



Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jeboIf I can do it, so can you! Peer effects on perseverance[☆]Berno Buechel^{a,1}, Lydia Mechtenberg^{b,*}, Julia Petersen^b^a Department of Economics, University of Fribourg, Bd. de Pérolles 90, Fribourg 1700, Switzerland^b Department of Economics, University of Hamburg, Von-Melle-Park 5, Hamburg 20148, Germany

ARTICLE INFO

Article history:

Received 19 March 2018

Revised 3 September 2018

Accepted 4 September 2018

Available online 25 September 2018

JEL-classification:

C91

D90

J24

Keywords:

Perseverance

Peer effects

Experiment

ABSTRACT

Successful performance – be it in school, at the job, or in sports activities – requires perseverance, i.e., persistent work on a demanding task. We investigate in a controlled laboratory experiment how an individual's social environment affects perseverance. We find evidence for two kinds of peer effects: being observed by a peer can postpone the decision to give up, while observing a peer can be informative and have more sustainable effects. In particular, we show that successful peers affect perseverance positively if they communicate their success in a motivating way and negatively otherwise. Our experimental results suggest that peers affect perseverance indirectly, via influencing self-confidence.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Motivation

Some people are more successful than others. Be it in school, at the job, or in sports activities, there are tasks which require painful effort at some point. While many people let themselves get distracted or fully give up, others stay focused and keep working on the successful completion, i.e., they *persevere*. Clearly, while perseverance is to some extent a matter of personality, it is also, and probably more importantly, shaped by the social environment, which includes peers, teachers, parents, and coaches. Personal interactions can have two very distinct effects on individual perseverance.¹ On the one hand, others who know about the task at hand and interact with the individual after the task can induce social pressure. This is the typical view that peer effects work through *being observed* by peers. On the other hand, *observing peers* before the

[☆] This paper is a substantial revision of an older version that circulated under the title “Peer Effects and Perseverance.” We would like to thank the editor and two anonymous referees for their helpful suggestions. Special thanks go to the WISO lab crew: Olaf Bock (leader), Thais M. Hamasaki, Jan Papemeier, Fenja Pauls, and Merlinda Claudia Tews; and to Willi Flemming and Lorenz Rumberger for supreme research assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author.

E-mail addresses: berno.buechel@unifr.ch (B. Buechel), lydia.mechtenberg@wiso.uni-hamburg.de (L. Mechtenberg), petersen.julia@t-online.de (J. Petersen).

¹ Web: www.berno.info

¹ In a field study with children, [Alan et al. \(2016\)](#) provide evidence for the view that perseverance is not a fixed personality trait but can be fostered if children learn to acknowledge the value of grit for achievement. The authors do not consider peer effects.

own task can affect an individual's ability to persevere. More experienced peers, as well as mentors and coaches, may foreshadow the value of perseverance and affect the confidence of the one aspiring to successfully complete a task.² It is important to understand the conditions for positive and negative peer effects on perseverance because they can have severe consequences. Applications range from students' educational success, over performance of sports athletes, to success within organizations.

In this paper, we test both categories of peer effects on perseverance in a controlled laboratory experiment. We find that both types of peer effects are present in some way. Peer pressure to justify the own performance after the task induces participants to try for a longer period before giving up. Communication with a successful peer before the task, on the other hand, can either induce participants to give up immediately or to persevere, depending on the content of the communication. This last result means that a successful peer can have a crucial effect on the perseverance of a less experienced individual, while unsuccessful peers are rather ignored, according to our data. Interestingly, both types of peer effects are more pronounced for women than for men.

We provide one of the first experimental studies on how peers influence individual perseverance by their own success or failure.³ We find that the effect depends on the way peers communicate about the difficulty of the task. Moreover, we describe effect heterogeneity between men and women.

1.2. Two categories of peer effects

Perseverance necessitates self-control, i.e., the strength to resist temptations to procrastinate.⁴ For instance, a student might want to study hard in the evening, in order to be well prepared for class the next day and to eventually perform better in the exam by the end of the semester. However, when evening actually arrives, the now immediate psychological costs from studying become much more salient than the still relatively far-away benefits, and the student is severely tempted to go out instead. More generally, if a disproportionately high weight is attached to whatever costs or benefits would accrue immediately, then plans implying that investments precede benefits will be likely to fail due to the temptation to avoid immediate costs. Since performance plans typically are of this kind, a present bias can impede perseverance and hence performance. This link is studied in the literature on present-biased preferences.⁵

The vast majority of the studies on peer effects suggest that being observed by peers can enhance performance, e.g. due to peer pressure.⁶ By imposing some social cost on a person whom they observe giving in to a temptation, they can help that person not to give in. An example for such an effect in the educational context is a group of students who do all their work in the library and who would feel ashamed if watched by the others surfing the internet instead of writing their term paper.⁷

The second category of peer effects on perseverance is elaborated in the theory of Battaglini et al. (2005), hereafter: BBT, which provides us with our main hypotheses. They stipulate that observing one's peers – rather than being observed by them – can affect own performance. The reason is that the behavior of others can be informative about how much the task at hand would challenge one's own ability to persevere.

In BBT, individuals have imperfect knowledge about their ability to persevere on a task (that may be anything from not drinking alcohol to preparing for an exam or working on a paper).⁸ Individuals face a number of subsequent decisions

² As an illustration of the two categories of peer effects, we may consider a leisure time runner who aspires to complete a marathon. Joining a group of runners might have a positive effect for the ability to persevere for multiple reasons. Committing to regularly practicing together and having announced the own plan to participate in a certain marathon to the group clearly fall into the category of peer effects through peer pressure. Learning about the challenges that running a marathon implies and increasing the own self-confidence by observing that others have succeeded to finish a marathon clearly fall into the category of peer effects through learning. Yet, to choose a group of runners, it is important to know which effect is at work, under which conditions they are working, and when there are actually negative peer effects.

³ The only other study we are aware of is the one by Gerhards and Gravert (2018).

⁴ The opposite need not be true: It is possible to be self-controlled and *not* to persevere, e.g., because the optimal decision is to give up a task that turned out to be too difficult.

⁵ For theoretical contributions see, e.g., Strotz (1955), Thaler and Shefrin (1981), Ainslie and Haslam (1992), Laibson (1997), O'Donoghue and Rabin (1999), Gul and Pesendorfer (2001), Bénabou and Pycia (2002), and Fudenberg and Levine (2006). Recent experimental contributions on time preferences include Ariely and Wertenbroch (2002), Andersen et al. (2008), Benhabib et al. (2010) and Andreoni and Sprenger (2012). Our paper studies perseverance rather than self-control.

⁶ See, e.g. Hoxby (2000), Hoxby and Weingarth (2005), Zimmerman (2003), Falk and Ichino (2006), Mas and Moretti (2009), Calvo-Armengol et al. (2009) who combine theory and empirics to investigate peer effects in an educational setting, and Rosaz et al. (2016) who investigate being observed and simultaneously observing a peer in the lab. Like our paper, Rosaz et al. (2016) implement a cognitive task and allow for communication. However, their focus is not on separating the effects of being observed and of observing a peer. Moreover, communication plays a different role in their setting (reducing social distance between peers).

⁷ This can also be a conscious strategy of individuals with a lack of self-control. Considering peers as commitment devices is an interesting variation of the general idea that commitment devices can help overcome self-control problems. See Ariely and Wertenbroch (2002), Houser et al. (2018), and Burger et al. (2011).

⁸ To be precise, in BBT individuals have imperfect knowledge of the costs (high or low) they accrue when resisting the urges to give up. In addition, they have limited self-control in the sense of being time-inconsistent which they are perfectly aware of. Hence, what they are only partially aware of is the *extent* of their self-control problem. This interpretation provided by BBT of the underlying ability to persevere – having a smaller or larger self-control problem – could easily be generalized to a composite ability, including self-control, but also physical and mental strength (i.e., having resources that are not easily depleted).

on whether to persevere on the task or to give up. At each decision node, the individual faces a trade-off between the immediate (and, due to time-inconsistency, overvalued) gratification of giving up, on the one hand, and the expected benefit of persevering until the next decision node is reached, on the other hand. Payoffs are highest if one perseveres until the end. Consider now an individual who faces her first decision between persevering and giving up. If this individual believes that she is likely to give up anyhow *later on*, then her expected payoff from persevering *now* is low, compared to an individual who believes her future self to persevere until the end. Hence, her incentive to persevere at any of the decision nodes except the last one crucially depends on her self-confidence, i.e., her confidence in her underlying ability to persevere until the end. This confidence, however, is affected by the behavior she observes in her peers: Individuals know that levels of the underlying ability to persevere are positively correlated within their peer group. Hence, observing how peers react to temptations provides individuals with additional information about their own level of ability.⁹ Intuitively, observing how peers can handle similar challenges to their ability to persevere can be encouraging or discouraging (“if he can do it, then so can I” or “if not even he can do it, then I do not even have to try”). Importantly, BBT find that these peer effects are frequent and strong only if peers are *similar* to each other in the sense that their prior probability to persevere until the end is of comparable size.¹⁰

Hence, the theory of BBT is complementary to the vast majority of the literature on peer effects which assumes that being observed by – rather than observing – one’s peers affects own performance.

