Because of you I did not give up – How peers affect perseverance*

Leonie Gerhards† [1] and Christina Gravert‡ [2]

[1]Universität Hamburg
[2]University of Gothenburg

September 2017

Abstract

Various empirical papers have shown that peers affect productivity and behavior at work. In the present experiment we focus on one specific mechanism. We consider a situation in which individuals look at their peers' behavior to motivate themselves to endure in a task that requires perseverance. We find that peers increase their observers' perseverance, while being observed does not significantly alter behavior. In a second experiment we investigate (i) the motives to self-select into the role of an observing subject or that of a subject that is observed and (ii) which kind of peers individuals deliberately choose. Our findings provide first insights into the perception of peer situations and offer new empirical evidence on how peer groups emerge.

Keywords: grit, perseverance, laboratory experiment, peer effects, real effort
JEL codes: C91, D03, M50, J24

*We are grateful for helpful comments from seminar and conference audiences at the Universities of Aarhus, Amsterdam, Copenhagen, Gothenburg, Hamburg, Lüneburg, Munich, Nottingham, Trier, Wellington, the Copenhagen Network of Experimental Economics, the Meeting of German Economists Abroad 2014, the Nordic Conference in Behavioral and Experimental Economics 2015, the Colloquium on Personnel Economics 2016, the annual meeting of the Verein für Socialpolitik 2016: In particular, we would like to thank Florian Englmaier, Gerd Mühlheußer and Uwe Sunde, who provided us with detailed feedback. This research was supported by funding from the Danish Council for Independent Research — Social Sciences (Det Frie Forskningsråd — Samfund og Erhverv), grant number 12-124835 and from the Swedish Research Council (Vetenskapsrådet) (2009-33472-67049-56).

†Corresponding author. Department of Economics, Universität Hamburg, Von-Melle-Park 5, 20146-Hamburg, Germany. Tel.: +49 (0) 40 42838 5573. Email: leonie.gerhards@wiso.uni-hamburg.de

‡Department of Economics, University of Gothenburg, Vasagatan 1, S-41124 Gothenburg, Sweden; christina.gravert@economics.gu.se, Tel.: +46 (0) 31 786 5252

1 Introduction

Peers affect productivity and behavior in the workplace. Both theoretical papers (for instance Kandel and Lazear, 1992 and Kräkel, 2016) as well as several empirical studies (such as the papers by Falk and Ichino, 2006; Mas and Moretti, 2009; Bandiera et al., 2010 and Bäker and Mechtel, 2013) show that individuals are influenced by others working besides them. However, the exact channels through which individuals influence each other are still largely unexplored (Charness and Kuhn, 2011). Peer effects could, for example, stem from imitation of behavior (Bandura and Walters, 1963; Mobius and Rosenblat, 2014), an increase in knowledge about the task (Banerjee, 1992; Guryan et al., 2009), competitive preferences (O'Keeffe et al., 1984), guilt and shame (Kandel and Lazear, 1992), a desire for conformity (Bernheim, 1994; Bellemare et al., 2010), self-perception in comparison to others (Deci and Ryan, 1985) or the pure joy of working with others (Bandiera et al., 2010).

In this paper we apply a novel experimental design to focus on one specific dimension of productivity that may be influenced by peers: We analyze if and how peers affect each others' perseverance on a task. We base our research design on the following idea: While workers might not know their co-workers' initial abilities and cannot assess the quality of the output before the termination of a project, they often have a good estimate of how much effort their colleagues exert and how long they remain working on the task at hand. Consider for instance graduate students working in the same office: They can observe on a daily basis when their peers come and go and whether they are working on their research projects or browsing the internet. But different fields of research and methods might make it difficult to compare the peers' progress and output to their own. Nevertheless, observing the other students working longer and more diligently likely affects a student's perseverance on his or her own project.

There exists a wide range of other situations, for instance in sports, social and educational contexts, in which perseverance plays an important role and diligent comrades may additionally affect their peers (see for instance Duckworth et al. (2007) and Tangney et al. (2004) for the former, and Battaglini et al. (2016) for the latter aspect). From a research perspective, however, field settings might be confounded by several factors that cannot be controlled for. Examples are self-selection out of task or treatment, (self-)selection into teams based on homophilistic preferences (which can cause spurious correlations (Manski, 1993)), "reflection problems" of peers simultaneously influencing each other or complementarities between individuals in case their work tasks or payoffs are interrelated. We therefore decided to conduct a controlled laboratory experiment that allows us to exclude these and other confounding factors.

Our experimental design is mainly based on the "grit task" introduced in Gerhards and Gravert (2017). Subjects are randomly assigned to treatments and work on a word play task in which they solve anagrams on an individual piece rate basis. We measure subjects' perseverance by considering the behavioral opposite of it: As a way to avoid working hard on the task, subjects can choose to skip individual anagrams, which comes at an explicit cost. As Gerhards and Gravert (2017) demonstrate, this avoidance behavior is significantly negatively correlated with subjects' scores on Duckworth

and Quinn's (2009) Short Grit scale, which, in turn, is indicative of behavior and success in various situations that require perseverance. We compare subjects' skipping behavior when working by themselves ("Baseline Treatment") to a peer situation ("Random Matching Treatment") where half of the subjects are randomly assigned the role of what we call an "Observer" who is informed about his or her randomly matched "Peer's" avoidance behavior. The Peers know that they are observed by another subject, but do not receive any further information about their Observer.

We chose this design for several reasons: First, providing only information about the Peer's avoidance behavior, rather than about his or her performance, is realistic in the sense that employees might not have a precise estimate of their coworker's productivity, but can well observe their coworker's (non-)perseverance. Second, by assigning distinct roles to subjects in the Random Matching Treatment, we can rule out the above mentioned reflection problems and separately study the effects of observing and being observed.

Our main findings are the following: (i) Observers (compared to subjects from the Baseline Treatment) skip individual anagrams significantly less often if they are presented with information about their Peers' respective behavior; (ii) Peers do not skip significantly more or less than subjects from the Baseline Treatment; and (iii) while the personal trait grit explains subjects' skipping behavior in the Baseline Treatment, it has neither a significant effect on skipping for Observers nor for Peers.

Taken as a whole, we conclude that staying perseverant on a task does not only depend on individuals' own grittiness, but hinges on the peer group composition at hand. In the conclusion we provide a range of domains in which we expect these peer effects in perseverance to affect peoples' performance and discuss resulting policy implications

Related literature. A handful of other papers have looked at peer effects and performance in laboratory settings. The first three papers that we briefly describe below consider peer effects in subjects' performance and output in general. The fourth and fifth paper focus on peer effects in perseverance in particular.[1]

Georganas et al. (2015) ask their subjects to work on Gill and Prowse's (2012) slider task and randomly assign them into three groups. Subjects in the first group are observed by another subject, those from the second group observe another subject and subjects in the third group work in isolation. The authors find no overall peer effects on performance in the first two groups when subjects work under individual piece rate incentives. They conjecture that this could be explained by a significant and dominant learning effect that conceals potential treatment effects.[2] Beugnot

[1]Furthermore, a number of field experiments study the effect of related personal attributes such as self-control on real-life outcomes (see for instance Kaur et al., 2010 who link self-control to success at work and Buechel et al., 2014 and Battaglini et al., 2016 who link self-control (and its interaction with peer effects) to academic success). Moreover, a number of laboratory experiments investigate the determinants of shirking in real-effort tasks (see for instance Corgnet et al., 2015 on diligence in summing up numbers or Bucciol et al., 2013 on the ability to withstand temptation in a counting task) and the effects of being observed by group members (see for example Charness et al., 2007 and Andreoni and Bernheim, 2009).

