Dynamic Incentive Effects of Assignment Mechanisms: Experimental Evidence *

Thomas Gall[†], Xiaocheng Hu[‡], and Michael Vlassopoulos[§]

March 2019

Abstract

Optimal assignment and matching mechanisms have been the focus of exhaustive analysis. We focus on their dynamic effects, which have received less attention, especially in the empirical literature: anticipating that assignment is based on prior performance may affect prior performance. We test this hypothesis in a lab experiment. Participants first perform a task individually without monetary incentives; in a second stage, they are paired with another participant according to a pre-announced assignment policy. The assignment is based on first-stage performance and compensation is determined by average performance. Our results are largely consistent with theory: pairing the worst performing individuals with the best yields 20% lower first stage effort than random matching and does not induce truthful revelation of types, which undoes any policy that aims to reallocate types based on performance. Perhaps surprisingly, however, pairing the best with the best yields only 5% higher first stage effort than random matching and the difference is not statistically significant.

Keywords: Matching, assignment games, truthful revelation, performance, dynamic incentives

JEL: C78, C91, M54, D23

*We greatly benefited from comments and suggestions received from an associate editor and two referees, as well as seminar participants at Southampton, Marburg, the London Experimental Workshop, the 12th Workshop Matching in Practice, the 2017 RES Annual Conference and the 2017 ESA European Meeting. Financial assistance from the School of Social Sciences, University of Southampton is gratefully acknowledged. All remaining errors are our own.

†Department of Economics, School of Social Sciences, University of Southampton, Southampton SO17 1BJ, UK; email: t.gall@soton.ac.uk.

‡European Centre for the Environment and Human Health, University of Exeter; email: x.hu@exeter.ac.uk.

§Department of Economics, University of Southampton; IZA, Bonn; email: m.vlassopoulos@soton.ac.uk.

1 Introduction

Individuals' payoffs depend on the economic environment they are placed in. For example, peers' attributes can affect an individual's payoff in the workplace or the classroom and one's spouse's attributes will affect marriage payoffs. However, the attributes that determine payoffs are likely to be the consequence of prior choices made, in full anticipation of the later assignment to other people, peers, tasks, or jobs. For instance, expectations about the future assignment into colleges, firms or teams may well provide powerful incentives for accumulating human capital. This raises an interesting question: does the manner of how individuals are assigned to each other or to tasks, jobs, schools etc. affect their prior choices, such as earlier stage investments and performance?

In the workplace a wide range of methods to assign workers to tasks and to each other is used in practice. 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 productivity gains (Hamilton et al., 2003). More importantly, this pattern may also be promoted by managers who pursue social goals such as fairness or equality in remuneration, with an aim to boost morale (see e.g. Bewley, 1999; Blinder and Choi, 1990; Pfeffer and Langton, 1993) or job satisfaction (Card et al., 2012) among workers. Such organisational choices may come at a cost, however, limiting an individual's desire to exert effort at an earlier stage. That is, there is a dynamic trade-off between equity ex post and efficiency ex ante. More subtly, distortions in early stage behavior will also make observations of earlier behavior less informative of future performance, jeopardising attempts to promote equity or fairness through assignment that is based on past choices. Conversely, if the best performers are assigned to better partners this will provide additional incentives for effort at an earlier stage. Depending on the degree of production complementarity (Franco et al., 2011) and the strength of incentives (Bandiera et al., 2013), this pattern will also arise when workers are allowed to choose their own teammates since workers will tend to match positively assortatively in ability. 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).

Also outside the workplace assignment mechanisms of individuals vary widely, sometimes as the result of an explicit policy, but often as the result of a decentralised market place. For example, in higher education there is a marked difference between the US and the UK where students self-select by academic ability into universities guided by detailed rankings, and continental Europe where students focus more on the city a university is located in. In secondary education, countries differ substantially in the degree to which they sort pupils by academic achievement, i.e. tracking (cf. Betts, 2011; Hanushek and Woessmann, 2011). Evidence on the marriage market suggests that mating is assortative in educational achievements (see e.g. Fernandez et al., 2005).

In all these examples, individuals who anticipate that their later assignments and outcomes depend on their earlier stage choices will therefore respond to the assignment mechanisms used at later stages. This reasoning has been considered in the theoretical literature, examining e.g. investments taken before marriage or business partnerships (see e.g. Cole et al., 2001; Felli and Roberts, 2016; Peters and Siow, 2002; Bidner, 2010), providing some insights into the incentive effects of different matching mechanisms (e.g. Booth and Coles, 2010; Gall et al., 2006, 2015), and mechanism design (Hatfield et al., 2017). However, there has been no comparable interest in examining empirically the dynamic effects of different assignment mechanisms.[1] The aim of the current paper is to fill this void.

We design a real effort experiment with two stages: in a first stage, participants perform a task 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, perform the task and receive compensation that depends on the average performance of the team. However, the tasks worked on permit learning by doing, introducing a dynamic complementarity by increasing individual productivity in the later stage. Given the novelty of examining the resulting dynamic incentive effects we opt for a clean design and shut down static complementarities or substitutabilities, i.e. peer effects within teams in the second stage. Thus the design will allow some extrapolation of the results for the presence of positive or negative peer effects.

The experimental variation stems from varying the rule that matches participants in the second stage. We examine three salient forms of team assignment: random matching (RAM), matching participants randomly ignoring their performance in the first stage as a baseline treatment, positive assortative matching (PAM), in which the best performer is matched with the second best and so on, and negative assortative matching (NAM), in which the best performer is matched with the worst and so on. Besides the practical relevance, mechanism design theory would suggest that these assignment policies are interesting for another reason: PAM rewards higher first stage performance with a better match, and thus has a positive dynamic incentive

[1] There has been considerable interest in examining experimentally the static properties of assignment mechanisms, especially strategy-proofness, (see e.g. Pais and Pinter, 2008; Calsamiglia et al., 2010; Braun et al., 2014).

effect on first stage effort. Under RAM, the second stage match is unrelated to first stage performance, thus shutting down the dynamic incentive effect. By contrast, under NAM higher first stage yields a partner with worse first stage performance, yielding a negative dynamic incentive effect. However, while PAM tends to induce investment behavior that is strictly monotone in productivity, thus revealing agents' types through their investment choices, this is not necessarily the case with NAM, as there is a tradeoff between one's own performance and the partner's performance in the second stage. Finally, as we are also 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 participants also receive an individual piece rate for their first stage performance and are randomly matched into teams.

We use a simple model to derive some theoretical pointers as to how outcomes in the individual work stage might differ across the treatments. Individuals who differ in their cost of effort, exert effort and invest through learning by doing, and then are assigned into teams of size two with payoffs increasing in their partner's effort. The results we obtain are largely consistent with the predictions. Intuitively, NAM leads to the lowest performance in the individual work stage, substantially lower than in the other treatments (20% reduction in mean performance compared to RAM, 30% compared to PAM). Perhaps surprisingly, RAM yields quantitatively similar performance outcomes as the positively incentivized treatments PAM and $R\&I$. The results point to an asymmetric effect: punishing effort appears to have a greater effect than rewarding effort, from a baseline of RAM. The evidence is also consistent with a more complex prediction of the model, namely, that NAM will not allow for truthful revelation of individual ability: since NAM precludes monotonicity of payoffs in one's choice, ex ante behavior will be characterised by bunching, i.e., different cost types will choose the same investment. Observed behavior under NAM is consistent with this prediction and suggests that measured performance before the assignment is not very informative about true productive and later performance. 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 that has not been highlighted before: a matching policy that aims to increase equity by equalizing average productivity and thus performance across groups (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. Far more worrying is the fact that NAM will fail to induce truthful revelation, reducing the informativeness of early stage behavior for later stage performance. That is, observed effort before the assignment will generally not allow reliable conclusions regarding individuals' innate abilities, which, however, is needed to effectuate the desired equitable assignment. Hence, achieved equality of the final allocation will likely fall considerably short of the policymaker's aspirations.

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 proofs, additional tables, and the experimental instructions.

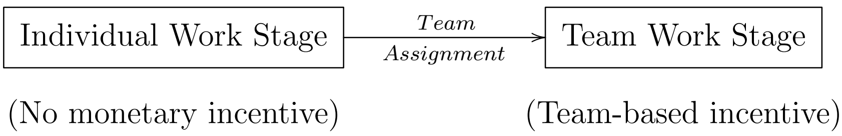
2 Experimental Design

To study the dynamic incentive effects of various assignment mechanisms we designed an experiment with the key feature that a participant's initial performance in a real effort task determines the partner that the participant is matched with in a later stage according to a known assignment rule. In this later stage compensation depends on the average performance of the pair. This provides an implicit incentive for a participant to exert effort in the earlier stage, in anticipation that this will lead to a more desirable match. This setup allows us to examine our main question of how early stage performance is affected by the mechanism, which varied across treatments, assigning participants to each other in a later stage.