1.3. Our contribution

In a large lab experiment with overlapping generations, we address the question whether observing or being observed affects own perseverance and performance. Hence, we test the major claims that BBT and the empirical literature on peer effects make about causality against each other. For this purpose, we match subjects who face a challenge to perseverance with subjects who faced the same challenge before and whose success or failure is observed by the former. We let the matched pairs talk to each other in bilateral free text chats about the challenge. Some of the subjects who chat *after* the task anticipate being observed by and chatting with the peer, and some do not. The former setting is a treatment targeted to reputation effects in which *being observed* by peers may levy an additional cost on giving up. Subjects who chat *before* the task can be encouraged or discouraged to persevere later on by *observing* an experienced peer. We find that knowing in advance that one has to justify one’s own performance toward one’s peers tends to prevent giving up immediately, although it does not necessarily have long-lasting effects on perseverance. Hence, there seems to be a reputational effect as stipulated by the empirical literature on peer effects. However, we also find effects on those unpracticed peers who observe how their advanced peers explain their own success. More specifically, both motivating and de-motivating messages from the advanced peer affect own perseverance if and only if the peer had been successful in her own task. A peer’s success or failure in itself, unmediated by communication, has no visible effect on perseverance in our data. Our findings are consistent with BBT and the empirical test of their theory in Battaglini et al. (2017).¹¹

In our lab experiment, communication is the channel through which information that is assumed to be public in BBT is made available. Communication addresses in particular information on how, according to the experienced peer’s belief, the subjective difficulty of the task at hand – or, alternatively, the ability to persevere on it – is distributed in the population. Hence, our lab experiment combines the approach of BBT with a communication channel similar to the one discussed in Bénabou and Tirole (2002, 2003) and Mechtenberg (2009). Subjects in our experiment can use communication to form beliefs about how similar they and their peer might be. Based on these beliefs and the observed behavior of their peer, they can form new beliefs about how able they themselves will be to persevere until the end. The presence of communication channels is also the main feature of our experimental design that essentially distinguishes our lab experiment from the one conducted by Gerhards and Gravert (2018). As opposed to them, we do not find that observing one’s peers – without the

⁹ A peer giving up hence generates “bad news” about the individual’s own ability to persevere until the end and thus reduces her incentive to persevere for one more time interval at the current decision node. Conversely, a peer who perseveres generates “good news” about the individual’s own ability to persevere until the end and thus enlarges her incentive to persevere once more. In different models, Bénabou and Tirole (2002, 2003) and Mechtenberg (2009) theoretically study effects of communication on self-confidence of subjects with incomplete self-knowledge. These studies find both positive and negative effects on self-confidence, i.e. learning from a peer can be both encouraging and discouraging.

¹⁰ To illustrate why dissimilar peers have weak effects, consider yourself and a peer who is much less likely to persevere until the end than you are. Then, the most frequent news that you can get from him are negative; i.e., you most often observe him giving up. However, these bad news have only a weak effect on your self-confidence since you know that your peer is not very informative about how able *you* are to persevere. Conversely, imagine that you observe a peer who is much more likely than you to persevere until the end. The news you will learn from him are most likely that he persevered. But again, this does not tell you much about how able you are to persevere, so you will not react strongly to these good news. A similar peer, however, is informative about your ability, and you will react strongly to the news he generates. That only similar peers affect individual beliefs and behavior resonates with a finding in the empirical literature on role model interventions. For instance, Porter and Serra (2017) show that female role models have a positive effect on the ambition of female but not male students; and Beaman et al. (2012) find that female leaders in India stimulated girls’ aspirations and enhanced their educational attainment via a pure role-model effect.

¹¹ Another empirical paper on peer effects and perseverance that explicitly relates to BBT is Patacchini and Arduini (2016). They find that only time-inconsistent individuals (so-called boomerang kids who move back to their parents after having moved out first) significantly react to decisions of their peers. Their paper confirms that peer effects are of specific importance in contexts of imperfect self-control, which is consistent with BBT and suggests that in our experiment, subjects have time-inconsistent preferences as well. However, we have not elicited an independent measure of time-inconsistency of our subjects.

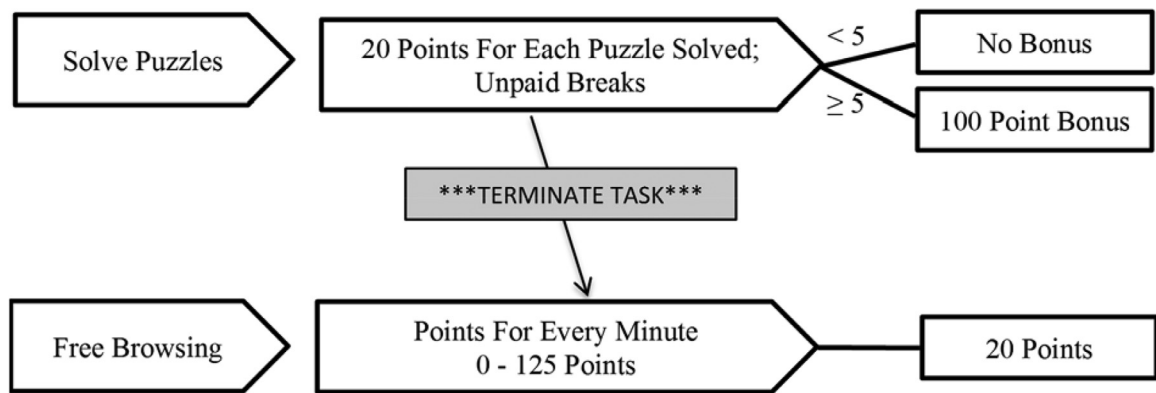


Fig. 1. Overview of payoff structure.

mediation of communication – has any effect on perseverance; in our lab experiment, it is the interaction between listening to communication and observing success that is effective.

Moreover, we find an interesting effect heterogeneity between men and women. In our data, it is the women who are more responsive to both types of peer effects. They increase perseverance substantially when observed by a peer and they increase perseverance substantially when having observed an encouraging peer.

2. Experimental design

We first describe the task environment and then turn to the different treatments and the overlapping-generations structure.

2.1. The task environment

To measure perseverance, we implemented the combination of a real-effort task with access to the internet (see, e.g., [Corngnet et al., 2015](#)). In order to create the necessary variation in perseverance, we chose a real-effort task that is (1) depleting, to require an effort to persevere, (2) meaningless, to reduce average intrinsic motivation, and (3) novel to our subject pool, to exclude experience from previous real-effort experiments. Subjects had to identify, among a number of meaningless strings of letters (like abCOAcAaZAgbCBZ) any such strings that contain the letter combination “abc” (in all combinations of small and capital letters) if that is followed by a vowel. (For such tasks and their depleting effect, see [Baumeister et al., 1998](#), and many others as surveyed in [Hagger, Wood, & Stiff, 2010](#).)

The task was divided into ten subtasks, each consisting of 99 strings displayed on the computer screen, with five input fields for the five strings that had to be identified. A subtask was counted as correctly solved if at least four of the five strings were identified correctly. The entire task was counted as correctly solved if five out of the ten subtasks were solved. We chose these standards (after piloting the real-effort task) to allow approximately one third of our subjects to solve the entire task correctly. For the correctly solved task, a bonus of 100 points was paid. Additional points were paid for each subtask that was correctly solved. The difficulty of the task was increasing in that more and more letters a, b, and c appeared in the strings, while still only five strings had to be identified. The points awarded for correct subtasks were fixed to 20. Subjects had 50 minutes to solve the task, but they could also give up on it whenever they wished. Those who wished to give up could click the “terminate task” button always visible on the task screen and spend the remainder of their time browsing the internet.

