

IZA DP No. 10393

## Dynamic Incentive Effects of Team Formation: Experimental Evidence

Thomas Gall  
Xiaocheng Hu  
Michael Vlassopoulos

November 2016

# Dynamic Incentive Effects of Team Formation: Experimental Evidence

**Thomas Gall**

*University of Southampton*

**Xiaocheng Hu**

*University of Southampton*

**Michael Vlassopoulos**

*University of Southampton  
and IZA*

Discussion Paper No. 10393  
November 2016

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### **Dynamic Incentive Effects of Team Formation: Experimental Evidence\***

Optimal team composition has been the focus of exhaustive analysis, academic and otherwise. Yet, much of this analysis has ignored possible dynamic effects: e.g., anticipating that team formation is based on prior performance will affect prior performance. We test this hypothesis in a lab experiment with two stages of a real effort task. Participants first work individually without monetary incentives and are then assigned to teams of two where compensation is based on team performance. Our results are consistent with a simple investment-cum-matching model: pairing the worst performing individuals with the best yields 20% lower first stage effort than random matching. Pairing the best with the best, however, yields 5% higher first stage effort than random matching. In line with the theory the latter result is more pronounced when the task has less scope for learning-by-doing. Moreover, pairing the best with the best achieves the same effort response as having explicit monetary incentives in the first stage.

JEL Classification: C78, C91, M54

Keywords: matching, team formation, performance, dynamic incentives

Corresponding author:

Michael Vlassopoulos  
Department of Economics  
Social Sciences  
University of Southampton  
Southampton, SO17 1BJ  
United Kingdom  
E-mail: [M.Vlassopoulos@soton.ac.uk](mailto:M.Vlassopoulos@soton.ac.uk)

---

\* We greatly benefited from comments and suggestions received from seminar participants at Southampton, Marburg, and the London Experimental Workshop. All remaining errors are our own.

# 1 Introduction

Team production as a mode of organizing the production of goods and services in firms and organizations has gained considerable traction over the last decades, in line with the rise of human resource management practices (see Ichniowski and Shaw, 2003; Bloom and van Reenen, 2011, and the studies reviewed there). For instance, in the US the share of Fortune 1000 companies that use team-based remuneration has gone up to 85% by 2005, an increase of 25 percentage points from 1990 (Merriman, 2008).

While much thought has gone into the analysis of how to optimally organize teams and how team composition may affect individual and aggregate performance in teams, by academics and practitioners alike,<sup>1</sup> there has been less emphasis, however, on possible effects on individual behavior *before* teams are formed. If individual payoffs depend on the teammates' attributes (through peer effects or group incentives) and teams are formed based on attributes that are the consequence of prior choice (such as performance), team formation will affect an individual's incentives for prior choice. The theoretical literature has considered investments before team formation or marriage (see e.g. Cole et al., 2001; Felli and Roberts, 2002; Peters and Siow, 2002) and has provided some insights into incentive effects of different matching mechanisms (e.g. Booth and Coles, 2010; Gall et al., 2006, 2012). However, there has been no comparable interest in examining empirically the dynamic effects of different team formation mechanisms. The aim of the current paper is to fill this void.

There is a wide range of methods used in practice to assign workers into teams. For instance, some firms assign workers into teams that are heterogeneous in ability, by partnering strong performers with weaker ones, in order to facilitate learning or to provide role models that lead to produc-

---

<sup>1</sup>The academic literature has focused, for instance, on the impact of group incentive schemes on productivity (Nalbantian and Schotter, 1997; Friebel et al., 2015), on the role of peer pressure among teammates (Kandel and Lazear, 1992; Falk and Ichino, 2006; Mohnen et al., 2008; Mas and Moretti, 2009; Babcock et al., 2015), the way that social connections among teammates affect performance (Bandiera et al., 2005), on whether having team members with different gender, age or ethnic background can be disruptive (Charness and Villeval, 2009; Apesteguia et al., 2012; Hjort, 2014), on the impact of financial incentives on task assignment within teams (Burgess et al., 2010), and on how the interaction of team composition and incentive schemes may affect performance (Hamilton et al., 2003; Bandiera et al., 2013; Delfgaauw et al., 2013; Chan et al., 2014).

tivity gains (Hamilton et al., 2003). This practice may limit an individual’s desire to exert effort at an earlier stage, i.e., there is an equity-efficiency trade-off. Conversely, if the best performers are assigned to better partners this will provide additional incentives for effort at an earlier stage. This pattern can also be the outcome when workers are allowed to choose their own teammates since workers will tend to match positively assortatively in ability (e.g. Bandiera et al., 2013). Of course, team formation may also be left to chance, for instance, if assignment is by sequence of arrival, follows a rotation system or is guided by alphabetic order of names (e.g. Bartel et al., 2014).

We design a real effort experiment with two work stages: in a first stage, participants work individually and do not receive compensation. In the second stage, they are assigned to teams of size two, based on their performance in the previous stage, and receive compensation that depends on the average performance of the team. The experimental variation arises from varying the rule that matches workers in the second stage. Firms’ team formation mechanisms in practice vary widely, ranging from those that generate teams that are as homogeneous as possible, for instance with an aim to foster excellence at the top, to those that implement as much heterogeneity as possible, possibly to enable the strong to support the weak. We adopt two salient, polar forms of team assignment policies for our treatment: positive assortative matching (*PAM*), in which the best worker is matched with the second best and so on, and negative assortative matching (*NAM*), in which the best worker is matched with the worst and so on. For comparison as a baseline treatment, we use an assignment policy that matches workers randomly (*RAM*), ignoring their performance in the first stage. As we are interested in comparing the efficacy of team formation policies as an incentive device in relation to explicit monetary incentives, we implement a fourth treatment (*R&I*), in which workers also receive an individual piece rate for their first stage performance and are randomly matched into teams.

We use a simple model of investment and matching to derive some theoretical pointers as to how performance in the individual work stage might differ across the treatments described above. The results we obtain are largely consistent with the predictions. In particular, *NAM* leads to the lowest performance in the individual work stage relative to the other treatments (20% reduction in mean performance compared to *RAM*). Moreover,

the evidence is consistent with the prediction of strategic behavior under *NAM* in the individual work stage, implying that measured performance before the assignment may be not very informative for the later performance. Both *PAM* and *R&I* yield slightly higher effort than *RAM* (about 5% higher mean individual work stage effort). Interestingly, *PAM* and *R&I* are not statistically distinguishable from each other in terms of effort, which suggests that in our experiment the dynamic implicit incentive is as strong as the within period monetary incentive, offering a possible avenue for efficiency gains in production. In line with the theory, the difference between *PAM* (*NAM*) and *RAM* is more (less) pronounced for a task with less scope for learning-by-doing. Finally, we do not find any differences in performance in the team work stage across all treatments.

Our results have an interesting efficiency-equity implication for human resources practices that has not been highlighted before. A matching policy that aims to increase equity by equalizing average performance across teams (*NAM*) will impose a dynamic cost, by discouraging effort before the assignment into teams. Although such policies may well be motivated by efficiency considerations at the team work stage they will lead to a substantial shortfall of effort in earlier stages.

The remainder of this paper is organized as follows. The next section describes the design of our laboratory experiment in more detail. Section 3 presents the theoretical predictions to be tested in our study. Section 4 presents and discusses our experimental results. Section 5 concludes, and the Appendix includes additional tables and the experimental instructions.

## 2 Experimental Design

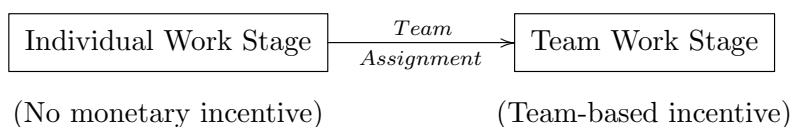
In order to study the dynamic incentive effects of team formation, we designed an experiment with a key feature that a participant's initial performance on a real effort task determines the partner that the worker is matched with in a later stage according to a known assignment rule. In this later stage, compensation is based on the average performance of the pair, which provides an implicit incentive for a participant to exert effort in the earlier stage in anticipation that this would lead to a desirable match. This setup allows us to examine our main question of how performance in the early stage is affected by the assignment rule - which varied across treatments -

that pairs workers in the subsequent stage.

## 2.1 The Stages of the Experiment

Specifically, participants in our experiment performed a real effort task in two stages. The first stage was an individual work stage, followed by a team work stage, with each stage lasting four minutes (see Figure 1). In addition, participants had two minutes to practice the task before the individual work stage. Subjects received live feedback about their score and the remaining time during practice. In addition, to help participants develop a better sense of their relative performance, they received feedback about their rank among all subjects based on the final scores achieved in practice.

**Figure 1:** The Two Stages of a Working Round



In the individual work stage, subjects were given four minutes to work on the task and their performance was rewarded implicitly or explicitly, depending on the treatment, which will be defined in the next subsection. At the end of the stage, we also elicited their belief of their performance relative to that of the other participants in their session in an incentive compatible way to measure participants' self-confidence.<sup>2</sup> Then individuals were informed about their true rank as well as the maximum and minimum scores and assigned to a partner for the team work stage. The assignment rule depends on the treatment and is explained in detail further on. Subjects were shown their partner's rank and score before beginning the team work stage.

In the team work stage, subjects had a final four minutes to work on the task. Performance was rewarded by monetary payment, which was based on both own and the partner's performance. Similar feedback was given as in the previous stage. At the end of the stage, subjects were informed about their partner's final score and their team score (the average of own and partner's score), which determined payment. Note that the team work

---

<sup>2</sup>They received £0.4 for correctly guessing in which quartile of the performance distribution they belong.

stage does not entail joint production per se, as workers perform the same task individually, however, there is a spillover across teammates arising from the compensation scheme.

This two-stage work was repeated three times (rounds), and in each round participants performed a different task. After the last round subjects answered a brief questionnaire, eliciting subjects' preferences over risk and time, as well as altruism and competitiveness, and their socioeconomic characteristics (including gender, age, nationality, and native speaking language) and educational achievements (major fields of study on university, academic level, and years of study in university).

The sequence of events, the assignment mechanism in place, and the payment rules were communicated very clearly to the subjects at the very beginning of the experiment in both written and spoken form. In particular, considerable effort was made to ensure that the instructions informed subjects about the payment rules while not emphasising that the individual work stage is not directly incentivised, in the three treatments where this was the case.

## 2.2 Treatments

To test possible effects of different team assignment mechanisms on the incentives to exert effort in the individual work stage prior to team formation, our experiment involved four treatments that were implemented in a between-subject design. We focus on random matching as a benchmark and two polar assignment regimes, in the hope that this will allow extrapolation to a general class of team formation mechanisms.

### **Random matching (*RAM*)**

Our benchmark treatment is random matching, that is, each individual is assigned to any other individual with equal probability. This assignment regime reflects both actual randomised assignments as well as situations where the assignment is based on markers that are orthogonal to prior performance (such as the alphabetical order of surnames or the sequence of arrivals).



### **Positive assortative matching (*PAM*)**

Positive assortative matching assigns individuals into teams based on their effort before the formation of teams. Specifically, the individual with the best performance is assigned to the individual with the second best performance, the third best individual to fourth best individual, and so on. This assignment mechanism rewards higher performance, corresponding to a higher effort, with a better teammate. This pattern endogenously arises in workplaces where workers are allowed to choose their own teammates since workers will tend to match positively assortatively in ability.

### **Negative assortative matching (*NAM*)**

Similarly to *PAM*, negative assortative matching assigns individuals into teams based on their effort before the formation of teams, but the higher one's own performance the lower the performance of one's match. Specifically, the individual with the best performance is assigned to the individual with the worst performance, the individual with the second best performance to the one with the second worst performance, etc. This assignment mechanism provides low performers with high performing teammates, thus generating balanced teams in terms of average performance of team members. Real life examples include the formation of balanced teams, or the assignment of better employees to support weaker colleagues.

### **Random matching with monetary incentives (*R&I*)**

The three treatments presented above do not use any explicit monetary incentives to encourage effort before individuals are assigned to teams; only effort within teams is rewarded by monetary payments. In contrast, the final treatment (*R&I*) rewards effort before team formation explicitly by a payment depending on performance. Assignment is by random matching, as under *RAM*.