2.1 The Stages of the Experiment

Participants in our experiment performed a real effort task over three rounds, consisting of two stages each. 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.[2]

Figure 1: The Two Stages of a Round



In the individual work stage subjects had four minutes to work on the task and their performance was not explicitly rewarded (except for treatment *R&I*, see below). At the end of the stage, we also elicited their belief about their performance relative to that of the other participants in their session in an incentive compatible way.[3] Then individuals were informed about their true rank as well as the maximum and minimum scores in the individual stage. After the individual work stage subjects were assigned to a partner for the team work stage. The assignment rule depends on the treatment and is explained in detail below. Subjects were shown their partner's rank and score before beginning the team work stage.

In the team work stage subjects had another four minutes to work on the task. Performance was rewarded by monetary payment, which was based on the average of 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 the average of own and partner's score, which determined payment. Note that the team work stage does not entail joint production per se, as individuals perform the same task individually, but the compensation scheme induces a local public good.

This pattern was repeated three times (rounds). In each round, participants performed a different task: first without a partner in the individual stage and then, assigned to a new partner, in the team work stage. 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

[2]Note that the information feedback is kept constant across treatments, ensuring that possible effects of the feedback, such as anchoring or intrisic motivation through relative performance feedback (see e.g. Charness et al., 2014; Kuhnen and Tymula, 2012; So et al., 2017), do not affect differences between treatments.

[3]They received £0.4 for correctly guessing in which quartile of the performance distribution they belong.

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 the hypothesis that incentives to exert individual effort in the first stage depend on the assignment mechanism employed in the second periods 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 assignment mechanisms. The first three treatments relied exclusively on the implicit incentives for first stage behaviour generated by different assignment mechanisms, while the fourth treatment added explicit monetary incentives in stage 1 to allow a quantitative comparison of the strength of explicit and implicit incentives.

Random matching (RAM)

Our benchmark treatment is random matching: 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).[4]

Positive assortative matching (PAM)

Positive assortative matching assigns individuals into groups based on their effort before the assignment. 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 partner. This pattern often endogenously arises in situations when individuals are allowed to choose their partners, since absent compensation payments individuals will tend to match positively assortatively (see e.g. Chen et al., 2015).

[4]Random matching could also capture the situation where the actual assignment is uncertain at the time of choosing effort, and RAM, PAM and NAM are chosen with equal probabilities, or if individuals use a Laplace heuristic.

Negative assortative matching (NAM)

Similarly to PAM, negative assortative matching assigns individuals into groups based on their effort before the assignment, 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 partners, thus generating balanced teams in terms of average performance of 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 each other; only effort within teams is rewarded by monetary payments. In contrast, the final treatment ($R\&I$) rewards effort before assignment 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 differ between 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 5×5 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 10×15 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 numbers range from 0 to 100, 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. (2018) who find that it reduces the scope for learning when repeating the task, but does not eliminate it.

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.[5] 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 to allow for only limited improvement due to learning (Benndorf et al., 2018).

2.4 Payments

In each of the treatments *RAM*, *PAM* and *NAM* subjects were paid only in the team work stage and according to the average of the scores of the two partners. At the end of a session, for each subject one of the three tasks was randomly chosen and

[5] An ideal test of the prediction would involve a task that does not allow for any learning by doing. However, this change would not only eliminate learning spill-overs, but also dynamic incentives altogether, as having a partner who performed better in stage 1 will not imply better performance in the task in stage 2.

the subject's payment was her average team work stage score in that task multiplied by a piece rate of £0.4 per score point.

Treatment *R&I* additionally rewards individual performance in the individual work stage, given by the subject's score in 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 per score point. This allows us to compare first stage performance in R&I to that in the other treatments where first stage performance is not financially rewarded, holding the second stage incentive scheme constant and the expected income comparable across the treatments.

2.5 Procedures

The experiment was conducted at the Social Sciences Experimental Lab (SSEL) of the University of Southampton in spring of 2016. We ran three sessions of each of the four treatments described above (RAM, *R&I*, PAM and NAM), for a total of 12 sessions. The order of treatments and the sequence of tasks within sessions was randomized, 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 16 student subjects from various departments, with 192 participants in total (104 females and 88 males).[6] We recruited the subjects from the SSEL subject pool, using ORSEE (Greiner, 2004). The experimental instructions were provided to each subject in written form and also read aloud. 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 they understood the rules and their treatment's assignment mechanism. Each subject earned a show-up fee of £4 and on average 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 θ, and

[6]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.

lasts for two stages. Suppose θ is distributed on an interval $[\underline{\theta}, \overline{\theta}]$ with $0 < \underline{\theta} < \overline{\theta}$. Suppose also that the distribution of productivity θ has full support.[7]

Production

In each stage t each agent can exert effort e_t to generate output y_t. Effort e_t comes at a utility cost c_t, however. To best reflect the experimental setup we impose the following formal assumptions on the cost function.

Assumption 1. *Stage 1 and 2 effort cost functions are given by:*

$$c_1(e_1) = \frac{e_1^2}{2\theta} \text{ and } c_2(e_1, e_2) = \frac{e_2^2}{2(\theta + \lambda e_1)},$$

where $\lambda > 0$. They have the following properties:

(i) $c_t(.)$ does not depend on other individuals (no peer effects in production),

(ii) $c_t(.)$ strictly increases and is strictly convex in e_t (real effort),

(iii) The marginal cost of effort strictly decreases in type θ (heterogeneity),

(iv) The marginal cost of stage 2 effort e_2 strictly decreases in stage 1 effort e_1 (learning-by-doing).

In the experiment, individual performance in a task only depends on the individual's own actions, in all stages. This will preclude peer effects, at least in performing the task, and is reflected in our first assumption. Since participants engage in a real effort task we impose increasing marginal cost, which will imply diminishing returns to effort. Heterogeneity of effort cost captures differences between individuals in their ability to perform the experimental tasks. Since performance in all experimental tasks can be enhanced by practice (even if it is only by familiarising oneself with the user interface), the cost of effort will decrease in past experience, i.e. there is learning-by-doing. The strength of the intertemporal learning spill-over is given by the parameter λ. The remainder of this section will use a fully parameterised cost function, but the results derived below are driven by the four qualitative assumptions given above, not by the functional form.

[7]Assuming a discrete set of agents or types will not change results for NAM, RAM and $R\&I$, in which one's equilibrium match is random. The assumption will preclude pure strategy equilibria under PAM. Using mixed strategies or introducing some noise in performance or the matching would ensure that an equilibrium exists and has similar properties as the one derived below. The intuition, namely, that dynamic incentives will boost stage 1 effort under PAM, reduce them under NAM, and be absent under random matching, will still be present.

Finally, suppose that individual output is simply given by effort:[8]

$$y_t = e_t.$$

Assignment

After stage 1 individuals are assigned into teams of size 2, with attributes (e_1, θ) and (e'_1, θ'). The assignment takes one of the four forms described above: RAM, $R\&I$, PAM, or NAM. Team output in stage 2 is the sum of individual stage 2 output y_2 and y'_2:

$$Y_2 = y_2 + y'_2.$$

Payoffs

An individual's monetary payoff is given by a piece rate 1 per unit of average team output in stage 2, $Y_2 = y_2 + y'_2$. Hence, an individual's overall payoff is given by

$$(y_2 + y'_2)/2 - c_1(e_1) - c_2(e_1, e_2).$$

Under $R\&I$, the individual receives a piece rate w both per unit of stage 1 performance y_1 and stage 2 average team output $y_2/2$, yielding payoff:

$$w(y_2 + y'_2)/2 + wy_1 - c_1(e_1) - c_2(e_1, e_2).$$

In the experiment we set $w = 1$, but only one of the two stages was selected to determine the payment, each with probability 1/2. Hence, if stage 1 and stage 2 output are of similar size, $R\&I$ will induce a similar aggregate wage bill as RAM.

Note that no peer effects in production implies that an individual's productivity does not depend on their match, but their payoff will. This also implies that, holding constant stage 1 effort e_1, aggregate output in stage 2 and surplus are independent of the assignment mechanism used. Hence, no peer effects in production implies that any differences in aggregate output and utility in stage 2 are entirely due to the dynamic incentives effects of the different assignment mechanisms.

