DOES RELATIVE THINKING EXIST IN MIXED COMPENSATION SCHEMES?

Ofer H. Azar[*]

Ben-Gurion University of the Negev

Abstract

Several earlier studies show that people exhibit "relative thinking": they consider relative price differences even when only absolute price differences are relevant. This article examines whether relative thinking exists when people face mixed compensation schemes that include both fixed and pay-for-performance components. Such compensation schemes are prevalent in many occupations (e.g., salespeople and managers). Surprisingly, the ratio between the pay-for-performance and the fixed compensation does not affect effort. Two additional experiments show that neither reciprocity nor financial incentives are the reason for the results. The results have implications for designing incentive schemes in firms and for designing experiments.

1. Introduction

Many experiments conducted over the last three decades suggest that people often consider relative price differences even when only absolute price differences should matter, a behavior that was recently denoted "relative thinking" (Azar, 2004). The seminal experiment in this literature is the one reported in Tversky and Kahneman (1981). They asked people whether they were willing to drive 20 minutes in order to save $5 on a calculator when they were going to buy a calculator and a jacket. When the calculator's price was $15 and the jacket's price was $125, 68 percent of the subjects were willing to drive, but when the calculator's price was $125 and the jacket's price was $15, only 29 percent were willing to drive.[1] Similar results were later obtained in a few additional experiments. Mowen and Mowen (1986) showed that not only students, but also business managers, exhibit this behavior. Ranyard and Abdel-Nabi (1993) varied the price of the second item (the jacket) and obtained similar results, and Frisch (1993) showed that the effect holds also when only a calculator is being purchased. Azar (in press) showed that when subjects can purchase a certain good either in a store they currently visit or in a remote store, the minimal price difference for which they are willing to travel to the remote store is an increasing function of the price of the good they want to purchase. In an experiment that included 9 different price-treatments, he found that people behave (on average) as if the value of their time is approximately proportional to the square root of the good's price.

---

[1] To see why such behavior does not fit standard economic theory, suppose, in line of the experimental results above, that one is willing to drive 20 minutes to save $5 on a certain cheap item, but then refuses to drive 20 minutes to save $8 on a more expensive item. This person could make the exact same effort of driving and yet be richer by $3 if he made the opposite choices, suggesting that his behavior is irrational.

While the studies mentioned above focus on the trade-off between spending time and finding a cheaper price of a certain good, recent research shows that a similar behavior exists when people consider differentiated goods or services and have a trade-off between the quality difference and the price difference. In scenarios in which only the absolute price difference should matter, people are affected also by the relative price difference. For example, Azar (2004) showed that consumers' willingness to add money for a high-quality good or service (over the price of a low-quality substitute good or service) is higher when the good's price is higher. The quality difference in his experiments was unrelated to the good's price and therefore the willingness to add should be independent of the good's price. The results were obtained both with undergraduate students and with economists participating in the 2003 North American Summer Meetings of the Econometric Society.

The literature discussed above suggests that there is a strong bias of relative thinking when people compare prices. An interesting question that was never explored before is whether people also exhibit relative thinking when they receive money rather than spend it. For example, when one receives a payment for a job he does, and he compares two possible payments (for example from two different employers, or the expected payment for low effort versus high effort), do the relative differences between the payments affect his decisions, just as relative price differences do? Intuitively, we can expect to observe a similar relative thinking effect. After all, compensation for work done is the price of labor, and if people exhibit relative thinking with respect to other prices, it seems reasonable that they also do so with respect to the price of labor.

One important implication of relative thinking when receiving payments is for mixed compensation schemes that include both a fixed component and a variable component that

depends on performance.[2] Such compensation schemes are very common in real life. For example, salespeople often receive a base salary plus a percentage (or another function) of their sales. Similarly, many workers in the investment banking and consulting industries, as well as managers at different levels throughout the economy, receive a base salary plus a performance-based bonus. In other cases, managers receive a salary and also have options or stocks of their firm, so their compensation consists of their salary plus the return on these options or stocks, which in turn depends on their performance. It is therefore of great importance to examine whether people exhibit relative thinking in the context of mixed compensation schemes. Better understanding of this issue, for example, can help firms choose the optimal mix of the fixed and variable components in the compensation schemes of millions of workers.

How does relative thinking relate to mixed compensation schemes? Relative thinking suggests that people consider relative magnitudes in addition to absolute magnitudes even when economic theory implies that only the latter should matter. When a person has to choose how much effort to exert in a certain task, relative thinking implies that he considers not only the absolute amount he can earn by exerting more effort, but also the relative increase in his earnings; that is, he might compare the additional earnings (due to higher effort) to his base salary (or to his total compensation). Consequently, a larger base salary may reduce effort because it makes the pay-for-performance bonus look smaller. This can have a dramatic effect on the optimal choice of incentive schemes by firms. Currently, the common wisdom is that increasing the fixed component of the salary can either have no effect on effort (because it does not depend on

---

[2] Increasing the sensitivity of pay to performance is generally believed to improve effort and productivity, although very high reward levels sometimes can have a detrimental effect on performance (Ariely, Gneezy, Loewenstein and Mazar, 2008).

performance), or it may increase effort due to worker's reciprocity[3] or because of efficiency-wage arguments.[4] Relative thinking suggests that an opposite effect might also be present, and that if this effect is strong enough to outweigh the reciprocity and efficiency-wage effects (if these are present), then increasing the fixed payment to workers might in fact reduce effort.

The rest of the article is organized as follows. The next section presents the main experiment. The experiment shows that when people receive a fixed payment and a bonus for each task completed successfully, the relative magnitude of the bonus (compared to the fixed payment) does not affect their effort. Section 3 presents the second experiment, which uses different levels of a fixed payment but without a pay-for-performance bonus to identify the effect of reciprocity. The joint analysis of the results from the two experiments suggests that people do not exhibit relative thinking in this context, and thus that the strong bias of relative thinking does not carry over from the domain of money paid (prices) to the domain of money received (compensation). In order to identify whether relative thinking disappears (compared to hypothetical-questions experiments about price differences) because of the different context or because of the introduction of financial incentives, Section 4 presents another experiment that is almost identical

---

[3] The idea is that the worker feels more grateful to his employer when he receives a higher salary, and because he wants to reciprocate, his effort is an increasing function of his salary (even though the salary is fixed and does not depend on performance). There is abundant experimental evidence for such behavior (that is sometimes described as "gift exchange"); for a review of some of this experimental research, see Fehr and Gachter (2000). Gneezy and List (2006), however, find in a field experiment that paying more increases worker effort in the first few hours on the job, but not afterwards.