## **2.3 Real Effort Tasks**

To measure participants' effort and possible effects of differential assignment mechanisms, we used three computerised real effort tasks: the slider task (Gill and Prowse, 2012), the counting zeros task (Abeler et al., 2011) and the word encryption task (Erkal et al., 2011). The use of different real

effort tasks was intended to provide subjects with a modicum of variety to maintain motivation through the 90 minutes duration of the experiment, and to account for the possibility that subjects' elasticity of effort provision to monetary incentives might be low in some tasks (Araujo et al., 2016). All tasks are simple to understand, do not require preexisting knowledge and offer little gains from guessing. Hence, the performance, or score, achieved in a real effort task is a good measure of individual effort.

In the *Slider Task* (as proposed by Gill and Prowse, 2012) forty-eight sliders appear on screen, each with a range of integer values from 0 to 100, initially positioned at 0, see Figure D3 in the appendix. Subjects were tasked to use their mouse to position the slider at 50, which requires a certain degree of manipulation. Subjects' performance in the task, i.e. their "score", was given by the number of sliders successfully positioned at exactly 50 within the allotted time.

The *Grid Task* consists of counting the number of 0's in a  $5 \times 5$  grid of randomly distributed 0's and 1's. Subjects were asked to enter the number on the screen, see Figure D4 in the appendix. If the number entered was correct, they continued to the next grid. The score in this task was the number of correctly counted grids within the allotted time. This task is similar to the task by Abeler et al. (2011), although they use  $10 \times 15$  grids and impose no time limit.

In the *Word Encryption Task* subjects were shown combinations of three letters (words) and tasked to transcribe them into numbers using an encryption table mapping letters uniquely to three digit numbers, see Figure D5 in the appendix. Once subjects entered the correct encryption, they were given a new random three letter combination to encode. The score in this task was the number of correctly encoded words. To limit training effects the encryption table was re-randomized before each stage (individual and team work), both changing the position of letters (not in alphabetical order) and the mapping from letters to numbers. Therefore subjects could not profit from memorising the encryption table nor the location of the keys. This task is similar to the task by Erkal et al. (2011), although they do not vary the encryption table. The double-randomisation of letters and numbers in the word encryption task was introduced by Benndorf et al. (2014) who find that it limits learning behavior when repeating the task.

We calibrated the difficulty of the tasks based on the results of pilot

sessions, so that the average performance is approximately a score of 9 per two minutes for all tasks. Importantly, the three tasks differ in their scope for learning-by-doing through improving hand-eye coordination allowing us to test our prediction (iii) in Proposition 1 which can be found in the next section. In particular, for both the slider and the grid task previous studies have found that subjects improve performance over time (Georganas et al., 2015; Vranceanu et al., 2013), whereas the version of the encryption task we employ has been shown not to allow for significant improvement due to learning (Benndorf et al., 2014).

## 2.4 Payments

In each of the treatments *RAM*, *PAM* and *NAM* subjects were paid based on their team’s performance in the team work stage measured by the team score. The team score is the arithmetic mean of the scores of the two teammates. At the end of a session, for each subject one of the three tasks was randomly chosen and the subject’s payment was her team score in that task multiplied by a piece rate of 0.4 pounds per score point.

Treatment *R&I* additionally rewards individual performance in the individual work stage, given by the subject’s score on that stage. In this treatment for each subject one of the three tasks and one of the individual and the team work stages is randomly chosen with equal probabilities and the subject’s payment will be the subject’s team, respectively individual score, in the selected task at the selected stage multiplied by a piece rate of 0.4 pounds per score point.

## 2.5 Procedures

The experiment was conducted at the Social Sciences Experimental Lab (SSEL) of the University of Southampton, in early March and late April 2016. We ran three sessions of each of the four treatments described above (*RAM*, *R&I*, *PAM* and *NAM*), for a total of twelve sessions. The order of treatments and the sequence of tasks within sessions was randomized, albeit under the condition that each of the three tasks was the first one to be performed in a session exactly once for each treatment. Each session had sixteen student subjects from various departments, with 192 participants

in total (104 females and 88 males).<sup>3</sup> The subjects were recruited from the SSEL subject pool, using ORSEE (Greiner, 2004). The experimental instructions were provided to each subject in written form and were also read aloud to the subjects. Seating positions were randomised and seat numbers were given in the order of arrival. To ensure subject-experimenter anonymity actions and payments were linked to seat number only. After reading the instructions and before performing the task, subjects completed a quiz to ensure understanding of the rules and the assignment mechanism in their treatment. Each subject was paid a show-up fee of £4 and earned an average of a further £10 during the experiment. Subjects were paid privately in cash at the end of each session. The experiment was programmed in z-Tree (Fischbacher, 2007).

### 3 Predictions

To organise thoughts consider the following model of investment and matching. An economy is populated by a continuum of agents, characterized by their types  $\theta$ , and lasts for two stages. Suppose  $\theta$  is distributed on an interval  $[\underline{\theta}, \bar{\theta}]$  with  $0 < \underline{\theta} < \bar{\theta}$ . Suppose also that the distribution of productivity  $\theta$  has full support.

In each stage  $t$  each agent can exert effort  $e_t$  to generate output  $y_t$  at a cost  $c_t$ . This cost depends on an agent's type  $\theta$ . In stage 1 the cost is

$$c_1(e_1) = \frac{e_1^2}{2\theta}.$$

In stage 2 the cost may additionally depend on stage 1 effort investment:

$$c_2(e_1, e_2) = \frac{e_2^2}{2(\theta + \lambda e_1)}.$$

The parameter  $\lambda > 0$  determines the degree of learning-by-doing manifest in the task that generates output. Notice that this cost function is separable in own type  $\theta$  and the degree of learning-by-doing  $\lambda$ , which greatly simplifies computations but does not drive our results stated below.

In stage 1 individuals generate output according to the technology:

$$y_1 = e_1.$$

---

<sup>3</sup>We invited 20 randomly selected subjects to each session. The first 16 subjects who showed up at the lab participated in the experiment. The other subjects received a show-up fee of £4 and were asked to leave the laboratory.

After stage 1 individuals are assigned into teams of size 2, with attributes  $(e_1, \theta)$  and  $(e'_1, \theta')$ . The assignment takes one of the four forms described above: *RAM*, *R&I*, *PAM*, or *NAM*. Matched individuals exert effort  $e_2$  in stage 2. Team production in stage 2 is separable and linear in individual stage 2 effort  $e_2$  and  $e'_2$ :

$$y_2 = e_2 + e'_2.$$

An individual's monetary payoff is given by half the team's output in stage 2,  $y_2/2$ . Under *R&I*, however, the individual receives additionally  $y_1$ . Separability of team payoffs implies that an individual's productivity does not depend on their match, and that, conditional on stage 1 effort  $e_1$ , aggregate output and utility are independent of the match. Hence, any differences in aggregate output and utility are entirely due to the dynamic incentives effects of the different assignment mechanisms.

We are interested in a Nash equilibrium in individual effort choice  $e_1$  and  $e_2$  in the two-stage game played by individuals. We omit general properties of, e.g., existence because this type of matching cum investment game with a continuum of players has been well explored elsewhere, for instance in the work by Cole et al. (2001) and the premarital investment game by Peters and Siow (2002), both imposing *PAM*, and by Booth and Coles (2010) and Gall et al. (2012) who also allow for *RAM*, respectively *NAM*.

### Stage 2 behavior

To solve for a Nash equilibrium we use backwards induction and start with behavior in stage 2. In stage 2 individuals are assigned into teams. Hence an individual chooses effort  $e_2$  to solve

$$\max_{e_2} \frac{e_2 + e'_2}{2} - \frac{e_2^2}{2(\theta + \lambda e_1)},$$

where  $e'_2$  denotes the effort of the individual's teammate. Hence, individual optimal stage 2 effort satisfies:

$$e_2^* = (\theta + \lambda e_1)/2.$$

Since this must be true for each teammate, under *RAM*, *PAM* and *NAM* an individual's overall payoff is given by:

$$u(e_1, \theta, e'_1, \theta') = \frac{\theta + \lambda e_1 + 2(\theta' + \lambda e'_1)}{8} - \frac{e_1^2}{2\theta},$$

which clearly increases in the attributes  $e'_1$  and  $\theta'$  of one's teammate. Under *R&I* an individual additionally obtains payoff  $y_1 = e_1$ , so that  $u(\cdot) = \frac{\theta + \lambda e_1 + 2(\theta' + \lambda e'_1)}{8} + e_1 - \frac{e_1^2}{2\theta}$ .

### Stage 1 behavior

Effort choice in stage 1 depends fully on the continuation payoff in stage 2, and thus on the possibility that own effort choice  $e_1$  determines the attributes  $e'_1$  and  $\theta'$  of one's partner in stage 2. An individual chooses  $e_1$  to solve

$$\max_{e_1} u(e_1, \theta, e'_1, \theta'),$$

taking into account that own stage 1 effort may change attributes of one's match  $e'_1$  and  $\theta'$ .

It is then straightforward to compute the Nash equilibrium for each assignment mechanism (details are in the appendix):

**Fact 1.** *Individual stage 1 effort  $e_1$  in a Nash equilibrium depends on the assignment mechanism as follows:*

- Under *RAM* effort is  $e_1^{RAM} = \frac{\lambda\theta}{8}$ .
- Under *R&I* effort is  $e_1^{R\&I} = \frac{\lambda\theta}{8} + \theta$ .
- Under *PAM* effort is  $e_1^{PAM} = \frac{3\lambda\theta}{16} + \sqrt{9\lambda^2 + 64} \frac{\theta}{16}$ .
- Under *NAM* there is  $\hat{\theta}$  such that  $e_1^{NAM} = 0$  for agents with  $\theta < \hat{\theta}$  and  $e_1^{NAM} = \frac{\lambda\theta}{8}$  for agents with  $\theta > \hat{\theta}$ .

Using these expressions for equilibrium effort investment in stage 1 allows us to compare to the different regimes in terms of observable outcomes, yielding testable predictions.

**Proposition 1** (Predictions). *Comparing equilibrium first stage effort levels under the different assignment mechanisms:*

- (i) *PAM and R&I induce higher effort for all types than RAM, which in turn induces higher effort than NAM, and strictly so for some types.*
- (ii) *R&I induces higher effort than PAM for low degrees of learning by doing and the opposite is true for high degrees of learning by doing.*

- (iii) *The percentage difference in effort between PAM (R&I) and RAM decreases, but the percentage difference in effort between RAM and NAM increases in the degree of learning by doing  $\lambda$ .*
- (iv) *Effort under RAM has higher mean and lower variance than under NAM.*

That is, in the equilibrium of our model average effort choices are lowest under *NAM*, intermediate under *RAM* and highest under *PAM* and *R&I*. Unfortunately, the order of the latter two is ambiguous and depends on the unobserved parameter  $\lambda$ , which precludes a precise prediction. The degree of learning-by-doing inherent in the task affects the comparison between the different mechanisms: outcomes under *RAM* become closer to outcomes under *PAM* and *R&I* as the degree of learning-by-doing increases and less close to those under *NAM*. Moreover, Fact 1 implies that effort choices are more dispersed under *NAM* than under *RAM*, since equilibrium efforts of high ability agents coincide under the two regimes, but low ability agents will all choose 0 under *NAM*, but a strictly positive effort, strictly increasing in type under *RAM*. The comparison of equilibrium efforts under *PAM* and *R&I* to those under *NAM* respectively *RAM* does not allow for a clear-cut characterisation (computations show that it depends on  $\lambda$  and the distribution of  $\theta$ , and on whether the cost function is separable in  $\theta$  and  $\lambda$ ).

## 4 Results

### 4.1 Sample

Summary statistics of participant characteristics are presented in Table C1. Using Chi-square test, t-test, and Mann-Whitney U test (M-W test), Table C2 in the appendix shows that participants' characteristics are balanced across treatments with the exception of academic level and age, which, though are most likely to be related to each other.