[2]In a team incentive treatment they find that observed participants increase their productivity a slightly more than the control group. However, this group catches up to them within 3 minutes of the experiment. Possibly, being

et al. (2013) test whether the strength of peer effects depends on the actual presence of a peer. In one of their treatments subjects are informed about a peer's outcome who has worked on the same real effort task (a math task) in an earlier session. In a second treatment subjects are continuously informed about the performance of a peer who is simultaneously working in the same room. In both treatments the authors find significant peer effects for men, but not for women. They conclude that this could be explained by rivalry between peers, which they find to be stronger for men than for women. Van Veldhuizen et al. (2015) set out to replicate the main features and findings of Mas and Moretti's (2009) seminal field study in the lab. Groups of four students work in a team to solve a predetermined number of math tasks. By design, subjects' payoffs do not depend on the other participants, but their workload does. Depending on their role in the team, subjects can either observe others, are observed by others, do both things simultaneously or cannot do either. In contrast to Mas and Moretti (2009), the authors find no difference between being observed and observing a coworker. They do, however, discover heterogeneous peer effects in productivity. When subjects are aware of the general productivity level of their peer, some subjects reciprocate a fast co-worker, while others free-ride on the faster co-worker's effort.

Similar to our study, Bonein and Denant-Boèmont (2015) study individuals' perseverance – more precisely their self-control – in the presence of a peer. In their experiment subjects work on the slider task and can commit themselves to an output level ex ante. The authors find that subjects choose higher penalties for not reaching their goal if they know that their output will be reported to another participant with whom they interact in a later stage of the experiment. However, they do not find any effects on increased perseverance on the actual task. Finally, in a recent laboratory experiment Rosaz et al. (2016) focus on peer effects in perseverance, measured at the extensive margin. In their experiment, subjects are free to stop working on their task whenever they want. The authors show that the difference in quitting times of actual peers is smaller than the difference in quitting time of two hypothetically matched individual workers. Working times, however, only significantly increase if peer subjects are allowed to communicate during the experiment. From this, the authors conclude that, in their study, peer effects are mainly driven by a "sociability effect" rather than by peer pressure or performance comparisons.

Besides the apparent design differences between the latter two experiments and our approach, it should moreover be noted that Bonein and Denant-Boèmont (2015) and Rosaz et al. (2016) measure perseverance as a zero/one decision. Once a subject decides to give in to temptation he or she cannot return to the experiment. In that sense, their perseverance measures are hence more "ultimate" than our measures of giving up in the form of skipping, which can happen multiple times during an experimental session. Moreover, compared to the five studies mentioned above, our design allows to analyze peer effects in perseverance in isolation, keeping biasing learning effects, competitive motives and social or monetary interdependencies between subjects at a minimum. First, we restrict our analysis of treatment effects to the final part of the experimental sessions when further learning can largely be ruled out. Second, we inform Observers only about their

observed leads to slightly faster learning in this experiment.

Peers' avoidance behavior and not about their performance or output. And third, we consider performance in individual tasks only. Communication is ruled out by design.

The remainder of the paper proceeds as follows: Section 2 describes our experimental design. In Section 3 we then discuss the results from our Baseline and Random Matching Treatment. In the last part of the paper, we widen our focus to the conscious decision of peer selection. First, we analyze which personal characteristics predict a preference for being observed (i.e. self-selecting into the Peer's role) rather than observing someone else (i.e. self-selecting into the Observer's role). Second, we investigate which type of Peer Observers choose to observe given information about that person's previous performance. We provide a detailed description of the experimental design and a discussion of our findings from this additional "Chosen Matching Treatment" in Section 4. The paper concludes in Section 5.

2 Experimental design

The task. During the experimental sessions subjects work on an anagram word play task in which they have to rearrange letters of English words to form new ones. As an example consider the word "teacher" that can be rearranged to "hectare" and "cheater". We accept all possible anagrams that can be built from a word as a correct solution. The subjects' main task is to solve "hard anagrams" that consist of 5 to 7 letters. Additionally, we provide them with an outside option, that entails working on "easy" anagrams, which comprise only 3 to 4 letters.

The experiment starts with a five minute practice round in which subjects are asked to solve easy anagrams. Performance in this part is not monetarily incentivized. It allows subjects to familiarize themselves with the experimental task. In the main part of the experiment subjects are paid based on their performance. This part is divided into two parts that consist of 10 rounds of 3 minutes each. In each first round subjects have to work on hard anagrams. We chose this set-up in order to stress the default character of working on hard anagrams and to make sure that the subjects get to know the level of difficulty of hard anagrams. At the beginning of each of the following nine rounds subjects can choose to "stay with the hard anagrams" or to "switch to the easy anagrams" for the coming three minutes.[3] We hence allow subjects to go back to solving hard anagrams after having switched to the outside option in the previous round (and vice versa).

Irrespective of treatment, within part 1 and part 2, all subjects are presented with the same anagrams in the same order. Each anagram is presented for up to 90 seconds. If a subject enters a correct solution, a new anagram is displayed immediately. If a subject does not manage to solve an anagram within the given time frame, the computer generates a new anagram free of cost. When working on hard anagrams, subjects, moreover, have the opportunity to "skip" individual anagrams. Then they do not have to wait until the end of the 90 seconds, but a new anagram is generated immediately.

[3]See `https://www.dropbox.com/s/qq838eaioo4yc3w/InstructionsAndScreenshots.pdf?dl=0` for a screen shot of this decision stage and examples of other screens that were displayed during the experimental sessions.

Incentives. For each correctly solved hard anagram subjects earn DKK 5.00, for each solved easy anagram they earn DKK 0.50.[4] Each skip comes at a cost of DKK 3.00. Also if subjects decide to switch to easy anagrams, they have to bear a cost of DKK 3.00. The total cost of switching to easy anagrams for one round hence consists of the explicit switching cost and the implicit cost of reduced earning opportunities. We chose this incentive structure to make switching to the outside option clearly monetarily unattractive. Subjects should be able to see without any formal calculation that even for less able individuals switching to easy anagrams is not monetarily optimal. Their choice can hence only be rationalized by a strong desire to avoid working hard.[5] We consider working on easy anagrams as similar to switching to a leisure outside option, which is often implemented in real-effort experiments.

Variables of interest. The number of skipped hard anagrams serves as our behavioral measure of subjects' perseverance and is our main variable of interest. As discussed in detail in Gerhards and Gravert (2017), subjects who skip fewer anagrams in the main task score significantly higher on Duckworth et al.'s (2007) Short Grit scale, classifying them as generally more perseverant. In both part 1 and part 2, an information box on the computer screen continuously informs subjects about the number of hard anagrams they have already skipped since the beginning of the current part and how many hard anagrams they have skipped in the current round. In rounds in which subjects work on the outside option of easy anagrams, no such information is displayed. This further emphasizes the outside option character of this task.

In order to control for subject's "tactics" in our later regression analysis, we consider subjects' average earnings from rounds in which they work on hard anagrams in part 1. These earnings reflect their ability to find their payoff maximizing mixture of solving and skipping hard anagrams: While one subject might yield a certain payoff by skipping and solving a large number of anagrams, another subject could generate the same payoff by solving, but also skipping a lower number of anagrams.

Moreover, we elicit a proxy for subject's task-related overconfidence. At the end of part 1 subjects receive feedback about their productivity on hard anagrams and payoffs in that part. Subsequently we ask them without previous announcement to make an incentivized guess which performance quintile they expect to belong to. In particular, the test on the computer screen asks them to make this guess based on their number of correctly solved hard anagrams in part 1. We reward correct answers with DKK 20.00. After subjects enter their guess they are informed about their actual performance quintile and part 2 starts.

Note that we use part 1 of the experiment mainly to elicit several control variables, such as subjects' general task ability, tactics and task-related overconfidence, defined as above. Part 2 varies across treatments and is hence the working period of interest in our data analysis.

[4] At the time of the experiment, the exchange rate of DKK 1 was US-Dollar 0.18 or Euro 0.13.

[5] Indeed, for only 11 out of our 152 subjects from the Baseline and Random Matching Treatment we find that their highest round earnings on easy anagrams in part 1 lay above their highest round earnings on hard anagrams. Only 4 out of these 11 subjects realize positive round earnings when working on hard anagrams. The remaining 7 earn maximum earnings of 0 and hence display irrational behavior by switching to easy.

The treatments. In the Baseline Treatment, the instructions presented in the beginning of part 2 simply ask subjects to continue working on the anagram task, as they did before in part 1.