[4] The efficiency-wage argument says that when a worker receives a higher salary, he becomes more afraid of losing his job (because the loss caused by losing the job and obtaining another job becomes higher), and therefore he exerts more effort on the job in order to reduce the risk that he will be fired.

to the first, but uses hypothetical questions without financial incentives. Relative thinking is not detected there either, suggesting that the different context, and not the introduction of financial incentives, is what eliminated relative thinking in the first experiment. The last section concludes.

2. Experiment 1: Mixed Compensation Schemes – The Incentive Condition

2.1. Experimental design

The main purpose of the experiment was to test whether in the context of task performance with mixed compensation schemes the bias of relative thinking exists. In order to create a decision problem in which the prediction of economic theory differs from that of relative thinking in a testable manner, the experiment involved a fixed payment and a variable payment. The variable payment was identical in all treatments – each correct answer increased the subject's earnings by 0.15 Shekels (1 Shekel was about $0.22 at the time the experiment was run). The fixed payment, however, differed in the two treatments. After subjects agreed to participate in the experiment, the amount of the fixed payment was randomly assigned: in the low-fee treatment, the subject received a participation fee of 5 Shekels, while in the high-fee treatment he received 15 Shekels.[5] In hypothetical questions about consumer decisions, a ratio of 1:3 in the price

---

[5] The range of possible fixed payments is limited for several reasons. A very low payment might create resentment and uncooperativeness of the subject. A payment too high might create an effect of willingness to stay in the experiment longer because of reciprocity issues (see the discussion below). In addition, research shows that very high reward levels have unexpected effects on behavior (see for example Ariely, Gneezy, Loewenstein and Mazar, 2008). The values of 5 and 15 Shekels were chosen so that they are in the reasonable range given the time the subject had to spend in the experiment, and yet they create a significant difference in the treatment in relative terms – one payment being three times higher than the other. The usual salary for students at the time was around 20 Shekels per

treatments is sufficient to produce a stark difference that is highly statistically significant in responses (see for example Azar, 2004).[6]

Because economic theory suggests that only incentives that depend on effort should affect the choice of effort, it follows that there should be no difference in effort in the two treatments. If relative thinking carries over from the domain of prices to the domain of payments for task performance, however, it suggests that the relative magnitude of incentives also plays a role in the choice of effort. When relative magnitudes are also considered, the same payment per correct answer seems larger when it is compared to a smaller fixed payment. This implies that subjects should make more effort in the low-fee treatment.

If the amounts involved were huge, economic theory could make a similar prediction because of wealth effects: when participation fee is higher the subject becomes richer, his marginal utility from money decreases, and therefore he makes less effort. But with the negligible amounts involved (the difference in the participation fee between the low-fee and high-fee treatments is slightly more than $2), it seems safe to assume that wealth effects can be ignored. Moreover, as we will see later, the results in fact indicate that subjects did not exert more effort in the low-fee treatment, and in addition, Part C of the experiment provides additional evidence that wealth effects do not seem to be an issue here.

One potential confounding factor in the experiment could be reciprocity. Many studies show that people want to reciprocate and are willing to pay a price to do so (for a review of some of

hour, and the students were recruited on campus and participated in the experiment on the spot (they did not have to come to a different place).

[6] For example, in the WTA version and the Econometric Society sample in Azar (2004), a ratio of 1:3 in prices yielded one-tailed t-test p-values of 0.002 or smaller in three different consumer decision problems, despite the small number of observations in this sample and version (13 in the high-price treatment and 13 in the low-price treatment).

this literature, see Fehr and Gachter, 2000). A subject who receives a higher participation fee might feel more willing to reciprocate, and if he thinks that the experimenter wants him to stay as much as possible in the experiment, this can lead to more effort in the high-fee treatment (in opposite direction to the possible effect of relative thinking). In order to avoid this potential confounding effect, two measures were taken.

The first measure was telling the subject several times, both in the consent form and in the instructions of Part B of the experiment (see Appendix A), that he can solve as many questions as he wants, he can quit anytime, and he is allowed not to solve the questions in Part B at all. This should eliminate the potential subject's belief that the experimenter wants him to stay in the experiment as much as possible and therefore that staying longer is a way to reciprocate.

The second measure was adding Part A at the beginning of the experiment. In this part, the subject was asked to answer four questions about various consumer decision problems. The participation fee of 5 or 15 Shekels was described as a payment for answering this part of the experiment. Consequently, after answering Part A subjects presumably felt that they completed the task for which they were paid (answering these four questions) and therefore they did not feel obligated to solve questions in Part B just in order to reciprocate and "pay with their time" for the participation fee.

In Part B the subjects were given the opportunity to solve up to 40 questions, where each correct answer increased their earnings by 0.15 Shekels. In order to give relative thinking its best chance, the earnings in all the parts of the experiment (Parts A, B, and the lottery in Part C if chosen) were paid together at the end. In addition, the earnings in Part B were described as a bonus in percentage of the amount earned in Part A (see Appendix A). It seems that if relative thinking exists in the context of task performance, it should be stronger under these conditions

8

than if each part was paid separately and the earnings in Part B were described as 0.15 Shekels for each correct answer.[7]

The purpose of the questions was to measure effort, and therefore they were designed as a particularly boring task: answering which letter appears on a certain page, in a certain line and a certain location. Each subject was given nine pages with letters (each subject received the same letters, but not the actual pages used before by another subject, to avoid the chance that a previous subject marked something on the pages). Three different measures of effort were employed. First, the number of attempted questions was recorded. Second, accuracy is presumably also positively correlated with effort, so the number of correct answers was a second measure, which takes also accuracy into account.

Finally, the research assistants also marked (without the subject paying attention to it) the time in which they gave the subject parts A and B (which were stapled together; the detailed description of the experimental procedure appears in Appendix A) and the time in which these parts were returned, thus allowing to compute the time the subject spent answering these two parts together. Because we are interested in the effort in Part B, this measure is a little noisy because it includes also the time it took the subject to answer part A. The variation in the time spent on Part A, however, is much smaller than the variation in time dedicated to Part B, because in Part A every subject answered the same questions, while in Part B some subjects did not answer any question while others solved up to 40 questions. Consequently, the amount of noise incorporated in this measure is not very large. Obviously, the three measures of effort are highly

---

[7] Paying for Part A immediately when it is finished and then paying separately for Part B will reduce the connection between the two parts and therefore the chances that the subject will think about the payment for Part B in comparison to the payment in Part A, thus encouraging less a behavior of relative thinking.

correlated; yet, looking at three different measures provides some additional robustness to the results.

To avoid the noise that can result from peer pressure and social influences among subjects (e.g., subjects continuing to solve questions as long as others do not leave the experiment, and stopping to solve when most others left), the experiment was conducted individually and not in large groups.[8] This design also allowed to give subjects their earnings as soon as they finished the experiment, without having to wait for others, thus creating a significant opportunity cost to solving more questions.[9] The cost of running the experiment individually was that it required many hours of research assistance, since the research assistants supervised the subjects during the experiment.