### 4.2 Effort in The Individual Work Stage

We use the score achieved in each task as a measure of an individual's effort. Table 1 summarizes performance in the individual work stage for

the whole sample and by each treatment and task separately.<sup>4</sup> Pooling across treatments and tasks we see that participants' mean score was 23. Comparing performance in the different tasks we see that mean scores in the slider and the grid tasks were very similar. However, the mean score in the word encryption was lower, and the differences with the other tasks were statistically significant.<sup>5</sup>

**Table 1:** Summary of Individual Work Stage Effort

Effort in the Individual Work Stage	Observations	Mean	SD	Minimum	Maximum
<i>Panel 0. All Treatments</i>					
All Tasks	575	22.98	7.449	0	39
Slider Task	191	24.62	9.178	0	39
Grid Task	192	24.02	7.027	0	37
Word Encryption Task	192	20.32	4.761	0	31
<i>Panel 1. RAM</i>					
All Tasks	144	23.58	7.123	0	39
Slider Task	48	25.23	9.911	0	39
Grid Task	48	25.38	5.060	16	37
Word Encryption Task	48	20.15	3.525	13	27
<i>Panel 2. NAM</i>					
All Tasks	144	18.93	9.515	0	39
Slider Task	48	19.77	11.40	0	39
Grid Task	48	19.56	9.700	0	35
Word Encryption Task	48	17.46	6.934	0	30
<i>Panel 3. PAM</i>					
All Tasks	144	24.70	5.878	0	39
Slider Task	48	26.60	7.454	0	39
Grid Task	48	25.60	5.127	15	37
Word Encryption Task	48	21.90	3.270	13	31
<i>Panel 4. R&amp;I</i>					
All Tasks	143	24.73	4.910	5	38
Slider Task	47	26.94	4.843	16	38
Grid Task	48	25.54	5.251	5	33
Word Encryption Task	48	21.77	2.800	14	27

We next address our main question of whether individual effort in stage 1 is affected by how teams are formed at stage 2. Figure 2 shows the mean individual performance levels by treatment. Observed patterns are indeed

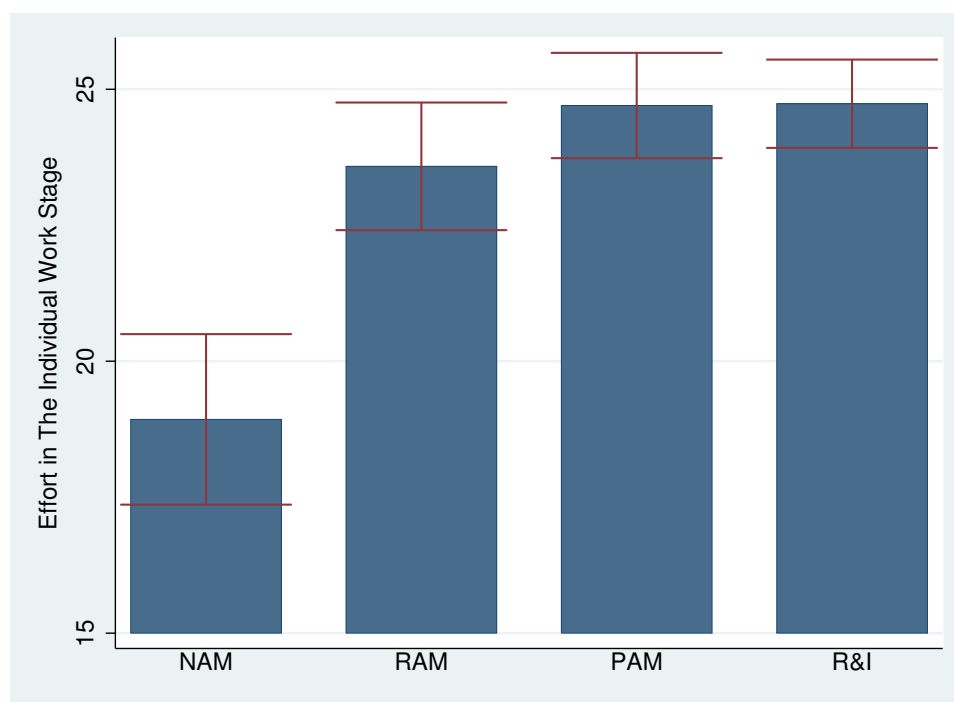
<sup>4</sup>Following the practice by Gill and Prowse (2012), we leave out of the analysis one participant (in the treatment *R&I*) who scored 0 in all three stages of the slider task. Our qualitative results do not depend on this sample selection and the quantitative results would change only marginally.

<sup>5</sup>Both the paired t-test (p-values < 0.001) and the Wilcoxon signed rank sum test (p-values < 0.001) reject equality of mean score in the word encryption and the other two tasks. Comparing mean scores in the slider and grid tasks the Wilcoxon test indicates a significant difference (p-value = 0.048), but not the t-test (p-value = 0.361).



very much in line with our expectations (prediction (i) in Proposition 1) stemming from the theoretical model in Section 3. In particular, the mean performance is low under *NAM*, intermediate under *RAM*, and high under both *PAM* and *R&I*. Mean performance under *PAM* and *R&I* is virtually indistinguishable. For a quantitative comparison, the mean score under *NAM* is about 20% less than under *RAM*, while *PAM* and *R&I* both induce 5% higher scores than *RAM*.

**Figure 2:** Effort in The Individual Work Stage



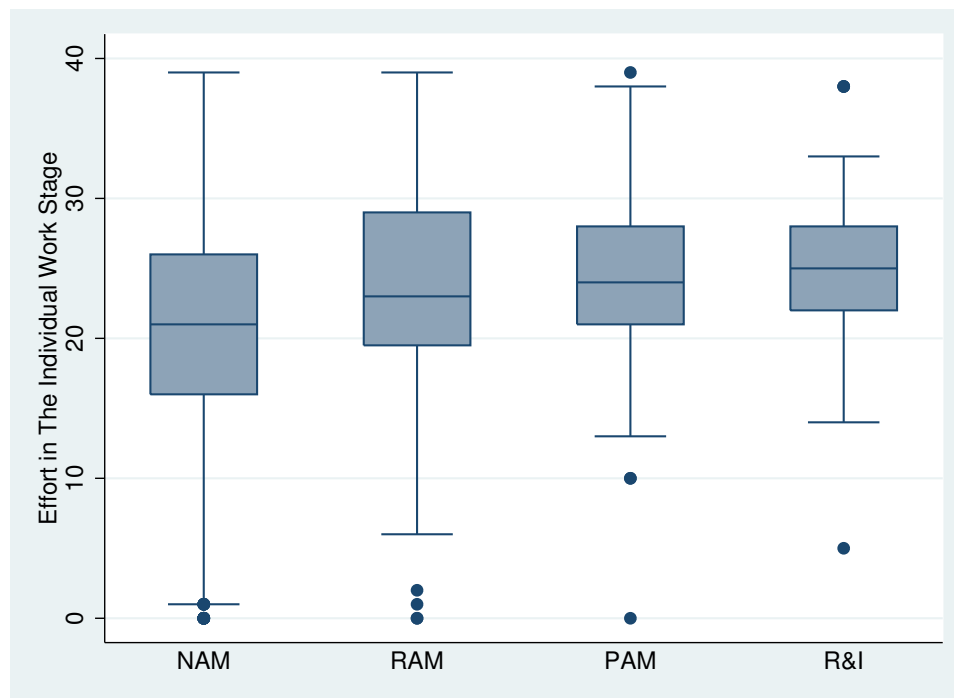
Notes: The top end of the bars indicate the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

Figure 3 shows box-plots of the distribution of individual stage efforts across treatments, indicating substantial differences in dispersion of effort across treatments. Indeed, moving across treatments from *NAM* to *R&I* in Figure 3 the average effort increases while its dispersion decreases, as does its standard deviation (see the fourth column in Table 1).<sup>6</sup> This is consistent with the theoretical results, in particular with testable prediction (iv) in Proposition 1, stating that equilibrium effort has higher mean and

<sup>6</sup>Pairwise F-tests of equality of variance across treatments reject equality in all cases.

lower dispersion under *RAM* than under *NAM*. However, our conclusion is hard to draw from comparing effort dispersion under *PAM* and *R&I* to that under *RAM* and *NAM*, since it depends on parameters and the properties of the cost function, see Section 3.

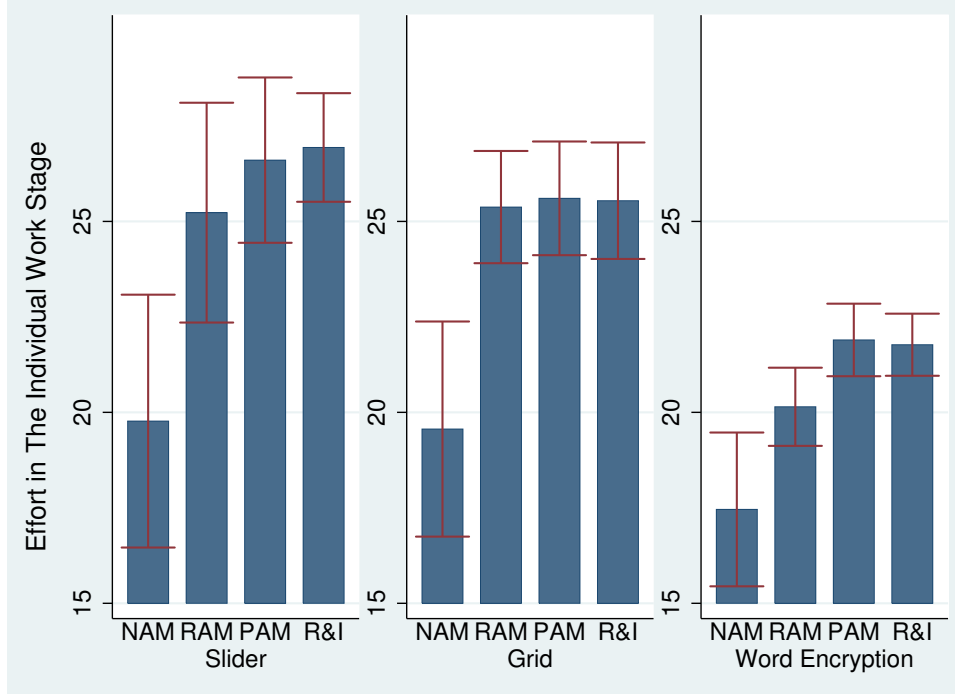
**Figure 3:** Dispersion of Effort in The Individual Work Stage



Since the experiment asked participants to work on different tasks, to account for possible differences in terms of learning-by-doing or sensitivity to explicit incentives, Figure 4 presents individual stage performance across treatments for each task. Disaggregating the results for the different tasks yields a nuanced picture: while the differences of outcomes between treatments are similar across tasks, the magnitudes of the differences vary considerably. While performance under *RAM* comes close to the one under *PAM* and *R&I* for both the grid and the slider task, this is not the case for the word encryption task in which performance under *NAM* is about 13% less than under *RAM*, while *PAM* and *R&I* both induce 8 – 9% higher performance than *RAM* (Table 1). Since the word encryption task offers less opportunity for learning-by-doing than the other tasks, this result is consistent with our prediction (iii) in Proposition 1.<sup>7</sup>

<sup>7</sup>One possible way to test for differences in learning-by-doing across tasks is to examine

**Figure 4:** Effort in The Individual Work Stage by Task



Notes: The top end of the bars indicate the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

Table 2 reports the results of the tests for possible differences between treatments. Summing up over the scores achieved in all tasks, both non-parametric (M-W) test and t-test yield significant differences between *NAM* and the other three treatments. On the other hand, we find no statistically significant differences between Treatments *RAM*, *PAM* and *R&I*.<sup>8</sup> Table 2 also contains the results for each task separately, confirming the picture in Figure 4: for the Word encryption task performance under *RAM* was significantly smaller than under each of *PAM* and *R&I*, whereas this was

the relative performance improvement between the two work stages in treatment *R&I*, since effort is incentivized with monetary payments in both stages. Doing this we find that the slider task has the largest improvement (mean = 8.4%), followed by the grid task (mean=3.5%) and the word encryption task (mean = 1.1%). A pairwise test rejects equality between the slider and the word encryption task (a paired t-test has a p-value of 0.024 and a Wilcoxon test 0.063), but fails to reject equality between the grid and the word encryption task (a paired t-test has p-value 0.313 and a Wilcoxon test 0.272).

<sup>8</sup>These results remain the same when using a weighted average of the scores instead of simply adding up. Detailed re-weighting methods and results are available upon request.

not the case for the other two tasks.