Browsing the internet is a common diversion at the work-place and hence a realistic obstacle to perseverance. However, it is to be expected that the laboratory setting activates the social norm to exclusively concentrate on the experimental task. Moreover, access to the internet is typically blocked in lab experiments to enhance control. Therefore, granting access to the internet may well create a demand effect; subjects might deduce from the design that browsing the internet would be encoded as lack of perseverance. Thus, simply granting access to the internet might not be sufficient to induce any usage of it. To prevent the demand effect and to overcome the social norm of “no surfing in the lab”, we paid a small bonus of 20 points for terminating the real-effort task and paid “surfing points” for browsing the internet afterwards, thereby providing a possible rationale for giving up the task. Browsing the internet in minute t yielded $t/10$ points to someone who terminated the task. For instance, if a subject clicked the “terminate-task” button in minute 12, she earned 1.2 points in minute 12; 1.3 points in minute 13, and so on. Hence, who only browsed the internet and did not even start working on the task received $125 + 20$ points in total. Generally, clicking the “terminate task” button at time T yielded $125 - \frac{1}{20}T^2 + 20$ points in total. 100 points correspond to 5 Euros. The payoff structure of the task environment is depicted in [Fig. 1](#). The respective amount of points possible to gain for continuing the task and for terminating it and surfing for the rest of the time was continuously announced on the experimental screen in z-tree.

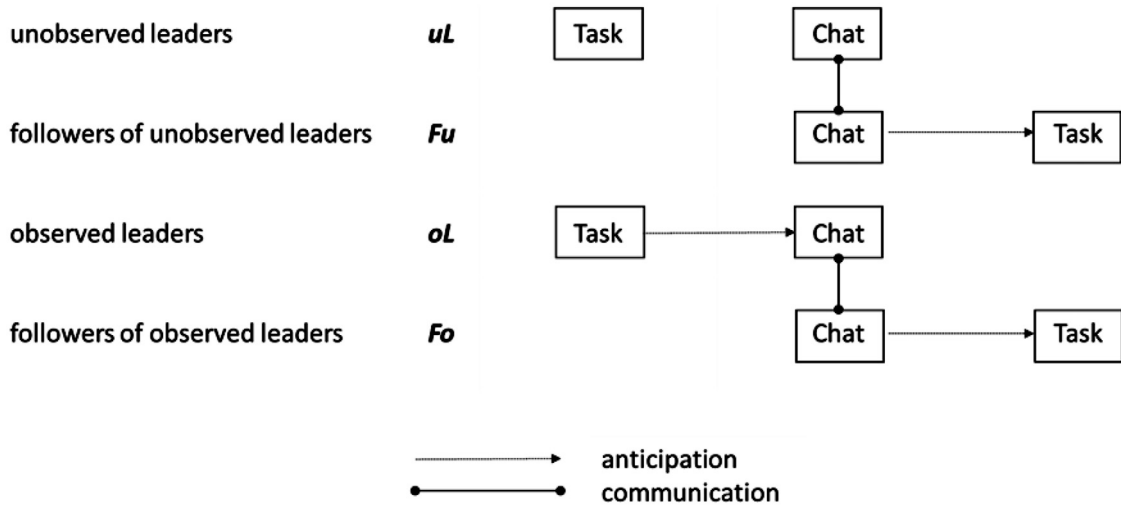


Fig. 2. Treatments: unobserved leaders (uL), followers of unobserved leaders (Fu), observed leaders (oL), and followers of observed leaders (Fo).

The payoff structure is such that the following holds in virtually any case: If one solves at least five of the ten subtasks correctly, i.e., if one is eligible for the task bonus, then it pays off to continue on the task until the end, that is, to never click the “terminate task” button. If one solves less than five of the ten puzzles correctly, then it pays off to quit the task, that is, to click the “terminate task” button. In this case, the best choice is to terminate right after the beginning of the task or right after solving at most one subtask. We did *not* communicate during the task whether the subject had reached the threshold for the bonus (this would strongly determine performance). Also, we took great care to explain the payoff structure and the choice between continuing and terminating the task in the instructions, using Fig. 1. Subjects were given an example of a subtask in the instructions for them to assess their own confidence in their performance. Also, in the course of their work at the task we elicited their beliefs about whether or not they solved the last subtask correctly. Hence, we could validate that their termination decision was optimal, given their beliefs, which is true for more than 90% of the participants. (More precisely, of those who believed that they had solved five subtasks or more correctly, no one gave up. Of the remaining participants, most (84%) gave up at some point.)

To familiarize subjects with the website that they could browse after task termination, we explicitly stated in the instructions that taking occasional breaks from the task and surfing during the breaks are allowed. However, we did not pay any points for browsing the internet during a break. Breaks could be scheduled freely by clicking the break button and switching to a specific website created by us for a deliberate amount of time.¹² Moreover, prior to each subtask, subjects had to go on a website to search a password with which they could unlock the next subtask in z-tree. The website offered a broad variety of pictures and articles in many categories such as politics, sports, diets, celebrities, cartoons, and news; and during the unlocking of each subtask, subjects were made aware of the possible diversions offered on the website.

2.2. Treatments

We conducted four different treatments in a between-subjects design, two *leader* treatments and two *follower* treatments. We now use the “leader” and “follower” terminology to easily refer to a more experienced peer versus an unexperienced peer. Subjects were invited in subsequent cohorts, with a leader cohort preceding a follower cohort. A cohort comprised twelve subjects, half as many as were cubicles in the lab. All treatments had two parts, a *task part* and a *chat part*. The task part was spent in the task environment as described above. In the chat part, subjects in a follower treatment were randomly matched one-to-one with the subjects of the preceding leader treatment and could chat with their matched partner for three minutes. The treatment structure is depicted in Fig. 2.

We now describe the chat part. A follower matched with a leader from the precedent cohort in the first (chat) part of a follower treatment learned the failure or success and the give-up decision of their matched partner, i.e., whether or not their matched partner had earned the bonus of 100 points and whether or not she had terminated the task. This was communicated to the followers on their computer screen and remained visible there throughout the next (task) part. In addition, followers could communicate with the matched leader via a chat window during the chat part. The chat opportunity lasted for three minutes. Chat format was free text; there were no prohibitions, except that anonymity must be kept. We asked the leaders from the preceding cohort to address the following three questions: 1) *Please estimate the fraction of participants in percent that manage the task, i.e., who reached the bonus of 100 points.* 2) *Does it pay off to strive for the bonus or should*

¹² Note that breaks need not signify lack of perseverance but might be part of an optimal working schedule since the task is depleting.

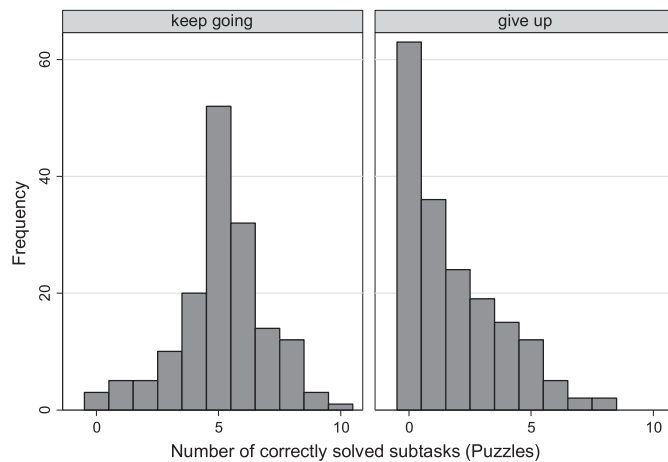


Fig. 3. Histogram of Puzzles by giving-up decision.

the task be quit immediately? 3) Please, explain your recommendation. Question 1, in particular, was asked to measure the similarity that a leader assumed between herself and the average person, regarding performance on the task – a variable of central importance in BBT. A successful (unsuccessful) leader who believes more than half of the other subjects to be successful (unsuccessful) as well deems herself rather *similar* to the average person; a successful (unsuccessful) leader who believes that no other is successful (unsuccessful) as well deems herself *exceptional*. We assume that a follower – who has no experience with the task – tends to perceive himself as an average person with regard to the task environment. Thus, the answer that his matched leader gives to question 1 provides him with information about how similar he himself is to his matched leader. Since BBT implies that only similar peers have effects on subjects' perseverance, we expect the answers given by the leaders to question 1 to set the conditions for any effects of their success or failure on the perseverance of the matched followers.