In order to avoid a situation where almost all subjects try to solve everything (or not to solve anything), the questions were presented in an increasing difficulty order (the font size became smaller and the pages became more congested as the subject proceeded in the questions). This design worked nicely and indeed there was significant heterogeneity in the number of questions attempted by different subjects (and therefore also in the number of correct answers and the time spent on parts A and B), as reflected in Table 1. The subjects were recruited on the campuses of two large Israeli universities, and were mostly undergraduate students in various academic fields.

---

[8] See Falk and Ichino (2006) for an experimental article documenting the effects of peer pressure in task performance.

[9] If many subjects participate simultaneously with everyone being paid together in the end, there is no significant opportunity cost to solving more questions, and in fact it might be less boring to solve questions than to wait for others to finish. Such an alternative design therefore undermines the purpose of measuring effort by the number of questions attempted or answered correctly or by the time dedicated to the task.

In total 227 subjects participated in this experiment, 118 in the low-fee treatment and 109 in the high-fee treatment (the numbers are unequal because treatment was randomized using a dice).

2.2. Results and discussion

[Insert Tables 1 and 2 about here]

Table 1 presents the distribution of the numbers of attempted questions, correct answers, and time devoted to Parts A and B. Table 2 reports the means, medians and standard deviations for these three effort variables, as well as the p-values of the t-test for difference in means and the p-values of the Mann-Whitney test. The tests suggest that there is no statistically significant difference in any effort variable between the two treatments.

As mentioned earlier, a potential reason for lower effort of subjects in the high-fee treatment could be wealth effects together with the concavity of the utility function. It was explained that the amounts involved are too small for wealth effects to play a role, and moreover, the results reveal that there is no significant difference in effort between the two base-fee treatments. Nevertheless, as a robustness check (and because beforehand it was not known that there would be no difference in effort), a fair lottery was incorporated in Part C of the experiment. Subjects could accept or reject a fair lottery in which they had a probability of 1/6 to win 5 Shekels, and a probability of 5/6 to lose 1 Shekel. Every subject who has a decreasing marginal utility from money even for changes of a few Shekels (which is the condition for wealth effects to be present in the choice of effort) should reject the lottery even if no transaction costs are involved (rejection should be even more common if transaction costs exist).[10] The result was that out of 227 subjects,

---

[10] Notice that the reverse is not true: a subject who rejects the lottery does not necessarily do so because his marginal utility from money is decreasing for changes of a few Shekels. He might reject the lottery but actually be indifferent

202 chose to accept the lottery. This large fraction of acceptances provides additional evidence that wealth effects do not seem to play a role in the decision of subjects about how many questions to solve.

Finally, in Part D subjects were asked to what extent the percentage increase and the absolute increase in earnings they could gain by solving more questions affected their decision when to stop solving the questions. We can mark the answer to the first question by p (importance of percentage increase) and to the second question by a (importance of absolute increase) and define a variable that measures relative thinking on a 0-1 scale as follows:[11]

$$R = [1.25p/(p + a)] - 0.125 .$$

The variable R satisfies several desired properties. It increases in p and decreases in a; it is equal to 1 in the extreme case that $p = 9$ and $a = 1$ and equal to 0 in the other extreme case (when $p = 1$ and $a = 9$); and it is equal to 0.5 when $p = a$. Thus, R can be interpreted as a measure of relative thinking expressed by the subject in the debriefing part (Part D). If there is a relationship between what people actually do and how they interpret their thoughts and actions in the debriefing part, then we should expect that the negative effect of high base fee on effort (if it exists) will be stronger for people with higher values of R.

[Insert Table 3 about here]

Table 3 presents the results of several regressions. The dependent variable is one of the three measures of effort: ATTEMPT (the number of questions attempted), CORRECT (the number of

---

whether to accept it or not, or he might reject it because of transaction costs (having to spend more time in the experiment).

[11] The sample means of p, a and R are 4.30, 3.83, and 0.54, and their standard deviations 3.03, 2.99, and 0.20, respectively.

correct answers), and TIME (the time in minutes the subject spent on parts A and B together). The independent variables include the dummy variables MALE (equal to 1 for males), HIGHBASE (equal to 1 in the treatment with the high base fee), and BGU (equal to 1 for subjects in Ben-Gurion University of the Negev and 0 for subjects in Tel Aviv University). Additional independent variables are R, which measures the extent of relative thinking reflected in the debriefing questions as explained above, and two interaction terms.

The results show that with all three dependent variables, the effect of the base fee is small and statistically insignificant. This implies that the difference in the relative magnitude of the payment for each correct answer between the two treatments did not affect the effort subjects made to solve questions. In addition, the interaction term between HIGHBASE and R is positive, although it is not statistically significant. Notice that if answers to Part D were consistent with behavior, a higher value of R would reflect more relative thinking, which in turn implies that this interaction term should be negative (the more relative thinking a subject exhibits, the larger the negative effect of high base-fee on effort). This suggests that asking people whether they thought about absolute or relative differences may be uninformative in measuring the extent of relative thinking of different people.

The results are intriguing. As mentioned in the introduction, many experiments show that people exhibit strong relative thinking when answering hypothetical questions about the effort they would make to find a lower price of a good they want to purchase, or about their willingness to pay for two differentiated goods. Surprisingly, however, we fail to detect here any relative thinking, despite giving it its best shot by integrating the payment on Parts A and B and by stating the bonus per correct answer as a percentage of the earnings in Part A rather than as an absolute amount. This seems to suggest that people do not exhibit relative thinking in the context of task performance with mixed compensation schemes. Before we can be confident in this

conclusion, however, there are two issues that should be addressed: possible reciprocity, and the introduction of financial incentives. The next two sections address these issues.

3. Experiment 2: Testing for Reciprocity

3.1. Experimental design

As was briefly discussed in the previous section, one concern in the experiment was that it might yield an effect of reciprocity. If subjects believe that the experimenter wants them to stay as long as possible in the experiment, it is possible that they will stay longer and solve more questions when they are paid more. This creates an effect that is in the opposite direction to relative thinking. The experiment tried to overcome the reciprocity issue by adding part A, so that when they get to part B subjects can feel that they already did what was expected of them and reciprocated for the participation fee they received. In addition, subjects were told that they can solve as many questions as they like in part B and can quit anytime, even without solving any question, to convey to them that the experimenter is not expecting them to stay as much as possible in the experiment.

If the results showed evidence of relative thinking, it would not be crucial to determine whether the measures taken to eliminate reciprocity worked or not, because reciprocity causes subjects to make more effort when paid the high participation fee whereas relative thinking results in more effort under the low participation fee treatment. So evidence for relative thinking in Experiment 1 captures the lower bound for the actual level of relative thinking, because it is the net effect of relative thinking and the possible opposite effect of reciprocity.