**Table 2:** Statistical Differences Across Treatments

Effort in Individual Work Stage	NAM vs RAM		NAM vs PAM		NAM vs R&I	
	t test (p-value)	M-W test (p-value)	t test (p-value)	M-W test (p-value)	t test (p-value)	M-W test (p-value)
All tasks	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Slider	0.014	0.003	< 0.001	0.002	< 0.001	0.001
Grid	< 0.001	0.005	< 0.001	0.001	< 0.001	< 0.001
Encryption	0.019	0.092	< 0.001	< 0.001	< 0.001	< 0.001

Effort in Individual Work Stage	RAM vs PAM		RAM vs R&I		PAM vs R&I	
	t test (p-value)	M-W test (p-value)	t test (p-value)	M-W test (p-value)	t test (p-value)	M-W test (p-value)
All tasks	0.147	0.207	0.112	0.179	0.959	0.841
Slider	0.444	0.956	0.291	0.528	0.798	0.687
Grid	0.826	0.797	0.874	0.488	0.953	0.564
Encryption	0.013	0.018	0.014	0.016	0.841	0.947

*Notes:* The null hypothesis for t-test/Mann-Whitney U (M-W) test is that the difference between the means/distributions of the two independent samples is zero.

To be able to control for variation at the task and round level, as well as individual characteristics, we complement the previous analysis with OLS regressions displayed in Table 3. Column (1) presents the results of a regression with only the treatment dummies as independent variables, column (2) adds task and round fixed effects to capture unobservable variation across tasks and rounds. Column (3) adds preference indicators, constructed from subjects' answers to the questions asked during the experiment. The preference indicators capture subjects' accuracy of beliefs about relative performance, competitiveness, altruism, time discounting and risk attitudes (see Appendix B for details on the construction of these variables and Table C1 for descriptive statistics.). Columns (4) and (5) add some demographic covariates, academic level and gender, respectively, to account for possible differences in sample composition, although the selection into treatments was fairly balanced on observables. Finally, column (6) adds controls for nationality and whether the participant studies for an economics-related degree.<sup>9</sup> The coefficients for the different treatments remain relatively stable

<sup>9</sup>Notice that individual age, years of study, and native speaking language are not included in the regressions as they are collinear with academic level and nationality, re-

across the different specifications.

**Table 3:** OLS Regression

	Dep. Var.: Effort in the Individual Work Stage					
	(1)	(2)	(3)	(4)	(5)	(6)
NAM	-4.653*** (0.827) [1.313]	-4.653*** (0.830) [1.317]	-4.612*** (0.579) [1.260]	-4.636*** (0.579) [1.249]	-4.678*** (0.601) [1.278]	-4.708*** (0.592) [1.262]
PAM	1.118 (0.699) [0.885]	1.118 (0.701) [0.889]	1.213* (0.577) [0.821]	0.974 (0.583) [0.793]	1.185* (0.589) [0.831]	0.997* (0.508) [0.795]
R&I	1.151*** (0.353) [0.849]	1.168*** (0.364) [0.851]	1.265** (0.481) [0.843]	1.127** (0.467) [0.822]	1.291** (0.468) [0.841]	1.003* (0.519) [0.840]
Constant	23.58*** (0.242) [0.683]	26.04*** (0.698) [0.910]	27.54*** (1.027) [1.177]	27.75*** (1.053) [1.167]	27.28*** (0.981) [1.191]	27.57*** (0.749) [1.362]
Observations	575	575	575	575	575	575
Participants	192	192	192	192	192	192
R-squared	0.103	0.177	0.226	0.229	0.227	0.242
Task and Round Fixed Effects:	NO	YES	YES	YES	YES	YES
Attitudes	NO	NO	YES	YES	YES	YES
Academic Level	NO	NO	NO	YES	NO	YES
Gender	NO	NO	NO	NO	YES	YES
Other	NO	NO	NO	NO	NO	YES

*Notes:* OLS Estimations. Dependent variable is the effort in the individual work stage. The omitted treatment is *RAM*. Robust standard errors clustered at session level and individual level are reported in brackets and square brackets below the estimates, respectively. (1) reports estimates for the baseline model without control variables. (2) adds task and round fixed effects. (3) adds elicited preferences (accuracy of beliefs about relative performance, competitiveness, time discounting, risk averse, and altruism). (4) adds academic level dummies. (5) adds gender dummy. (6) controls for all individual demographics (gender, academic level, nationality, and degree). \*\*\* Significant at 1% level, \*\* significant at 5% level, \* significant at 10% level.

Overall, the regression analysis confirms the results above, indicating that *NAM* was associated to a decrease in score of 4.6 – 4.7 (about 20%) relative to *RAM*, while *PAM* and *R&I* were associated to an increase in score of 1.0 – 1.3 (about 5%) each. The drop in performance under *NAM* is statistically significant throughout all specifications, as is the increase under *R&I*, although the significance level drops to 10% as we saturate the model with controls. The performance increase under *PAM* is only significant in some specifications, however, and only at the 10% level. Finally, the

coefficients of *PAM* and *R&I* are statistically indistinguishable across all specifications.<sup>10</sup>

Absent from this analysis is a proxy for individual ability at a task. One could, however, use individual performance at the team work stage to proxy for individual ability in a given task. It is, however, plausible that the team work stage effort choice could be affected by the treatment, i.e., by the team composition, which would generate an endogeneity problem. Nevertheless, adding team work performance into our specifications does not change our conclusions above.

### 4.3 Subgroup Analysis

There are some plausible reasons to expect that some subgroups of individuals would be affected more by our treatments than others. For instance, since participants in the experiment received feedback on their relative performance, highly competitive individuals could be expected to be motivated intrinsically and to respond less to extrinsic incentives. Similarly, our theoretical model assumes that players are fully informed about the type distribution, i.e., that participants have accurate expectations on the relative position of their performance among all subjects in the session. Therefore, participants who have more accurate beliefs about their relative performance are more likely to be affected by our treatments.

In this subsection, we explore possible differences of treatment effects across subsamples split along two dimensions: first, with respect to how accurately they were able to predict their relative position in the performance distribution and, second, with respect to their intrinsic ability or motivation as measured by their actual performance in the team work stage.

---

<sup>10</sup>The results are very similar when using an individual random effects estimation approach (which may be warranted as individuals are not independent within each session). The same is true when including observations from the one subject dropped because of a failure to score at all in the slider task. Results are also qualitatively unchanged when using the logarithm of the dependent variable, although the treatment effect size increases. Regression results are available upon request.

### 4.3.1 Splitting The Sample Based on Accuracy of Beliefs about Relative Performance

Recall that at the end of the individual work stage of each round we elicited participants' beliefs as to their relative standing in terms of their performance. More specifically participants predicted the quartile, in which they believed their performance to lie, and received a reward if they were correct. Indeed, a sizable fraction of participants (41% to 47%) accurately predicted their quartile, about 20% underestimated it, and 33%-39% overestimated it. Overall, 42.6% of participants were able to accurately predict their relative performance in their session in at least two of the three rounds, with 11.8% correctly predicting their quartile in all three rounds. 20.9% did not predict correctly in any of the rounds.<sup>11</sup>

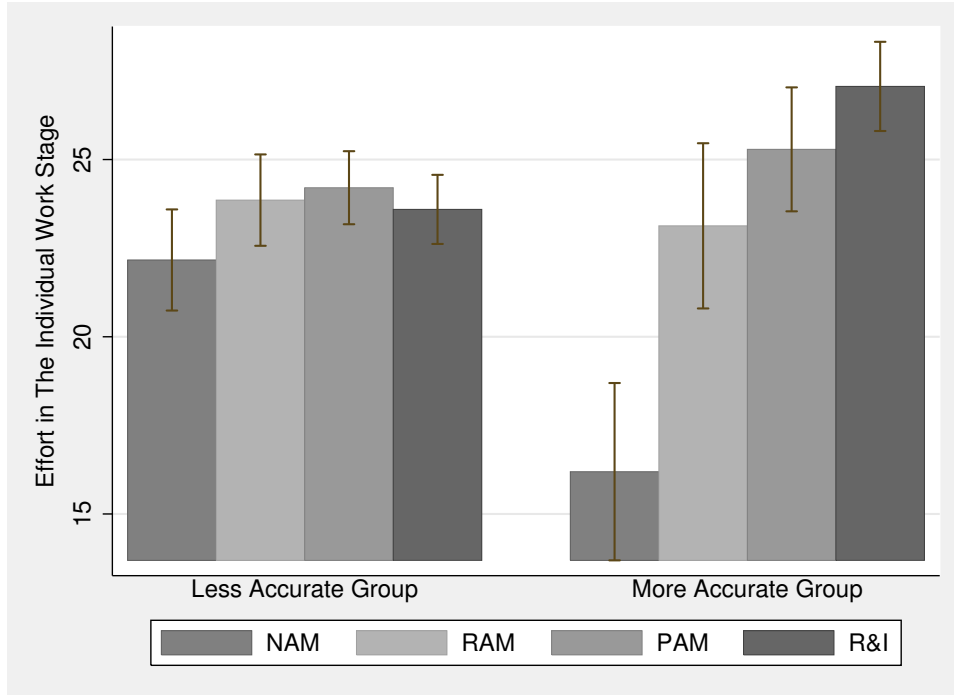
Since strategic behavior requires individual expectations of relative performance to be reasonably accurate (at least under *NAM* and *PAM*) one would expect that the treatment differences are more pronounced among the group of participants that well predicted their own relative performance. To explore this possibility, we split the sample into two similarly sized groups: one group ( $n = 110$ ) that predicted their ranking correctly in at most one round, and the other group ( $n = 82$ ) that correctly predicted their ranking in at least two rounds. Figure 5 shows mean performance by treatment separately for each of the two groups. For the subject group that predicted their relative performance more accurately the pattern of treatment effects mirrors closely the theoretical predictions: performance was very low under *NAM*, while *PAM* and *R&I* were significantly higher than *RAM*. On the other hand, the performance of the subjects who predicted less accurately does not differ much across treatments, except for *NAM*, which yields slightly lower performance, albeit significantly higher than under *NAM* for the other group.<sup>12</sup>

---

<sup>11</sup>See Table C3 in the appendix for the demographic composition of the groups. 56.3% of the participants who had correct beliefs about their relative performance are male, and there seems to be a higher incidence of them under *NAM*, although the difference is only statistically significant at 10% (regression results are available upon request).

<sup>12</sup>In particular, both t-test and Mann-Whitney U test indicate statistically significant differences in subjects' performances between the two subgroups within treatment *NAM* (mean difference 5.974, p-value of t-test  $< 0.001$ , and p-value of Mann-Whitney U test  $< 0.001$ ) and *R&I* (mean difference  $-3.470$ , p-value of t-test  $< 0.001$ , and p-value of Mann-Whitney U test 0.005). On the other hand, the differences are statistically insignificant for *RAM* (mean difference 0.726, p-value of t-test 0.556, and p-value of Mann-Whitney

**Figure 5:** Subgroup Analysis by Accuracy of Beliefs about Relative Performance



Notes: The top end of the bars indicate the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

These observations carry over to a regression analysis similar to the one reported above for the whole sample. The results, shown in Table 4, indicate that for the more accurate group treatment effects are much greater in magnitude than for the other group ( $-7\%$  vs.  $-28\%$  for *NAM*,  $1.5\%$  vs.  $9\%$  for *PAM* and  $0\%$  versus  $17.5\%$  for *R&I*), and the increase under *R&I* is statistically significant throughout.<sup>13</sup>

#### 4.3.2 Splitting The Sample Based on Team Stage Performance

Our second subgroup analysis addresses a possible concern of any real effort experiment: subjects might exert substantial effort regardless of the exper-

U test 0.666) and *PAM* (mean difference  $-1.083$ , p-value of t-test 0.272, and p-value of Mann-Whitney U test 0.118).

<sup>13</sup>The results are robust to clustering standard errors at the individual level, with the exception that the coefficients for treatment *NAM* become statistically insignificant for the less accurate group.