Solution Concept

The type of assignment mechanisms may affect participants' stage 1 behaviour. Individual stage 2 payoff increases in the effort of their partner. If individuals anticipate

[8]An equivalent formulation could specify output as a strictly increasing and concave function of effort and use a linear cost, for instance, and impose assumptions on the output function analogous to Assumption 1.

that the quality of their partner in stage 2 depends on their own stage 1 performance (in the non-random assignments), stage 1 effort will be rewarded (or punished) with a better (worse) partner. Of course, stage 1 performance may or may not be informative of stage 2 behaviour, as both are part of equilibrium behaviour. To identify equilibrium behaviour we use therefore a sub-game perfect Nash equilibrium in individual effort choice e_1 and e_2 in the two-stage game played by individuals. We omit general properties, e.g., existence because this type of matching cum investment game with a continuum of players has been 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. (2015) who also allow for RAM, respectively, NAM.

Stage 2 behavior

Since stage 1 equilibrium behaviour will depend on anticipated stage 2 equilibrium outcomes, we use backward induction to derive an equilibrium and start with behavior in stage 2. In stage 2, individuals are assigned into teams and choose individual effort e_2 given their assignment and their stage 1 effort choice e_1. That is, 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 partner. Note that the cost of stage 1 effort is sunk. Hence, individual optimal stage 2 effort satisfies:

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

and $e_2^* = w(\theta + \lambda e_1)/2$ under *R&I*. Since this must be true for each individual in all teams, an individual's overall payoff from both stages under RAM, PAM and NAM is given by:

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

This payoff clearly increases in the attributes e_1' and θ' of one's partner in stage 2. Under *R&I* an individual additionally obtains payoff $y_1 = we_1$, so that $u(.) = w^2 \frac{\theta+\lambda e_1+2(\theta'+\lambda e_1')}{8} + we_1 - \frac{e_1^2}{8\theta}$, which also increases in the stage 2 partner's characteristics. This implies the following fact.

Fact 1. *An individual's payoff strictly increases in the quality θ' and stage 1 effort e_1' of their stage 2 partner.*

This property relies on the stage 2 reward scheme, introducing a local public good, and learning-by-doing for the assertion on stage 1 efforts.

Stage 1 behavior

In stage 1 an individual chooses only effort e_1. This choice depends on the continuation payoff in stage 2 through two possible channels: higher stage 1 effort will reduce the cost of stage 2 efforts through learning-by-doing (the intertemporal spill-over channel), and it may affect the attributes e_1' and θ' of one's stage 2 partner through the assignment mechanism in place (dynamic incentive channel). Moreover, under $R\&I$, stage 1 effort choice will additionally depend on the reward for stage 1 performance w directly. That is, 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 θ'.

We start by examining the two random assignment mechanisms RAM and $R\&I$. Under both mechanisms e_1' and θ' do not depend on an agent's choice of e_1. That is, both RAM and $R\&I$ shut down the dynamic incentive channel, and $R\&I$ introduces direct piece rate incentives for stage 1 effort on top of the intermporal spill-over channel. The first order conditions become:

$$\frac{e_1^{RAM}}{\theta} = \frac{\lambda}{8} \text{ and } \frac{e_1^{R\&I}}{\theta} = w^2\frac{\lambda}{8} + w.$$

Under PAM and NAM both channels are present, but will have opposite signs. Under PAM e_1' increases in e_1, and θ' will also depend on e_1. The individual optimization problem becomes now:

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

If, as under RAM, e_1 increases in type θ, then higher e_1 also implies being matched to a higher type θ'; see the appendix for a proof that this is indeed the case. In this case $e_1^{PAM} > e_1^{RAM}$ because PAM will reward stage 1 effort through the dynamic incentive channel in addition to the positive intertemporal spill-over through learning-by-doing. See the appendix for the full derivation.

Under NAM the stage 1 effort of one's partner e_1' (weakly) decreases in own effort e_1. But this means that the two channels are in conflict: a higher stage 1 effort e_1 will be rewarded by lower effort cost in stage 2 (intertemporal spill-overs), but punished by receiving a partner with lower e_1' (dynamic incentives). Hence, strategies need not increase in type, and a semi-pooling equilibrium will result (see appendix for details): lower productivity participants will choose effort $e_1 = 0$, while higher productivity participants will choose $e^{NAM} = e^{RAM}$. This is because individuals at the top will be matched with a random partner among all individuals who set $e_1 = 0$, and increasing effort even marginally at the bottom would imply a discrete loss in the quality of stage 2 assignment.

To sum up, the Nash equilibrium for each assignment mechanism has the following properties (details are in the appendix):

Fact 2. *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 $e_1^{R\&I} = w^2\frac{\lambda\theta}{8} + w\theta$.*
- *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 in stage 1 allows us to compare 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, if $\lambda < 8w/(1-w^2)$, $R\&I$ induce higher effort for all types than RAM, which in turn induces higher effort than NAM, and strictly so for some types.

(ii) PAM induces higher effort than $R\&I$ if the degree of learning-by-doing λ is sufficiently high and the difference increases in λ, i.e. there is $\bar{\lambda}(w) \geq 0$ such that $e_1^{PAM} > e_1^{R\&I}$ for all $\lambda > \bar{\lambda}(w)$, where $\bar{\lambda}(1/2) = 0$ and $\bar{\lambda}$ increases in w.

(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 λ.

Very intuitively, dynamic incentives under PAM boost stage 1 effort, while those under NAM reduce it, compared to RAM where no dynamic incentives are present, as one's stage 2 partner does not depend on stage 1 effort. Consequently our main result, the comparison between RAM, PAM, and NAM, relies on the assumption of learning-by-doing (NAM and RAM coincide if $\lambda = 0$) and no peer effects in production combined with splitting the surplus in a pair. Allowing for positive peer effects, that is, letting efforts be complements, would generate a game with increasing differences and increase equilibrium effort under PAM in stage 2 and thus also in stage 1, decrease both under NAM and not affect expected effort in the random protocols. The opposite will hold for negative peer effects. The comparison to *R&I*, which uses monetary incentives in stage 1 as well as in stage 2 depends on the power of incentives. For $w > 1/2$ stage 1 effort is higher than RAM for all λ. This is because under RAM stage 1 effort is only rewarded through the intertemporal spill-over, while *R&I* rewards stage 1 effort directly through a piece rate. Comparing PAM to *R&I* generates some ambiguity: PAM induces higher effort than *R&I* for sufficiently high degrees of learning-by-doing; note, however, this is the case for all $\lambda > 0$, if $w = 1/2$. In the experiment, in treatment *R&I* only one of the two stages was chosen randomly for payment, using the same piece rate as in the other treatments for both stages.

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* and less close to those under NAM, and effort under PAM increases faster than under *R&I*, as the degree of learning-by-doing increases.

Incentive Compatibility and Truthful Revelation

Since stage 2 assignments are based on stage 1 outcomes, a social planner wishing to alter the assignment of productive types into teams relies on the equilibrium choices of stage 1 effort to reveal individual types.[9] Truthful revelation of individual types is implicitly guaranteed under PAM, RAM and *R&I* since equilibrium stage 1 effort choices strictly increase in type θ_i.

Fact 2 states that this is not true for NAM, however, since it induces partial bunching (or pooling): all agents with types $\theta_i \leq \hat{\theta}$ will choose the same effort level 0 and higher types will choose the same effort as under RAM. This implies that effort choices are more dispersed under NAM than 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 λ and the distribution of θ, and on whether the cost function is separable in θ and λ). Partial pooling under NAM also implies that stage 1 effort choices are not strictly monotone in type, and therefore stage 1 performance does not reveal types. Hence, using NAM conditional on stage 1 performance will not induce NAM conditional on true types θ_i.

[9]Note, however, that the predictions in Proposition 1 carry over to a setting where assignment is in terms of productivity types θ, in contrast to e.g. the ratchet effect (see Cooper et al., 1999; Charness et al., 2011).

Recalling that stage 2 effort choices strictly increase in type under all assignment mechanisms, this reasoning generates another prediction on the correlation of individuals' ranks in stage 1 and stage 2 performance.