We can now define the treatments. In the treatment “observed leaders” (short: *oL*), subjects are leaders: They first do the task part and then the chat part. During the task part they know the content of the second part, i.e., that they will be matched with followers (after carrying out the task), that these followers will learn their success or failure and their perseverance, and that they will chat for three minutes. By contrast, in the other leader treatment “unobserved leaders” (short: *uL*), they do not anticipate the details of the chat part since instructions for that are only given after completion of the task part. Hence, subjects do not anticipate that they will be matched with followers of the subsequent cohort and that information about their own perseverance and performance will be provided to their matched partners.

In the “followers of observed leaders” treatment (short: *Fo*) subjects are followers: They first do the chat part and then the task part. They are matched with leaders from *oL* (observed leaders) in the chat part. In the other follower treatment “followers of unobserved leaders” (short: *Fu*) subjects are matched with leaders from *uL* (unobserved leaders) during the chat part.

2.3. Implementation

In total, we have $N=336$ participants which amounts to 84 subjects per treatment.¹³ Cohorts are overlapping: always two consecutive cohorts were chatting with each other after the first of the two had finished the task and before the second started with it. Hence, for logistic reasons we conducted multiple sequences of sessions per day. We distributed these sequences of sessions across different days and made sure that treatments do not differ much in the days and time slots at which they were run. The experiment was run at the WISO-laboratory of the University of Hamburg. A different room outside the computer lab was reserved for reading of instructions and a subsequent quiz to guarantee understanding of the decision and payoff structure.¹⁴ Another room outside the lab was reserved for payment (in cash and in private). Communication other than required in the chat part of the experiment was prohibited, and compliance with this rule was monitored by the staff of the lab throughout the entire experiment. In the time span between the end of the task and leaving (in treatments *Fo* and *Fu*) or chatting with the subsequent cohort (in treatments *oL* and *uL*), subjects answered a questionnaire. This questionnaire elicited age and sex of the subject and other control variables. The chat data were analyzed following the procedure in Brandts and Cooper (2007). A research assistant unaware of the purpose of the study and the different treatments did the transcription and coding. The reliability of the coding was checked and confirmed by having

¹³ There was only one case of a system failure within a regular session, which reduces the number of observations in *Fo* to 83.

¹⁴ The instructions and the quiz are provided as an online appendix.

another, equally unaware research assistant, independently coding the same chat protocols. The coding of the three questions that had to be addressed in the chat into answer categories was straight-forward.

3. Results

Before we turn to the effects of observing peers and to the effects of being observed by peers, we first provide some summary statistics for the main variables.

3.1. Some summary statistics

The main outcome variables are the number of subtasks solved correctly (*Puzzles*) and the dummy whether the bonus is reached or not (*Bonus*). Table A.1 in the Appendix reports summary statistics of these main outcome variables and also reports the payoff earned in EUR (*Payoff*) and the minutes participants stayed in the task before clicking the “terminate task” button (*MinutesTask*).¹⁵ The average number of correctly solved puzzles is below four, but still roughly every third participant (34%) reached the bonus of 100 points, which requires solving five or more. There was indeed one person who solved 10 puzzles correctly, each time believed that the puzzle is solved correctly, and received the bonus of 100 points. (This leads to the maximal number of points of $10 \cdot 20 + 10 \cdot 1 + 100 = 310$, which yields the maximal payoff of EUR $15.50 + 8.00 = 23.50$.) The average payoff was around 16 EUR.

In Table A.1 there are separate outcome statistics for those participants who gave up, i.e., clicked the “terminate task” button at some point, and those who stayed in the task, reported in the last two columns. Separating these two groups, we observe that the bonus was reached by 73% of those who stayed in the task. The histogram below shows that those who stayed in the task solved typically five or six puzzles correctly. Very few managed to solve nine or ten tasks. Most of those who gave up solved none or less than three tasks. Table A.2 reports the timing of giving up, i.e., the minutes before clicking the “terminate task” button, separated by the number of puzzles solved. Those who solved three or four puzzles gave up after 44 of 50 minutes, which indicates that they strived for the bonus until they realized that they would not reach it. Participants who solved two puzzles or less gave up earlier.

Since each participant was asked after each subtask whether she believed that she had solved it correctly, we can analyze the participants' beliefs about the number of tasks they had solved correctly. The correlation of this belief with the actual amount solved correctly (*Puzzles*) is 0.88. Looking at the difference between these two variables, it turns out that a majority of participants, 72%, has exactly accurate beliefs, while 24% overestimate the amount, and 4% underestimate it. Hence, beliefs are quite accurate, but there is a bias consistent with overconfidence. Most overconfident participants overestimate the amount solved by one or two.

We now first analyze the effects of observing peers and communication, i.e., we focus on the follower treatments. Afterwards, we turn to the effects of being observed, i.e., we focus on the observed-leader (*oL*) treatment.

3.2. Effects of observing peers and communication

Consider the follower treatments, *Fo* and *Fu*. From the perspective of the followers, no treatment variable changes between these two treatments. Hence, when analyzing the followers' reaction to what they observe from their matched leaders, we pool the data from *Fo* and *Fu*.

The main aspect observed by a follower is whether her matched leader did or did not receive a bonus (captured by the dummy variable *BonusPeer*). This piece of information is highlighted during the chat and on each subtask in the follower treatments. Moreover, in the chat the leaders are asked to provide a guess about the fraction of participants that was able to receive the bonus. The answers range from 0% to 100%, and we call this continuous variable *FractionSuccess*. As the histograms in Fig. 4 show, leaders who do not receive the bonus more often guess that the number of participants receiving the bonus is small, but there is still a considerable variation across leaders such that our subjects in the follower treatments get very different signals about the difficulty of the task and their personal likelihood of being able to receive the bonus.

FractionSuccess provides information on the leader's belief about how likely it is that the follower can succeed. Combined with the leader's success or failure, *FractionSuccess* provides information on the peer's belief on how similar she is to the general population and hence to the follower. If it follows from *FractionSuccess* that most will succeed or fail like the peer, this signals high similarity. Moreover, in the chat leaders gave explicit recommendations whether to give up or to strive for the bonus, and we elicited how difficult the task is described as. These two variables *Recommendation* and *Difficulty* are highly correlated with the variable *FractionSuccess*, but coarser measures since they are binary, respectively ordinal.

3.2.1. Basic regressions

In the following, we report the results of regressing the main outcome variables (*Puzzles*, *Bonus*) and *Payoff* on *BonusPeer* and *FractionSuccess*, which are the two most important variables of observing and communicating with a peer. They cover the objective information provided by the experimenter that the peer (leader) has or has not received the bonus, as well as

¹⁵ If the “terminate task” button was not clicked, *MinutesTask* was 50 which was measured by the computer as 49.98.

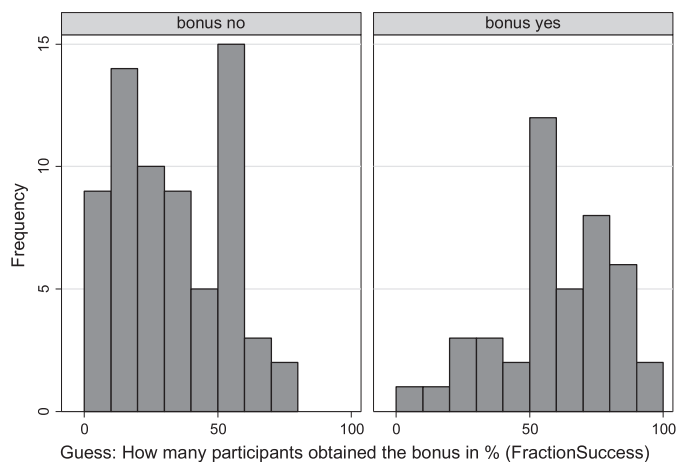


Fig. 4. Histogram of *FractionSuccess*, i.e., the leader's estimate concerning how many participants manage to reach the bonus, differentiated by whether the leader has received the bonus herself (right panel) or not (left panel).

Table 1
Basic regressions.