**Table 4:** Subgroup Analysis by Accuracy of Beliefs about Relative Performance: OLS Regression

	Less Accurate Group <sup>a</sup>		More Accurate Group <sup>b</sup>	
	(1)	(2)	(3)	(4)
NAM	-1.689*** (0.445)	-1.552** (0.559)	-6.937*** (1.997)	-6.244*** (1.417)
PAM	0.350 (0.297)	0.344 (0.337)	2.158 (1.816)	2.018 (1.754)
R&I	-0.262 (0.922)	0.0478 (0.939)	3.934*** (1.223)	4.259*** (1.307)
Constant	23.86*** (0.284)	26.55*** (0.998)	23.13*** (1.051)	26.78*** (1.887)
Observations	330	330	245	245
Participants	110	110	82	82
R-squared	0.018	0.248	0.208	0.352
Task and Round Fixed Effects	NO	YES	NO	YES
Other Controls	NO	YES	NO	YES

*Notes:* OLS Estimations. Dependent variable is the effort in the individual work stage. The omitted treatment is *RAM*. Robust standard errors clustered at session level are reported in brackets below the estimates. Columns (1) and (3) report estimates for the baseline model without control variables. Columns (2) and (4) add task and round fixed effects, elicited preferences, and individual demographics. \*\*\* Significant at 1% level, \*\* significant at 5% level, \* significant at 10% level.

<sup>a</sup>Less Accurate Group: subjects who did not correctly predict their relative standings more than once.

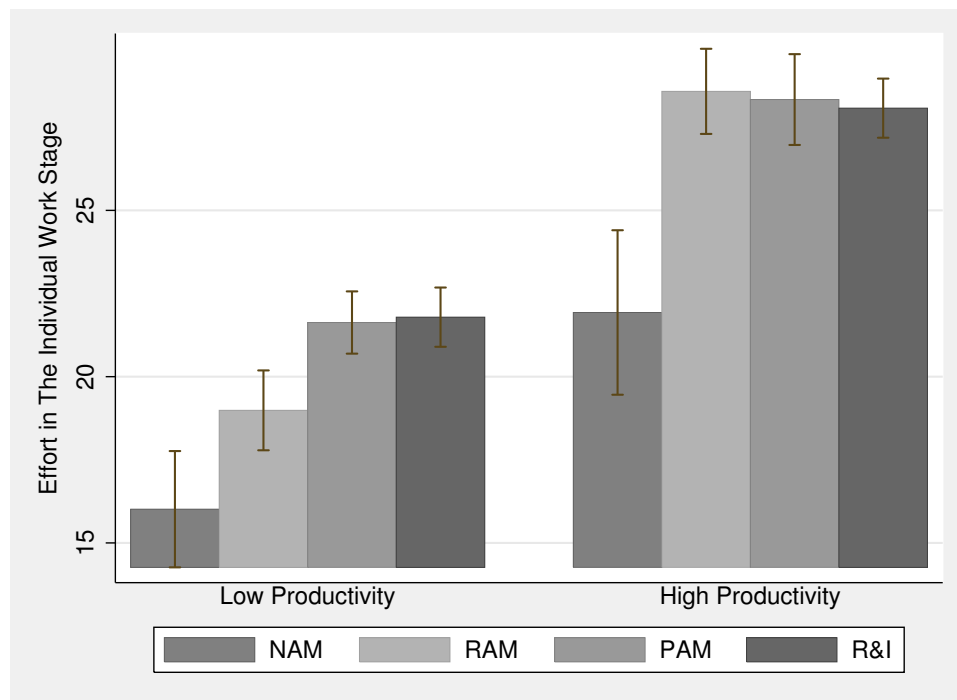
<sup>b</sup>More Accurate Group: subjects who correctly predicted their relative standings at least twice.

imental treatment because of intrinsic motivation. This might be due to a desire to perform well, either because participants enjoy working on the task, or because they feel challenged and enjoy overcoming this challenge. Alternatively, subjects may feel a moral obligation to exert effort knowing that they will receive a compensation for participating in the experiment. Finally, the feedback on their relative performance that participants received may already offer substantial non-monetary incentive for status-concerned individuals to exert effort. That is, some participants may already be exerting effort close to their capacity and thus make it difficult to detect variations across treatments. On the other hand, participants that lack such intrinsic motivation will perform well below their full potential and may be more susceptible to respond to our treatments.

To examine this possibility we split the sample into two similarly sized

groups on the basis of their performance in the team work stage. Subjects with less than the median performance form a low productivity group (n=101) and those with higher than the median performance form a high productivity group (n=91).<sup>14</sup> Recall that the team work stage is the only stage where individual effort is explicitly incentivised (with a team piece rate). Hence, individual performance in the team work stage is arguably a reasonable proxy for individuals' intrinsic motivation, in particular since average team work stage performance did not vary across treatments (see Section 4.4).

**Figure 6:** Subgroup Analysis by Team Stage Performance



Notes: The top end of the bars indicate the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

Figure 6 depicts the performance for each group by treatment and shows marked differences between the two groups in all treatments. What we find is that for the low productivity group observed treatment effects closely

<sup>14</sup>See Table C4 in the appendix for details on the composition of the two groups. The main difference appears to be that the high productivity group has a higher share of UK nationals and, reassuringly, of participants who prefer competitive settings (regression results available upon request).

**Table 5:** Subgroup Analysis by Team Stage Performance: OLS Regression

	Low Productivity Group		High Productivity Group	
	(1)	(2)	(3)	(4)
NAM	-2.973*** (0.490)	-3.132*** (0.524)	-6.650*** (1.382)	-5.934*** (1.445)
PAM	2.642*** (0.678)	3.020*** (0.534)	-0.246 (1.072)	-0.370 (1.011)
R&I	2.803*** (0.690)	2.347*** (0.734)	-0.505* (0.281)	0.386 (0.531)
Constant	18.99*** (0.409)	19.41*** (1.207)	28.58*** (0.224)	31.53*** (0.895)
Observations	302	302	273	273
Participants	101	101	91	91
R-squared	0.162	0.237	0.148	0.263
Task and Round Fixed Effects	NO	YES	NO	YES
Other Controls	NO	YES	NO	YES

*Notes:* OLS Estimations. Dependent variable is effort in the individual work stage. The omitted treatment is *RAM*. Robust standard errors clustered at session level are reported in brackets below the estimates. Columns (1) and (3) report estimates for the baseline model without control variables. Columns (2) and (4) add task and round fixed effects, elicited preferences, and individual demographics. \*\*\* Significant at 1% level, \*\* significant at 5% level, \* significant at 10% level.

mirror the theoretical predictions. On the other hand, for the high productivity group, only treatment *NAM* is distinguishable from other treatments. These observations can be further seen in a regression analysis by productivity group. The regression results in Table 5 indicate that for the low productivity group (columns 1 and 2) there are statistically significant differences across treatments except for the difference between *PAM* and *R&I*. For the high productivity group (columns 3 and 4), however, only treatment *NAM* shows the expected drop in performance compared to the other treatments, while performances in *PAM* and *R&I* are not statistically different from those in *RAM*.

#### 4.4 Effort in The Team Work Stage

While the main focus of this paper lies on individual performance before the formation of teams, it is of interest to examine whether different mechanisms of team formation affected participants' performance once they were assigned to a team. Table 6 presents the average individual performance in the team work stage by treatment. Mean performance is very similar across

treatments, and this remains true when examining the three different tasks separately.<sup>15</sup> Thus, the treatment in form of assignment mechanism has no effect on the individual performance after the assignment to teams.

**Table 6:** Summary of Team Work Stage Effort

Effort in the Team Work Stage	Observations	Mean	SD	Minimum	Maximum
<i>Panel 0. All Treatments</i>					
All Tasks	575	25.48	5.981	5	45
Slider Task	191	28.70	6.648	5	45
Grid Task	192	26.00	5.185	5	38
Word Encryption Task	192	21.75	3.468	13	30
<i>Panel 1. RAM</i>					
All Tasks	144	25.31	6.338	10	45
Slider Task	48	28.79	7.377	10	45
Grid Task	48	25.88	4.858	15	36
Word Encryption Task	48	21.25	3.829	13	28
<i>Panel 2. NAM</i>					
All Tasks	144	25.12	5.844	14	43
Slider Task	48	28.04	6.633	14	43
Grid Task	48	25.48	5.165	17	38
Word Encryption Task	48	21.85	3.673	14	30
<i>Panel 3. PAM</i>					
All Tasks	144	25.82	6.186	5	43
Slider Task	48	29.10	7.051	5	43
Grid Task	48	26.38	5.354	16	38
Word Encryption Task	48	21.98	3.411	14	30
<i>Panel 4. R&amp;I</i>					
All Tasks	143	25.65	5.560	5	43
Slider Task	47	28.85	5.525	20	43
Grid Task	48	26.27	5.457	5	38
Word Encryption Task	48	21.90	2.955	15	28

#### 4.4.1 Peer Effects

One particular reason why team composition may be related to individual performance of team members are peer effects, which have been the subject of a considerable attention in the literature (e.g. Falk and Ichino, 2006; Mas and Moretti, 2009, among many others). In our experiment the information participants received about their teammate was limited to the absolute and relative performance in the individual work stage and no live feedback was given. This setup allows for a possible peer effect through the knowledge

<sup>15</sup>Results of statistical tests are available upon request. In addition, an F-test confirms that there are no statistically significant differences in the standard deviations (column SD in Table 6) across treatments in the team work stage.

of being paired with a better or worse performing peer. The possible effect is ambiguous: on one hand a better peer may make free riding more attractive, but on the other hand reciprocity may induce higher effort anticipating higher effort of one's peer. To identify the peer effect, we estimate an OLS regression of individuals' performance in the team work stage on their partner's performance in the individual work stage. We constrain our sample to treatments *RAM* and *R&I* since both treatments used the same assignment mechanism, random matching. The regression results (see Table C5 in the appendix) suggest modest effects: the teammate's performance in the individual work stage was associated negatively with own performance in the team, but not significantly so. The negative sign of the coefficient is consistent with a free-riding effect.

#### 4.4.2 Inequality

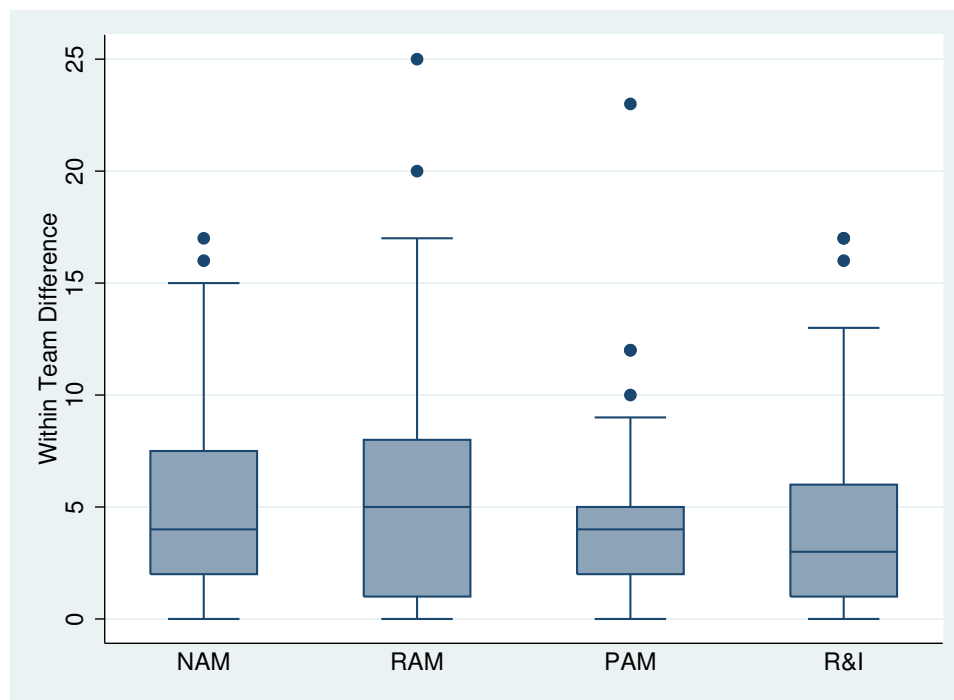
One possible motivation for the use of the different assignment mechanisms may be a concern for inequality in the organisation. For instance, *PAM* will induce very little inequality *within* teams in terms of individual attributes (i.e., past performance) but substantial inequality *across* teams. The converse will be the case for *NAM*: there will be considerable inequality within teams (matching the best to the worst performers, etc.), but very little inequality across teams. If past performance reflects individual ability this difference in within and across teams inequality of past performance should be mirrored by the performance in the team work stage.