Subjects in the Random Matching Treatment are presented a new set of instructions on their computer screens. These inform them that half of the subjects in their session will be randomly assigned the role of an Observer and the other half will assume the role of a Peer. Always one Observer and one Peer are randomly matched and stay together for the entire duration of part 2. Both are presented with the same anagrams in the same order. In order to prevent framing effects Observers are called "Person A" and Peers "Person B" during the experiment.

Observers learn their Peer's performance quintile from part 1. Moreover and more importantly, the instructions on both Observers' and Peers' computer screens make clear that Observers are constantly informed about their Peer's skipping and switching behavior during part 2: When Observers work on hard anagrams their computer screens display the total number of hard anagrams the matched Peer has skipped since the beginning of that part and whether the Peer is working on hard or easy anagrams in the current round. This information is updated at the beginning of each round, that is, every three minutes. To prevent the emergence of so called "rat races" we do not provide them with further information on the number of correctly solved hard or easy anagrams. Peers receive no information about their Observer's performance, but are informed about the type of information that the Observer receives about them.

Questionnaire. After having finished working on the real-effort task we ask the subjects to fill out a short questionnaire that comprises two parts. In the task-specific part we ask them how they perceived working on the task. In the survey part we elicit a number of non-cognitive skills and personality traits through non-incenitivized survey questions. In particular, we administer the Short Grit scale (Duckworth and Quinn, 2009) which comprises eight items that are ranked on a 5 point scale and measure the tendency to sustain effort and maintaining interests over a long period of time. The full list of items is presented in Appendix B. In the Baseline Treatment we randomize the order of experimental task and the survey-part of the questionnaire at the session level. This does neither significantly affect behavior on the task nor answers provided in the questionnaire (see Gerhards and Gravert (2017) for details).

Procedural details. We ran 16 experimental sessions in spring to fall 2014 at the Cognition and Behavior (Cobe) Lab at Aarhus University. Subjects were recruited via the laboratory's online recruiting website from a subject pool of mostly undergraduate students from all faculties. In total 152 subjects participated, out of which 69 (45 percent) were female. We observe 62 subjects in 7 Baseline Treatment sessions. In one of these only 6 subjects participated, in the remaining ones we observed at least 9 participants. Note that the experimental data from this treatment is also studied in Gerhards and Gravert (2017), with the aim of developing a behavioral measure of grit that is correlated with Duckworth et al.'s (2007) questionnaire measure. To, furthermore, study peer effects we compare behavior of subjects from the Baseline Treatment to that of additional 90 subjects (45 Observers, 45 Peers) who participated in the Random Matching Treatment. The

latter treatment is exclusively analyzed in the present paper. All sessions in this treatment involved exactly 10 subjects. Subjects in our sample had on average 9 years of English at school. Most of them are Danish undergraduates.

At the beginning of each session we distributed consent forms and detailed instructions about the experimental task to the subjects. Subjects had 10 minutes to read the instructions. Afterwards the experimenter asked if there were any questions or whether anyone needed more time to read. When all subjects denied, the experiment (programmed in z-Tree (Fischbacher, 2007)) started and all further instructions were provided on the computer screens. At the end of each session either part 1 or part 2 was randomly selected for payment. This was done to counter potential income effects from the first part.[6] Average earnings for the 90 minutes sessions amounted to DKK 139.64 (minimum: DKK 44.50, maximum: DKK 697), including the reward for a correct guess about a subject's own performance rank and a DKK 40.00 show-up fee. The payments were directly transferred to the subjects' bank accounts.

3 Results

We start the discussion of our results with a short summary of the descriptive statistics of the main variables of interest. The first part of Table 1 provides an overview of the most important performance measures from part 1 of the experiment. As becomes evident, the distributions of correctly solved easy anagrams in the practice round as well as the subjects' propensity to switch to easy anagrams in part 1 do not differ significantly across treatments. However, we do observe significant differences in the number of solved hard anagrams and skipping: Subjects in the Baseline Treatment solve slightly fewer hard anagrams correctly than Peers from the Random Matching Treatment (Mann-Whitney ranksum test result: $p = 0.088$). Moreover, they skip hard anagrams significantly more often than Observers from the Random Matching Treatment ($p = 0.037$). This gap cannot be caused by any treatment differences since up to this point the instructions are identical in both treatments. It can hence only be explained by individual differences of the participants. We will therefore control for these differences by including an *ability* as well as a *tactics* control variable in our later regression analysis. We will explain the details below.

Moving to the summary statistics of behavior in part 2, when considering the raw performance variables, we do not find any significant overall treatment differences in the non-parametric tests. Nevertheless, we consider the figures from Table 1 as encouraging in the sense that they do not show a "selection out of treatment" either: Subjects from the Random Matching Treatment do neither switch to the outside option more often than those from the Baseline Treatment nor do they switch more often in part 2 than in part 1 of the experiment. On the contrary, Observers and Peers actually slightly reduce their switching in part 2 (Wilcoxon sign rank test results: $p = 0.09$ and $p =$, respectively).

[6] This fear, however, proved unfounded. Subjects' earnings even slightly increased over the course of the experiment (mean earnings in part 1: DKK 89.97 (s.d. 79.74), mean earnings in part 2: DKK 92.33 (s.d. 85.97)).

And finally, as presented in the bottom part of Table 1 the median grit level in our subject pool amounts to 3 (mean = 3.06) and is hence comparable to the ones found in previous studies (Duckworth and Quinn (2009): mean: 3.4, s.d.: 0.7; Alaoui and Fons-Rosen (2016): mean: 3.38, s.d.: 0.54). Moreover, the anagram task is perceived as rather challenging and strenuous. Additional tests further show that it does not favor any of the sexes in the sense that men's and women's performance does not differ significantly.[7]

Table 1: Summary statistics (medians, min and max values and Mann-Whitney ranksum test results) for selected variables

		Random Matching	
	Baseline (62 obs.)	Observers (45 obs.)	Peers (45 obs.)
Performance measures			
Average in practice round:			
Solved easy anagrams	7.50 [0.00, 32.00]	9.00 [0.00, 45.00]	9.00 [0.00, 55.00]
Averages per round in part 1:			
Solved hard anagrams	2.00 [0.00, 7.40]	1.70 [0.00, 9.00]	2.00 [0.00, 12.33]*
Skipped anagrams	0.28 [0.00, 2.71]	0.10 [0.00, 2.00]**	0.10 [0.00, 3.78]
Propensity of switching to easy	0.11 [0.00, 0.89]	0.11 [0.00, 0.89]	0.11 [0.00, 1.00]
Guessed performance quintile	3.00 [1.00, 5.00]	3.00 [1.00, 5.00]	3.00 [1.00, 5.00]
Guessed performance quintile correct	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]
Averages per round in part 2:			
Solved hard anagrams	1.60 [0.00, 5.60]	1.60 [0.00, 14.40]	2.00 [0.40, 12.11]
Skipped anagrams	0.05 [0.00, 3.50]	0.00 [0.00, 2.10]	0.10 [0.00, 4.78]
Propensity of switching to easy	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]
Questionnaire measures			
Grit	3.00 [1.75, 4.50]	3.00 [1.88, 5.00]	3.00 [1.63, 4.50]
Perceived task as a challenge	4.00 [1.00, 5.00]	4.00 [1.00, 5.00]	4.00 [2.00, 5.00]
Exerted effort on task	5.00 [2.00, 5.00]	5.00 [1.00, 5.00]	5.00 [2.00, 5.00]
Enjoyed working on task	3.00 [1.00, 5.00]	3.00 [1.00, 5.00]	3.00 [1.00, 5.00]
Perceived task as exhausting	4.00 [1.00, 5.00]	4.00 [2.00, 5.00]	4.00 [1.00, 5.00]

Note that the reported averages of solved and skipped hard anagrams per round are calculated conditional on the number of rounds in which subjects actually worked on hard anagrams. Stars indicate the results of Mann-Whitney ranksum tests comparing observations from the Baseline Treatment to the respective group from the Random Matching treatment:* p<0.10; ** p<0.05. The fractions of subjects guessing their performance quintile correctly were compared in a Fisher exact test (both tested differences are not significant at any conventional level).