The absence of evidence for relative thinking makes the interpretation of the results somewhat difficult. One possibility is that subjects exhibit neither relative thinking nor reciprocity. Another possibility is that both relative thinking and reciprocity are present, but cancel out each other. Therefore, it seems helpful to design another experiment that will allow us to distinguish between these two interpretations.

One alternative is to conduct an experiment that reduces even further the possibility of reciprocity. The problem is that to do this one needs to separate more between the payment of the participation fee and the pay-for-performance component. For example, by having the subjects participate in an experiment with a fixed participation fee (that differs between the two treatments) on one day, and come to an experiment with a pay-for-performance payment (and no fixed payment) on the next day. The problem is that such separation, while reducing reciprocity, also reduces the chance that subjects will exhibit relative thinking. To give relative thinking a chance, the fixed payment and the pay-for-performance payment should be relatively close in time and context, but such closeness also makes reciprocity more likely.

Consequently, another approach was taken to address the reciprocity issue. The experiment was re-run with the same two levels of participation fees, but without any pay-for-performance component. Subjects solved Part A as before, and in Part B they were given the same questions as before but without any bonus for solving them. Because there is no pay-for-performance component here, the issue of relative thinking – subjects viewing the pay-for-performance component as more substantial when the participation fee is lower – is irrelevant. However, the possibility of reciprocity still exists in the same way as in the original experiment. If subjects want to reciprocate for the participation fee they receive and therefore solve more questions when the participation fee is higher, we should observe this also in this experiment.

After the subjects returned Parts A and B they received Part C in which they were asked several questions, one of which asks them directly how many questions they would have solved if they had received a different payment (5 Shekels if they received 15 and vice versa). Appendix B presents additional details about the instructions in the experiment. Experiment 2 was run with 99 subjects in the same universities as in Experiment 1 (Ben-Gurion University of the Negev and Tel Aviv University), who were randomly assigned to one of the two participation fee treatments.

3.2. Results and discussion

[Insert Table 4 about here]

The results of Experiment 2 show no significant difference between the two participation fee treatments in all three effort variables: attempted questions, correct answers, and the time dedicated to Parts A and B. Table 4 reports the means, medians and standard deviations for these three variables, as well as the p-values of the t-test for difference in means and the p-values of the Mann-Whitney test. The tests suggest that there is no statistically significant difference in any effort variable between the two treatments. This implies that the different participation fee did not trigger a behavior of reciprocity according to which subjects who received a higher participation fee made more effort to solve questions in Part B. It follows that in Experiment 1, the correct interpretation of the results is that subjects show neither relative thinking nor reciprocity.

[Insert Table 5 about here]

Further support for this conclusion is obtained in regressions of the effort measures. Let us denote the pay-for-performance experiment by a dummy variable PFP, which equals 1 in Experiment 1 and equals 0 in Experiment 2. The three independent variables that capture the treatment effects in the regressions are HIGHBASE, PFP, and the interaction of the two

16

variables, HIGHBASE*PFP. The coefficient of HIGHBASE measures the effect of the participation fee treatment that is unrelated to the question whether the pay-for-performance component exists or not; that is, HIGHBASE captures the possible effect of reciprocity. Table 5 shows that this coefficient is never statistically significant.

The coefficient of PFP captures the impact of adding a bonus for each correct answer, apart from the impact that is related also to HIGHBASE. Not surprisingly, this coefficient is always positive, large, and statistically significant: when subjects are paid a bonus for correct answers they solve more questions than when they are only paid their fixed participation fee. The data show that in Experiment 1 subjects attempted on average 22.7 questions (the first and third quartiles being 10 and 40) and solved correctly on average 20.1 questions (the first and third quartiles being 9 and 33). In Experiment 2 subjects attempted on average 12.5 questions (the first and third quartiles being 6 and 16) and solved correctly on average 10.5 questions (the first and third quartiles being 5 and 13). Thus, on one hand, subjects solved a significant number of questions even when no bonus was offered for doing so. On the other hand, the bonus did increase significantly the number of questions attempted (by 82%) and solved correctly (by 91%).

The coefficient of the interaction effect, HIGHBASE*PFP, captures the effect of jointly being paid a bonus for correct answers and of being in the high participation fee treatment. This joint effect is the possible impact of relative thinking: if relative thinking exists, it should result in a negative coefficient of HIGHBASE, but only in the mixed-compensation scheme, i.e., when PFP is equal to 1. The coefficient of HIGHBASE already takes away the effect of reciprocity, and thus the interaction term HIGHBASE*PFP captures the net effect of relative thinking. Table 5 shows that this interaction term has a negative coefficient – the direction predicted by relative thinking – but that this coefficient is never statistically significant. Thus, the regressions also

17

support the same conclusion that Table 4 suggested: the combination of Experiments 1 and 2 do not provide evidence for either relative thinking or reciprocity.

Finally, it is also interesting to see what Part C of the experiment (see Appendix B) tells us about reciprocity. As opposed to the rest of the experiment, which attempts to find out the effect of reciprocity using a between-subject experiment, Part C tries to elicit a within-subject measure of reciprocity by asking the subject directly how many fewer or additional questions he would have solved if he had been paid differently (5 Shekels instead of 15 or vice versa). Of the 99 subjects, 80 marked option A, which means that they would have solved the same number of questions even if they had been paid differently for their participation. Of the other 19 subjects, three exhibited preferences that are in opposite direction to reciprocity, i.e., according to their answers they solve more questions when being paid a smaller participation fee. Of the remaining 16 subjects whose answers matched the direction of reciprocity (solving more questions when being paid more as a participation fee), the extent of reciprocity was rather small. Only two subjects stated that the higher participation fee would cause them to solve more than seven additional questions compared to the lower participation fee. Thus, the results of Part C further reinforce the conclusion that reciprocity did not play an important role in the experiment, from which it follows that the results in Experiment 1 can be interpreted as showing no relative thinking.

4. Experiment 3: Mixed Compensation Schemes – The Hypothetical Condition

4.1. Experimental design

By now it is clear that the results of Experiment 1 show no evidence of relative thinking. However, it is not clear whether relative thinking disappeared here because people do not exhibit

relative thinking in the context of mixed compensation schemes, or because of the introduction of financial incentives.[12] In psychology it is very common and acceptable to conduct experiments that lack financial incentives, but in economics some researchers are skeptic about results that come from experiments without financial incentives.[13] Since all the previous studies showing relative thinking used hypothetical questions, while the experiment here included financial incentives, it might be the case that the introduction of financial incentives, and not the different context, is what eliminated relative thinking here. In order to determine whether relative thinking was not detected because of the different context or the introduction of financial incentives, another experiment was conducted by replicating Experiment 1 but without financial incentives (see Appendix C for more details). Subjects went through the same steps as in Experiment 1, but instead of actually solving the questions for real money, they were asked hypothetically (after solving three questions of different difficulty levels to get a feeling of the task) how many questions they think they would have chosen to solve if they had received payment for each correct answer (the payment described was identical to that used in the incentive condition experiment).

