a race against themselves.

[1]After completing the IQ test, participants in this treatment had an opportunity to earn higher rewards from answering other questions.

[2]The difference between these two treatments is not statistically significant.

This could be expected to generate strong intrinsic motivation, as confidence in ability is a key determinant of self-esteem. In this interpretation, our results show that intrinsic motivation of this kind can be very powerful, and is largely unaffected by monetary incentives. An alternative explanation is that performance on an IQ test like Raven depends only on ability, not "effort", making monetary incentives simply irrelevant.

If this second interpretation were correct, we should find that if the *same* subjects participate in a follow-up experiment, where they are asked to complete another IQ test but under conditions that could be expected to weaken their intrinsic motivation, once again performance will not vary significantly with monetary incentives. In our second experiment, subjects (who had participated in the first experiment) once again took an IQ test, consisting of 12 questions from Raven's Advanced Progressive Matrices that had not been used previously. This time each subject received a payment per correct answer determined by another subject, the "Principal". Moreover, the Principal received one euro for each correct answer given by the subject taking the test (the "Agent"). Every participant knew this, and learned his piece rate before taking the test. We expected intrinsic motivation to be lower in this second experiment, because instead of participating in a race against himself, each subject would be essentially working for an employer (the Principal).

We found that in the second experiment the *piece rate had a significant positive effect on performance*, and that this overall effect was *due primarily to the positive impact* of monetary incentives *on the more talented participants* (those who had performed better on the IQ test in the first experiment). Moreover, our subjects behaved as if they expected this: when asked to behave as "Principals" and set piece rates for their "Agents", they offered significantly *higher piece rates to the more talented individuals*.

Our combined findings from the two experiments suggest that strong intrinsic motivation was driving performance in the first experiment, and was not crowded out by monetary incentives. The results were not simply due to monetary incentives being irrelevant for performance on IQ tests, as shown in the second experiment where the piece rate did have a significant effect on performance. In this second experiment, intrinsic motivation was weakened by making the task more like working for an employer. Moreover, intrinsic motivation was weakened more for subjects who had performed well in the first experiment, since they had already proved to themselves that they could achieve a high score in terms of IQ. This is consistent with the finding that the positive overall impact of the piece rate was due primarily to its effect on the performance of individuals who had an above-average score on the first test.

We have found no evidence that monetary incentives crowd out intrinsic motivation - at least for interesting and challenging tasks. When intrinsic motivation is sufficiently strong, monetary incentives are not needed. They do play a (positive) role again as soon as intrinsic motivation is weakened

in other ways. The optimal design of incentives must therefore consider carefully the factors that can enhance or weaken intrinsic motivation, and how this interacts with monetary rewards.

The remainder of the paper is organized as follows. The final part of this section relates our work to the existing literature in economics and psychology. Section 2 describes in detail the experimental design and procedures. Our results are presented in section 3, and our conclusions in section 4.

1.1. **Interesting tasks and motivation.** Our work is related to previous studies in both psychology and economics. In psychology, Deci ([2]) showed that external rewards could undermine intrinsic motivation for interesting tasks. The approach taken in this and most subsequent studies in psychology was to identify "intrinsic motivation" as the extent to which participants return to and persist on tasks after completion of the experimental phase in which external rewards were applicable. The reason is that intrinsic and extrinsic motivation would both be (potentially) relevant during the phase where external rewards are applicable, making it difficult to disentangle the two. However, these studies typically focus on the amount of time devoted to tasks in different phases, not performance. Moreover, they investigate how behavior of the same individuals evolves over time, as the reward condition is varied experimentally. Behavior over time may change for different reasons though, including learning effects.

GR ([6]) rely instead on a between-subject, single-phase design. This makes it possible to examine the effect on performance of varying rewards. Intrinsic motivation can then be identified with the performance in the baseline treatment, where no payment per correct answer is offered. We build on this approach in our first experiment, as discussed above. We also exploit the value of a within-subject design in our second experiment, where we are able to distinguish between the effects of monetary incentives on above-average and below-average talented individuals, based on their test scores in the first experiment.