The focus of the following empirical analysis lies on studying peer effects in perseverance, which we (reversely) measure as the number of skips on hard anagrams in part 2. Figure 1 gives a first overview of how the average number of skips per round evolve over time. In general, the graphs paint a rather positive picture of peer effects in perseverance: In each of the 10 rounds, Observers skip on average fewer anagrams than the other groups of subjects.

In the following we will investigate these peer effects in more detail. Since, as noted above, subjects display a relatively large variance in their performance, we apply a regression analysis that allows us to control for individual abilities and tactics. We define tactics as the average round

[7]Considering the performance of all 152 participants in part 1 (which does not involve any exogenous treatment variations), we do not find a significant gender difference in the number of hard anagrams solved per round worked on the main task ($p = 0.84$)).

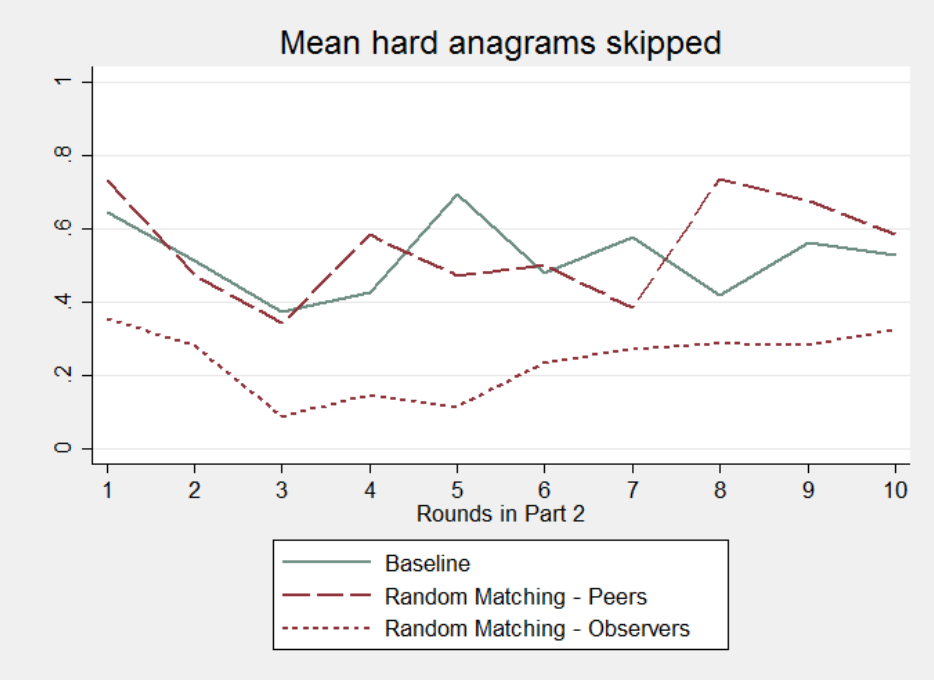

Figure 1: Skipping behavior across treatments

earnings that subjects realize when working on hard anagrams in part 1. These earnings depend on the number of correctly solved and skipped anagrams per 'hard' round. They indicate the subjects' capability to find their individual payoff-optimal mix of skipping and solving anagrams, which may well vary across subjects. Some subjects might skip individual hard anagrams that they cannot solve in reasonable time, deliberately accepting the costs of DKK 3.00, in order to try their luck on the next word that might yield them a payoff of DKK 5.00. Other subjects might refrain from skipping entirely in order not to reduce their earnings. Our measure allows us to control for these individual tactics in a unified way. Ability is defined as the average number of solved anagrams when working on hard anagrams in part 1. Since our tactics and ability measures are, by design, highly correlated (Spearman's $\rho = .95$, $p < .01$), we control for them only in separate regressions. Additionally, we account for potential learning effects by including round dummies in the regression models in order to isolate the true treatment effects from the effect of these confounds. We begin our analysis by comparing Observers' and Peers' behavior to that of subjects from the Baseline Treatment. Subsequently, we move on to study how individuals' general grit levels interact with treatment effects. Lastly, we will consider the question of how Observers are affected by specific attributes of their matched Peers.

3.1 Observers react to their Peers, but Peers don't seem to care

Using the pooled OLS models in Table 2, we test for general treatment differences and in particular role effects. We regress our main variable of interest, that is, subjects' number of skipped hard anagrams on role dummies that take the value 1 for subjects in the Observer and Peer role,

respectively. We hence treat the subjects from the Baseline Treatment as reference group.

As becomes evident from Table 2, Observers are more perseverant than subjects who work in isolation: Observers skip significantly fewer hard anagrams than subjects from the Baseline Treatment. The coefficient of the Peer dummy is comparably smaller and in none of the regressions significant. These results are robust to controlling for individual ability or tactics (see Models (2) and (3)) and when including subjects' grit scores from the questionnaire (see Models (4) and (5)). We summarize our findings as follows:

Finding 1 *Being able to observe another subject significantly increases the Observers' perseverance when working on hard anagrams. In contrast, being aware of the fact that another subject observes one's actions does not significantly affect Peers' performance.*

This is in contrast to both Mas and Moretti (2009) who find stronger effects (in effort) for observed than for observing workers, and van Veldhuizen et al. (2015), who replicate Mas and Moretti's (2009) setup in the lab and do not find significant differences in the behavior of observed and observing subjects. A potential explanation for the conflicting findings is the difference in the incentive structure of these two studies and our experiment. In these studies the observed and observing workers' workload is interdependent. If one person free-rides, its peer has to work harder. Thus, when workers are observed they might work harder out of guilt or fear of upsetting their team. Especially in the case of Mas and Moretti (2009) where workers are not anonymous, slacking workers might fear negative reciprocity from observers. In our study, we abstract from these behavioral drivers by making the payoff and workload independent.

A further, noteworthy finding from Table 2 is that individual grit, measured using Duckworth et al.'s (2007) Short Grit scale seems to matter significantly less when working in a peer situation. The main effects of grit for Observers and Peers (considering *Observer* $\times$ *Self-reported grit* + *Self-reported grit* and *Peer* $\times$ *Self-reported grit* + *Self-reported grit*) in Models (6) and (7) are both far from significant ($p > 0.557$). This result nicely complements Gerhards and Gravert's (2017) findings. Being "gritty" seems to help subjects if they work in isolation – like in the Baseline Treatment. If, conversely, subjects can observe someone else's perseverance or know that they are observed, they do not require to be generally gritty themselves in order to stay perseverant.

Lastly, Figure A.1 and Table A.1 in Appendix A consider role and treatment differences in further outcome variables. They corroborate and complement the positive impression from our preceding analysis. Figure A.1 suggests that Peers and Observers in the Random Matching Treatment realize higher mean earnings than their counterparts from the Baseline Treatment who work in isolation. In particular they solve a larger number of anagrams correctly and switch on average less often to the outside option than subjects from the Baseline Treatment do. However, while the estimated coefficients in Table A.1 largely have the expected signs, the role differences from Figure A.1 are not statistically significant once one controls for subjects' tactics and ability.[8]

[8]It is of note that we find significant peer effects in skipping, but not in switching to the outside option, although one could argue that switching constitutes an avoidance behavior, too. We suspect that the high explicit and implicit

Table 2: Skipping in part 2: Observers and Peers compared to Baseline subjects

	Hard anagrams skipped						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Observer	−0.283**	−0.366***	−0.385***	−0.371***	−0.390***	−2.226***	−2.202**
	(0.144)	(0.138)	(0.143)	(0.136)	(0.141)	(0.794)	(0.860)
Peer	0.029	−0.088	−0.082	−0.086	−0.078	−1.957**	−1.896**
	(0.198)	(0.154)	(0.164)	(0.155)	(0.164)	(0.801)	(0.861)
Ability: Average no. of solved anagrams per 'hard' round in part 1		0.253*** (0.058)		0.244*** (0.057)		0.264*** (0.058)	
Tactics: Average earnings per 'hard' round in part 1			0.041*** (0.015)		0.039*** (0.014)		0.044*** (0.014)
Self-reported grit				−0.144	−0.186*	−0.535***	−0.565***
				(0.091)	(0.100)	(0.192)	(0.211)
Observer × Self-reported grit						0.602***	0.586**
						(0.232)	(0.251)
Peer × Self-reported grit						0.604***	0.585**
						(0.233)	(0.251)
Constant	0.661***	0.149	0.312*	0.611*	0.903**	1.774***	2.026***
	(0.127)	(0.139)	(0.163)	(0.344)	(0.390)	(0.636)	(0.710)
R^2	0.025	0.238	0.142	0.246	0.157	0.281	0.189
Number of clusters	152	152	152	152	152	152	152
Observations	1232	1232	1232	1232	1232	1232	1232

Pooled OLS regressions, dependent variable: Number of of skipped anagrams when working on hard anagrams. Bootstrapped standard errors are given in parentheses: * p<0.10 ** p<0.05 *** p<0.01. These are obtained from 700 iterations of the estimations. All regressions include round dummies. The number of observations is smaller than the maximum possible number of 1520 observations since some of the 152 subjects switched to the outside option in some of the ten rounds.