	(1) Puzzles Poisson	<i>p-values</i>	(2) Bonus Logit	<i>p-values</i>	(3) Payoff OLS	<i>p-values</i>
BonusPeer	−0.892* (0.476)	<i>0.061</i>	−2.956** (1.358)	<i>0.030</i>	−3.586** (1.455)	<i>0.015</i>
FractionSuccess	−0.006 (0.005)	<i>0.230</i>	−0.018 (0.013)	<i>0.177</i>	−0.025 (0.018)	<i>0.155</i>
FractionSuccess x BonusPeer	0.018** (0.008)	<i>0.020</i>	0.064*** (0.024)	<i>0.008</i>	0.075*** (0.027)	<i>0.007</i>
Constant	1.294*** (0.157)	<i>0.000</i>	−0.444 (0.443)	<i>0.316</i>	16.367*** (0.657)	<i>0.000</i>
N	110		110		110	
χ^2	6.282		7.566			
P	0.099		0.056		0.060	
Pseudo R ²	0.021		0.065			
F					2.543	
R ²					0.058	

Notes: Dependent variables are *Puzzles*, i.e., the number of subtasks solved correctly, in model (1), *Bonus*, i.e., whether the bonus has been reached or not, in model (2), and *Payoff*, i.e., the experimental payout in EUR, in model (3). Variable *FractionSuccess* is the fraction of participants who receive the bonus as estimated by the matched peer of the observed subject. The variable *BonusPeer* is equal to 1 if the matched peer has received the bonus and zero otherwise. “*FractionSuccess* x *BonusPeer*” stands for the interaction term between *FractionSuccess* and *BonusPeer*. Robust standard errors are reported in parentheses, the corresponding *p-values* are reported in italics. *, **, *** indicate significance at the 10%, 5% and 1% level.

the subjective message of how hard it is to reach the bonus. Since it must be expected that these two variables interact with each other, we also include the interaction term. The outcome variable *Puzzles*, i.e., the number of subtasks solved correctly, is a count variable which we estimate by a Poisson regression. The outcome variable *Bonus*, i.e., whether the bonus has been reached or not, is a binary variable, for which we use a logistic regression. Table 1 presents the three basic regressions.

We first observe that the coefficients have the same sign in each of the three models and that significance levels are similar across the three models. Thus, the results seem to be robust to the definition of the outcome variable. Strikingly, there is a highly significant interaction effect between *BonusPeer* and *FractionSuccess*, which means that the effect of observing a partner who has reached the bonus (*BonusPeer*) depends on what this partner communicates concerning how many are able to reach the bonus (*FractionSuccess*).

To investigate this interaction effect and to test for significance of the difference between having a peer with bonus and a peer without bonus, we run a sequence of regressions each time modifying the variable *FractionSuccess* by subtracting a constant before it is used in the interaction term. More precisely, we generate the variables *FractionSuccess0*, *FractionSuccess10*, *FractionSuccess20*, ..., *FractionSuccess100*, with *FractionSuccess_k* defined as *FractionSuccess* minus *k*; and use them to

Table 2
Effect of *BonusPeer* at different levels of *FractionSuccess*.

	FS0	FS10	FS20	FS30	FS40	FS50	FS60	FS70	FS80	FS90	FS100
Puzzles											
Poisson	−0.892*	−0.708*	−0.524	−0.340	−0.156	0.028	0.212	0.396	0.580*	0.764**	0.948**
<i>p-values</i>	0.061	0.083	0.128	0.239	0.526	0.901	0.356	0.127	0.060	0.037	0.028
IRR	0.410***	0.493***	0.592**	0.712*	0.856	1.028	1.236	1.486**	1.786**	2.147***	2.581***
Bonus											
Logit	−2.956**	−2.318**	−1.680*	−1.042	−0.404	0.234	0.872	1.509**	2.147**	2.785**	3.423***
<i>p-values</i>	0.030	0.043	0.076	0.177	0.529	0.691	0.165	0.043	0.019	0.012	0.010
odds ratios	0.052**	0.098**	0.186*	0.353	0.668	1.263	2.391	4.524*	8.561**	16.202**	30.661**
Payoff											
OLS	−3.586**	−2.841**	−2.095**	−1.350	−0.605	0.140	0.885	1.631*	2.376**	3.121**	3.866**
<i>p-values</i>	0.015	0.022	0.041	0.109	0.405	0.843	0.265	0.088	0.041	0.025	0.019

Notes: Dependent variables are *Puzzles*, i.e. the number of subtasks solved correctly, in the first block, *Bonus* in the second block, and *Payoff*, i.e. the experimental payout in EUR, in the third block. The columns stand for the effect of the bonus of the matched peer when her estimated fraction of subjects who reach the bonus is 0, 10, ..., 100. As detailed in the main text, we obtain these effects by using the variable *FractionSuccess_k*, which is defined as *FractionSuccess* minus constant *k*, and considering the coefficient for *BonusPeer*. The first block reports the coefficients and *p-values* of the Poisson regressions that correspond to model (1) of Table 1 along with the incident rate ratios (IRR) that correspond to model (1') of Table A.3 in the Appendix. The second block reports the coefficients and the *p-values* of the logit regressions that correspond to model (2) of Table 1 along with the odds ratios that correspond to model (2') in Table A.3. The third block reports the coefficients and the *p-values* of the ordinary least squares (OLS) regressions that correspond to model (3) in Table 1. The *p-values* are based on robust standard errors. *, **, *** indicate significance at the 10%, 5% and 1% level.

create the interaction term by multiplying them with *BonusPeer*. The results of these regressions are all identical apart from the coefficient of *BonusPeer*. Consider the interaction term, which is the product of the binary variable *BonusPeer* and the continuous variable *FractionSuccess_k*. When the variable *FractionSuccess_k* is zero, the interaction term is always zero such that the coefficient of *BonusPeer* picks up the whole effect. We use this fact to provide an interpretation for the effects. All coefficients of *BonusPeer*, which are the only coefficients that differ from the ones in Table 1, are reported in Table 2. Hence, the entries in column FS0 of Table 2 correspond to the first row (*BonusPeer*) of model (1) in Table 1.

We observe that in all models coefficients are significant for low values and for high values of *FractionSuccess* but not in between. Moreover, the sign of the effects changes from low to high values. We provide an interpretation for each outcome variable, starting with *Payoff* since OLS coefficients can be directly interpreted.

3.2.2. Effects on payoff

At the upper end of the spectrum, OLS coefficients are significant for *FractionSuccess* in the range from 70 to 100. For example, the coefficient 2.376 (see column 'FS80') is the increase in *Payoff* when the matched leader received the bonus, i.e., was successful, while holding the message *FractionSuccess* at the level of 80%. Those who receive the news that their matched leader has obtained the bonus are hence expected to have a payoff that is EUR 2.38 higher than that of subjects that receive the opposite news, while the average payoff is EUR 15.73. This is the good-news effect; it seems to transport the message "I could do it and I am similar to the average person, hence so can you". It seems encouraging to hear that many of the subjects are able to receive the bonus from a person who did it as well.

At the lower end of the distribution, OLS coefficients are significantly negative for *FractionSuccess* in the range from zero to around thirty. For instance, if the peer received a bonus and reported that only 20% were successful (i.e., at *FractionSuccess*20) *Payoff* decreases by EUR 2.01 (see column 'FS20'). Thus, a low *FractionSuccess* is bad news, in particular if the matched leader has received the bonus; it seems to transport the message "I did it but it was so hard, you should not try it". Low *FractionSuccess* of a matched leader who did not receive the bonus is in principle bad news as well, but much less so. The statement that only few reached a bonus does not seem highly credible if it comes from a person who did not manage to receive the bonus herself. Maybe subjects believe that their unsuccessful peer did not even try and only wants to justify her own behavior.

3.2.3. Effects on Puzzles and Bonus

Turning to the outcome variables *Puzzles* and *Bonus*, we also observe that coefficients are significantly positive for high values of *FractionSuccess* and significantly negative for low values. The interpretation is analogous to the interpretation for the outcome variable *Payoff* that we estimated by OLS. Since pure coefficients of both the Poisson model and the Logit model cannot be interpreted as easily, we first transform the coefficients and we finally illustrate the results by plotting the predicted outcome.

To quantify the effects on *Puzzles* we use the incidence rate ratio (IRR) of the Logit model.¹⁶ An IRR below one indicates a negative effect, an IRR above one indicates a positive effect. As Table 2 reports, the IRR coefficients at the upper end show

¹⁶ The basic Poisson regression with the IRR option is reported as model (1') in Table A.3 in the Appendix. The significance levels slightly differ when comparing the robust and the IRR - option.