Intrinsic motivation for interesting tasks has recently attracted attention in the theoretical literature in economics. Bénabou and Tirole ([1]) develop a Principal-Agent model where the Principal has superior information about the characteristics of the Agent (e.g. his talent) or of the task (e.g. whether it is interesting or dull, whether it is hard or easy). In this kind of setting, the offer of a performance-contingent payment by the Principal can signal "bad news" to the Agent: for example, that the task is uninteresting, or very hard. In our experiment, the first inference is unlikely, as we clearly explain to participants that the task is an IQ test, and we use Raven's picture-based test, which requires the ability to detect a logical pattern but no computational or language skills, or prior knowledge[3]. Subjects could, on the other hand, interpret the offer of a piece rate as indicating that

[3]Inferences concerning the individual participant's characteristics are clearly ruled out by the experimental setting.

the task is hard - in the first experiment, where the offer is made by the experimenter, who has superior information about the nature of the task (unlike the Principals in the second experiment). As discussed earlier, we find no evidence that such an inference, if it was made, had a negative impact on performance.

Our paper is also related to a different literature in economics, which studies gender differences in responses to performance incentives. Much of this literature has focused on how men and women respond differently to the degree of competitiveness of the environment. For example, Gneezy, Niederle and Rustichini ([7]) found that women perform less well in more competitive environments, with a significant gender gap emerging in tournaments, while no such gap was present when piece rates were used. Dohmen and Falk ([3]) find that women are less likely to select into variable-pay schemes (piece rates, tournaments, and revenue-sharing) relative to fixed-pay schemes. We focus instead on how monetary incentives affect motivation to perform well on interesting and challenging tasks. In section 3.2, we examine gender differences in performance for the different treatments in our first experiment. Our main finding is that there is no difference in performance in the baseline treatment, where no payment per correct answer is offered and performance is driven by intrinsic motivation. A significant gender gap emerges instead in two of our treatments, where subjects are given a piece rate, and is particularly marked for the treatment with the low (one-cent) piece rate, where men's average performance is higher and women's average performance is lower than in the baseline treatment. Our results therefore suggest that men and women's intrinsic motivation respond differently to monetary incentives.

## 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 ([5]).

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. We refer to this test as *Raven 27* 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 *Raven 12*. Subjects (who had all participated in the first experiment a month

earlier) were simply told they would be given 12 questions from another Raven test, similar to the first one, but not the same.

2.2. **First Experiment: Design.** Subjects were all given the same test (*Raven 27*). 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). 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 euros 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.* This condition differed from Treatment 1 in one respect: subjects were asked whether they wished to receive, in addition to the 15 euros 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. Importantly, for all but one treatment[4] 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 website. 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: Performance and Monetary Incentives.** All subjects in this experiment took an IQ test consisting of 27 questions from

[4]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.

Raven's Advanced Matrices (*Raven 27*). 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 *monetary incentives do not appear to crowd out intrinsic motivation*: subjects offered a very small piece rate (one cent) answered correctly an average of 18.19 questions, as compared with an average of 18.04 correct answers for subjects in the treatment with no mention of a payment per correct answer. Thus our subjects do not appear to have felt insulted by the offer of a one-cent piece rate, or to have interpreted it as a negative signal about the nature of the task they were facing. Among those who were given the option to choose whether to receive one cent per correct answer or not, 86% chose to receive the one cent piece rate, suggesting that they were not worried about signaling their willingness to work for a low wage.

Why do we not replicate the crowding-out effect found by GR ([6]), where a very small piece rate induced a worse performance than no piece rate? Table 2 reveals that our results differ from theirs more broadly, in the following way: there are *no significant differences in performance across treatments*,

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 *Raven 27*

| Test score | Mean | Observations |
|---|---|---|
| treatment 1 | 18.04 | 51 |
| treatment 2 | 18.19 | 36 |
| treatment 3 | 19.35 | 105 |
| treatment 4 | 18.48 | 50 |
| treatment 4 one cent | 18.77 | 43 |
| treatment 4 no cent | 16.71 | 7 |
| treatment 5 | 18.19 | 36 |
| treatment 1 versus 2 | | $p = 0.779$ |
| treatment 1 versus 3 | | $p = 0.133$ |
| treatment 2 versus 4 one cent: $p = 0.560$ | | |
| treatment 1 versus 5: $p = 0.976$ | | |

with pairwise Mann-Whitney tests failing to reject the hypothesis that the data comes from the same population[5]. This finding is confirmed by the results of a Tobit regression[6] 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[7].