3.2 Which Peers affect Observers' performance most?

Next, we focus on the extent to which the Peer affects his or her Observer's skipping behavior in the Random Matching Treatment. In the OLS regressions presented in Table 3 we regress the Observer's average number of skipped anagrams per round in part 2 on his or her matched Peer's respective behavior in part 2. Since the impact of this information might well depend on the similarity (or dissimilarity) of the Observer and Peer, we add a further dummy variable that takes the value one if the Peer was from a better performance quintile (i.e. solved more hard anagrams correctly in part 1) than the Observer. Furthermore, we interact this dummy with the Peer's avoidance behavior to allow for heterogeneous treatment effects.[9] Lastly, as in our previous regressions, we separately add controls for ability and tactics as well as for the Observer's self-reported grit level that was elicited in the final questionnaire.

Table 3: Determinants of Observers' skipping behavior

	Skips in hard			
	(1)	(2)	(3)	(4)
Peer from better performance quintile	−0.441*	−0.528**	−0.446*	−0.534**
	(0.240)	(0.243)	(0.242)	(0.247)
Peer's skips in part 2	−0.014*	−0.014**	−0.014*	−0.015**
	(0.007)	(0.007)	(0.007)	(0.007)
Peer's skips × Peer from better performance quintile	0.024**	0.023**	0.026**	0.027**
	(0.009)	(0.009)	(0.012)	(0.012)
Ability: Average no. of solved anagrams per 'hard' round in part 1	0.053		0.050	
	(0.093)		(0.091)	
Tactics: Average earnings per 'hard' round in part 1		0.002		0.001
		(0.020)		(0.019)
Self-reported grit			−0.048	−0.067
			(0.124)	(0.131)
Constant	0.410	0.553	0.558	0.760
	(0.335)	(0.338)	(0.454)	(0.480)
R^2	0.213	0.193	0.216	0.198
Observations	45	45	45	45

OLS, dependent variable: Mean number of of skipped anagrams per round when working on hard anagrams. Performance quintile 1 contains those subjects who solve the most hard anagrams. Robust standard errors are given in parentheses: * p<0.10 ** p<0.05 *** p<0.01.

Models (1) and (2) reveal that being able to observe a Peer from a better performance quintile – who does not skip any anagram – significantly lowers the Observer's skipping. Moreover, the number of skipped anagrams – of a Peer from a similar or worse performance quintile – has a

costs of switching to easy anagrams reduce subjects' tendency to switch in general, leaving less room for peers to have an effect.

[9]Note that subjects from different performance quintiles are indeed of significantly different ability. Mann-Whitney ranksum tests reveal significant performance differences for all relevant quintile group comparisons (number of solved anagrams per 'hard" round in part 1 of subjects in quintile 1 vs. quintile 2, ..., quintile 4 vs. quintile 5, all $p < 0.02$).

significantly negative impact on the Observer's skipping. This could be driven by the fact that the negative effects of skipping are more salient to the Observer when being presented with the Peer's number of skips. Lastly, the positive interaction term of *Peer's skips* $\times$ *Peer from better performance quintile* indicates that the Peer's skipping behavior has less of an effect if the Peer is from a better performance quintile. We summarize our findings from Table 3 as follows:

Finding 2 *Both being able to observe a more able Peer (irrespective of his or her skipping behavior), and observing the actual skipping behavior of a similar or less able peer has a significant negative effect on the Observer's own skipping. Both types of Peer-Observer matchings hence lead to more perseverant behavior of the Observer.*

In models (3) and (4) we extend the previous regressions by controlling for Observers' self-reported grit scores as elicited in the final questionnaires. As evident, this leaves our findings unaffected.

4 Self-selection into the Observer and Peer Roles

The findings from our Baseline and Random Matching Treatments raise further questions. The first one relates to the deliberate choice of peers: What type of Peer would Observers like to be matched to if they had a choice? A recent theory by Battaglini et al. (2005) suggests that in situations that require perseverance, the ideal Peer is of similar or slightly worse ability than the Observer. This makes own successes more encouraging and failures less discouraging. Also according to the social comparison model by Falk and Knell (2004), which is based on the social comparison theory by Festinger (1954), Observers should on average choose Peers that are similar to them. This inference is deduced from balancing two competing motivations: On the one hand, an upwards comparison can inspire individuals to work harder (self-improvement). On the other hand, a downwards comparison can make them feel better about themselves (self-enhancement). Thus, as one's own ability rises, the reference level for social comparison rises as well in order to stay close to one's own level. Based on these theoretical considerations we conjecture that in our experiment, Observers would like to choose Peers who are equally or slightly less able than they are themselves.

A further question relates to the detection of the factors and personal characteristics that drive subjects' desire to voluntarily self-select into the Observer and Peer role. The questionnaire data from the Random Matching Treatment provide a first hint. Since we exogenously and randomly assigned Observers and Peers to their roles in this treatment, this data allows us to cleanly identify potential differences in task perception depending on role. In particular, we asked all subjects from the Random Matching Treatment (1) how much they enjoyed working on the task, how (2) challenging and (3) exhausting they perceived the task and (4) how much effort they provided during the task. Answers were given on 5-point scales. At the aggregate level, ranksum tests do not reveal significant differences between the answers provided by Observers and Peers ($p = .431$,

$p = .246$, $p = .469$ and $p = .251$, respectively). But the picture changes if we restrict the group of Observers to those 22 subjects who were matched to a Peer from a higher performance quintile. These Observers report to having enjoyed the task significantly less ($p = .069$) and having perceived the task as a greater challenge ($p = .030$) than the 45 Peers. From this, one could conjecture that subjects will exhibit a tendency to avoid the Observer role – in particular if they expect to be matched with a Peer from a better performance quintile, for instance, due to their own relative low ability.

On the other hand, one could argue that subjects dislike the position in which they are observed by others and hence rather avoid the Peer role. The fact that only avoidance behavior and not positively connoted productivity is reported to the Observer might reinforce this effect. It is hence an empirical question which of the two effects prevails, i.e. which role is preferred on average.

In order to tackle the questions of peer selection and role choice, we designed a third treatment, the Chosen Matching Treatment, that we will describe below.

4.1 Experimental Design: Chosen Matching Treatment

In the summer and fall of 2014 we invited 90 additional subjects to the lab using the same recruiting procedure as for the Baseline and Random Matching Treatment. During the experimental sessions the only difference to the Random Matching Treatment is the intervention that we introduce between part 1 and part 2: After guessing and being informed about their performance quintile in part 1 all subjects have to make two choices for part 2. First, we ask them to indicate whether they want to assume the role of an Observer or that of a Peer (labeled Person A and Person B, similar to the Random Matching Treatment). We inform subjects that the computer allocates the desired roles under the premise that within each performance quintile always one of the subjects assumes either role. This procedure is strategy-proof, as the subjects' best strategy is still to state their preferences truthfully, and yet it gives us the flexibility to allocate the experimental roles in an efficient way.[10] Subsequently, we ask the subjects to state from which performance quintile they want to observe a Peer in case they are assigned the role of an Observer. Using these procedures it is possible that more than one Observer is matched to one Peer.[11]

4.2 Results: Chosen Matching Treatment

In the following, we first consider subjects' peer choices. That is, in the words of Falk and Knell (2004) we focus on the question of whether on average the self-enhancing (downward) or the self-improving (upward) comparison motivation determines subjects' peer choices.