Corollary 1 (Correlation)**.** *The correlation of individual ranks in equilibrium stage* 1 *and stage* 2 *performance is* 1 *under* RAM, PAM *and R&I, and strictly less than* 1 *under* NAM *with random tie-breaking.*

4 Results

4.1 Sample

Summary statistics of participant characteristics are presented in Table 4. Using Chi-square, t-, and Mann-Whitney-U-tests (M-W test) Table 7 in the appendix shows that participants' characteristics are balanced across treatments with the exception of academic level and age (which are plausibly highly related).

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.[10] The mean score across treatments and tasks was 23. While the mean scores in the slider and the grid tasks were very similar, the mean score in the word encryption task was significantly lower.[11]

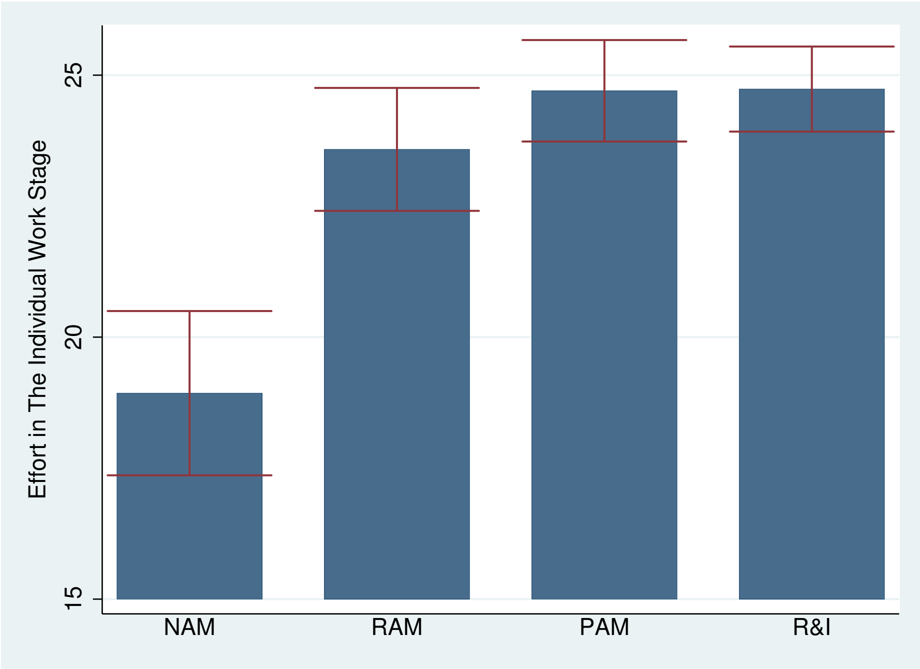
To address our main question of whether individual effort in stage 1 is affected by the assignment mechanism used in stage 2, Figure 2 shows the mean individual performance levels by treatment. In particular, the point estimate of the mean score under NAM is about 20% less than under RAM, while PAM and *R&I* are virtually indistinguishable and both induce 5% higher scores than RAM. However, mean performance under RAM, PAM and *R&I* are not statistically different at conventional levels. Observed patterns in terms of point estimates, though not necessarily in terms of significance of the differences, are indeed in line with our expectations stemming

[10]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.

[11]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 $= 0.361$).

from the theoretical model in Section 3, reflecting that the dynamic incentive effect will have a positive impact on first stage effort under PAM, be absent under RAM and $R\&I$ and negative under NAM. That RAM generates higher first stage effort than NAM is consistent with a presence of the intertemporal spill-over. Adding direct monetary rewards under $R\&I$ was expected to yield higher first stage performance than RAM.

Figure 2: Effort in The Individual Work Stage

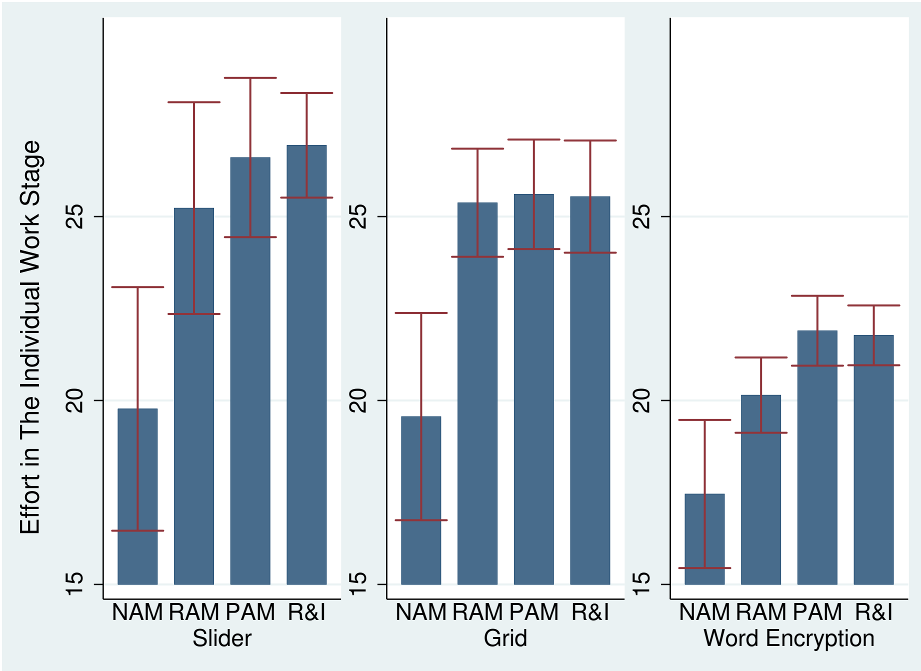


Notes: The top end of the bars indicates the mean effort in the individual work stage, and the line segments represent the 95% confidence intervals.

As mentioned above, the three different tasks may have differed in terms of learning-by-doing or sensitivity to explicit incentives. To assess possible differences in outcomes, Figure 3 shows individual stage performance across treatments for each task and 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.[12]

[12]One possible way to test for differences in learning-by-doing across tasks is to examine the

Figure 3: Effort in The Individual Work Stage by Task



Notes: The top end of the bars indicates 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 nonparametric test (M-W) 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 $R\&I$.[13] Table 2 also contains the results for each task separately, confirming the picture in Figure 3: for the word encryption task performance under RAM was significantly smaller than under each of PAM and $R\&I$, whereas this was

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).

[13]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. Restricting the analysis to subjects teams with similar stage performance ranks under NAM and PAM, to control for a measure of ability, yields the same picture: stage 1 scores under NAM are significantly lower than under PAM, but the difference is less and not significant for stage 2 performance.

not the case for the other two tasks.

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 4 for descriptive statistics.). Columns (4) to (6) add demographic covariates (academic level, gender, and finally controls for nationality and an economics-related degree subject) to account for possible differences in the sample composition, although the selection into treatments was fairly balanced on observables.[14] The coefficients for the different treatments remain relatively stable across the different specifications.

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 for 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 in all specifications.[15]

This analysis does not include 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.

[14] 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, respectively.

[15] 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 Truthful Revelation

As noted above in Section 3, individual work stage effort choice may be strategic and not reflect individuals' true productivities. The theory predicts that individual work stage effort choices are indeed strictly monotone in productivity type under PAM. This 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 a positive measure 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.

Indeed, we find that under RAM, PAM and *R&I* about 1% of subjects score 0 in the first stage, but 8% of subjects under NAM. More generally the prediction of bunching under NAM would imply that first stage performance is more dispersed under NAM than under RAM (recall that the comparison between NAM and either PAM or *R&I* is ambiguous). Figure 4 shows box-plots of the distribution of individual stage efforts across treatments, indicating that moving across treatments from NAM to *R&I* the average effort increases while its dispersion decreases, as does its standard deviation (see the fourth column in Table 1).[16]