While the incentive condition experiment differs from the previous literature both in its context (task performance vs. price comparisons) and in providing financial incentives, in the hypothetical condition only the context is different, because like the previous literature, this experiment also lacks financial incentives. Consequently, the results can suggest what eliminates

---

[12] "Financial incentives" here and below should be interpreted as incentives that depend on the subject's performance, not just a fixed participation fee.

[13] Several articles discuss the issue of financial incentives and how they affect experimental results, see for example Jenkins et al. (1998), Camerer and Hogarth (1999), Hertwig and Ortmann (2001a, 2001b, 2003) and Harrison and List (2004).

relative thinking in Experiment 1. If relative thinking is detected in the hypothetical condition, while it was not documented in the incentive condition, this will suggest that providing financial incentives eliminates the relative thinking bias. If relative thinking is not detected here, this will imply that the different context is what eliminates relative thinking, and that people do not exhibit relative thinking in the context of task performance with mixed compensation schemes.

The subject pool in the hypothetical condition experiment was similar to that in the incentive condition, and consisted of 89 students in the same two universities as before. Appendix C includes the wording of the experiment in this condition when it differs from that of the incentive condition. Three different measures of effort were used: how many questions the subject believed he would attempt, how many questions he thought he would solve correctly, and how much time he thought it would take him. Notice that this last measure is not directly comparable to that in the incentive condition, because there the time recorded includes the time spent on answering Part A, reading the instructions in Part B, and answering the questions in Part B, and here the question refers only to how much time solving the questions in Part B would take. This does not prevent us from using this effort measure, however, because the goal is to compare the responses between the low-fee and high-fee treatments, and not to measure whether subjects correctly evaluate how much time it will take them to solve questions.

4.2. Results and discussion

[Insert Table 6 about here]

Table 6 presents some summary statistics and the results of the t-tests for difference in means and the Mann-Whitney tests that examine whether the three effort measures are different between the low-fee and the high-fee treatments. It is easy to see that the means are almost identical in the

two treatments and that none of the tests performed shows a statistically significant difference between the two treatments, indicating that subjects exhibit no relative thinking.

[Insert Table 7 about here]

To check the robustness of this conclusion, the data from Experiments 1 and 3 were combined, and a dummy variable HYPOTHETICAL which equals 1 in the hypothetical condition and 0 in the incentive condition was defined.[14] The regressions that include both the hypothetical and incentive conditions (Experiments 1 and 3) are reported in Table 7. The coefficient of the interaction between HIGHBASE and HYPOTHETICAL is very close to zero and is statistically insignificant. If relative thinking existed in task performance, and financial incentives were the reason it disappeared in the incentive condition, this interaction coefficient should be negative.

Thus, the results in Tables 6 and 7 suggest that subjects in the hypothetical condition did not exhibit relative thinking. Since this condition lacks financial incentives and its conclusion of no relative thinking is similar to the conclusion in the incentive condition, we can infer that what eliminates relative thinking in these experiments compared to the previous literature is the different context and not the introduction of financial incentives.

The statistically significant coefficient of HYPOTHETICAL when the dependent variable is TIME is not surprising, because in the incentive condition TIME includes the time it took the subject to answer Part A, to read the instructions of Part B, and to answer the questions in Part B, whereas in the hypothetical condition TIME only refers to how much time the subject thought it

---

[14] Because Experiment 2 does not include a pay-for-performance component it is irrelevant for the purposes of this section and therefore is not included in the analysis reported here.

would take him to solve the questions in Part B. Other than HYPOTHETICAL, none of the independent variables is statistically significant at the 5% level.

The two conditions also allow us to examine two other questions about the differences between the behavior people think they will choose and what they actually choose. The separation between the subjects in the two conditions is an advantage: if the same person is first asked what he thinks he will do (e.g., how many questions he will attempt solving) and then asked to actually do something (e.g., solve questions), his response in the first part can influence what he chooses to do in the second part (for example because he wants to be consistent with what he said earlier). Consequently, it will not be surprising in such an experiment to find that what people think they will do is similar to what they later do. Here, however, the people who answer the hypothetical questions are not the ones who actually make choices, so it is less obvious whether the answers in the two conditions are going to be similar. Because the characteristics of the subject pool in the two conditions are similar, if people are good at predicting their behavior (at least in contexts similar to the one in the experiments) then the answers to the hypothetical questions should be similar to the actual behavior in the incentive condition.

We can see in Table 7 that the coefficient of HYPOTHETICAL in the regression of ATTEMPT is negative but not statistically significant, indicating that people slightly underestimate the number of questions they will solve. A closer look at the data, however, reveals a striking difference in the percentage of people who do not solve any question: in the hypothetical condition, 18 out of 86 subjects (20.9%) who answered that question indicated that they would not attempt to solve any questions. In the incentive condition, only 7 out of 227 subjects (3.1%) did not try to solve any question. A possible reason for this difference may be that the subjects in the incentive condition wanted at least to get a feeling of what the task

requires, while the subjects in the hypothetical condition were asked anyway to solve three questions first in order to give them a feeling of the task so that their responses for the hypothetical questions would be more meaningful.

The coefficient of HYPOTHETICAL in the regression of CORRECT shows that subjects were quite accurate in their prediction of how many questions they would solve correctly. Notice, however, that CORRECT is affected both by how many questions the subject attempts and by how accurate he is. To examine accuracy only, we can consider the percentage of questions solved correctly, CORRECT / ATTEMPT, as a measure of accuracy. Are people optimistic, thinking they will make less mistakes than they actually do? It turns out that this effect is very small and not statistically significant: average accuracy in the incentive condition is 88.4% and in the hypothetical condition it is 90.1%, and the p-value of the t-test for difference in means is 0.328.

5. Conclusion

Several previous studies show that when people compare prices of the same good in different stores or prices of differentiated goods, the relative price differences affect their decisions even when economic theory implies that only absolute price differences should matter, a phenomenon that was recently denoted "relative thinking." This article presents experiments that examine whether relative thinking also exists in the context of task performance with mixed compensation schemes that include both a fixed and a pay-for-performance components. The results show that the ratio between the pay-for-performance compensation and the fixed compensation does not affect effort. A second experiment is conducted to test whether this result indicates no relative thinking, or opposite effects of relative thinking and reciprocity cancelling out each other. The

results suggest that reciprocity does not play a role here, and therefore that the results in the first experiment imply that people do not exhibit relative thinking in this context.