Notice that individual stage effort choice may be strategic, however, and not reflect individuals' true productivities. The theoretical model in Section 3 predicts that individual stage effort choices are indeed strictly monotone in productivity type under *PAM*. That means that individuals with better performance in the individual stage can be expected to perform better in the team work stage, so that inequality of individual performance in teams should be low within teams, but high across teams. For *NAM* the model predicts that more than half of the population will choose the same effort level (zero) in the individual work stage. That is, in mechanism design terms, *NAM* will induce bunching and is not incentive compatible: individual stage performance is not necessarily informative about true productivity. Strategic behavior in the individual work stage would imply, of course, that performance rankings of individuals will differ between the individual and

the team work stage under *NAM*, but not in the rest of the treatments. The data offer support for this prediction: the rank correlation of individual and team work stage performance is significantly lower under *NAM*.<sup>16</sup>

Therefore, matching in *NAM* will not be strictly assortative in true productivity, but involve some randomness, and thus inequality in *actual team stage performance* will be lower within teams and higher across teams than inequality in individual stage performance. These considerations appear to be consistent with the data from the experiment. Figures 7 and 8 show the performance difference both within and across teams in the different treatments. *PAM* is clearly distinguishable from the other treatments and shows both low within team and high across teams inequality of actual team work stage performance, while *NAM* does not appear substantially different from the two treatments that match randomly.<sup>17</sup>

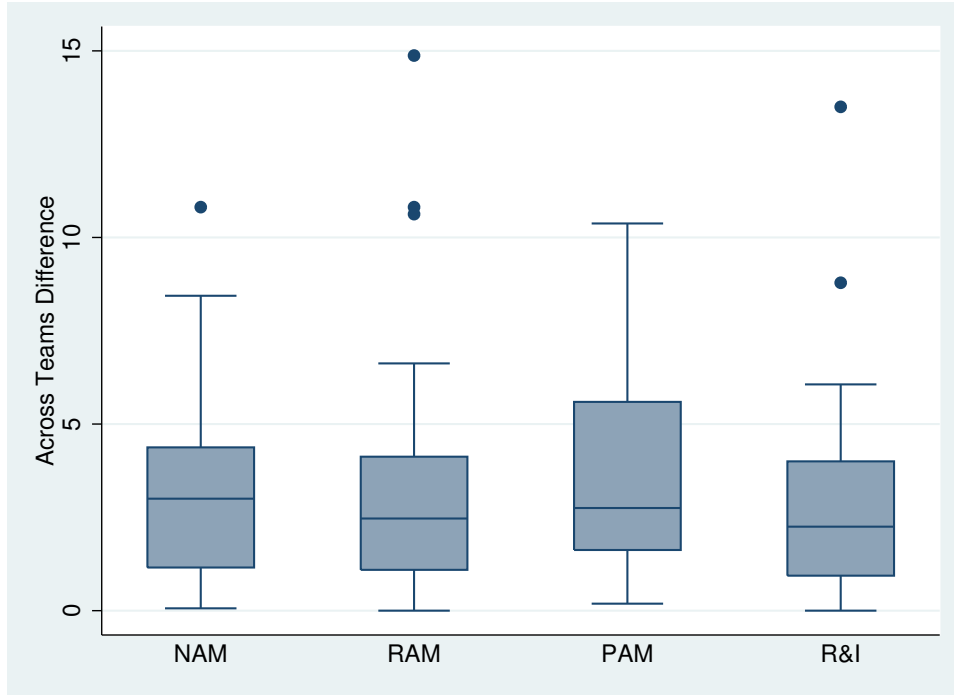
**Figure 7:** Within Team Difference in The Team Work Stage by Treatment



<sup>16</sup>Indeed, the rank correlation of individual performance in the two stages is 0.37 under *NAM* compared to 0.86, 0.77 and 0.80 under *RAM*, *PAM* and *R&I*, respectively, and the difference between *NAM* and any other treatment is highly significant, with p-values of less than 0.001.

<sup>17</sup>The differences between *PAM* and the other treatments are statistically significant in some but not all of the comparisons (results are available upon request).

**Figure 8:** Across Teams Difference in The Team Work Stage by Treatment



## 5 Conclusion

Does the manner of how individuals are assigned to teams affect prior effort choice? Our results from a real effort task experiment strongly suggest that the answer is in the affirmative. Specifically, we find that subjects substantially reduce prior effort under a team formation rule that matches high performers with low performers relative to a scenario where workers are randomly matched. The evidence is consistent with strategic behavior in the early stage, rendering measured performance in earlier stages a poor predictor of later performance. This finding confirms expectations of an equity-efficiency trade-off in team formation, as a process that yields teams similar in average prior performance of their members comes at the cost of reducing effort *ex ante*, much as Ramsay logic would suggest. While our results give a possible reason for caution when using matching on attributes based on prior choice in experiments, perhaps more importantly in practice are adverse implications for personnel policies that are designed to implement heterogeneity in terms of markers correlated with prior performance,

such as race. Further research on this matter would appear highly desirable in order to inform policy.

Assignment policies that match better performing subjects with better partners or explicitly reward early stage effort with monetary payments tend to outperform random matching, but the effect is relatively small. More interesting is perhaps the finding that effort choices under both explicit (monetary) and implicit (team assignment) incentives are statistically indistinguishable. That is, using some form of positive assortative matching can replace costly monetary payment in earlier stages (perhaps reminiscent of the use of low or unpaid internships before workers are promoted to full-paid positions).

The analysis we present in this paper is a first pass at bringing an investment and matching framework to the lab. There are several directions in which the analysis could be extended. For instance, team formation could be made endogenous, allowing participants to submit preference rankings over peers and then employing tried and tested matching algorithms. Theoretical results suggest outcomes will depend on whether side payments can be used and will resemble *PAM* if not. Moreover, while our results suggest the presence of learning-by-doing, effort in our experiment was not explicitly designed as an investment. Explicitly incorporating investment before assignment could be a potentially valuable approach to model educational policies in the lab. Moreover, many effort and investment decisions are taken in a team environment, potentially subject to peer effects. Hence, a repeated team formation and effort choice setup may shed some more light on productive processes.

Despite the effects found for performance before team formation we do not find significant differences in effort across treatments at the team work stage. This is not entirely unexpected as the real effort task performed in teams is independent across team members and the payoff additively linear in individual performance. Hence, there are no peer effects by design and teams matter only through group incentives. Corresponding to the latter we do find some evidence for mild free-riding at the team work stage. Of course, a potentially fruitful direction for further research could be to incorporate complementarities at the team work stage, for instance by tweaking the payments to reflect increasing or decreasing differences of joint production in individual output. In particular, when weaker individuals profit more



from stronger teammates than stronger individuals (decreasing differences) a tension will arise between static optimisation (favoring *NAM*-like policies) and dynamic considerations in terms of crowding out earlier stage effort (favoring *PAM*-like policies).

## References

- Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.
- Apestequia, J., Azmat, G., and Iriberry, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science*, 58(1):78–93.
- Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W., and Wilson, A. J. (2016). The slider task: an example of restricted inference on incentive effects. *Journal of the Economic Science Association*, 2(1):1–12.
- Babcock, P., Bedard, K., Charness, G., Hartman, J., and Royer, H. (2015). Letting down the team? social effects of team incentives. *Journal of the European Economic Association*, 13(5):841–870.
- Bandiera, O., Barankay, I., and Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962.
- Bandiera, O., Barankay, I., and Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5):1079–1114.
- Bartel, A. P., Beaulieu, N. D., Phibbs, C. S., and Stone, P. W. (2014). Human capital and productivity in a team environment: Evidence from the healthcare sector. *American Economic Journal: Applied Economics*, 6(2):231–59.
- Benndorf, V., Rau, H. A., and Sölch, C. (2014). *Minimizing Learning Behavior in Experiments with Repeated Real-Effort Tasks*.
- Bloom, N. and van Reenen, J. (2011). Chapter 19 - human resource management and productivity. volume 4, Part B of *Handbook of Labor Economics*, pages 1697 – 1767. Elsevier.
- Booth, A. and Coles, M. (2010). Education, matching, and the allocative value of romance. *Journal of the European Economic Association*, 8(4):744–775.

- Burgess, S., Propper, C., Ratto, M., Kessler Scholder, S. v. H., and Tominey, E. (2010). Smarter task assignment or greater effort: The impact of incentives on team performance. *The Economic Journal*, 120(547):968–989.
- Chan, T. Y., Li, J., and Pierce, L. (2014). Compensation and peer effects in competing sales teams. *Management Science*, 60(8):1965–1984.
- Charness, G. and Villeval, M.-C. (2009). Cooperation and competition in intergenerational experiments in the field and the laboratory. *American Economic Review*, 99(3):956–78.
- Cole, H. L., Mailath, G. J., and Postlewaite, A. (2001). Efficient non-contractible investments in large economies. *Journal of Economic Theory*, 101(2):333 – 373.
- Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31(2):305–326.
- Erkal, N., Gangadharan, L., and Nikiforakis, N. (2011). Relative earnings and giving in a real-effort experiment. *American Economic Review*, 101(7):3330–48.
- Falk, A., Becker, A., Dohmen, T., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. Working Papers 2016-003, Human Capital and Economic Opportunity Working Group.
- Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57.
- Felli, L. and Roberts, K. (2002). Does competition solve the hold-up problem? *CEPR Discussion Paper Series*, 3535.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Friebel, G., Heinz, M., Krüger, M., and Zubanov, N. (2015). Team incentives and performance: Evidence from a retail chain. Technical report, CEPR Discussion Paper No. DP10796.

- Gall, T., Legros, P., and Newman, A. (2012). Mismatch, rematch, and investment. Discussion paper series in economics and econometrics, University of Southampton, Economics Division, School of Social Sciences.
- Gall, T., Legros, P., and Newman, A. F. (2006). The timing of education. *Journal of the European Economic Association*, 4(2-3):427–435.
- Georganas, S., Tonin, M., and Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior and Organization*, 117:223 – 232.
- Gill, D. and Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1):469–503.
- Greiner, B. (2004). *An Online Recruitment System for Economic Experiments*.
- Hamilton, B. H., Nickerson, J. A., and Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3):465–497.
- Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 129(4):1899–1946.
- Ichniowski, C. and Shaw, K. (2003). Beyond incentive pay: Insiders' estimates of the value of complementary human resource management practices. *Journal of Economic Perspectives*, 17(1):155–180.
- Kandel, E. and Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, 100(4):801–817.
- Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–45.
- Merriman, K. (2008). Low trust teams prefer individualized pay. *Harvard Business Review*, 86(11):32.
- Mohnen, A., Pokorny, K., and Sliwka, D. (2008). Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence. *Journal of Labor Economics*, 26(4):693–720.

- Nalbantian, H. R. and Schotter, A. (1997). Productivity under Group Incentives: An Experimental Study. *American Economic Review*, 87(3):314–41.
- Peters, M. and Siow, A. (2002). Competing premarital investments. *Journal of Political Economy*, 110(3):592–608.
- Vranceanu, R., El Ouardighi, F., and Dubart, D. (2013). Coordination in Teams: A Real Effort-task Experiment with Informal Punishment. ESSEC Working Papers WP1310, ESSEC Research Center, ESSEC Business School.

## A Appendix: Proof for Section 3

### Proof of Fact 1

Start with the benchmark case *RAM*. Then  $e'_1$  and  $\theta'$  does not depend on an agent's choice of  $e_1$  and the optimal stage 1 effort (and Nash equilibrium effort) is given by:

$$e_1^{RAM} = \frac{\lambda\theta}{8}.$$

If an individual is additionally paid a piece rate of 1 for output in the first stage,  $y_1 = e_1$ , the Nash equilibrium effort increases to:

$$e_1^{R\&I} = \frac{\lambda\theta}{8} + \theta.$$

Under PAM  $e'_1$  increases in  $e_1$ . Suppose that strategies are strictly monotone increasing and differentiable in type.<sup>18</sup> Since  $\theta$  has full support by assumption, so does  $e_1$  and the positive assortative assignment satisfies  $e'_1 = e_1$ . Moreover, since  $\theta'$  is a function of  $e'_1$ , anticipating the matching outcome  $\theta'$  is a function of  $e_1$ . The individual optimization problem becomes thus:

$$\max_{e_1} \frac{\theta + \lambda e_1 + 2(\theta'(e_1) + \lambda e'_1(e_1))}{8} - \frac{e_1^2}{2\theta}. \quad (1)$$

Since the optimisation problems are the same for any two individuals of the same type  $\theta$ , equilibrium strategies  $e_1^*(\theta)$  will be the same and thus  $\theta'(e'_1) = \theta(e_1) = (e_1^*)^{-1}(\theta)$ . Hence an optimal choice of  $e_1$  satisfies

$$e_1^* = \frac{3}{8}\lambda\theta + \frac{\theta}{4} \frac{\partial\theta^*(e_1^*)}{\partial e_1}.$$

If  $\lambda = 0$ ,  $e_1^{PAM}(\theta) = \frac{\theta}{2}$  will solve this equation. Solving the differential equation for  $\lambda > 0$  yields  $e_1^{PAM}(\theta) = \frac{3\lambda + \sqrt{9\lambda^2 + 64}}{16}\theta$ , however.

