Salience of Ability Grouping and Biased Belief Formation*

Mira Fischer†

14th November 2017

Abstract

Recent research has found that a higher ordinal rank within one's class affects subsequent skill acquisition positively and has linked this finding to the "big-fish-little-pond-effect", a popular proposition in educational psychology claiming that assignment to a peer group with lower skills increases one's confidence in academic ability. Findings from a lab experiment suggest that salience of the group assignment mechanism matters for how ability grouping affects ability beliefs. If the assignment mechanism is non-salient, it does not matter for subjects' confidence whether they are assigned to the weaker or the stronger group, however, when the group assignment mechanism is salient, weaker group assignment makes people less confident. Subjects are on average less confident when the group assignment mechanism is salient than when it is non-salient. This is found to be the case due to weaker group assignment making people more underconfident than stronger group assignment making people overconfident, indicating that people overweigh negative information as compared to positive information. These findings may help to understand the effects of ability grouping in the field and may inform the design of educational and workplace environments.

Keywords: confidence in abilities; ability grouping; salience of group assignment; motivation

JEL Codes: D83, I21, J24

*Financial support of the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

†University of Cologne, Albertus Magnus Platz, 50923 Köln, Germany, phone: +49 221 470 6311, email: mira.fischer@uni-koeln.de

1 Introduction

In recent years, economists have recognized the importance of ability beliefs and social identity for explaining the motivation of individuals to invest in their human capital and to sort into different career paths (Akerlof and Kranton, 2002; Benabou and Tirole, 2002, 2016; Heckman et al., 2006; Dohmen and Falk, 2010, 2011). Whether someone decides to pursue a college degree or to apply for a demanding job depends on how they judge their academic and work-related abilities. In these situations our abilities affect our chances of success and thus our beliefs about them influence the expected payoff of our decisions. Two people with the same abilities may have very different beliefs about them and thus make very different decisions and have very different outcomes in life. While individual characteristics, such as gender (Reuben et al., 2017) and family background (Filippin and Paccagnella, 2012) are known to be correlated with confidence in abilities, the mechanisms bringing about these differences are not well understood. Situational factors, such as the presence of good or bad feedback have been found to influence people's beliefs about their abilities but the effects of more complex social influences, such as the abilities of people in one's immediate environment have only recently attracted the interest of economists.

When people judge their own ability, they may infer their ability level from comparisons with people in their peer group. For example, someone who finds out that he can do better math than most of his peers may be led to think that he is a good at mathematics and may enjoy it more. However, the person at some point likely encounters another group of people who are on average better at mathematics than he and he might learn that membership in the two groups depends in some way on their mathematics ability. Is it still beneficial for the individual's confidence to be in a weaker group, or not, when both the own as well as other groups can be observed? In other words: Do individuals assign correct weights to ability signals that come from within-group and between-group comparisons?

These questions are important because in different areas of life, such as work and

education, groups of different abilities are deliberately formed, often with the intention of improving overall individual performance. However, the empirical evidence suggests that ability grouping may have negative effects on performance (Hanushek and Wößmann, 2006; Malamud and Pop-Eleches, 2011; Guyon et al., 2012; Kerr et al., 2013), although experimental studies that control for environmental factors have found positive effects (Duflo et al., 2011; Booij et al., 2017). More recently, Murphy and Weinhardt (2014) as well as Elsner and Isphording (2017) identified positive effects of having weaker students within one's group on one's long-term academic outcomes and suggest that higher confidence in abilities due to favorable within-group comparisons are the driving force behind this finding. Additionally, experimental studies have shown that between-group comparisons may matter for academic performance. If a person is a member of a group that stereotypically is worse at a given task, salience of this fact may have a negative effect on this person's outcomes (Coffman, 2014; Dee, 2014). In many real-world situations people may have some idea about both their standing within their group and how their group compares to other groups (cf. Trautwein et al., 2006), however the interaction of within-group and between-group information on ability beliefs is not yet well explored. The net effect of assignment to a weaker group on confidence may be negative or positive, depending on the information available to people as well as how they interpret it.