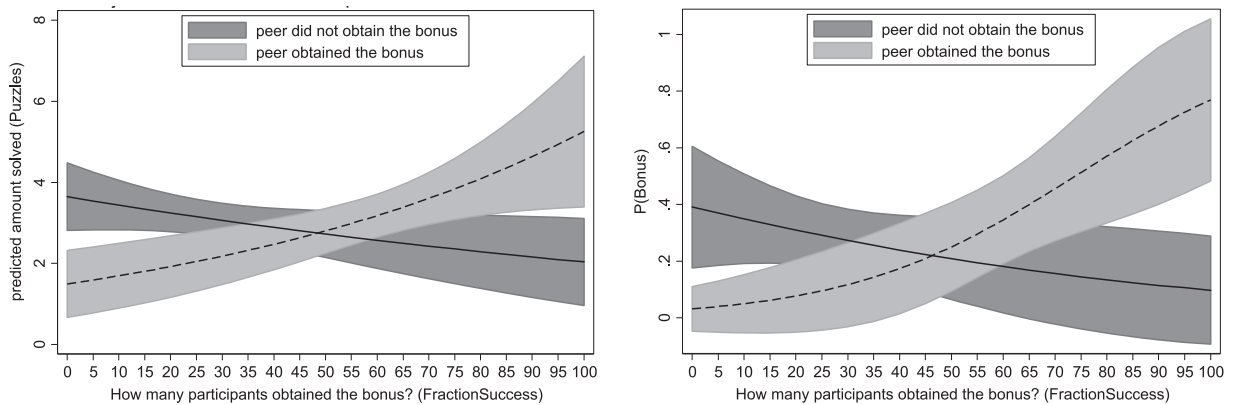


Fig. 5. Marginsplots of *Puzzles* and *Bonus* by *FractionSuccess* and *BonusPeer*.

positive effects. For instance, holding *FractionSuccess* at a high level of 80%, i.e., when observing peers who state that 80% have received the bonus, the IRR coefficient is 1.786 (column 'FS80' of Table 2): Those who receive the news that their matched leader has obtained the bonus are expected to have an incidence rate for *Puzzles* 1.786 times that of subjects that receive the opposite news, which corresponds to an increase of almost 79%. This is again the good-news effect. At the lower end of the distribution, IRR coefficients are below one indicating the discouraging effect of observing a successful leader who communicates that only few are able to receive the bonus.

To quantify the effects on the outcome variable *Bonus* we use the log odds from the logistic regression.¹⁷ As reported in Table 2, we again have the positive effect for high values of *FractionSuccess* and the negative effect for low values of it. For instance, if the peer received a bonus and reported that only 20% were successful (i.e., at *FractionSuccess*20), the odds ratio of receiving a bonus is only 0,186 to 1. Hence, there is a sizeable effect, not only on *Payoff*, and on the number of puzzles solved, but also on the likelihood to receive a bonus.

We illustrate the results with respect to the outcome variables *Puzzles* and *Bonus* by plotting the predicted outcome values by different levels of *FractionSuccess* for both subjects with matched leaders who received a bonus and subjects with matched leaders who did not. We obtain the predicted number of puzzles and the predicted probability of a bonus for values of *FractionSuccess* that range from 0 to 100 with increments of 5.

Fig. 5 shows how the outcome variables are increasing in *FractionSuccess* for a matched leader who received a bonus and rather decreasing in *FractionSuccess* for a matched leader who did not. The graphs illustrate that the best performance is predicted for those who observe that their matched leader obtained the bonus and who receive the message that a very high fraction is believed to be able to reach the bonus. Lowest performance is obtained if the matched leader received a bonus but reports her belief that only very few are able to do so. In essence, it seems that a message from a peer who has been successful can be both particularly encouraging and particularly discouraging.

3.2.4. Robustness

These results remain valid when we include the socio-demographic variables *Gender* and *Age* (linear and squared) that we collected. As another test of robustness, we reran the regressions including all variables from the questionnaire that survived the stepwise reduction procedures. The magnitude and significance levels are virtually the same.¹⁸

As another test of robustness, we exchange the variable for the message *FractionSuccess* with the binary variable *Recommendation*, which is one if the matched leader recommends staying in the task and zero if she recommends giving up at some point.¹⁹ Table A.4 in the Appendix reports the average number of puzzles solved and the percentage of subjects who received the bonus in dependence of whether the matched leader has recommended giving up or staying in the task. The table gives an indication that the recommendation to stay has generally a positive effect on the outcome. This strongly indicates that it is indeed perseverance which is mainly required to succeed in the task. It also confirms the strong interaction between the content of the chat and the success of the peer. Recommendations of peers who received the bonus heavily affect outcome variables. In particular, the most encouraging leaders are those who have received the bonus and recommend striving for the bonus. Recommendation to stay in the task leads to significant increase of *Puzzles* (Mann Whitney U test, $p < 0.05$) and a significant increase of the fraction of bonus reached (Fisher exact test, $p < 0.05$), given that the matched leader has obtained the bonus.

¹⁷ The basic Logit regression with the odds ratios option is reported as model (2') in Table A.3 in the Appendix. The significance levels are highly similar when comparing the robust and the odds ratios - option.

¹⁸ The magnitudes are even a bit higher and the p -values are even a bit smaller when additional variables are included. These results can be requested from the authors.

¹⁹ The coding of the chat for the recommendation entails three categories: "stay in the task", "work first on the task and quit later", and "quit immediately". The binary variable *Recommendation* simply distinguishes between the first and the two latter categories.

Summarizing we find both a good-news effect and a bad-news effect. Thereby, the signal about the peer's performance alone is not sufficient since it is put in relation to what the peer communicates. Good news are obtained from a successful peer who communicates in an encouraging way. Bad news are obtained from a successful peer who communicates in discouraging ways.

3.3. Effects of being observed

We now compare the observed-leader treatment (*oL*), in which subjects know that their performance will be shown to a peer and that they are going to chat with this peer, to the other treatments. Table A.5 in the Appendix reports summary statistics of the variable *Puzzles* for *oL* and for all the other treatments.²⁰ Two-sample Mann-Whitney U (MWU) tests reveal: *Puzzles* is significantly higher under *oL* than under both follower treatments ($p < 0.1$), and significantly higher under *oL* than under all other treatments pooled ($p < 0.05$). The effect of being observed on *Puzzles* is not so strong at the mean or the median as it is at the lower end of the distribution. While the null hypothesis is not rejected for *oL* vs. *uL* (the unobserved-leader treatment) when using the MWU test (with $p = 0.264$), it is rejected when considering the fraction of participants who solved at least one, respectively at least two, puzzles, as illustrated in Fig. 7. Fisher exact tests reveal that a fraction of subjects who correctly solved at least one (or at least two) puzzles is significantly higher under *oL* than under each of the three other treatments ($p < 0.05$), as well as higher under *oL* than under all other treatments pooled, of course. Hence, subjects who know that their result will be observed by a peer and that they have to justify themselves in a chat, tend to solve more puzzles on average and less often solve zero or only one puzzle.

It seems that the fact that there will be an observer to whom the subject has to justify her performance motivates the participants to try the puzzles at least for a while. If subjects manage to solve five puzzles this behavior pays off with a bonus of 100 points. Otherwise, this is not so clear because at the same time the subjects could have earned points for surfing in the internet. In particular, subjects who solve two, three or four puzzles can neither accumulate many points in the surfing mode nor receive the bonus of 100 points. Correspondingly, we do not find that the observed-leader treatment *oL* significantly increases the likelihood of receiving the bonus. In the *oL* treatment the fraction is 36% and in the other treatments 33%. Hence, it is not surprising that the difference of the observed-leader treatment *oL* to the other treatments in the number of puzzles solved does not materialize in higher payoffs. The average payoffs are 15.81 for *oL* and 15.70 EUR for all other treatments (these numbers are not repeated in the Appendix).

To summarize, we find that knowing in advance that one has to justify one's own perseverance toward one's peer motivates not to give up immediately. Hence, there seems to be a reputational effect as stipulated by the empirical literature on peer effects. Our overall findings therefore provide evidence both for the direction of causality assumed by the empirical literature on peer effects and the direction of causality suggested by BBT.

3.4. Gender differences in peer effects

We have so far uncovered two basic effects: First, observing a successful peer can have both an encouraging and a discouraging effect on perseverance with respect to the outcome variables. Second, being observed by a peer increases perseverance in the short run, i.e., it increases the fraction of participants who solve at least one or two subtasks. We investigate gender differences in these peer effects. For this purpose, we reran each analysis separately for men and women and compared the coefficients. As before, we start with the follower treatments, i.e., the effects of observing.

Table A.6 in the Appendix reports the basic regression, separately estimated for men and women. We observe that the sign of the crucial coefficients (*BonusPartner* and the interaction effect) is always the same for both sexes. Despite the fact that we only look at subpopulations, which reduces our statistical power, many effects of the baseline model are also significant for women and men separately. The *p*-values are generally lower for women than for men. Moreover, all coefficients are larger in absolute terms for women than for men. Hence, the effects are more pronounced for women. Women respond more strongly to encouraging and discouraging news.