[10] In fact, 29 out of 90 subjects in that treatment were allocated their non-preferred roles: 6 out of 28 subjects who wanted to be an Observer had to assume the Peer role, 23 out of 62 subjects who stated a preference for the Peer role had to take on the role of an Observer. The remaining 61 subjects could assume the roles they initially stated as their preferred choice.

[11] At the beginning of part 2, when Peers do not yet know how many subjects are going to observe them in the following rounds, we elicit their belief on somebody choosing to observe them. Only at the very end of part 2 they receive information on whether their beliefs were correct and whether they will earn additional DKK 20.00 for their correct guess.

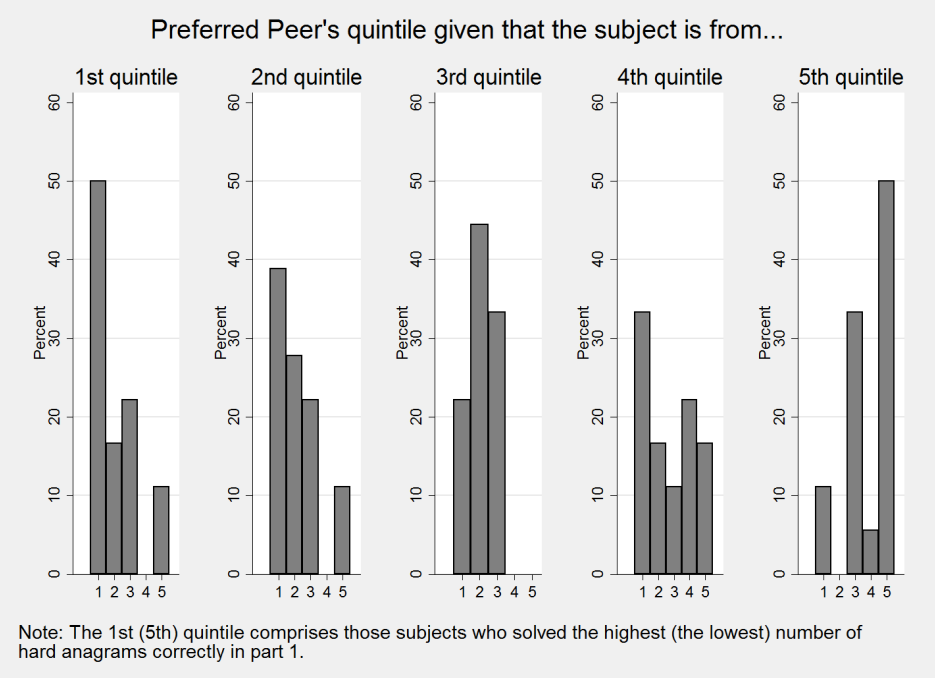

Figure 2: Choice of Peer quintile

Figure 2 shows for each of the subjects' performance quintile separately, their preferred Peer's performance quintile. As evident, in all five quintiles (the first quintile comprising the subjects with the highest performance in part 1) at least fifty percent of the subjects prefer to be informed about the avoidance behavior of a Peer from their own or a better performance quintile. This suggests that, on average, the self-improvement comparison motivation seems to prevail in our experiment.

However, there is also a non-negligible fraction of subjects in each of the five quintiles that wishes to observe the avoidance behavior of a (slightly) worse peer, which is in line with Battaglini et al.'s (2005) model.

Next we consider subjects' role choices. Interestingly, we find that 62 out of the 90 subjects want to assume the Peer role, that is, they want to be the ones being observed. In order to study the determinants of this choice, we run individual Probit models in which we regress the subjects' decision to become a Peer (= 1, vs. an Observer = 0) on their average performance on hard anagrams in part 1, their amount of skipping and switching in part 1, a dummy variable indicating overconfidence (regarding their performance quintile in part 1) as well as on their self-reported levels of narcissism, grit, altruism and gender that we elicit in the questionnaire. It turns out that only two of these variables have significant effects on the desire to assume the Peer role. Subjects' average performance on hard anagrams is positively correlated with their desire to be observed, while subjects' inclination to switch to easy anagrams is negatively correlated with it. The corresponding regression Table A.2 is reported in Appendix A.

Still puzzled by the finding that about two thirds of the subjects wanted to take on the Peer role, we further consulted their statements in the final questionnaire where we asked them to explain

the reasons why they selected a specific role. Interestingly, 41 out of the 62 would-be Peers stated to have chosen that role in order to avoid any information that might distract them from their work. One of those subjects, for instance, wrote: *"I did not want to observe anybody else, as I thought it would be annoying while working."*. These subjects were hence consciously self-selecting out of a situation in which they would receive information about a peer. Since we did not provide them with a neutral "No peer option", selecting the Peer role was their only option to avoid peer information.

Only three of the remaining 21 subjects who wanted to assume the Peer role stated that they wanted to do so in order to provide a good example for their Observer such that he or she could learn from observation. One of them, for instance, wrote that she *"felt very confident in the experiment, so [she] figured that someone else, might benefit from seeing, what/how [she] was doing."*.

Taken together, our results from the Chosen Matching Treatment indicate that individuals would rather like to avoid receiving information about a peer's performance in a cognitively demanding task such as the anagram task. However, when being forced into the role of an Observer, they seem to be guided by the motive of self-improvement.

5 Conclusion

The literature on peer effects has been growing over the last years. However, the precise mechanisms how peer effects arise are still a black box. In this paper we uncover new puzzle pieces that help us to better understand how individuals are affected by peers. In particular, instead of simply measuring how peers influence produced output, we study how the presence of a peer affects subjects' behavior in a rather qualitative dimension, that is, how peers affect their co-workers' perseverance.

We find that Observers' skipping behavior in our word play task decreases significantly when they receive information about a Peer. This effect is independent of the Peer's actual behavior if the Peer is known to be from a better performance quintile. If, conversely, the Peer is from a similar or lower performance quintile, the more often the Peer skips, the more perseverant the Observer behaves him- or herself. For Peers, knowing to be observed does not affect their behavior. On average, they do not behave significantly different than subjects from our Baseline Treatment who work in isolation.

The fact that on average all Observers, independent of their ability, increase their perseverance in the presence of a Peer has policy implications for diverse situations in which peer effects can be expected, including, but not limited to sport, social, work and educational contexts. Being conscious about being in a peer situation increases a group's average perseverance. This allows for the formation of teams with varying ability and grit levels. In the organizational context superiors are well advised to arrange working conditions that permit workers to observe each other regardless of their individual tasks, in order to profit from perseverance-increasing peer effects. Examples being open plan offices or open factory work spaces. In a similar vein, our findings suggest that students could benefit from studying in the library or study rooms instead of working individually

at home. In the domain of sports, athletes might benefit from training with other athletes – even if they compete in different sports – in order to reap motivational effects by observing peer athletes staying persistent in their training.

Although our findings from the Random Matching Treatment indicate that peers increase their co-workers' perseverance, the results from our Chosen Matching Treatment, suggest that a significant part of the workforce might prefer to entirely avoid the peer situation. When given the choice between the Observer and Peer role, the majority of subjects prefers the Peer role. Subjects' survey responses suggest that they use this as a means to avoid being distracted by information about another subject's actions, even at the expense of their own avoidance behavior being observed by others. This finding warrants further study, as it shines new light on how to interpret peer effects. Individuals might prefer and even deliberately choose to work alongside a peer in a tedious task such as Falk and Ichino's (2006) folding letters into envelopes task. However, if they are confronted with a cognitively more challenging task that requires a high level of concentration, the findings from the present study indicate that they might prefer to work in isolation in order to not get distracted by peer information. In a similar vein, also Rosaz et al. (2016) note that peer effects, in particular sociability effects, might be more pronounced in cognitively less demanding tasks. It hence seems worthwhile to gain further understanding of how the work environment and specificities of the work task at hand influence the effectiveness of peer effects.

References

Alaoui, L., Fons-Rosen, C., 2016. Know when to fold'em: The grit factor. Working paper, Universitat Pompeu Fabra and Barcelona.