In this paper, we study the effects of assignment to a weaker group versus a stronger group on confidence and subsequent performance in a laboratory experiment. In our setting, group assignment depends imperfectly on ability so that the ability distributions of the two groups overlap. This implies that the ability signal from group assignment is noisy, which, on the one hand, increases uncertainty that leaves room for interpretation by the subjects and, on the other hand, generates randomness of group assignment that allows for the causal identification of the effect of group assignment on ability beliefs and subsequent performance. We randomly vary whether subjects only receive information about their performance relative to their

group or whether they learn additionally whether they were assigned to a weaker or a stronger group and that group assignment depends imperfectly on ability. This allows us to study the causal effects of assignment to a weaker or a stronger group, and its interaction with salience of ability grouping, on confidence in ability and subsequent test outcomes.

We find, first, that the effect of assignment to a weaker group on confidence depends on the salience of ability grouping. When ability grouping is non-salient, it does not matter for subjects' confidence whether they were assigned to the weaker or the stronger group. However, when ability grouping is salient, assignment to the weaker group makes people less confident in their abilities. Second, subjects on average gave quite correct estimates of their ability rank, when grouping was non-salient. However, when grouping was salient, subjects who were assigned to the stronger group were significantly overconfident while subjects who were assigned to the weaker group were significantly underconfident, indicating that people *overweighed* ability signals coming from *between-group* information. Also, subjects who learned they were assigned to the weaker group were more underconfident than subjects who learned they were assigned to the stronger group were overconfident. This difference cannot be explained by lower ability subjects reporting less correct beliefs, rather, it shows that people *overweighed negatively surprising* information as compared to positively surprising information. Third, results also suggest that higher ability subjects perform worse if they learn they were assigned to a weaker group, while lower ability subjects perform better when learning that they were assigned to a weaker group. We do not find this difference when ability grouping is non-salient. These findings indicate that when people are sorted into different ability groups, within-group and between-group information interact in complex ways to affect ability beliefs and subsequent performance.

To our knowledge, this is the first study to show causal effects of ability grouping on ability beliefs. It shows that both within-group and between-group information, which may not be processed symmetrically, matter for people's beliefs about their

abilities. The results of this study demonstrate that the effects of one's group's abilities on beliefs in own ability and subsequent performance are sensitive to information about the group assignment process. For this reason, one should be careful when interpreting effects of peer group ability on performance from field experiments in which the rules determining group assignment are non-salient (as e.g. in Duflo et al. 2011; Carrell et al. 2013; Booij et al. 2017) as these effects may not hold once people understand that groups of different abilities were formed deliberately.

The paper is structured as follows: Section 2 summarizes the related literature, Section 3 describes the experimental design, Section 4 presents and discusses the results and Section 5 concludes.

2 Related Literature

Higher confidence in one's abilities has been found to have beneficial effects on one's educational and labor market outcomes (Heckman et al., 2006; Cebi, 2007; Heineck and Anger, 2010). Recent evidence also suggests that confidence in one's abilities may be influenced by the abilities of people in one's peer group. Murphy and Weinhardt (2014) find that, controlling for own ability as measured by standardized test scores at age 11, an increase in rank during one's primary school class has a large and significant positive effect on test scores at age 14. The authors also find that the development of subject-specific confidence is the most likely driver of these effects. Similarly, Elsner and Isphording (2017) find that, controlling for own ability, students who have a higher rank within their cohort in high school perceive their intelligence to be higher, have higher expectations about their future careers and are more likely to go to college and complete a degree. These studies run counter to the received wisdom from the peer effects literature that better peers are better for academic performance but provide evidence in favor of the so called "big-fish-little-pond effect" (Marsh, 1987), a popular proposition claiming that assignment to

a peer group with lower skills increases one's confidence in ability[1] that is based on theories of social comparison processes (Festinger, 1954).