The first experiment included financial incentives, whereas previous studies on relative thinking used hypothetical questions. This raises the question what is the reason that relative thinking disappears here: is it the different context (payments received for task performance instead of price comparisons), or the introduction of financial incentives? To address this question, a hypothetical experiment was run. This experiment was almost identical to the first experiment, but instead of making actual decisions about how much time to stay and how many questions to solve, the subjects were asked hypothetically about these decisions. No relative thinking was documented in this experiment either. This suggests that in the context of task performance with mixed compensation schemes there is no relative thinking, regardless of whether financial incentives are provided or not.

This is an important finding given that it is the first examination of relative thinking in this context, and given the prevalence of mixed compensation schemes in the remuneration of managers, salespeople, and other workers in various industries. It may suggest that firms need not fear that increasing the fixed payment to workers will reduce the perceived magnitude of the pay-for-performance component.

In addition to the implications for worker compensation schemes, the results are also relevant when designing the incentives in other experiments. In particular, this study shows that varying the fixed participation fee does not affect the perceived magnitude of the pay-for-performance component. This suggests that experimenters can choose a participation fee that will bring the average subject's earnings to the desired level, without worrying that this payment will erode the impact of incentives provided by the pay-for-performance component of the experiment.

Finally, the experiment shows that the behavior of relative thinking, according to which people consider relative differences in addition to (or instead of) absolute differences, is limited in scope and does not apply in every context. Although people exhibit strong relative thinking in the context of comparing prices of goods, they turn out not to exhibit relative thinking when considering amounts of money that they receive as compensation.

Appendix A: Experiment 1 (Mixed Compensation Schemes – The Incentive Condition) [15]

Procedure: First the subject received an informed-consent form to sign. After returning it, the base-fee treatment was chosen randomly and the subject was handed the appropriate version of Part A and Part B stapled together and the 9 pages containing the letters on which the questions in Part B are based. When he returned them, he received Part C and Part D stapled together. After filling Part C, if he was interested, the lottery was conducted. At the end he received his payment in cash according to the amount he earned in all the parts of the experiment together.

<u>Consent form</u>

The purpose of this experiment is to examine various aspects of decision making. In the first part you will be asked to answer a few short questions. The second part is optional: you can end it and return the questionnaire anytime, and you are even allowed not to answer any question in this part. Afterwards you will be able to participate in a lottery if you want, and you will be asked to

---

[15] This is a translation of the Hebrew original. The brackets indicate the numbers used in the low base-fee treatment (the bonus percentage is higher when the base fee is lower because the payment for each correct answer is identical in both treatments).

answer two additional short questions. At the end you will receive the amount that you earned in the experiment in cash.

The questionnaire is anonymous. Participation in the experiment is voluntary and you can leave the experiment anytime.

Participant's declaration: I agree to participate in this experiment.

Signature and date:

Part A

Thank you very much for your participation in the experiment. The questions in this part have no right or wrong answer. Try to imagine yourself in the situations described in the questions before you answer. For your participation in this part of the experiment you will receive 15 [5] Shekels at the end of the experiment.

[The subject was then given four questions in which he was asked to provide matching prices in various consumer decision problems; the questions are omitted for the sake of brevity and are available from the author upon request].

Part B

This part is optional. You can end it and return the questionnaire any time, even if you chose not to answer any question. For your participation in this part you can earn an additional bonus. If you solve all 40 questions in this part correctly, you will get a bonus of 40% [120%] (in relation to the amount you earned for part A, i.e., you can receive in total for the two parts 1.4 [2.2] times the payment for part A). If you solve some of the questions correctly, you will receive a pro-rated bonus according to the number of questions you solve correctly. For example, if you solve 20 questions correctly, you will receive a bonus of 20% [60%]. As you will see, the questions

become more and more difficult, so it is recommended to proceed according to their order, up to the point in which you decide that you do not want to continue solving additional questions.

On the following pages different letters are written. In the questions below you are asked to write which letter appears on a certain page, in a certain line, and in a certain location. <u>The location of the letter in the line is counted from left to right</u>. For example, the letter on page 1, line 1, column 3 is s.

    1. Page 1, line 1, location 2:_____

    2. Page 1, line 5, location 6:_____

    …

    40. Page 9, line 31, location 68:_____

Part C – Lottery (Optional)

You can now participate in the following lottery: you guess a number between 1 and 6, and we throw a dice. The participation in the lottery will cost you 1 Shekel (which will be reduced from your earnings in parts A and B), but if the number you guessed is equal to the number that will be on the dice, you will earn 6 Shekels (i.e., together with the cost of the lottery, you have in fact a probability of 5/6 to lose 1 Shekel and a probability of 1/6 to earn 5 Shekels). If you want to participate in the lottery, please write the number you guess: ____

Part D

In order to compute the amount you earned, I need to check your responses in part B. I will be grateful if you can answer in the meantime the following questions (they do not have a right or wrong answer, simply write whatever is valid for you):

Rank on a scale of 1 to 9 (circle the answer you choose) to what extent each of the following reasons affected your decision when to stop solving the questions in Part B (1: this reason was not relevant at all; 9: this reason was my main consideration).

I compared the time and effort required to solve the questions to the fact that I could increase my earnings by 40% [120%] by solving additional questions: 1 2 3 4 5 6 7 8 9

I compared the time and effort required to solve the questions to the fact that I could increase my earnings by 6 Shekels by solving additional questions: 1 2 3 4 5 6 7 8 9

Please write below additional comments, if you have, about the experiment and the manner in which you made decisions in it:

[Space provided but omitted here]

Thank you very much for your participation in the experiment!

Appendix B: Experiment 2 (Testing for Reciprocity)

Procedure: First the subject received an informed-consent form to sign. After returning it, the base-fee treatment was chosen randomly and the subject was handed the appropriate version of Part A and Part B stapled together and the 9 pages containing the letters on which the questions in Part B are based. When he returned them, he received Part C.[16] At the end he received his payment in cash.

---

[16] The lottery was omitted in this experiment to simplify it and because its only purpose in Experiment 1 was to examine whether behavior in Part B was affected by a concavity of the utility function and wealth effects created by the different participation fee. In experiment 2 this is irrelevant because the subjects receive no payment for answering the questions in Part B. Because the lottery in Experiment 1 was offered only after the subjects handed out

The purpose of this experiment is to examine various aspects of decision making. In the first part you will be asked to answer a few short questions. The second part is optional: you can end it and return the questionnaire anytime, and you are even allowed not to answer any question in this part. Afterwards you will be asked to answer three additional short questions. At the end you will receive the amount that you earned in the experiment in cash.

[The rest of the consent form is identical to Experiment 1.]

Part A: Identical to Experiment 1 (see Appendix A).

Part B

This part is optional. You can end it and return the questionnaire any time, even if you chose not to answer any question. As you will see, the questions become more and more difficult, so it is recommended to proceed according to their order, up to the point in which you decide that you do not want to continue solving additional questions.

[The rest of this part is identical to Experiment 1.]