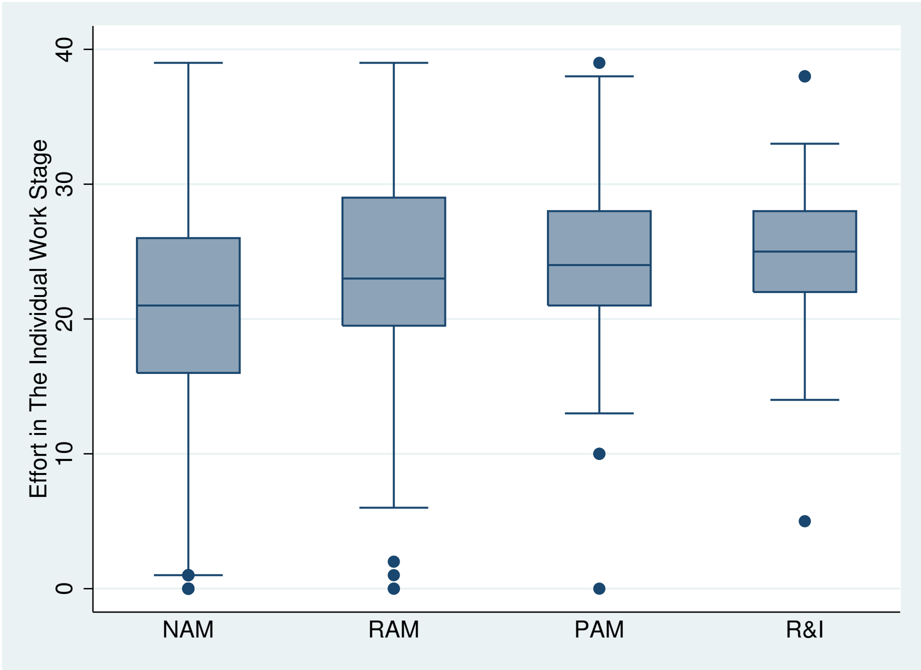
Strategic behavior in the individual work stage would also imply that performance rankings of individuals differ between individual and team work stage under NAM, but not under the other treatments. The data support this prediction: Table 5 shows the rank correlation of individual and team work stage performance, which is significantly (at conventional levels) lower under NAM than under the other treatments.

All this strongly suggests that NAM does not truthfully implement a negative assortative matching of true productivity types, but will involve some randomness. This is relevant since one possible motivation for the use of NAM may be a concern for inequality. 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.

A failure to truthfully implement negative assortative matching in true ability types implies that inequality in *team stage performance* is lower within teams and

[16]Pairwise F-tests of equality of variance across treatments reject equality in all cases.

Figure 4: Dispersion of Effort in The Individual Work Stage



higher across teams than inequality in individual stage performance. This implication seems consistent with the data from the experiment. Figure 5 shows 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 to differ substantially from the two treatments that match randomly.[17]

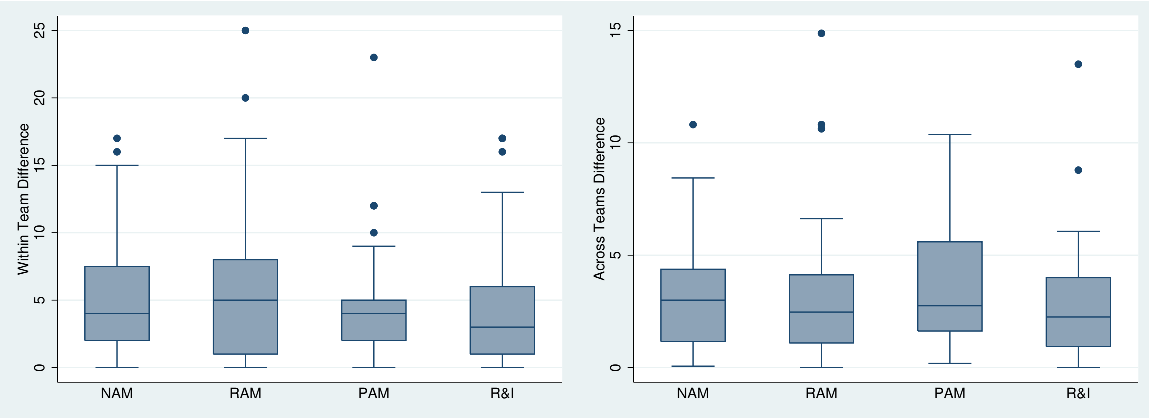
4.4 Subgroup Analysis

There are some plausible reasons why certain subgroups of individuals might be expected to react more to our treatments than others. First, our theoretical model predicts a degree of strategic behavior that is informed by individuals' knowledge about the type distribution. Hence, we would expect that behavior of participants who are well informed about their relative performance among all subjects conforms more closely to the theory. Second, performance in the real effort tasks may be subject to ceiling effects, which would imply that high performers would respond less to extrinsic incentives.

Therefore, we explore possible differences of treatment effects between subgroups

[17]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 5: Team Difference in The Team Work Stage by Treatment



(a) Within Team Difference **(b)** Across Teams Difference

by splitting the sample 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 actual performance in the team work stage.

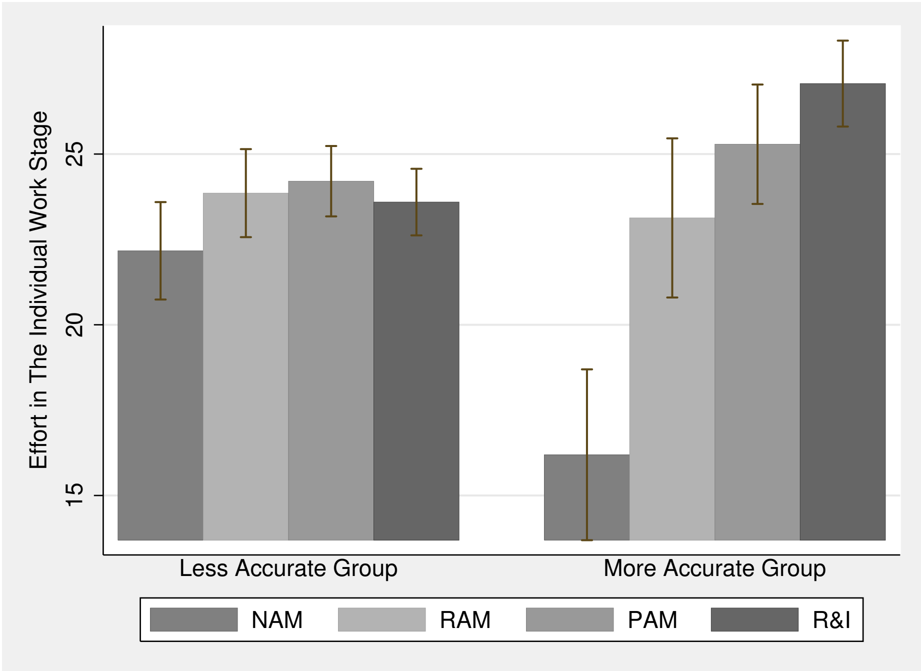
4.4.1 Splitting The Sample Based on Accuracy of Beliefs about Relative Performance

Recall that at the end of the individual work stage of each round we elicited participants' beliefs about their relative 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.[18]

Since strategic behavior requires individual expectations of relative performance to be reasonably accurate (at least under NAM and PAM) one would expect treatment differences to be more pronounced among those participants that predicted their own relative performance well. To explore this hypothesis we split the sample into two similarly sized groups: one group ($n = 110$) that predicted their ranking correctly in

[18]See Table 8 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).

Figure 6: Subgroup Analysis by Accuracy of Beliefs about Relative Performance



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

at most one round, and the other group ($n = 82$) that correctly predicted their ranking in at least two rounds. Figure 6 shows mean performance by treatment separately for each of the two groups. For the group that predicted their relative performance more accurately the treatment effects mirrors the theoretical predictions: performance was very low under NAM, but significantly higher under PAM and $R\&I$ than under RAM. By contrast, 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.[19]

These observations carry over to a regression analysis similar to the one for the whole sample reported above. The results, shown in Table 9 in the appendix, indicate that for the more accurate group treatment effects are much greater in magnitude

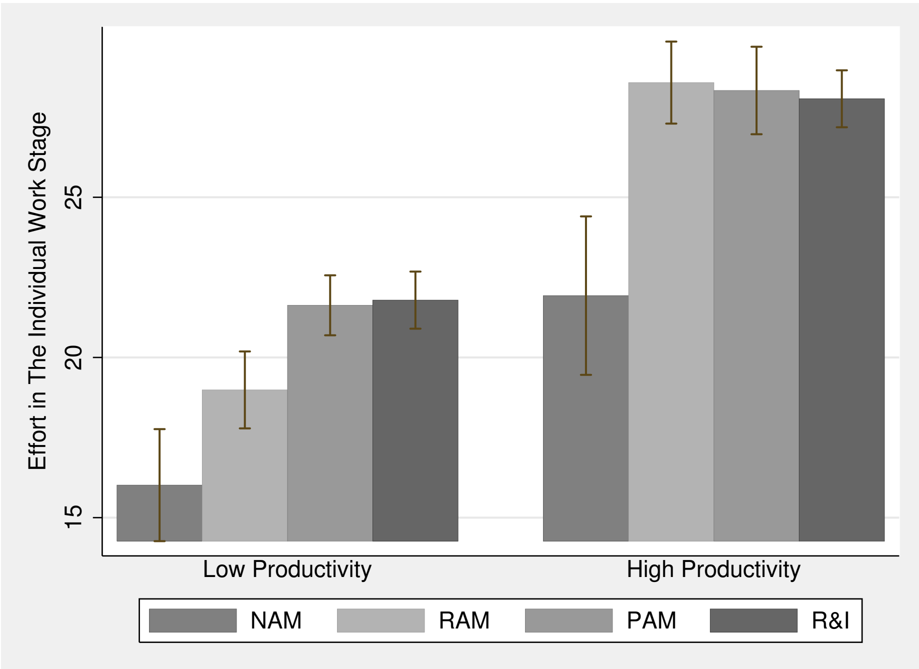
[19] In particular, both t-test and Mann-Whitney U test indicate statistically significant differences in subjects' performances between the two subgroups within treatment NAM (mean difference 5.974, p-value of t-test < 0.001, and p-value of Mann-Whitney U test < 0.001) and $R\&I$ (mean difference -3.470, 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 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).