On the contrary, the experimental literature highlights the importance of between-group comparisons. For example, people infer individual characteristics from group characteristics, which may lead to self-stereotyping (Coffman, 2014; Dee, 2014). While the traditional economic approach assumed that people form rational expectations about a group member in terms of the aggregate distribution of the characteristics of his group (e.g. Phelps 1972; Arrow 1973, for an overview of the literature see Fang and Moro, 2011), the social cognition approach, which has influenced behavioral economics, holds that people form intuitive generalizations that allow them to save mental resources but which may lead to biases in beliefs. The generalizations are based on real differences between groups and as such contain a "kernel of truth" but they are selective and may exaggerate between-group differences while tending to underweigh within-group differences (Schneider, 2004). Several studies have provided evidence in support of this hypothesis. Recently, Dee (2014) presents empirical evidence from a framed field experiment that self-stereotyping effects can be relevant in an education context: Students at a selective college were randomly assigned to a treatment that primed their awareness of a negatively stereotyped identity (here: a student-athlete). This social-identity manipulation reduced the test performance of athletes relative to non-athletes in spite of causing them to attempt to answer more questions. Similarly, Coffman (2014) finds that, conditional on measured ability, individuals are less willing to contribute ideas in areas that are stereotypically outside of their gender domain, which is largely driven by self-assessments rather than by fear of discrimination, and cannot be easily corrected by

[1]Trautwein et al. (2006) qualify this statement based on correlations between confidence in mathematics ability and mathematics test scores of students in German secondary schools. In their study, schools are either in the high, middle, or low ability track or comprehensive schools that incorporate all three tracks. Controlling for math ability, within tracked schools, students' confidence is higher in schools of lower ability tracks. However, in comprehensive schools where different ability tracks can be found under the same roof, making ability tracking highly observable for students every day, controlling for ability, students' confidence in the higher and the lower tracks did not differ significantly. These observations support the central assumption of this study that both within-group and between-group comparisons of abilities as well as the salience of ability tracking should matter for students' confidence in their abilities.

providing contrary feedback. Furthermore, Albrecht et al. (2013) show that individuals from groups that perform badly on average receive low evaluations, even when it is known that the individuals themselves perform well. This shows that people incorporate group information when evaluating individuals even in cases where it is irrelevant. However, Fryer et al. (2008) cannot reproduce the standard finding that female performance declines in absolute terms when the experimental instructions include a passage emphasizing that men outperform women on a given test.

There is a trade-off between a favorable within-group comparison and a favorable between-group comparison of abilities as the the within-group effect ("big-fish-little-pond-effect") runs counter to the between-group effect, also called the effect of "stereotype threat": One can either be "a bigger fish in a smaller pond" or "a smaller fish in a bigger pond" and it is not ex-ante clear which is better for confidence in abilities. When assigning correct weights to within-group and between-group ability signals, it should not matter for one's confidence whether one is assigned to the weaker or the stronger group as between-group information would counterbalance within-group information. However, subjects could possibly place a greater weight on within-group or between-group information, on positive (Eil and Rao, 2011; Mobius et al., 2011; Grossman and Owens, 2012; Wiswall and Zafar, 2015) or negative (Ertac, 2011) information, or exhibit other forms of biased belief formation (see e.g. Albrecht et al. 2013; Butler 2016).

Furthermore, negative information about one's abilities could both induce higher (Kuhnen and Tymula, 2012; Azmat et al., 2016; Fischer and Wagner, 2017) or lower (Buser, 2016) subsequent performance, depending on how subjects' effort depends on their ability beliefs. Overall, there is mixed evidence on the association between feedback and performance (Kluger and DeNisi, 1998; Hattie and Timperley, 2007), possibly because the relationship between ability beliefs and effort is complex. In recent years, a number of studies has highlighted the importance of distinguishing between confidence in abilities that are complements and confidence in abilities that are substitutes to effort (Santos-Pinto, 2008; Ederer, 2010; Caliendo et al., 2015;