Andreoni, J., Bernheim, B. D., 2009. Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. Econometrica 77 (5), 1607–1636.

Bäker, A., Mechtel, M., Jun. 2013. The Role of Task Meaning on Output in Groups: Experimental Evidence. IAAEU Discussion Papers 201308, Institute of Labour Law and Industrial Relations in the European Union (IAAEU).

Bandiera, O., Barankay, I., Rasul, I., 2010. Social Incentives in the Workplace. Review of Economic Studies 77 (2), 417–458.

Bandura, A., Walters, R. H., 1963. Social learning and personality development.

Banerjee, A. V., 1992. A Simple Model of Herd Behavior. The Quarterly Journal of Economics 107 (3), 797–817.

Battaglini, M., Benabou, R., Tirole, J., 2005. Self-control in peer groups. Journal of Economic Theory 123 (2), 105–134.

Battaglini, M., Díaz, C., Patacchini, E., February 2016. Self-Control and Peer Groups: An Empirical Analysis. Journal of Economic Behavior & Organization 134, 240–254.

Bellemare, C., Lepage, P., Shearer, B., 2010. Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace. Labour Economics 17 (1), 276–283.

Bernheim, B. D., 1994. A theory of conformity. Journal of Political Economy 102 (5), 841–877.

Beugnot, J., Fortin, B., Lacroix, G., Villeval, M. C., Jul. 2013. Social networks and peer effects at work. IZA Discussion Papers 7521, Institute for the Study of Labor (IZA).

Bonein, A., Denant-Boèmont, L., 2015. Self-control, commitment and peer pressure: a laboratory experiment. Experimental Economics 18 (4), 543–568.

Bucciol, A., Houser, D., Piovesan, M., 01 2013. Temptation at work. PLoS ONE 8 (1), 1–5.

Buechel, B., Mechtenberg, L., Petersen, J., 2014. Peer effects and students' self-control. MPRA Paper No. 53658, Munich Personal RePEc Archive.

Charness, G., Kuhn, P., 2011. Lab labor: What can labor economists learn from the lab? Handbook of Labor Economics 4, 229–330.

Charness, G., Rigotti, L., Rustichini, A., 2007. Individual behavior and group membership. American Economic Review 97 (4), 1340–1352.

Corgnet, B., Gonzalez, R. H., Mateo, R., 2015. Cognitive reflection and the diligent worker: an experimental study of millennials. PloS one 10 (11), e0141243–.

Deci, E. L., Ryan, R. M., 1985. Intrinsic motivation and self-determination in human behavior. Springer Science & Business Media.

Duckworth, A. L., Peterson, C., Matthews, M. D., Kelly, D. R., 2007. Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology 92 (6), 1087–1101.

Duckworth, A. L., Quinn, P. D., 2009. Development and validation of the short grit scale (GRIT-S). Journal of Personality Assessment 91 (2), 166–174.

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

Falk, A., Knell, M., October 2004. Choosing the Joneses: Endogenous goals and reference standards. Scandinavian Journal of Economics 106 (3), 417–435.

Festinger, L., 1954. A theory of social comparison processes. Human Relations 7 (2), 117–140.

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

Georganas, S., Tonin, M., Vlassopoulos, M., 2015. Peer pressure and productivity: The role of observing and being observed. Journal of Economic Behavior and Organization 117, 223–232.

Gerhards, L., Gravert, C., 2017. Grit Trumps Talent? An experimental approach. Working Paper, Universität Hamburg.

Gill, D., Prowse, V., 2012. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. American Economic Review 102 (1), 469–503.

Guryan, J., Kroft, K., Notowidigdo, M. J., 2009. Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. American Economic Journal: Applied Economics 1 (4), 34–68.

Kandel, E., Lazear, E. P., 1992. Peer pressure and partnerships. Journal of Political Economy, 801—817.

Kaur, S., Kremer, M., Mullainathan, S., 2010. Self-control and the development of work arrangements. American Economic Review 100 (2), 624–628.

Kräkel, M., May 2016. Peer effects and incentives. Games and Economic Behavior 97, 120–127.

Manski, C. F., 1993. Identification of endogenous social effects: The reflection problem. The Review of Economic Studies 60 (3), 531–542.

Mas, A., Moretti, E., 2009. Peers at work. American Economic Review 99 (1), 112–145.

Mobius, M., Rosenblat, T., 2014. Social learning in economics. Annual Review of Economics 6 (1), 827–847.

O'Keeffe, M., Viscusi, W. K., Zeckhauser, R. J., January 1984. Economic Contests: Comparative Reward Schemes. Journal of Labor Economics 2 (1), 27–56.

Rosaz, J., Slonim, R., Villeval, M. C., 2016. Quitting and peer effects at work. Labour Economics 39, 55–67.

Tangney, J., Baumeister, R., Boone, A., 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality 72 (2), 271–322.

van Veldhuizen, R., Oosterbeek, H., Sonnemans, J., Apr. 2015. Peers at work: From the field to the lab. Tinbergen Institute Discussion Papers 14-051/I, Tinbergen Institute.

A Additional figures and tables

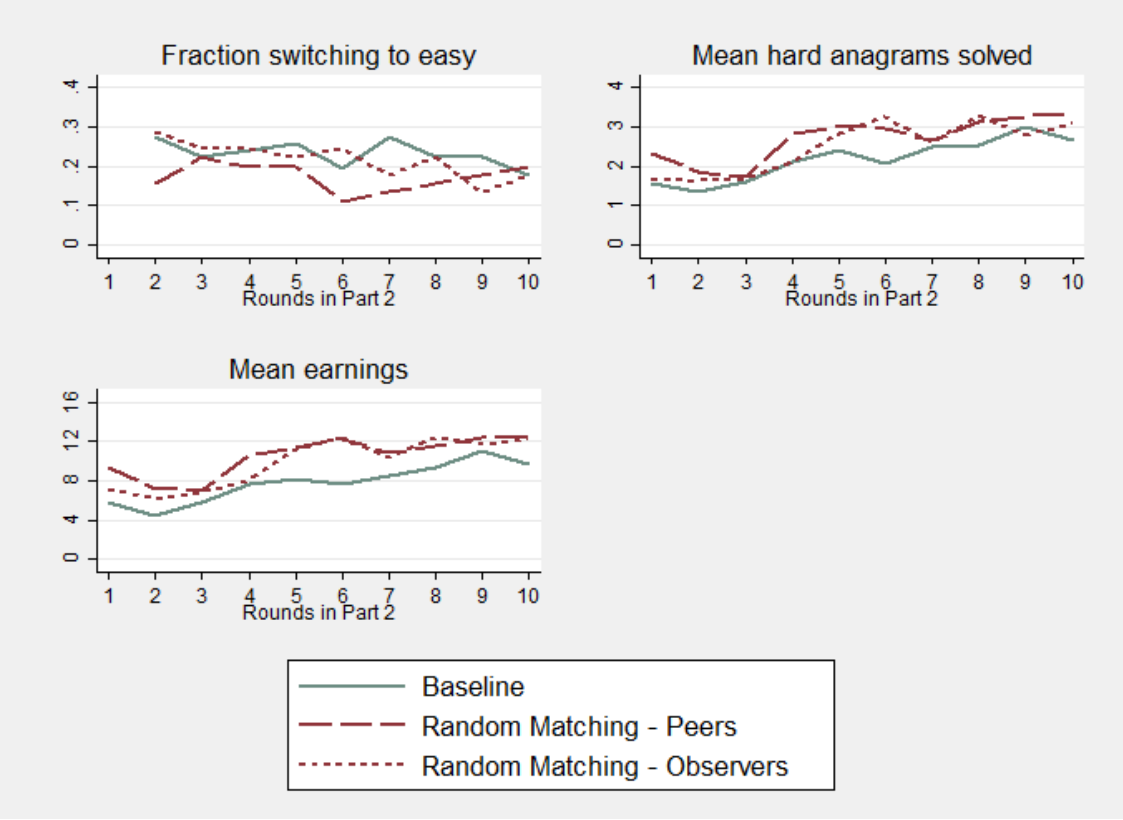

Figure A.1: Further outcome variables across treatments

Table A.1: Switching, Performance and Earnings in part 2: Observers and Peers compared to Baseline subjects