Table 3: Tobit regression for IQ test score in *Raven 27*.

| Variable | Coefficient | $p-value$ |
|---|---|---|
| treatment2 | 0.129 | 0.884 |
| treatment3 | 0.948 | 0.181 |
| treatment 4 one cent | 0.667 | 0.431 |
| treatment 4 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

These results are quite striking, and surprising: *monetary rewards seem to have hardly any impact on performance*, with an average of 18 correct answers out of 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).

[5]We highlight the more interesting tests in Table 2: treatment 1 versus treatment 2 (comparing the treatment with a one-cent piece rate with the treatment with no mention of a piece rate), and treatment 1 versus treatment 3 (comparing the treatment with a one-euro piece rate with the treatment with no mention of a piece rate). The reported p-values indicate that the differences are not significant. 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 also insignificant ($p = 0.102$).

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

[7]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.

What could explain the apparent insensitivity to monetary rewards? Looking more closely at the nature of the task suggests a possible explanation. Subjects were given questions from Raven's Advanced Matrices and were clearly told before starting that this "is often used as a test of intelligence quotient (IQ)". The task therefore represented an opportunity to test one's *ability*, independent of prior knowledge or preparation.[8] This could be expected to generate considerable *intrinsic motivation*, since confidence in one's ability is a very important determinant of self-esteem. Moreover, each subject's attention during the test was focused on their own performance - how many correct answers they could give to the 27 questions in the time available (66 seconds per question). This could be expected to be particularly motivating for those who enjoy "racing against themselves": the goal was to perform to the best of their abilities in the allotted time. Our results suggest that this form of intrinsic motivation can be very powerful. This is important, not least because in many settings it can play a role quite distinct from the motivation induced by tournaments - indeed, the relationship between the two merits further investigation.

3.2. **First Experiment: Gender Differences.** The results from the Tobit regression for performance on the IQ test, presented in Table 3, showed a significant gender difference: women performed less well than men. In view of our suggested interpretation of the overall results in terms of intrinsic motivation, we now explore this gender difference further. In particular, could it be the case that men's and women's intrinsic motivation respond differently to monetary incentives? Table 4 shows average performance in each treatment, separately for men and women.

Table 4: Mean number of correct answers on *Raven 27*, by gender.

| Treatment | Men | Women |
|---|---|---|
| treatment 1 | 18.04 | 18.04 |
| treatment 2 | 20.36 | 16.82 |
| treatment 3 | 19.91 | 18.77 |
| treatment 4 | 17.82 | 19.00 |
| treatment 5 | 18.77 | 17.68 |
| treatment 4 one cent | 17.53 | 19.58 |
| treatment 4 no cent | 18.80 | 11.50 |

** Wilcoxon Mann Whitney rank-sum test, $p - value 0.028$

[8] In this respect our task differed from the one used in GR ([6]): in their experiment, students from the University of Haifa were told "they would be asked to answer a quiz consisting of 50 problems taken out of a psychometric test used to scan applicants to the university". Moreover, "The problems in the quiz were chosen to make the probability of a correct answer depend mostly on effort" (p.796).

The table shows that average performance (the mean number of correct answers) is identical for men and women in treatment 1, the one where subjects are simply paid a participation fee to take part in the experiment and there is no mention of a payment per correct answer on the IQ test. Men perform better than women, on average, in treatments 2, 3 and 5, i.e. the treatments where subjects are told they will be given a payment per correct answer on the test (the payment being either one cent or one euro). Women, on the other hand, perform better in treatment 4, the one where subjects are asked to choose whether they wish to receive a payment per correct answer (one cent). The result for treatment 4 is driven by the performance of subjects who choose to receive the one cent per correct answer (as shown in Table 1, only seven participants chose not to receive the one cent per correct answer).