Spinnewijn, 2015; Fischer and Sliwka, 2017). In a setting of human capital investment, Fischer and Sliwka (2017) distinguish between confidence in learning ability – the belief that one can raise one's probability of being successful by exerting effort – and confidence in prior knowledge – the belief that one's probability of being successful is already high prior to investing any additional effort. The authors show experimentally that the use of feedback that raises confidence in learning ability increases motivation to strive towards a better outcome. However, the use of feedback that raises confidence in prior knowledge decreases motivation to strive towards a better outcome for individuals for whom success was more likely at baseline. Fischer and Sliwka's notion of confidence in the effectiveness of effort is equivalent to Benabou and Tirole's (2002, 2003) notion of confidence as an agent's (rational) belief in her own marginal product of effort and possibly captures ability beliefs *positively* related to educational outcomes, as e.g. in Heckman et al. (2006), Cebi (2007), Heineck and Anger (2010), Murphy and Weinhardt (2014), and Elsner and Isphording (2017). In contrast, their notion of confidence in the baseline probability of success possibly describes the kind of belief measured in studies that find higher confidence to have *negative* effects on people's outcomes (see e.g. Camerer and Lovallo (1999), Malmendier and Tate (2005), and Niederle and Vesterlund (2007)). The current study uses within- and between-group information to manipulate people's confidence in their learning ability which, according to theory, is complementary to effort. We therefore expect feedback that bolsters this ability belief to positively influence effort and in turn performance.

3 Experimental Procedure

The experiment was conducted in November 2016 at the Cologne Laboratory of Economic Research[2] using the experimental software z-tree (Fischbacher, 2007). Participants were recruited using the software ORSEE (Greiner, 2004) and, upon arrival, were randomly assigned to one of 32 terminals that were divided by panels.

[2]Financial support of the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

Before the experiment started, participants received instructions that communication with each other and the use of mobile phones or pens was not permitted and that compliance with this rule would be monitored during the whole experiment. These, and all of the following instructions were given on-screen. Participants were also informed that they would receive 4 euros for participating in the experiment and that they could earn additional money by correctly answering questions in several "learning tests". They then received instructions for the first learning test, including the task and the reward scheme, and had to correctly solve a sample question before they could proceed to the test.

First test Each task consisted in assigning to a city name the first digit of its corresponding four digit fictitious "city code". The test consisted of 36 tasks and subjects earned 0.10 euros for each correctly solved task. Before the test phase, there was a 10 minutes learning phase during which subjects could study the city name and code pairs. As shown in Figure 7 in Appendix A.1, during the learning phase city names were listed alphabetically in three columns and the corresponding city codes were displayed next to them for three seconds when the button with the respective name was pressed.[3] Subjects could press these buttons as often as they wanted, without incurring any costs, and in quick succession such that several codes were be displayed at once. Subjects who did not want to study could leave the study screen and spend time looking at comics but could return to studying at any time without this having any implications for them beyond the loss of time. This element was introduced to allow for opportunity costs of studying. During the 6 minutes test phase (see Figure 8 in Appendix A.1), city names were displayed in random order and the correct digit had to be filled in next to them.

Feedback stage (treatment randomization) After the first test, subjects were informed that they would receive feedback about their "learning ability" relative to the other participants based on their result in the learning test. On the next screen,

[3]This feature was meant to capture subjects' intensity of learning.

subjects received their feedback. The assignment mechanism of the feedback was as follows: Subjects were randomly assigned to one type of "ability grouping system", which was either "Non-salient Grouping" or "Salient Grouping". Next, the experimental software assigned each subject either to the "Stronger Learners" or the "Weaker Learners" group. Here, the probability of assignment differed depending on a person's performance in the first test. Those who in the first test performed in the upper half relative to the other participants in the session (percentile rank <0.5 relative to all) were assigned to "Stronger Learners" with a probability of 2/3 and were assigned to "Weaker Learners" with a probability of 1/3. On the contrary, those who in the first test performed in the lower half (percentile rank >0.5 relative to all) were assigned to "Stronger Learners" with a probability of 1/3 and were assigned to "Weaker Learners" with a probability of 2/3. Depending on the group someone was assigned to, the experimental software then computed a person's rank within her group and determined whether this rank was in the upper (percentile rank <0.5 relative to group) or the lower half (percentile rank >0.5 relative to group).

As summarized in Table 1 subjects received different information, depending on the treatment (i.e. "ability grouping system") they were assigned to (The messages displayed to subjects in each treatment can be found in Table A.1 in Appendix A.1.):

Non-salient Grouping: Subjects received feedback relative to their group, which they knew was half the session's participants and did not learn anything about the characteristics of the group.