than for the other group (-7% vs. -28% for NAM, 1.5% vs. 9% for PAM and 0% vs. 17.5% for $R\&I$), and the increase under $R\&I$ is statistically significant throughout.[20]

4.4.2 Splitting The Sample Based on Team Work Stage Performance

Our second subgroup analysis addresses a possible concern of any real effort experiment: some participants may already be exerting effort close to their capacity and thus make it difficult to detect variations across treatments, that is, a ceiling effect may be operative. This would primarily apply to high performers and potentially render the measurement of treatment effects for this group less precise.

Figure 7: Subgroup Analysis by Team Stage Performance



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

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).[21] Recall that the team work stage is the only stage where individual effort is explicitly incentivised

[20]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.

[21]See Table 10 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,

(with a team piece rate). Hence, individual performance in the team work stage is arguably a reasonable proxy for individuals' innate ability, in particular since average team work stage performance did not vary across treatments (see Section 4.5).

Figure 7 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 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 11 in the appendix 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.

4.5 Effort in The Team Work Stage

While the main focus of this paper lies on individual performance before the assignment into teams, it is of interest to examine whether different assignment mechanisms affected participants' performance once they were assigned a partner. 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.[22] Thus, the treatment in form of assignment mechanism has no effect on the individual performance after the assignment.

4.5.1 Peer Effects

Though not at the focus of our analysis, later stage performance could have been affected by the assignments into teams through the presence of peer effects, which have received considerable attention in the literature (e.g. Eisenkopf, 2010; Falk and Ichino, 2006; Mas and Moretti, 2009, among many others). In our experiment, subjects received information on their own and their partner's absolute and relative performance, but no live feedback was given. This setup allows for a possible peer

reassuringly, of participants who prefer competitive settings (regression results available upon request).

[22]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.

effect through the knowledge 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 or inequity aversion may induce higher effort anticipating higher effort of one's peer.

To examine possible peer effects we estimate an OLS regression of individual performance in the team work stage on two dummy variables indicating whether an individual's partner had performed better, respectively, worse than that individual in the individual stage. We constrain our sample to treatments *RAM* and *R&I*, because both treatments used the same assignment mechanism (random matching), thus excluding selection bias. In this subsample, 45.3% of participants were paired with a better partner, 45.6% were paired with a worse partner, and 9.1% had the same individual stage performance as their partner. The results (see columns (1) and (2) of Table 12 in the appendix) suggest peer effects were modest: being assigned a better (worse) partner (compared to one's own individual work stage performance) is associated negatively (positively) with own performance in the team work stage, but not significantly so. The coefficients for the two dummies have the expected (i.e. different) signs and differ significantly from each other. The negative sign of the coefficient for a better partner is consistent with a free-riding effect. This is corroborated by regressing individual team work performance on the continuous individual work stage performance of one's partner (instead of the dummy indicating a better or worse partner), see columns (3) and (4) of Table 12 in the appendix.[23]

5 Conclusion

Does the manner of how individuals are assigned to each other 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 an assignment rule that matches high performers with low performers relative to a scenario where individuals are randomly matched. The evidence is consistent with strategic behavior in the early stage because *NAM* is not incentive compatible so that early stage performance does not reveal true productivity types – i.e. the usual disclaimer applies: past performance is not indicative of future results.

This finding confirms expectations of an equity-efficiency trade-off: an assignment rule that yields teams similar in average prior performance of their members comes at

[23]When standard errors are clustered at the individual level the coefficients for the two dummies become weakly significant and the implications remain the same for the continuous individual work stage performance of one's partner. Results are available upon request.

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 important in practice are adverse implications for policies e.g. in school admission or personnel organisation 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 individuals with better partners or explicitly reward early stage effort with monetary payments tend to outperform random matching in terms of early stage performance, but the estimated effect is relatively small and, for PAM, statistically significant only in some specifications and only at the 10% level. More interesting is perhaps the finding that effort choices under both explicit (monetary) and implicit (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 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, the assignment could be made endogenous, allowing participants to submit preference rankings over peers and then employing tried and tested matching algorithms. 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 members and the payoff additively linear in individual performance. Hence, there are no peer effects by design and one's peer matters 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).

Another policy design question, in particular when equality is a policy goal, is whether and when subjects should be given the opportunity to select into an assignment mechanism: before they receive feedback on their performance (and thus ability) or after. Mirroring the lessons from insurance markets, a veil of ignorance may be needed when selecting into an assignment mechanism to achieve equitable outcomes.

References

Abeler, J., Falk, A., Goette, L., and Huffman, D. (2011). Reference points and effort provision. *American Economic Review*, 101(2):470–92.

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.

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. (2018). Minimizing learning behavior in repeated real-effort tasks. Discussion paper, University of Göttingen cege Discussion Paper No. 343.

Betts, J. R. (2011). The economics of tracking in education. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, chapter 7, pages 341–381. Elsevier, Amsterdam.

Bewley, T. F. (1999). *Why wages don't fall during a recession*. Harvard University Press, Cambridge, MA.

Bidner, C. (2010). Pre-match investment with frictions. *Games and Economic Behavior*, 68(1):23–34.

Blinder, A. S. and Choi, D. H. (1990). A shred of evidence on theories of wage stickiness. *The Quarterly Journal of Economics*, 105(4):1003–1015.

Booth, A. and Coles, M. (2010). Education, matching, and the allocative value of romance. *Journal of the European Economic Association*, 8(4):744–775.

Braun, S., Dwenger, N., Kübler, D., and Westkamp, A. (2014). Implementing quotas in university admissions: An experimental analysis. *Games and Economic Behavior*, 85:232–251.

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010). Constrained school choice: An experimental study. *American Economic Review*, 100:1860–1874.

Card, D., Mas, A., Moretti, E., and Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6):2981–3003.

Charness, G., Kuhn, P., and Villeval, M. C. (2011). Competition and the ratchet effect. *Journal of Labor Economics*, 29(3):513–547.

Charness, G., Masclet, D., and Villeval, M. C. (2014). The dark side of competition for status. *Management Science*, 60(1):38–55.

Chen, Y., Fehr, E., Fischbacher, U., and Morgan, P. (2015). Decentralized matching and social segregation. *Games and Economic Behavior*, 90:17 – 43.

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.

Cooper, D. J., Kagel, J. H., Lo, W., and Gu, Q. L. (1999). Gaming against managers in incentive systems: Experimental results with chinese students and chinese managers. *American Economic Review*, 89(4):781–804.

Eisenkopf, G. (2010). Peer effects, motivation, and learning. *Economics of Education Review*, 29(3):364–374.

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. Discussion paper, Netspar Discussion Paper Series No. 01/2016-003.

Falk, A. and Ichino, A. (2006). Clean evidence on peer effects. *Journal of Labor Economics*, 24(1):39–57.

Felli, L. and Roberts, K. (2016). Does competition solve the hold-up problem? *Economica*, 83(329):172–200.

Fernandez, R., Guner, N., and Knowles, J. (2005). Love and money: A theoretical and empirical analysis of household sorting and inequality. *Quarterly Journal of Economics*, 120(1):273–344.

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

Franco, A. M., Mitchell, M., and Vereshchagina, G. (2011). Incentives and the structure of teams. *Journal of Economic Theory*, 146(6):2307 – 2332.

Gall, T., Legros, P., and Newman, A. (2015). College diversity and investment incentives. Discussion paper, CEPR Discussion Paper Series No. DP10337.

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.