To see which of these differences are significant when we control for age and education, we estimate Tobit regressions for the IQ test score separately for each treatment. Table 5 presents the estimated coefficients for the three variables, age, education and gender, in each regression. [9]

Table 5: Tobit regressions for IQ test score (*Raven 27*), by treatment

| | Treatment 1 | Treatment 2 | Treatment 3 | Treatment 4 | Treatment 5 |
|---|---|---|---|---|---|
| Graduate | 0.135 | 2.158 | 0.839 | 2.712 | - |
| Age | -0.363* | 0.011 | -0.073 | -0.249 | 0.019 |
| Female | -0.403 | -3.729** | -1.183* | 1.119 | -1.076 |

Table 5 shows that the gender difference is significant, when we control for age and education, in two cases, treatment 2 and treatment 3, where subjects are given, respectively, a one-cent and a one-euro piece rate. These results suggest that indeed, men's and women's intrinsic motivation respond differently to monetary incentives. In particular, we find no gender difference in performance when there is no mention of monetary incentives, and subjects are simply paid a fixed participation fee for taking part in the experiment (treatment 1). Here, presumably, performance is driven entirely by intrinsic motivation. In contrast, we do find a significant gender difference in performance when subjects are offered a payment per correct answer. Relative to the case with no monetary incentives, average performance for men increases. For women, average performance increases when the piece rate is high (less than for men), while it decreases when the piece rate is low. As a consequence, performance differs significantly between men and women once monetary incentives are introduced.

[9]The education variable in the regression for treatment 5 is omitted because of collinearity.

Gender differences in average performance are small nevertheless: intrinsic motivation appears to be powerful for both men and women, with and without monetary incentives.

3.3. **Second Experiment: Performance and Monetary Incentives.** We have suggested that performance in our first experiment was driven primarily by strong intrinsic motivation. An alternative explanation for the relative insensitivity of performance to monetary rewards could be the following: performance in an IQ test depends purely on ability and not on "effort", making monetary incentives irrelevant. Subjects had to wait until the end of the experiment before leaving the lab, so they might as well respond to the questions - the more clever ones answered correctly more often than the less clever ones.

This alternative explanation would have the following implication: if we were to run another experiment, in which participants take a similar IQ test under similar time conditions, we should again find that monetary incentives have no significant impact on performance. Our second experiment tests this implication. The participants had all taken part in the first experiment. The following month they were invited to participate in another experiment studying decision-making. Of the 81 who registered to take part, 46 were women, and the mean age was 20.86.

In this second experiment, all subjects took a new IQ test, consisting of 12 new questions from Raven's Advanced Matrices (*Raven 12*). They had, once again, 66 seconds to answer each question. While the test and time conditions were similar to those in the first experiment, the new experimental setting, described in detail in section 2, was intended to generate less intrinsic motivation. In particular, each participant knew that his performance this time would benefit another participant, chosen randomly and anonymously (the "Principal"), who would receive one euro for each correct answer. Moreover, every subject knew the piece rate chosen for him by his "Principal", i.e. the amount he would receive himself for every correct answer. Instead of participating in a "race against himself" as in the first experiment, each subject taking the test in the second experiment was essentially working for another subject, in return for a piece rate chosen by his "employer". We conjectured that this would weaken intrinsic motivation.

3.3.1. *Performance on the IQ test (Raven 12).* Our data from the first experiment provides a rough proxy for an individual's talent (ability): the number of correct answers on the first test (*Raven 27*).[10] We can therefore study how performance on the second test varied with the piece rate, controlling for talent (performance on the first test). Table 6 presents the estimates from Tobit regressions for the test score on *Raven 12* (number of correct answers).

[10]Indeed, under the hypothesis that performance on an IQ test depends only on ability and not on "effort", which is the competing explanation we are trying to test with our second experiment, the previous test score should be a very good proxy for ability.

The first column shows the results for the whole sample. As discussed earlier, if the relative insensitivity to monetary incentives found in the first experiment were due to performance on IQ tests being determined only by ability (and not "effort"), we should find that the piece rate has no significant effect on performance in the second experiment.

Column 1 shows instead that *the piece rate has a highly significant positive effect on performance, controlling for talent.* This is consistent with our suggested interpretation, that performance in the first experiment was driven mainly by strong intrinsic motivation, which is weakened in the second experiment. Columns 2 and 3 present the results for the same regression estimated separately for the subsamples of individuals who had performed better than the average on the first experiment (talent $> 18$), and those who had an average or below-average performance (talent $\leq 18$). We find that the result for the whole sample in Column 1 was due to the positive impact of the piece rate on the performance of *talented* individuals.