Under negative assortative matching the stage 1 effort of one's partner (weakly) decreases in own effort. Hence, strategies need not increase in type. The individual optimization problem becomes:

$$\max_{e_1} \frac{\theta + \lambda e_1 + 2(\theta'(e_1) + \lambda e'_1(e_1))}{8} - \frac{e_1^2}{2\theta}.$$

Hence, an optimal choice of  $e_1$  satisfies

$$\frac{e_1}{\theta} = \frac{\lambda}{8} + \frac{\lambda}{4} \frac{\partial e'_1}{\partial e_1} + \frac{1}{4} \frac{\partial\theta}{\partial e_1}.$$

<sup>18</sup>While strict monotonicity will be guaranteed when stage 1 effort decreases effort cost in stage 2, there may be a ‘‘pooling’’ equilibrium when there is no learning (i.e.,  $\lambda = 0$ ).

Note first that  $e_1 < 0$  if  $\frac{\partial e'_1}{\partial e_1} < 0$  and  $e_1 = \lambda\theta/8$  if  $\frac{\partial e'_1}{\partial e_1} = 0$ . That is, a positive measure of agents will choose  $e_1^{NAM} = 0$ , i.e., there is bunching. On the other hand, agents matched to  $e_1 = 0$  agents will choose  $e_1 = \lambda\theta/8$ , since increasing  $e_1$  will still yield a match with  $e_1 = 0$  and the same expected type  $\theta'$  (supposing uniform rationing of  $e_1 = 0$  agents). Hence, under NAM an equilibrium is

$$e_1^{NAM} = 0 \text{ if } \theta < \theta^* \text{ and } e_1^{NAM} = \lambda\theta/8 \text{ if } \theta > \theta^*,$$

where  $\theta^*$  is a cutoff type who is just indifferent between investing  $e_1 = \lambda\theta/8$  and investing  $e_1 = 0$ . The intuition is that investing in the first stage, although profitable in isolation, is made unprofitable, as investment is punished by obtaining a worse match in expectation (both in terms of  $e_1$  and  $\theta$ ).

### Notes for Proposition 1

For Proposition 1 note that while  $e_1^{PAM} > e_1^{RAM}$ ,  $e_1^{PAM} > e_1^{R\&I}$  only for  $\lambda$  sufficiently high, and  $e_1^{PAM} < e_1^{R\&I}$  otherwise. Moreover, the ratios  $e_1^{PAM}/e_1^{RAM}$  and  $e_1^{R\&I}/e_1^{RAM}$  are both strictly decreasing in  $\lambda$ . Finally, the ratio  $e_1^{NAM}/e_1^{RAM}$  is either 0 or 1 depending on the type  $\theta$ , so that the ratio of aggregate effort investment must be less than unity.

## B Variable Definitions

*Accuracy Of Beliefs About Relative Performance:* qualitative response to the question “How do you think your individual score ranks among the other participants?” Participants could choose between the following options: “Bottom 25%”, “Between 25% and 50%”, “Between 50% and 75%”, and “Top 25%.” In our analysis, this variable is redefined into a dummy variable which equals to 0 if individuals did not manage to predict their relative standings more than once and equals to 1 if successfully predicted their relative standings at least twice.

*Time Discounting:* we elicited subjects’ time discounting preferences using simple hypothetical choices, similar to Falk et al. (2016). Subjects in our experiment were shown a table with 11 rows. In each row they had to decide whether they preferred an early payment “today” (100 pounds) or paying a varying delayed payment “in 12 months” (100 / 103 / 106 / 109 / 112 / 115 / 118 / 121 / 124 / 127 / 130 pounds). In our analysis, subjects who accepted to receive more than 115 pounds in 12 months (the mean of overall amounts offered) are regrouped as “impatient”, and for the subjects who accepted to receive 115 pounds and lower are regrouped as “patient”. However, for those who misunderstood the question (either switched preferences more than once or chose to receive payment today against high payments in 12 months while chose low payments in 12 months against receiving payment today) are recategorised into the third group - “misunderstand”.

*Risk Attitude:* we elicited subjects’ risk preferences using simple lottery choices as used in Falk et al. (2016). Subjects in our experiment were shown a table with 9 rows. In each row, they had to decide whether they preferred a safe option or playing a lottery. In the lottery, they could receive either 10 pounds or 6 pounds with 50 percent probability. The lottery was the same in each row, but the safe option decreased from row to row. In the first row, the safe option was 10 pounds; in the second it was 9.5 pounds, and so on down to 6 pounds in row 9. Similar to the changes in time discounting, the cutting (re-grouping) point is set at the mean of all certain pay offers (which is paying 8 pounds for certain against the lottery). Therefore, 0 indicates the subjects are risk lovers while 1 means risk averse and 2 identifies those who misunderstood the question.

*Competitiveness:* we used a simple hypothetical choice question to elicit



subjects' competitive preferences. Subjects were asked the choices between a tournament payment (16 pounds per score if the score is the highest, otherwise 0) and a piece-rate payment (1 pound per score).

*Altruism:* To elicit information about subjects' altruistic preferences, we first asked them how much of a prize (100 pounds) he/she would like to share with the other participants if he/she was the lucky winner. Subjects could choose any amount between 0 and 100. In an alternative way, namely by asking individuals to indicate their willingness to share with others without expecting anything in return when it comes to charity on an 11-point scale, with zero indicating completely unwilling to share, and ten indicating complete willingness to share. We use the same wording of the question as in Falk et al. (2016). For altruism, we introduce the product of the two indicators and categorise it into three groups. The first group has the value of 0 implies that the subject is completely unwilling to share. The second group shares the values between 0 and 250 including 250 (where 250 is given by the product of the medians of the two indicators). This group indicates subject's willingness to share is either equal or below the median. Finally, the last group includes all subjects valuing more than 250 which implies these subjects are strongly willing to share.

## C Further Tables

**Table C1:** Descriptive Statistics of Other Variables

	Participants	Mean	SD	Minimum	Maximum	Fractions (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Played Slider Task Before	192	0	0	0	0	
Played Grid Task Before	192	0.016	0.124	0	1	
Played Word Encryption Task Before	192	0.037	0.188	0	1	
Accurately Predicted Relative Standings At Least Twice	192	0.426	0.495	0	1	
Competitive	192	0.130	0.337	0	1	
Patient	182	0.450	0.498	0	1	
Risk Averse	186	0.838	0.368	0	1	
Female	192	0.541	0.499	0	1	
Degree is Econ-related	192	0.405	0.491	0	1	
Final Earning	192	14.75	2.301	6.40	21.60	
Accuracy of Beliefs about Relative Performance	192			1	3	
1 = Accurate						44.52
2 = Underestimate						19.65
3 = Overestimate						35.83
Altruism	192			0	2	
0 = Completely Unwilling to Share						49.39
1 = Willing to Share (Below Average)						31.30
2 = Willing to Share (Above Average)						19.30
Nationality:						
1 = UK						44.79
2 = EEA						13.02
3 = Others						40.62
4 = Prefer Not to Say						1.56
Native Speaking Language is English:						
1 = Yes						48.96
0 = No						49.48
2 = Prefer Not to Say						1.56
Academic Level:						
1 = Undergraduate						79.69
2 = Postgraduate						19.27
3 = Prefer Not to Say						1.04
Years of Study:						
0 = Less Than 1 Year						60.42
1 = 1 Year						9.38
2 = 2 Years						12.50
3 = 3 Years						13.02
4 = More Than 3 Years						4.17
5 = Prefer Not to Say						0.52

**Table C2:** Tests of Sample Balance on Demographics

	RAM (%) (1)	NAM (%) (2)	PAM (%) (3)	R&I (%) (4)	Chi-square test (p-value) (5)	t-test (p-value) (6)	M-W test (p-value) (7)
Gender:					0.536		
Male	26.14 {47.92}	20.45 {37.50}	25.00 {45.83}	28.41 {52.08}			
Female	24.04 {52.08}	28.85 {62.50}	25.00 {54.17}	22.12 {47.92}			
Degree:					0.968		
Econ-related	23.68 {43.75}	25.44 {39.58}	25.44 {39.58}	25.44 {39.58}			
Not Econ-related	26.92 {56.25}	24.36 {60.42}	24.36 {60.42}	24.36 {60.42}			
Nationality:					0.239		
UK	23.26 {41.67}	20.93 {37.50}	23.26 {41.67}	32.56 {58.33}			
EEA	16.00 {8.33}	32.00 {16.67}	28.00 {14.58}	24.00 {12.50}			
Others	30.77 {50.00}	28.21 {45.83}	24.36 {39.58}	16.67 {27.08}			
Prefer Not to Say	0.00	0.00	66.67	33.33			
Native Speaking Language is English:					0.114		
Yes	26.32 {52.08}	21.05 {41.67}	21.05 {41.67}	31.58 {62.50}			
No	24.47 {47.92}	29.79 {58.33}	27.66 {54.17}	18.09 {35.42}			
Prefer Not to Say	0.00	0.00	66.67	33.33			
Academic Level:					0.031		
Undergraduate	22.88 {72.92}	22.88 {72.92}	27.45 {87.50}	26.80 {85.42}			
Postgraduate	35.14 {27.08}	35.14 {27.08}	10.81 {8.33}	18.92 {14.58}			
Prefer Not to Say	0.00	0.00	100.00	0.00			
Years of Study:					0.233		
Less Than 1 Year	25.86 {62.50}	28.45 {68.75}	22.41 {54.17}	23.28 {56.25}			
1 Year	22.22 {8.33}	33.33 {12.50}	33.33 {12.50}	11.11 {4.17}			
2 Years	16.67 {8.33}	8.33 {4.17}	41.67 {20.83}	33.33 {16.67}			
3 Years	24.00 {12.50}	24.00 {12.50}	16.00 {8.33}	36.00 {18.75}			
More Than 3 Years	50.00 {8.33}	12.50 {2.08}	12.50 {2.08}	25.00 {4.17}			
Prefer Not to Say	0.00	0.00	100.00	0.00			
Age:							
RAM vs NAM						0.013	0.313
RAM vs PAM						0.885	0.000
RAM vs R&I						0.097	0.070
NAM vs PAM						0.097	0.007
NAM vs R&I						0.885	0.474
PAM vs R&I						0.013	0.051

*Notes:* The null hypothesis for t-test/Mann-Whitney U (M-W) test is that the difference between the means/distributions of the two independent samples is zero. The Chi-square test is used to check if there is a relationship between the demographical variables and treatments. Notice that curly bracket indicates the fraction of the corresponding group within that treatment.