Hanushek, E. A. and Woessmann, L. (2011). The economics of international differences in educational achievement. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 3, chapter 2, pages 89–200. Elsevier, Amsterdam.

Hatfield, J. W., Kojima, F., and Kominers, S. D. (2017). Strategy proofness, investment efficiency, and marginal returns: An equivalence. Working paper, Becker Friedman Institute for Research in Economics, University of Chicago.

Kuhnen, C. M. and Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, 58(1):94–113.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–45.

Pais, J. and Pinter, A. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, 64(1):303–328.

Peters, M. and Siow, A. (2002). Competing premarital investments. *Journal of Political Economy*, 110(3):592–608.

Pfeffer, J. and Langton, N. (1993). The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly*, 38(3):382–407.

So, T., Brown, P., Chaudhuri, A., Ryvkin, D., and Cameron, L. (2017). Piece-rates and tournaments: Implications for learning in a cognitively challenging task. *Journal of Economic Behavior and Organization*, 142:11 – 23.

Vranceanu, R., El Ouardighi, F., and Dubart, D. (2013). Coordination in Teams: A Real Effort-task Experiment with Informal Punishment. Working paper, ESSEC Research Center Working Paper Series No. WP1310.

A Appendix: Proof for Section 3

Proof of Fact 2

RAM and $R\&I$ are shown in the text.

Under PAM e_1' increases in e_1 by assumption. Suppose that strategies are strictly monotone increasing and differentiable in type.[24] Since θ has full support by assumption, so does e_1 and the positive assortative assignment satisfies $e_1' = e_1$. Moreover, since θ' is a function of e_1', anticipating the matching outcome θ' is a function of e_1. The individual optimization problem is given by (1).

Since the optimisation problems are the same for any two individuals of the same type θ, 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

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

Solving the differential equation (through guess and verify) yields $e_1^{PAM}(\theta) = \frac{3\lambda+\sqrt{9\lambda^2+64}}{16}\theta$ (which would become $\frac{\theta}{2}$ If $\lambda = 0$).

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

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

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 θ' (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 θ^* 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 θ).

[24] 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., $\lambda = 0$).

Notes for Proposition 1

For Proposition 1 note that $e_1^{PAM} > e_1^{RAM}$ and $e_1^{PAM} > e_1^{R\&I}$ follow directly from comparing the expressions in Fact 2. Existence and the properties of $\bar{\lambda}$ follow directly from comparing e_1^{PAM} and $e_1^{R\&I}$ as given in the fact. Moreover, the ratios e_1^{PAM}/e_1^{RAM} and $e_1^{R\&I}/e_1^{RAM}$ are both strictly decreasing in λ. Finally, the ratio e_1^{NAM}/e_1^{RAM} is either 0 or 1 depending on the type θ, 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 Tables

Table 1: Summary of Individual Work Stage Effort

Effort in Individual Work Stage	Observations	Mean	SD	Minimum	Maximum
Panel 0. All Treatments					
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
Panel 1. RAM					
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
Panel 2. NAM					
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
Panel 3. PAM					
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
Panel 4. R&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

Table 2: Statistical Differences Across Treatments

Effort in	NAM vs RAM		NAM vs PAM		NAM vs R&I	
Individual	t test	M-W test	t test	M-W test	t test	M-W test
Work Stage	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(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	RAM vs PAM		RAM vs R&I		PAM vs R&I	
Individual	t test	M-W test	t test	M-W test	t test	M-W test
Work Stage	(p-value)	(p-value)	(p-value)	(p-value)	(p-value)	(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.

Table 3: OLS Regression

	Dep. Var.: Effort in the Individual Work Stage					
	(1)	(2)	(3)	(4)	(5)	(6)
NAM	-4.653***	-4.653***	-4.612***	-4.636***	-4.678***	-4.708***
	(0.827)	(0.830)	(0.579)	(0.579)	(0.601)	(0.592)
	[1.313]	[1.317]	[1.260]	[1.249]	[1.278]	[1.262]
PAM	1.118	1.118	1.213*	0.974	1.185*	0.997*
	(0.699)	(0.701)	(0.577)	(0.583)	(0.589)	(0.508)
	[0.885]	[0.889]	[0.821]	[0.793]	[0.831]	[0.795]
R&I	1.151***	1.168***	1.265**	1.127**	1.291**	1.003*
	(0.353)	(0.364)	(0.481)	(0.467)	(0.468)	(0.519)
	[0.849]	[0.851]	[0.843]	[0.822]	[0.841]	[0.840]
Constant	23.58***	26.04***	27.54***	27.75***	27.28***	27.57***
	(0.242)	(0.698)	(1.027)	(1.053)	(0.981)	(0.749)
	[0.683]	[0.910]	[1.177]	[1.167]	[1.191]	[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.

Table 4: Descriptive Statistics of Other Variables

	Participants (1)	Mean (2)	SD (3)	Minimum (4)	Maximum (5)	Fractions (%) (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 5: Spearman's Rank Correlation Coefficients of Stage 1 and Stage 2 performances across Treatments

	NAM	RAM	PAM	R&I
All tasks	0.368	0.865	0.774	0.797
Slider	0.303	0.731	0.634	0.548
Grid	0.284	0.860	0.799	0.832
Word Encryption	0.292	0.883	0.770	0.708

Table 6: Summary of Team Work Stage Effort

Effort in the Team Work Stage	Observations	Mean	SD	Minimum	Maximum
Panel 0. All Treatments					
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
Panel 1. RAM					
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
Panel 2. NAM					
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
Panel 3. PAM					
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
Panel 4. R&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

Table 7: Tests of Sample Balance on Demographies

	RAM (%)	NAM (%)	PAM (%)	R&I (%)	Chi-square test (p-value)	t-test (p-value)	M-W test (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Gender:					0.536		
Male	26.14	20.45	25.00	28.41			
	{47.92}	{37.50}	{45.83}	{52.08}			
Female	24.04	28.85	25.00	22.12			
	{52.08}	{62.50}	{54.17}	{47.92}			
Degree:					0.968		
Econ-related	23.68	25.44	25.44	25.44			
	{43.75}	{39.58}	{39.58}	{39.58}			
Not Econ-related	26.92	24.36	24.36	24.36			
	{56.25}	{60.42}	{60.42}	{60.42}			
Nationality:					0.239		
UK	23.26	20.93	23.26	32.56			
	{41.67}	{37.50}	{41.67}	{58.33}			
EEA	16.00	32.00	28.00	24.00			
	{8.33}	{16.67}	{14.58}	{12.50}			
Others	30.77	28.21	24.36	16.67			
	{50.00}	{45.83}	{39.58}	{27.08}			
Prefer Not to Say	0.00	0.00	66.67	33.33			
Native Speaking Language is English:					0.114		
Yes	26.32	21.05	21.05	31.58			
	{52.08}	{41.67}	{41.67}	{62.50}			
No	24.47	29.79	27.66	18.09			
	{47.92}	{58.33}	{54.17}	{35.42}			
Prefer Not to Say	0.00	0.00	66.67	33.33			
Academic Level:					0.031		
Undergraduate	22.88	22.88	27.45	26.80			
	{72.92}	{72.92}	{87.50}	{85.42}			
Postgraduate	35.14	35.14	10.81	18.92			
	{27.08}	{27.08}	{8.33}	{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	28.45	22.41	23.28			
	{62.50}	{68.75}	{54.17}	{56.25 }			
1 Year	22.22	33.33	33.33	11.11			
	{8.33}	{12.50}	{12.50}	{4.17}			
2 Years	16.67	8.33	41.67	33.33			
	{8.33}	{4.17}	{20.83}	{16.67}			
3 Years	24.00	24.00	16.00	36.00			
	{12.50}	{12.50}	{8.33}	{18.75}			
More Than 3 Years	50.00	12.50	12.50	25.00			
	{8.33}	{2.08}	{2.08}	{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 8: 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 9: Subgroup Analysis by Accuracy of Beliefs about Relative Performance: OLS Regression

	Less Accurate Group[a]		More Accurate Group[b]	
	(1)	(2)	(3)	(4)
NAM	-1.689***	-1.552**	-6.937***	-6.244***
	(0.445)	(0.559)	(1.997)	(1.417)
PAM	0.350	0.344	2.158	2.018
	(0.297)	(0.337)	(1.816)	(1.754)
R&I	-0.262	0.0478	3.934***	4.259***
	(0.922)	(0.939)	(1.223)	(1.307)
Constant	23.86***	26.55***	23.13***	26.78***
	(0.284)	(0.998)	(1.051)	(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.