	Switch to easy		Hard anagrams solved		Earnings	
	(1)	(2)	(3)	(4)	(5)	(6)
Observer	−0.409	−0.066	1.061	0.127	8.550	4.328
	(1.323)	(1.306)	(1.508)	(1.545)	(5.804)	(5.502)
Peer	−0.282	0.010	−0.331	−1.091	4.398	1.142
	(1.236)	(1.205)	(1.024)	(1.129)	(4.055)	(3.971)
Ability: Average no. of solved	−0.214***		0.971***		3.909***	
anagrams per 'hard' round in part 1	(0.081)		(0.098)		(0.497)	
Tactics: Average earnings per		−0.064***		0.209***		0.879***
'hard' round in part 1		(0.020)		(0.022)		(0.099)
Self-reported grit	−0.155	−0.106	−0.260	−0.470	0.447	−0.353
	(0.312)	(0.303)	(0.261)	(0.288)	(0.879)	(0.898)
Observer × Self-reported grit	0.130	0.034	−0.359	−0.121	−2.539	−1.430
	(0.429)	(0.421)	(0.461)	(0.472)	(1.786)	(1.705)
Peer × Self-reported grit	0.063	−0.007	0.135	0.345	−1.313	−0.435
	(0.398)	(0.386)	(0.314)	(0.344)	(1.241)	(1.212)
Constant	0.055	−0.012	0.400	1.241	−3.240	−0.354
	(0.948)	(0.917)	(0.822)	(0.907)	(2.883)	(2.840)
R^2			0.498	0.479	0.389	0.409
Pseudo R^2	0.051	0.079				
Number of clusters	152	152	152	152	152	152
Observations	1368	1368	1232	1232	1520	1520

Regressions 1 and 2: Pooled Probit, dependent variable: Decision to switch to easy anagrams; Regressions 3 and 4: pooled OLS, dependent variable: Number of solved hard anagrams (given worked on hard); Regressions 5 and 6: pooled OLS, dependent variable: Earnings. Bootstrapped standard errors are given in parentheses: * p<0.10 ** p<0.05 *** p<0.01. These are obtained from 700 iterations of the estimations. All regressions include round dummies. In Regressions 1 and 2 the number of observations is 1368 since the 152 subjects could switch to the outside option only from round 2 (out of 10) onward. In Regressions 3 and 4 the number of observations is smaller than the maximum possible number of 1520 observations since some of the 152 subjects switched to the outside option in some of the ten rounds.

Table A.2: Chosen Matching Treatment: Determinants of Peer role choice

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solved hard anagrams in part 1	0.183*							
	(0.100)							
Skipped hard anagrams in part 1		0.019						
		(0.037)						
Switches to easy anagrams in part 1			−0.108**					
			(0.053)					
Self-reported narcissism				0.102				
				(0.528)				
Self-reported grit					−0.073			
					(0.234)			
Self-reported altruistic inclination						0.016		
						(0.052)		
Overconfident regarding performance quintile in part 1							−0.279	
							(0.294)	
Female								−0.111
								(0.280)
Constant	0.114	0.446***	0.810***	0.445	0.724	0.379	0.587***	0.541***
	(0.242)	(0.164)	(0.213)	(0.283)	(0.759)	(0.407)	(0.172)	(0.186)
Pseudo R^2	0.033	0.003	0.037	0.000	0.001	0.001	0.008	0.001
Observations	90	90	90	90	90	90	90	90

Probit regressions with robust standard errors that are given in parentheses: * p<0.10 ** p<0.05 *** p<0.01.

B Questionnaire

Question sets 2-9 were asked either at the beginning of the experiment or at the end. Question sets 1 and 10 were always asked at the end of the experiment. The headlines for each set of questions were neutral in the questionnaire.

1. Perception of the task
   (a) How much did you enjoy working on the task?
   (b) How challenging did you perceive the task?
   (c) How much effort did you provide during the task?
   (d) How exhausting did you perceive the task?
2. General questions
   (a) Please tell us your age
   (b) What is your gender?
   (c) What is your field of study?
   (d) How many years of English did you have in school?
   (e) How often do you play scrabble or solve crossword puzzles?
3. Questions on time and risk preferences and optimism (Scored from 0-10)
   (a) Are you generally an impatient person or someone who always shows great patience?
   (b) Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?
   (c) Are you generally an optimistic person or do you expect things to go wrong?
4. Short Grit scale (Scored from 1 ("Not at all like me") to 5 ("Very much like me"). Items 1, 3, 5 and 6 are reversely coded.)
   (a) New ideas and projects sometimes distract me from previous ones.
   (b) Setbacks don't discourage me.
   (c) I have been obsessed with a certain idea or project for a short time but later lost interest.
   (d) I am a hard worker.
   (e) I often set a goal but later choose to pursue a different one.
   (f) I have difficulty maintaining my focus on projects that take more than a few months to complete.
   (g) I finish whatever I begin.
   (h) I am diligent.

5. Brief-Self-Control-Scale (Scored from 1 ("Not at all like me") to 5 ("Very much like me"). Items b, c, d, e, g, i, j, l and m are reversely coded.)

   (a) I am good at resisting temptation.

   (b) I have a hard time breaking bad habits.

   (c) I am lazy.

   (d) I say inappropriate things.

   (e) I do certain things that are bad for me, if they are fun.

   (f) I refuse things that are bad for me.

   (g) I wish I had more self-discipline

   (h) People would say that I have iron self- discipline.

   (i) Pleasure and fun sometimes keep me from getting work done.

   (j) I have trouble concentrating.

   (k) I am able to work effectively toward long-term goals.

   (l) Sometimes I can't stop myself from doing something, even if I know it is wrong.

   (m) I often act without thinking through all the alternatives.

6. Rotter's 4-Item Locus of Control Scale (Choose A or B and on 2 levels)

   (a) A. What happens to me is my own doing.
   B. Sometimes I feel that I don't have enough control over the direction my life is taking.

   (b) A. When I make plans, I am almost certain that I can make them work.
   B. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune.

   (c) A. In my case getting what I want has little or nothing to do with luck.
   B. Many times we might just as well decide what to do by flipping a coin.

   (d) A. Many times I feel that I have little influence over the things that happen to me.
   B. It is impossible for me to believe that chance or luck plays an important role in my life.

7. Narcissism Scale (Choose A or B)

   (a) A. I know that I am good because everybody keeps telling me so.
   B. When people compliment me I sometimes get embarrassed.

   (b) A. I like having authority over people.
   B. I don't mind following orders.

   (c) A. I really like to be the center of attention.
   B. It makes me uncomfortable to be the center of attention.

(d) A. People always seem to recognize my authority.
B. Being an authority doesn't mean that much to me.

(e) A. I find it easy to manipulate people.
B. I don't like it when I find myself manipulating people.

(f) A. I have a natural talent for influencing people.
B. I am not good at influencing people.

(g) A. I see myself as a good leader.
B. I am not sure if I would make a good leader.

(h) A. I would prefer to be a leader.
B. It makes little difference to me whether I am a leader or not.

(i) A. I am a born leader.
B. Leadership is a quality that takes a long time to develop

8. Altruism

(a) How would you assess your willingness to share with others without expecting anything in return, for example your willingness to give to charity?

(b) Imagine the following situation: You unexpectedly receive 7000 kr. How much of this would you donate to charity? Give a value between 0 and 7000.

(c) How well does the following statement describe you as a person? I do not understand why people spend their lifetime fighting for a cause that is not directly beneficial for them.

9. Personal Questions

(a) Do you smoke cigarettes?

(b) Do you exercise regularly (at least once a week or more)?

(c) Are you on a sports team?

(d) How many siblings do you have?

(e) How many younger siblings do you have?

(f) Do you currently volunteer anywhere?

(g) Do you have any official function in any student association or club?

(h) Do you usually study in a study group?

(i) If you study in a study group are you the person who encourages others do better or does studying with others rather help you to be better?

10. Last question

(a) What do you think this experiment was about?