**Table C3:** Individuals Who Predicted Their Relative Standings At Least Twice

	Observations	Fraction (%)
Treatment:	245	42.61
RAM		22.0
NAM		31.8
PAM		26.9
R&I		19.2
Female	245	43.7
Studied More Than 1 Year	242	37.2
Speak English Natively	239	45.2
From UK	240	41.2
Postgraduate	239	17.6
Degree is Econ-related	245	35.1
Competitive	245	15.9
Patient	245	38.8
Risk Averse	245	81.6

**Table C4:** Individuals Who Belong to High Productivity Group

	Observations	Fraction (%)
Treatment:	273	47.48
RAM		25.27
NAM		26.01
PAM		24.18
R&I		24.54
Female	273	54.9
Studied More Than 1 Year	272	39.0
Speak English Natively	268	51.9
From UK	271	48.0
Postgraduate	270	17.4
Degree is Econ-related	273	38.1
Competitive	273	16.5
Patient	273	42.9
Risk Averse	273	84.2

**Table C5:** Peer Effects: OLS Regression

	Dep. Var.: Effort in The Team Work Stage	
	(1)	(2)
Own Effort (Individual Work Stage)	0.693*** (0.0683)	0.690*** (0.0717)
Partner's Effort (Individual Work Stage)	-0.0333 (0.0570)	-0.0297 (0.0523)
R&I	-0.411 (0.399)	-0.420 (0.419)
Constant	11.21*** (2.820)	10.12*** (2.764)
Observations	287	287
Participants	96	96
R-squared	0.677	0.698
Task and Round Fixed Effects	YES	YES
Other controls	NO	YES

*Notes:* OLS Estimations. Dependent variable is the effort in the team work stage. The omitted treatment is *RAM*. Robust standard errors clustered at session level are reported in brackets below the estimates. Notice that using robust standard errors clustered at individual level will not change our implications in subsection 4.4. Column (1) reports estimates which controlled task and round fixed effects. Column (2) further adds for all other individual characteristics. \*\*\* Significant at 1% level, \*\* significant at 5% level, \* significant at 10% level.

## D Experimental Instructions

### Instructions [All Treatments]

Thank you for participating in this session. Please raise your hand if you want to ask a question. Apart from asking questions in this way, you must not communicate with anybody in this room. Please now turn off mobile phones and any other electronic devices. These must remain turned off for the duration of this session.

You have been allocated to a computer booth according to the number on the card we gave you as you came in. You must not look into any of the other computer booths at any time during this session. To ensure anonymity, your actions in this session are also linked to this number. From now on, please keep it safe as this card will be required for payment at the end.

You will be paid a show up fee of £4, plus any earnings you accumulate during this session. The amount of money you accumulate will depend partly on your actions, partly on the actions of others and partly on chance. All payments will be made in cash. None of the other participants will see how much you have been paid.

### The Setup [All Treatments]

This session consists of three rounds in which you will work on three different tasks. You will perform only one of the tasks in each round and for each task you will get a score based on your performance. The order in which you will perform each task is random.

Each round is divided into three stages: a practice stage, an individual work stage, and a team work stage. The practice stage lasts for 2 minutes and allows you to familiarise yourself with the tasks. Both work stages, individual and team work, last for 4 minutes. Your performance in the individual work stage will be ranked against all other participants. The computer will assign to you another participant as a partner for the team work stage according to a rule explained below [**RAM and R&I**] (Based on this ranking the computer will assign to you another participant as a partner for the team work stage according to a rule explained below [**NAM and PAM**]).

Further details of the payment, the pairing rule and the tasks will be explained below.

### **Payment [RAM, NAM, and PAM]**

In each round your team performance at the team work stage will affect your earnings. In particular, for your team work you earn CREDITS. Your CREDITS are given by the average score of your team.

For example, if player A's score is 38 and player B's score is 28 in the team work stage, each of them earns  $\frac{38+28}{2} = 33$  CREDITS.

At the end of the experiment the computer will randomly choose one out of the three rounds to determine your earnings. In other words, all rounds (or tasks) are equally important to you regarding the payment. The CREDITS that you earned from the selected round will determine your payment from performing the tasks: the CREDITS will be exchanged into pounds and the exchange rate will be: 1 CREDIT = £0.40.

As an example, suppose that in the round that is randomly chosen for payment at the end you earned 38 CREDITS. Then your total earnings from performing the tasks will be as follows:

$$\text{Total Earnings} = 38 * £0.40 = £15.20$$

### **Payment [R&I]**

In each round your performance will influence your earnings. In particular, for your work you earn CREDITS. In the individual work stage your CREDITS are equal to your score. In the team work stage your CREDITS are given by the average score of your team.

For example, if player A's score is 30 in the individual work stage, player A earns 30 CREDITS. If player A is working in a team with player B in the team work stage, player A's score is 38 and player B's score is 28, each of them earns  $\frac{38+28}{2} = 33$  CREDITS.

At the end of the experiment the computer will randomly choose one round (out of the three rounds) and one stage (out of individual work stage and team work stage) to determine your earnings. In other words, both work stages in all rounds (or tasks) are equally important to you regarding the payment. The CREDITS that you earned from the selected round and the selected stage will determine your payment from performing the tasks: the CREDITS will be exchanged into pounds and the exchange rate will be: 1 CREDIT = £0.40.

As an example, suppose that in the round that is randomly chosen for payment at the end you earned 38 CREDITS at the selected stage. Then

your total earnings from performing the tasks will be as follows:

$$\text{Total Earnings} = 38 * \text{£}0.40 = \text{£}15.20$$

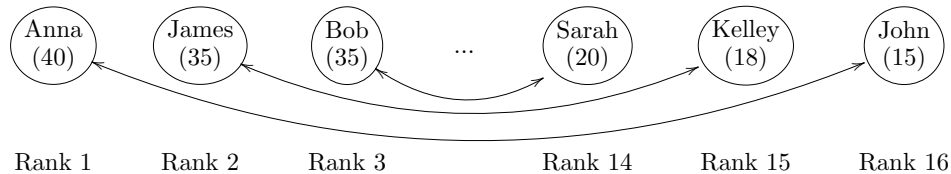
### Pairing Rule [RAM and R&I]

The computer will randomly assign to you another participant as a partner for the team work stage. Each team consists of 2 partners.

### Pairing Rule [NAM]

The computer will rank all participants according to their scores in the individual work stage. Each team consists of 2 partners. Teams are formed by pairing participants based on their scores in the individual work stage: the best performing participant will be working in a team with the worst performing one, the second best will be working in a team with the second worst, and so on and so forth (see the example in the figure below). If some participants share the same score their rank will be drawn randomly to avoid ties. For instance, Bob and James who have a score of 35 each, have each a chance of 50% to be assigned rank 2, respectively rank 3.

**Figure D1:** Team assignment with 16 participants (individual scores are shown in brackets).



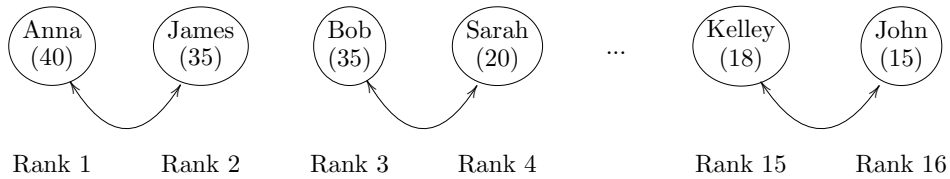
### Pairing Rule [PAM]

The computer will rank all participants according to their scores in the individual work stage. Each team consists of 2 partners. Teams are formed by pairing participants based on their scores in the individual work stage: the best performing participant will be working in a team with the second best performing one, the third will be working in a team with the fourth, and so on and so forth (see the example in the figure below). If some participants share the same score their rank will be drawn randomly to avoid ties. For



instance, Bob and James who have a score of 35 each, have each a chance of 50% to be assigned rank 2, respectively rank 3.

**Figure D2:** Team assignment with 16 participants (individual scores are shown in brackets).

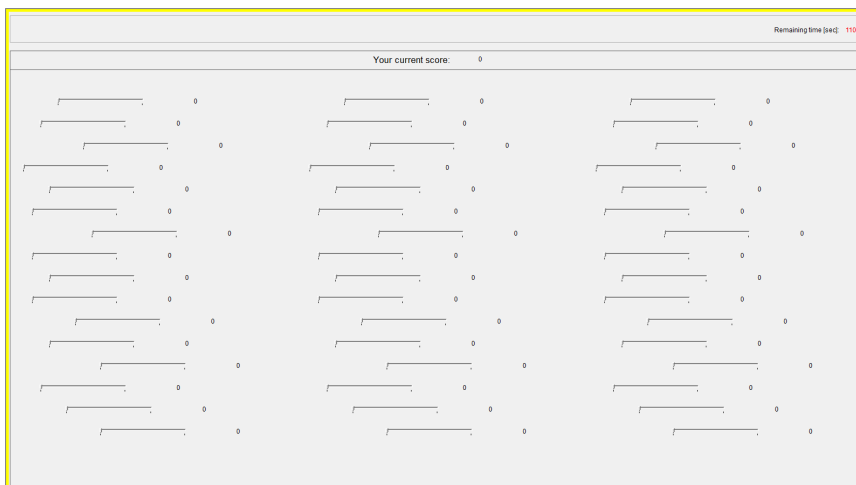


## The Tasks [All Treatments]

### Slider

The task will consist of a screen with 48 sliders. Each slider is initially positioned at 0 and can be moved as far as 100. Each slider has a number to its right showing its current position. You can use the mouse in any way you like to move each slider. You can re-adjust the position of each slider as many times as you wish. Your task is to position each slider at 50. Your score in the task will be the number of sliders positioned at exactly 50 within 4 minutes. The decision screen is seen in the figure below.

**Figure D3:** The Slider Task



## Grid

5 by 5 grids with randomly distributed 0's and 1's will appear on the screen. Your task is to count the number of 0's. Once you count a table correctly, the computer will prompt you with another table which you will be asked to count 0's. Once you count that table, you will be given another table and so on. Your score in the task will be the number of grids with a correct count of 0's entered within 4 minutes. The decision screen is seen in the figure below.

**Figure D4:** The Grid Task

The screenshot shows a web-based interface for the Grid Task. At the top, it displays "Your current score: 0" and "Remaining time (sec): 117". The main area contains a 5x5 grid of numbers. To the right of the grid, there is a text prompt "How many zeros are in the table:" followed by a text input field containing the number "1" and a "Next Table" button.

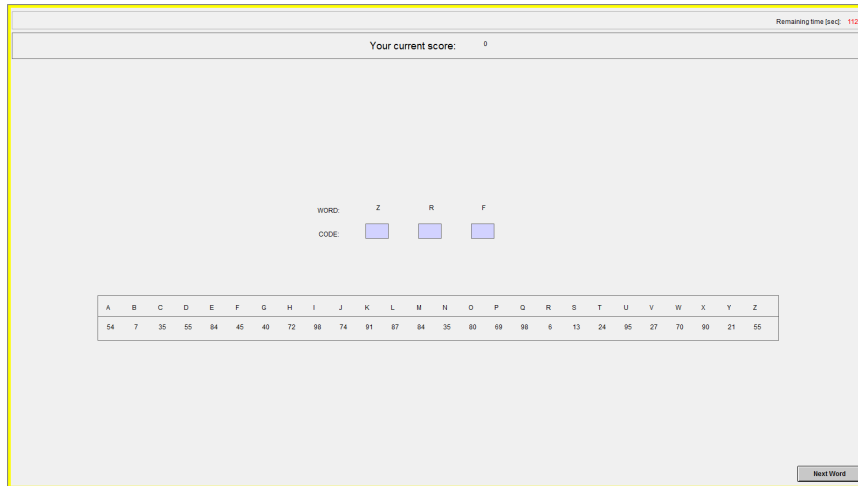
1	0	1	0	0
0	0	0	1	1
0	0	1	0	1
1	0	1	1	0
0	0	0	0	1

## Word Encryption

This task consists of encoding words into numbers. Each word is a combination of three letters. You have to allocate a number (0-100) to each letter. The encryption code can be found in a table below the corresponding word. Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. Your score in the task will be the number of words encoded correctly within 4 minutes. As an example, the decision screen can be seen in the figure below.

Note that the encryption table during the experiment will be different from the given example. Before each stage of this task, the computer first selects in the table a new set of random numbers (0-100) to be used for the encoding of the capital letters. Then, the computer program shuffles the

**Figure D5:** The Word Encryption Task



position of the capital letters in the table. Note that the encryption table will differ between practice, individual, and team work stages.

### Other Information [All Treatments]

During each task, some information will appear at the top of your screen, including the time remaining and your score in the task. After successfully generating all possible teams, the computer will first show you your score, your rank, the highest score and the lowest score among all participants in the individual work stage and then your partner's rank and score. At the end of the team work stage, you will see a summary screen showing your score, your partner's score, and your team's score.

At the end of the session your total cash payment, including the £4 show up fee, will be displayed on your screen. Please leave the computer booth one by one when asked to do so to receive your payment. Please leave all other material on your desk. Thank you for participating. Are there any questions?