Part C

I will be grateful if you can answer the following questions (they do not have a right or wrong answer, simply write whatever is valid for you):

Parts A and B, taking the lottery out in Experiment 2 cannot affect the behavior in Part B, which is the part compared to Experiment 1.

Do you think that you would have solved a different number of questions in part B if the payment for part A had been 15 Shekels instead of 5 Shekels? ["5 Shekels instead of 15 Shekels?" in the high participation fee treatment] (Circle A, B, or C; if you chose B or C please write the number of questions that you would have solved less or more)

A. I would have solved the same number of questions

B. I would have solved _____ fewer questions

C. I would have solved _____ additional questions

How did you decide when to stop solving the questions in part B?

[Space provided but omitted here]

Please write below additional comments, if you have, about the experiment and the manner in which you made decisions in it:

[Space provided but omitted here]

Thank you very much for your participation in the experiment!

Appendix C: Experiment 3 (Mixed Compensation Schemes – The Hypothetical Condition)

Procedure: Identical to Experiment 1 (see Appendix A).

<u>Consent form</u>

The purpose of this experiment is to examine various aspects of decision making. In the first and second parts you will be asked to answer a few short questions. Afterwards you will be able to

participate in a lottery if you want, and you will be asked to answer two additional short questions. At the end you will receive the amount that you earned in the experiment in cash.

[The rest of the consent form is identical to Experiment 1.]

Part A: Identical to Experiment 1 (see Appendix A).

Part B

On the following pages different letters are written. In the questions below you are asked to write which letter appears on a certain page, in a certain line, and in a certain location. The location of the letter in the line is counted from left to right. For example, the letter on page 1, line 1, column 3 is s.

First, you are asked to solve the three questions 10, 20 and 30 in order to get a feeling about what is required in the questions. After you finish these three questions, please answer the following questions:

Suppose that you were told the following:

This part is optional. You can end it and return the questionnaire any time, even if you chose not to answer any question. For your participation in this part you can earn an additional bonus. If you solve all 40 questions in this part correctly, you will get a bonus of 40% [120%] (in relation to the amount you earned for part A, i.e., you can receive in total for the two parts 1.4 [2.2] times the payment for part A). If you solve some of the questions correctly, you will receive a pro-rated bonus according to the number of questions you solve correctly. For example, if you solve 20 questions correctly, you will receive a bonus of 20% [60%]. As you will see, the questions become more and more

31

difficult, so it is recommended to proceed according to their order, up to the point in which you decide that you do not want to continue solving additional questions.

How many questions out of the 40 questions below do you think you would have tried to solve? _____

How many of them do you think you would have solved correctly? _____

How much time do you think it would have taken you to solve the number of questions you indicated? _____

Comment: The description above is hypothetical only, please do not solve the 40 questions below.

    1. Page 1, line 1, location 2:_____

    2. Page 1, line 5, location 6:_____

    …

    40. Page 9, line 31, location 68:_____

Part C: Identical to Experiment 1 (see Appendix A).

<div align="center">Part D</div>

I will be grateful if you can answer the following questions (they do not have a right or wrong answer, simply write whatever is valid for you):

Rank on a scale of 1 to 9 (circle the answer you choose) to what extent each of the following reasons affected your decision about how many questions you would choose to solve in Part B (1: this reason was not relevant at all; 9: this reason was my main consideration).

[The rest of this part is identical to Experiment 1.]

References

Ariely, Dan, Uri Gneezy, George Loewenstein and Nina Mazar (2008): "Large Stakes and Big Mistakes," *The Review of Economic Studies* 76, 451-469.

Azar, Ofer H. (2004): "Psychological Motivations and Biases in Economic Behavior and their Effects on Markets and Firm Strategy," Ph.D. dissertation, Northwestern University.

Azar, Ofer H. (in press): "Do Consumers Make Too Much Effort to Save on Cheap Items and Too Little to Save on Expensive Items? Experimental Results and Implications for Business Strategy," *American Behavioral Scientist*.

Camerer, Colin F. and Robin M. Hogarth (1999): "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty* 19, 7-42.

Falk, Armin and Andrea Ichino (2006): "Clean Evidence on Peer Effects," *Journal of Labor Economics* 24:1, 39-58.

Fehr, Ernst and Simon Gachter (2000): "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives* 14:3, 159-181.

Frisch, Deborah (1993): "Reasons for Framing Effects," *Organizational Behavior and Human Decision Processes* 54, 399-429.

Gneezy, Uri and John List (2006): "Putting Behavioral Economics to Work: Field Evidence on Gift Exchange," *Econometrica* 74:5, 1365-1384.

Harrison, Glenn W. and John A. List (2004): "Field Experiments," *Journal of Economic Literature* 42, 1009-1055.

Hertwig, Ralph and Andreas Ortmann (2001a): "Experimental Practices in Economics: A Challenge for Psychologists?" *Behavioral and Brain Sciences* 24, 383-403.

Hertwig, Ralph and Andreas Ortmann (2001b), "Money, Lies, and Replicability: On the Need for Empirically Grounded Experimental Practices and Interdisciplinary Discourse," *Behavioral and Brain Sciences* 24, 433-451.

Hertwig, Ralph and Andreas Ortmann (2003): "Economists' and Psychologists' Experimental Practices: How They Differ, Why They Differ, and How They Could Converge," in Brocas, Isabelle and Juan D. Carrillo (Eds.), *The Psychology of Economic Decisions, Volume I: Rationality and Well-Being*, Oxford University Press, Oxford, pp. 253-272.

Jenkins, G. Douglas, Jr., Atul Mitra, Nina Gupta and Jason D. Shaw (1998): "Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research," *Journal of Applied Psychology* 83, 777-787.

Mowen, Maryanne M. and John C. Mowen (1986): "An Empirical Examination of the Biasing Effects of Framing on Business Decisions," *Decision Sciences* 17, 596-602.

Ranyard, Rob and Deborah Abdel-Nabi (1993): "Mental Accounting and the Process of Multiattribute choice," *Acta Psychologica* 84, 161-177.

Tversky, Amos and Daniel Kahneman (1981): "The Framing of Decisions and the Psychology of Choice," *Science* 211, 453-458.

Table 1: Distribution of Effort Measures in the Incentive Condition

	Attempted questions		Correct answers		Time in minutes	
	Low fee	**High fee**	**Low fee**	**High fee**	**Low fee**	**High fee**
0-10	25.4%	29.4%	30.5%	33.9%	15.3%	21.1%
11-20	27.1%	27.5%	26.3%	27.5%	34.7%	33.9%
21-30	10.2%	11.0%	11.9%	11.0%	11.0%	19.3%
31-40	37.3%	32.1%	31.4%	27.5%	23.7%	16.5%
41+					15.3%	9.2%
Total	100%	100%	100%	100%	100%	100%

The table shows the division of subjects in each treatment (118 in the low fee, 109 in the high fee) to the various effort levels.