3.3.2. *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 increase the Principal's costs with an insignificant 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 Principal, 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).

Table 6: Second experiment: Tobit regressions for IQ test score (*Raven 12*). Dependent variable: Test score on *Raven 12*. Score: number of correct answers. Talent: Test score for *Raven 27*. Pay: piece rate for *Raven 12*.

| | 1 | 2 | 3 |
|---|---|---|---|
| | Whole sample | Talent $> 18$ | Talent$\leq 18$ |
| pay | 0.077*** | 0.104*** | 0.046 |
| (SD) | (0.026) | (0.038) | (0.034) |
| talent | 0.132 | 0.516 | 0.199 |
| (SD) | (0.087) | (0.283) | (0.161) |
| female | 0.567 | 0.023 | 1.252 |
| (SD) | (0.775) | (1.099) | (1.038) |
| age | -0.180 | -0.179 | -0.063 |
| (SD) | (0.182) | (0.247) | (0.257) |
| Observations | 81 | 49 | 32 |

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 (*Raven 27*). This is true across treatments, and the differences are statistically significant (using the matched pairs Mann-Whitney-Wilcoxson test). No other differences are significant.

Table 7: Second Experiment: Principals' piece rate offers

| Treatment | Agent's score in first test | Mean offer (in 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

## 4. Conclusions

Monetary incentives, appropriately designed, are a powerful, stable, reliable motivation, and are the essential condition for orderly running of everyday activity in modern western societies. On the other hand, an important part of our history would have been different, if motivation had to rely exclusively on monetary incentives; or even worse, if monetary incentives had perverse effects on the most creative, risk taking, innovative activity. Still, small monetary incentive may have a negative effect on motivation. The discussion in recent years has introduced and tested several possible explanations of this phenomenon; and a common thread for all these theories is the idea that acceptance of small monetary incentives may act through signaling of the agent that he is willing to work for small compensations. Which precisely is the type that is being signaled is object of debate; but surely the inclination to avoid cheapening oneself with accepting a small reward is a large part of the effect.

We introduce the idea of strong incentive motivation, that is intrinsic motivation that is resistant to the negative effect of small compensations, and the negative signal attached with accepting them. We test the idea experimentally and find that indeed crowding out is absent. We emphasize that our negative results, which are the crucial ones, are not driven by a special feature of our setup. In a slightly different design, where the intrinsic motivation is weaker, we find results consistent with standard results.

Since the negative implication on one's type attached to performing well for a small pay is still active, we may wonder what drives our results. One

reason may be simply that some intrinsic motivation is too compelling to be discouraged by considerations about the signal on one's type. In this view, strong intrinsic motivation is resistant because it taps on fundamental motivations like pride and social competition. There might be an additional reason for the failure of crowding out: a possible interesting conjecture is that there is a widespread social recognition that when tasks are a tests of skill, and socially perceived to be so, a positive performance with small rewards enhances the skill signal, at the expense of the signal on the willingness to provide effort for small pay. Since performance is approximately a function of the product of skill and effort, having a good performance in spite of the small reward increases the weight in the posterior on the skill component.

## References

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

[2] Deci, Edward L., (1971) "Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105-115.

[3] Dohmen, Thomas, and Armin Falk. (2011). "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review*, 101(2): 556-90.

[4] Falk, Armin, and Michael Kosfeld. (2006). "The Hidden Costs of Control." *American Economic Review*, 96(5): 1611-1630.

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

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

[7] Uri Gneezy, Muriel Niederle and Aldo Rustichini, (2003), Performance in Competitive Environments: Gender Differences, *The Quarterly Journal of Economics*, 118 (3): 1049-1074.

[8] John A. List and Imran Rasul, (2011) *Field Experiments in Labor Economics*, Chapter 2 in Handbook of Labor Economics Volume 4a, O. Ashenfelter and D. Card (editors), Elsevier, 104–228.

## 5. Appendix

5.1. **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.

5.2. **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.

(Roberta Dessí) Toulouse School of Economics and CEPR.
*E-mail address*: roberta.dessi@tse-fr.eu

(Aldo Rustichini) Department of Economics, University of Minnesota, 1925 4th Street South 4-101, Hanson Hall, Hanson Hall, Minneapolis, MN, 55455
*E-mail address*: aldo.rustichini@gmail.com