As in subSection 3.2 we also assess robustness of this finding by replacing the variable *FractionSuccess* with the explicit recommendation that followers have received. As reported in Fig. 6, the recommendation to stay in the task increases the average number of puzzles solved from 1.75 to 3.7 for both genders pooled. In the split-sample analysis sample sizes are small. For men the recommendation to stay changes the mean from 3.2 to 3.25, while for women it changes the mean from 0.71 to 4.24 (as reported in Table A.4). Despite the small sample size, for women the effect is highly significant ($p = 0.002$, MWU), while it is not so for men. Similar results are obtained for the outcome variable *Bonus* (cf. Table A.4). This confirms that the effects of encouraging leaders are more pronounced for women than for men, but it does not imply that all these effects are zero for men.²¹ In fact, in Table A.6 we report significant effects in the Logit and the OLS regressions for men as well.

We now turn to the effect heterogeneity with respect to being observed by peers. With Fig. 7 we have illustrated that being observed increases the fraction of participants who solve at least one puzzle or at least two puzzles. Differentiating

²⁰ For the follower treatments, we have seen that there are positive and negative effects, depending on the peer. These effects may well cancel each other out such that no treatment effects may be apparent when comparing the follower treatments with the *uL* treatment.

²¹ For followers (men or women) of unsuccessful leaders, there is no significant difference.

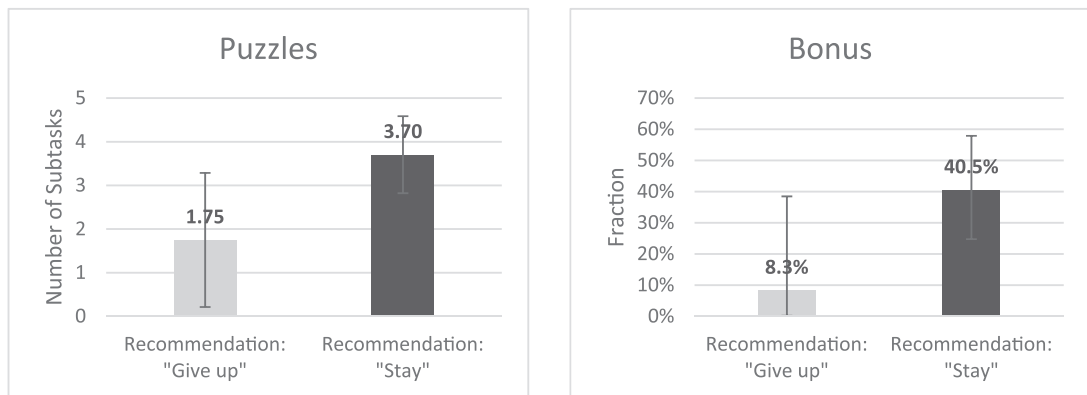


Fig. 6. Recommendation from a peer who has reached the bonus positively affects both the number of correctly solved subtasks (*Puzzles*, left panel, $p < 0.05$) and whether the bonus has been reached (*Bonus*, right panel, $p < 0.05$).

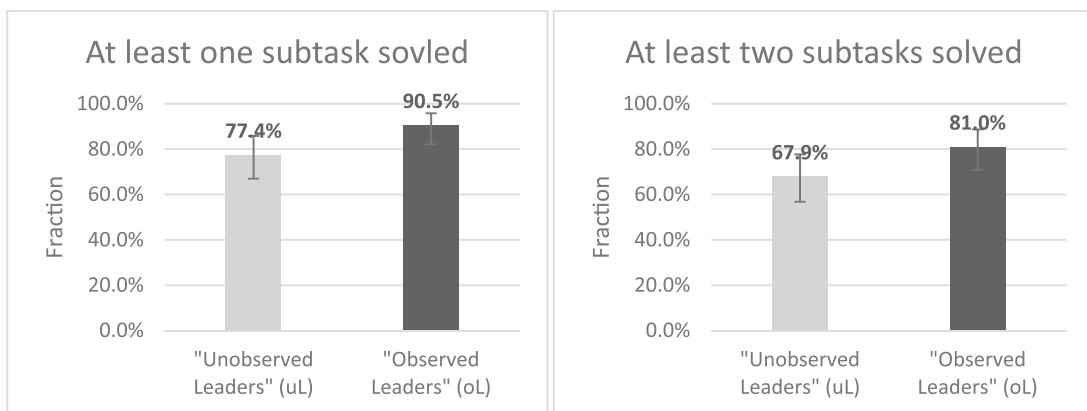


Fig. 7. Fraction who solved at least one subtask, respectively at least two subtasks, is higher in the observed-leader treatment than in the unobserved-leader treatment (Fisher exact tests: $p < 0.05$).

between men and women, we see that it is again the women for whom the peer effects are strong (cf. Table A.5 in the Appendix, last rows). The fraction of women who solve at least one puzzle increases from 73.5% to 91.7% when observed and similarly for two puzzles (Fisher exact test, $p < 0.05$, for each of the two comparisons). In contrast to the pooled analysis in [subSection 3.3](#), we now also find a significant effect on the average number of puzzles solved: Women increase the number from 2.74 to 3.97 when being observed (MWU $p < 0.05$). For men these effects are either weaker or not present since none of them is significant and the average number even slightly reduces.

Summarizing, it is the women who respond more heavily to both types of peer effects, being observed by a peer and observing a peer. This observation is consistent with the literature on educational interventions; see, e.g., [Angrist et al. \(2009\)](#) who find that offering study support services (including a form of mentoring and facilitated study groups) fosters educational performance mainly for females.

4. Discussion and conclusion

In our lab experiment, we investigated the two causal channels through which peers might influence individual perseverance. One channel ("being observed by peers") is stipulated by the empirical literature on peer effects, the other channel ("observing peers") is stipulated in a theoretical model (BBT). Importantly, our findings on the interaction of communication and observed success corroborate the approach of BBT who argue that some assumed similarity between one's own and one's peers' ability to persevere is necessary for one's own perseverance to be affected by the peer's success or failure. In our lab experiment, such beliefs are formed through communication when the advanced peers express their beliefs about how widespread their own success or failure may be among the experimental subjects (and, thus, how similar they believe themselves to be to the average student). Our results confirm the prediction of BBT that a peer's success enhances own perseverance the more, the more likely it is that her type – i.e., her ability to persevere – is similar to one's own: Only successful peers who expressed the belief that more than half of the experimental subjects were successful, too, had

a significant stimulating effect on the perseverance of the subjects matched with them. Intuitively, their success could be interpreted as the message “Since I am an average person, and I did it, you can do it, too”.

However, our experiment also shows that, unpredicted by BBT, subjects can be affected negatively by their peer's success if they are told that their peer believes herself to be the exception rather than the rule. In this case, BBT would predict no effect rather than a negative effect. Psychologically, we believe our finding to be intuitive: The success of an – according to herself – exceptionally perseverant peer may be interpreted as the message “I did it, but since I am an exceptional person, you do not even have to try”. While this demotivating effect of exceptional peers is not prominently discussed in the older empirical literature on peer effects or in the literature on mentoring and role-model interventions, a related finding is reported in newer papers on peer effects in the classroom. Carrell et al. (2013), Booji et al. (2017), and Eyring and Narayanan (2018) report the findings of field experiments that provide strong support for discouraging effects of better-performing peers on low performers.²² This suggests that distance in abilities indeed interacts with information about the peer's performance in a way that is consistent with the findings of our lab experiment.²³

Moreover, our experiment suggests that one important channel for both the encouraging and the discouraging effects is communication among peers (or mentors and mentees). Hence, our study is a first step to bridge the gap between BBT and the literature on how communication affects self-confidence. It confirms the important role that Bénabou and Tirole (2002, 2003) and Mechtenberg (2009) ascribe to the way advanced peers communicate with beginners whose self-confidence is still under formation; and it shows that – and how – one and the same fact (the success of an advanced peer) can be turned into a motivating or a de-motivating message.