Table 2: Treatment Effect in the Incentive Condition

	Attempted questions	Correct answers	Time in minutes
Low-fee mean	23.3	20.6	24.6
High-fee mean	22.1	19.7	22.2
Low-fee median	20	18	20.5
High-fee median	20	18	19
Low-fee standard deviation	13.8	12.6	13.9
High-fee standard deviation	13.6	12.5	13.0
p-value of the t-test for difference in means	0.533	0.587	0.182
p-value of the Mann-Whitney test	0.591	0.676	0.201

The t-test for difference in means and the Mann-Whitney test are testing the hypothesis that the distribution of the relevant variable (mentioned at the top of the column) is the same in the low-fee and high-fee treatments. The p-values show that this hypothesis cannot be rejected at any conventional level of significance for any variable.

Table 3: Regression Results – The Incentive Condition

Dependent var. Independent var.	ATTEMPT	CORRECT	TIME
Constant	23.90** (4.29)	22.51** (3.97)	23.75** (4.26)
HIGHBASE	-0.45 (5.90)	-1.15 (5.45)	-0.43 (5.86)
R	-0.49 (6.79)	-2.32 (6.28)	-2.40 (6.75)
MALE	-2.17 (2.56)	-2.16 (2.37)	1.03 (2.54)
BGU	1.90 (1.91)	1.22 (1.77)	2.88 (1.90)
HIGHBASE*R	0.07 (9.44)	2.50 (8.74)	1.86 (9.38)
HIGHBASE*MALE	-2.08 (3.70)	-2.45 (3.42)	-5.21 (3.68)
N	221	221	221
R^2	0.03	0.03	0.04

Standard errors are reported in parentheses. The number of observations is 221 because 6 subjects did not answer the questions in part D and therefore are omitted from the regressions. Levels of statistical significance are denoted by asterisks: [*] represents 5%-level and [**] represents 1%-level. ATTEMPT is the number of attempted questions. CORRECT is the number of correct answers. TIME is the time in minutes the subject spent on parts A and B together. HIGHBASE is a dummy variable equal to 1 in the high-fee treatment. R is a measure of relative thinking based on part D and is explained in more detail in the text. Male and BGU are dummy variables equal to 1 for males and Ben-Gurion University students, respectively.

Table 4: Treatment Effect in Experiment 2 (Testing for Reciprocity)

	Attempted questions	Correct answers	Time in minutes
Low-fee mean	12.1	9.3	13.0
High-fee mean	12.9	11.3	13.5
Low-fee median	11	8	10
High-fee median	9.5	8.5	10
Low-fee standard deviation	8.9	6.9	8.7
High-fee standard deviation	10.6	9.5	10.1
p-value of the t-test for difference in means	0.692	0.249	0.795
p-value of the Mann-Whitney test	0.862	0.576	0.635

The t-test for difference in means and the Mann-Whitney test are testing the hypothesis that the distribution of the relevant variable (mentioned at the top of the column) is the same in the low-fee and high-fee treatments. The p-values show that this hypothesis cannot be rejected at any conventional level of significance for any variable.

Table 5: Regression Results – Experiments 1 and 2 Together

Dependent var. Independent var.	ATTEMPT	CORRECT	TIME
Constant	12.86** (2.41)	10.17** (2.18)	12.21** (2.37)
HIGHBASE	0.27 (2.89)	1.76 (2.62)	1.51 (2.84)
PFP	11.47** (2.29)	11.52** (2.07)	11.68** (2.25)
MALE	-3.80 (2.01)	-3.57 (1.83)	-0.98 (1.98)
BGU	1.68 (1.43)	1.36 (1.30)	1.97 (1.41)
HIGHBASE*PFP	-1.80 (3.08)	-2.71 (2.79)	-2.51 (3.03)
HIGHBASE*MALE	1.31 (2.82)	0.62 (2.55)	-1.82 (2.77)
N	326	326	326
R^2	0.14	0.16	0.14

Standard errors are reported in parentheses. Levels of statistical significance are denoted by asterisks: * represents 5%-level and ** represents 1%-level. ATTEMPT is the number of attempted questions. CORRECT is the number of correct answers. TIME is the time in minutes the subject spent on parts A and B together. HIGHBASE is a dummy variable equal to 1 in the high-fee treatment. PFP stands for pay-for-performance and is a dummy variable equal to 1 in experiment 1 and to 0 in experiment 2. Male and BGU are dummy variables equal to 1 for males and Ben-Gurion University students, respectively.

Table 6: Treatment Effect in the Hypothetical Condition

	Attempted questions	Correct answers	Time in minutes
Low-fee mean	20	17.8	14.0
High-fee mean	19.6	17.9	13.6
Low-fee median	20	16.5	17.5
High-fee median	20	13.5	10
Low-fee standard deviation	16.7	15.3	11.1
High-fee standard deviation	15.3	14.4	14.1
p-value of the t-test for difference in means	0.913	0.980	0.899
p-value of the Mann-Whitney test	0.936	0.846	0.453

The t-test for difference in means and the Mann-Whitney test are testing the hypothesis that the distribution of the relevant variable (mentioned at the top of the column) is the same in the low-fee and high-fee treatments. The p-values show that this hypothesis cannot be rejected at any conventional level of significance for any variable.

Table 7: Regression Results – Incentive and Hypothetical Conditions Together (Experiments 1 and 3)

Dependent var. Independent var.	ATTEMPT	CORRECT	TIME
Constant	23.31** (2.24)	20.75** (2.07)	23.15** (2.07)
HIGHBASE	0.41 (2.68)	1.07 (2.47)	1.07 (2.48)
HYPOTHETICAL	-3.22 (2.74)	-2.73 (2.53)	-10.27** (2.51)
MALE	-2.30 (2.34)	-1.91 (2.16)	-0.34 (2.16)
BGU	1.96 (1.69)	1.39 (1.56)	2.57 (1.56)
HIGHBASE* HYPOTHETICAL	0.06 (3.71)	0.43 (3.42)	0.86 (3.41)
HIGHBASE*MALE	-1.99 (3.25)	-2.87 (3.00)	-5.18 (3.00)
N	313	313	314
R^2	0.03	0.03	0.13

Standard errors are reported in parentheses. Levels of statistical significance are denoted by asterisks: [*] represents 5%-level and [**] represents 1%-level. ATTEMPT is the number of attempted questions. CORRECT is the number of correct answers. TIME is the time in minutes the subject spent on parts A and B together. HIGHBASE is a dummy variable equal to 1 in the high-fee treatment. HYPOTHETICAL is a dummy variable equal to 1 in experiment 3 and to 0 in experiment 1. Male and BGU are dummy variables equal to 1 for males and Ben-Gurion University students, respectively.