[a]Less Accurate Group: subjects who did not correctly predict their relative standings more than once.

[b]More Accurate Group: subjects who correctly predicted their relative standings at least twice.

Table 10: 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 11: Subgroup Analysis by Team Stage Performance: OLS Regression

	Low Productivity Group		High Productivity Group	
	(1)	(2)	(3)	(4)
NAM	-2.973***	-3.132***	-6.650***	-5.934***
	(0.490)	(0.524)	(1.382)	(1.445)
PAM	2.642***	3.020***	-0.246	-0.370
	(0.678)	(0.534)	(1.072)	(1.011)
R&I	2.803***	2.347***	-0.505*	0.386
	(0.690)	(0.734)	(0.281)	(0.531)
Constant	18.99***	19.41***	28.58***	31.53***
	(0.409)	(1.207)	(0.224)	(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.

Table 12: Peer Effects: OLS Regression

	Dep. Var.: Effort in The Team Work Stage			
	(1)	(2)	(3)	(4)
Matched With a More Productive Partner	-2.250	-1.943		
	(1.327)	(1.501)		
Matched With a Less Productive Partner	1.811	1.946		
	(1.410)	(1.562)		
Own Effort (Individual Work Stage)			0.693***	0.690***
			(0.101)	(0.0995)
Partner's Effort (Individual Work Stage)			-0.0333	-0.0297
			(0.0190)	(0.0180)
R&I	0.333	0.278	-0.411	-0.420
	(0.573)	(0.433)	(0.348)	(0.264)
Constant	28.88***	28.51***	11.21**	10.12**
	(2.075)	(2.462)	(3.400)	(2.933)
Observations	287	287	287	287
R-squared	0.360	0.407	0.677	0.698
Task and Round Fixed Effects	YES	YES	YES	YES
Other controls	NO	YES	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 slightly improve the statistical significance of the coefficients for the two dummies (matched with a more productive partner and matched with a less productive partner) in columns (1) and (2), but not change our implications in columns (3) and (4). Columns (1) and (3) report estimates with task and round fixed effects. Columns (2) and (4) further add 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:

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

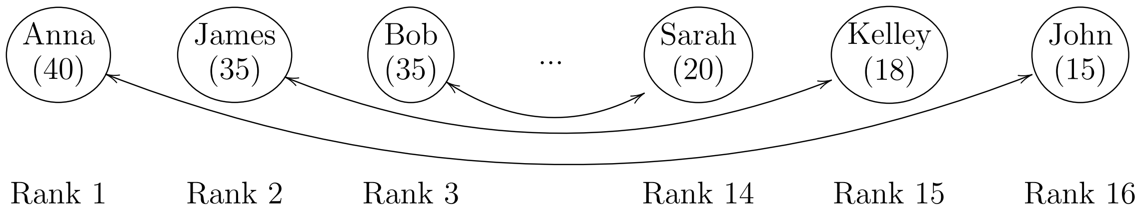
Pairing Rule [RAM and R&I]

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]

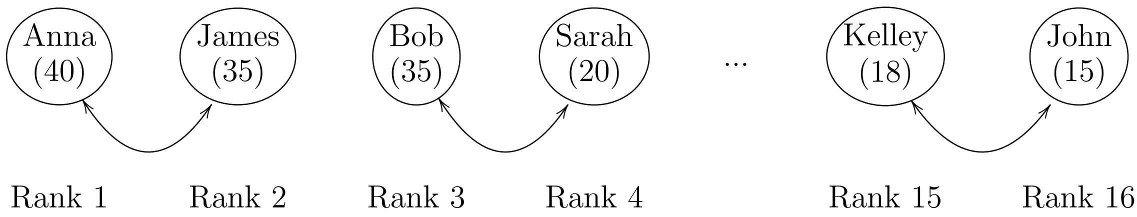
The computer will rank all participants according to their scores in the individual work stage. Each team consists of 2 partners. Teams are formed by pairing participants based on their scores in the individual work stage: the best performing participant will be working in a team with the second best 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.

The Tasks [All Treatments]

Slider

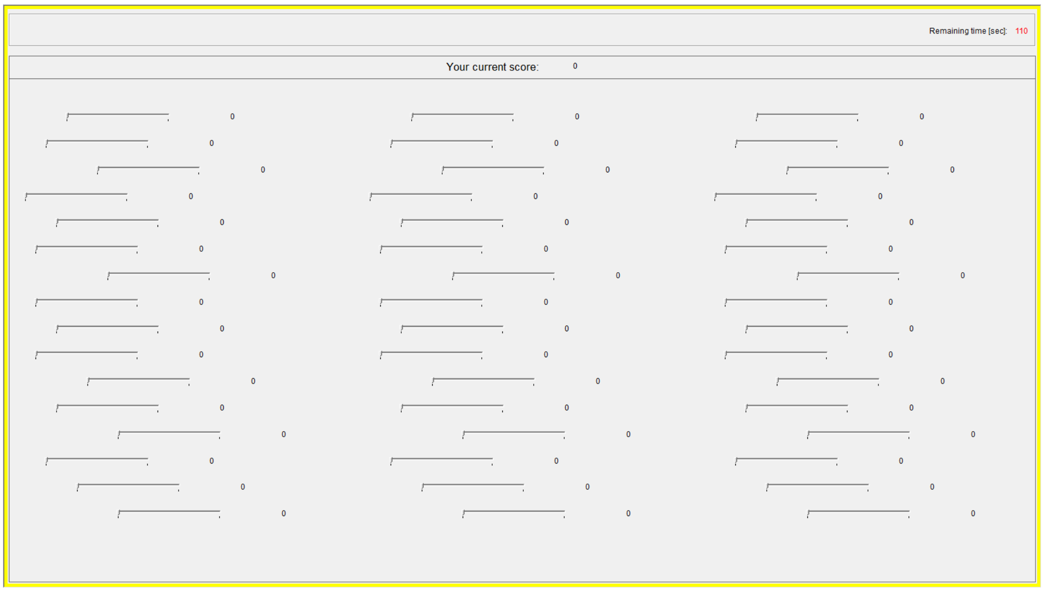
The task will consist of a screen with 48 sliders. Each slider is initially positioned at 0 and can be moved as far as 100. Each slider has a number to its right showing its current position. You can use the mouse in any way you like to move each slider.

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



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.

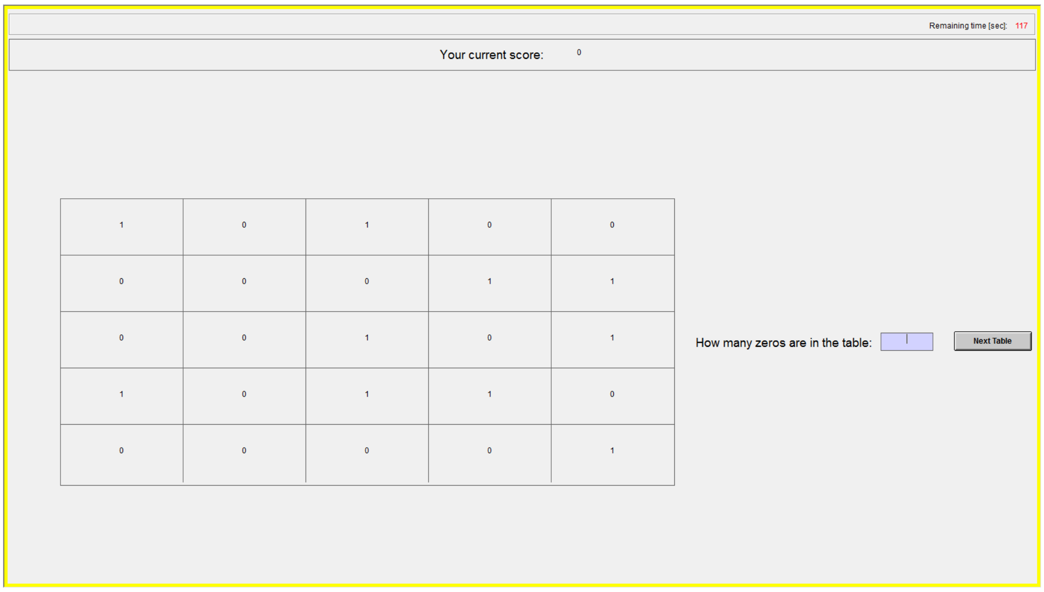
Figure D3: The Slider Task



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



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 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

Figure D5: The Word Encryption Task

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?