Surprisingly, we find that a peer's failure, even if justified by this peer herself as likely to be widespread among the experimental subjects, has no significant effect on own perseverance. At first glance, this seems contrary to BBT, who would predict a negative effect. However, BBT assume *objective public* information about the similarity between levels of the ability to persevere among peers whereas we, in our lab experiment, implement *subjective private information* (through peer-to-peer communication) about such similarities. Hence, it seems that in our lab experiment, only successful peers are taken to be credible in their communication about how similar they believe themselves to be to the average person. Therefore, the results of our lab experiment can be used to test the predictions of BBT about the effects of successful (i.e., highly perseverant) peers on own perseverance but may reveal nothing regarding BBT's predictions about the effects of unsuccessful peers.

We conclude that both channels of causality can be active. Consequently, they may even complement each other and lead to a virtuous circle between the ability to persevere and positive peer effects.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jebo.2018.09.004](https://doi.org/10.1016/j.jebo.2018.09.004).

References

- Ainslie, G., Haslam, N., 1992. Hyperbolic discounting. In: Loewenstein, G., Elster, J. (Eds.), *Choice Over Time*. Russell Sage Foundation, New York, pp. 57–92.
- Angrist, J., Lang, D., Oreopoulos, P., 2009. Incentives and services for college achievement: Evidence from a randomized trial. *Am. Econ. J. Appl.* 1 (1), 136–163.
- Alan, S., Boneva, T., and S. Ertac (2016): Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. HCEO Working Paper.
- Andersen, S., Harrison, G., Lau, M., Rutström, E., 2008. Eliciting risk and time preferences. *Econometrica* 76 (3), 583–618.
- Andreoni, J., Sprenger, C., 2012. Risk preferences are not time preferences. *Am. Econ. Rev.* 102 (7), 3333–3356.
- Ariely, D., Wertenbroch, K., 2002. Procrastination, deadlines, and performance: self-control by pre-commitment. *Psychol. Sci.* 13 (3), 219–224.
- Bandiera, O., Barankay, I., Rasul, I., 2010. Social incentives in the workplace. *Rev. Econ. Stud.* 77 (2), 417–458.
- Battaglini, M., Bénabou, R., Tirole, J., 2005. Self-control in peer groups. *J. Econ. Theory* 123, 105–134.
- Battaglini, M., Díaz, C., Patacchini, E., 2017. Self-control and peer groups: an empirical analysis. *J. Econ. Behav. Org.* 134, 240–254.

²² More specifically, Carrell et al. (2013) report the backfiring of a policy intervention that they based on the older literature (including their own): When they assigned the lowest-ability students at the United States Airforce Academy to mixed groups with the highest-ability students, the performance of the former was significantly reduced, while the absence of both the lowest- and highest-ability students affected the medium-range ability students positively. The highest-ability students were unaffected by the intervention. These results are corroborated by Booji et al. (2017). They find that lowest- and medium-range-ability students in an undergraduate economics program would profit from homogenous rather than mixed groups while the highest-ability students are unaffected by group composition. Relatedly, in a field experiment with students enrolled in a large online course, Eyring and Narayanan (2018) provide their subjects with a reference point against which to assess their own achievement. Varying the reference point from the median to the top-quartile achievement, they observe, when comparing the latter with the former treatment, higher performance of the students between the median and the top-quartile but, at the same time, lower performance for the students below the median.

²³ The absence of similar discouraging effects of exceptional peers from the empirical literature on mentoring might hence be due to the complexity of mentoring relationships in the real world: Mentors do not only give advice or stand as an example but also help their mentees in various ways, passing on human and social capital. For recent empirical papers on mentoring, see, e.g., Blau et al. (2010), Lyle and Smith (2014), and Muschalik and Pull (2016). All find significant positive effects of mentoring on productivity (Blau et al. 2010, Muschalik and Pull 2016) and success (Lyle and Smith 2014). As regards peer effects, all studies that we know of find that better performing peers affect their friends' or colleagues' performance positively, and worse performing peers negatively. For a recent field study with this finding, see Bandiera et al. (2010). They find that peers seem to adjust their effort to the level of their friends' in order to be assigned to work areas in close proximity to their friends. The positive peer effect prevails in their study; i.e., social incentives are productivity enhancing on average. No such social incentives are at work in our experiment where the peer matching is exogenous to tightly control directions of causality.

- Beaman, L., Duflo, E., Pande, R., Topalova, P., 2012. Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science* 335 (6068), 582–586.
- Bénabou, R., Pycia, M., 2002. Dynamic inconsistency and self-control: a planner-doer interpretation. *Econ. Lett.* 77 (3), 419–424.
- Bénabou, R., Tirole, J., 2002. Self-confidence and personal motivation. *Q. J. Econ.* 117 (3), 871–916.
- Bénabou, R., Tirole, J., 2003. Intrinsic and extrinsic motivation. *Rev. Econ. Stud.* 70 (3), 489–520.
- Benhabib, J., Bisin, A., Schotter, A., 2010. Present bias, quasi hyperbolic discounting, and fixed costs. *Games Econ. Behav.* 69 (2), 205–223.
- Blau, F.D., Currie, J.M., Crosron, R.T., Ginther, D.K., 2010. Can mentoring help female assistant professors? Interim results from a randomized trial. *Am. Econ. Rev.* 100 (2), 348–352.
- Brandts, J., Cooper, D.J., 2007. It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *J. Eur. Econ. Assoc.* 5 (6), 1223–1268.
- Booij, A.S., Leuven, E., Oosterbeek, H., 2017. Ability peer effects in university: Evidence from a randomized experiment. *Rev. Econ. Stud.* 84 (2), 547–578.
- Burger, N., Charness, G., Lynham, J., 2011. Field and online experiments on self-control. *J. Econ. Behav. Org.* 77 (3), 393–404.
- Calvo-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. *Rev. Econ. Stud.* 76, 1239–1267.
- Carrell, S.E., Sacerdote, B.I., West, J.E., 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica* 81 (3), 855–882.
- Corgnet, B., Hernán-González, R., Schniter, E., 2015. Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experiment. Econ.* 18 (2), 284–301.
- Eyring, H., Narayanan, V.G., 2018. Performance effects of setting a high reference point for peer-performance comparison. *J. Account. Res.* 56 (2), 581–615.
- Falk, A., Ichino, A., 2006. Clean evidence on peer effects. *J. Labor Econ.* 24 (1), 39–58.
- Fudenberg, D., Levine, D., 2006. A dual-self model of impulse control. *Am. Econ. Rev.* 96 (5), 1449–1476.
- Gerhards, L., Gravert, C., 2018. Because of You I did Not Give Up – How Peers Affect Perseverance. University of Hamburg, Mimeo.
- Gul, F., Pesendorfer, W., 2001. Temptation and self-control. *Econometrica* 69 (6), 1403–1435.
- Houser, D., Schunk, D., Winter, J., Xiao, E., 2018. Temptation and commitment in the laboratory. *Games Econ. Behav.* 107, 329–344.
- Hoxby, C. (2000): Peer effects in the classroom: learning from gender and race variation. NBER Working Paper 7867.
- Hoxby, C., Weingarth, G., 2005. Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Harvard University, Mimeo.
- Laibson, D., 1997. Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112 (2), 443–478.
- Lyle, D.S., Smith, J.Z., 2014. The effect of high-performing mentors on junior officer promotion in the US Army. *J. Labor Econ.* 32 (2), 229–258.
- Mas, A., Moretti, E., 2009. Peers at work. *Am. Econ. Rev.* 99 (1), 112–145.
- Mechtenberg, Lydia, 2009. Cheap talk in the classroom: how biased grading at school explains gender differences in achievements, career choices and wages. *Rev. Econ. Stud.* 76 (4), 1431–1459.
- Muschallik, J., Pull, K., 2016. Mentoring in higher education: does it enhance mentees' research productivity. *Edu. Econ.* 24 (2), 210–223.
- O'Donoghue, T., Rabin, M., 1999. Doing it now or later. *Am. Econ. Rev.* 93 (2), 186–191.
- Patacchini, E., Arduini, T., 2016. Residential choices of young Americans. *J. Hous. Econ.* 34, 69–81.
- Porter, C., and D. Serra (2017): Gender differences in the choice of major: The importance of female role models. Working paper.
- Rosaz, J., Slonim, R., Villeval, M.C., 2016. Quitting and peer effects at work. *Labour Econ.* 39, 55–67.
- Strotz, R., 1955. Myopia and inconsistency in dynamic utility maximization. *Rev. Econ. Stud.* 23 (2), 165–180.
- Thaler, R., Shefrin, H., 1981. An economic theory of self-control. *J. Polit. Econ.* 89 (2), 392–406.
- Zimmerman, D., 2003. Peer effects in academic outcomes: evidence from a natural experiment. *Rev. Econ. Stat.* 85 (1), 9–23.