Salient Grouping: Subjects received both feedback relative to their group and, on the same screen, they also received information about whether they were assigned to the "Stronger Learners" or the "Weaker Learners" group, which they knew consisted of half the session's participants. They also knew that their assignment depended imperfectly on their ability as they were told that "a better result makes it much more likely to be assigned to the stronger learners". Table 1 summarizes the information provided in each treatment.

Table 1: Information by Treatment

Treatment:	**Non-salient grouping**	**Salient grouping**
Information:	upper/lower half in group	upper/lower half in group + stronger/weaker group

Belief elicitation After receiving feedback, subjects were asked to estimate their rank with respect to their performance and their effort (in terms of clicks on city names in the learning phase) in the first test relative to the other participants in the room (session). They knew that for each of the two rank estimates they would earn one euro if it was correct.

Second test and questionnaire After indicating their beliefs the next screen informed subjects that the second test was of the same type, length and duration as the first test but that this time they would earn 0.20 euros (as compared to 0.10 euros in the first test) for each correctly solved task. They were also informed that, unlike after the first test, they would not be able to earn any money by estimating their performance or effort rank relative to other participants. After having read these instructions subjects proceeded to the learning stage of the second test. As can be seen in Figure 12 and Figure 13 in Appendix A.1, the second test looked identical to the first test, it only contained other city names and numbers. When the test was designed, the questions were randomly assigned to test 1 and test 2 in order to create "parallel" tests of the same difficulty. After the second test, subjects were asked to indicate in which of the two tests they believed they performed better and in which they had invested more effort. They could earn 0.50 euros for each correct answer. They then filled in a short demographic survey and learned their earnings from each stage of the experiment.

4 Experimental Results

The experiment lasted approximately 50 minutes and participants on average earned 11.41 euros. In total 7 sessions were conducted, which were orthogonal to treatments to rule out self-selection. All participants were university students, who were on average in their 6th semester of study. 49 percent of participants were female. On average, 19.8 out of 36 questions in the first test and 22.7 out of 36 questions in the second test were answered correctly. There were 79 participants in the non-salient grouping treatment and 78 participants in the salient grouping treatment.[4]

In Section 4.1 we will analyze the effect of salience of ability grouping and assigned group on confidence. Separately for salient and non-salient grouping, we will then explore the response of people with higher and lower ability to higher and lower group assignment. In Section 4.2 we will then shed light on the mechanisms underlying the observed results. In particular, we will (i) address the question to what extended the observed responses are rational given the information provided to people and (ii) investigate whether information processing is affected differently by positive and negative within-group and between-group information. In order to do so, we will derive rank predictions conditional on feedback and will then study how well different groups match their predicted ranks. Finally, in Section 4.3 we will analyze the effects of group assignment and salience of ability grouping on test outcomes.

We expect, based on prior research (Murphy and Weinhardt, 2014; Elsner and Isphording, 2017), that when subjects only learn about their standing within their group, they become more confident when they are assigned to the weaker group. However, when learning about both their standing within their group and their

[4]A treatment where participants were not assigned to a group was also conducted to check whether these two treatments lead to an overall distortion of beliefs. In this benchmark treatment subjects received feedback about whether their performance was in the upper or lower half relative to the whole session. 63 subjects originally participated in this treatment, however only 36 observations are usable due to a programming error. This error affected participants randomly, so that this treatment is still completely balanced to the other two treatments, as can be seen in Table 6 in Appendix A.2. It may thus, as intended, serve to benchmark the distortions caused by the two treatments relevant to our research question.

group's standing relative to another group, this effect disappears if subjects assign correct weights to within-group and between-group ability signals, as in this case between-group information counterbalances within-group information (cf. Trautwein et al., 2006). However, if subjects overweigh between-group information, the effect of weaker group assignment is negative, while if they overweigh within-group information the effect of weaker group assignment is still positive. Furthermore, the current study gives people feedback about their "learning ability" in order to influence people's beliefs in the marginal productivity of learning effort, which according to theory (e.g. Fischer and Sliwka, 2017), is positively related to learning effort. We therefore expect feedback that strengthens this ability belief to positively influence performance.

4.1 Effects of Salience of Ability Grouping and Group Assignment on Confidence

Our first variable of interest is confidence, which we define as

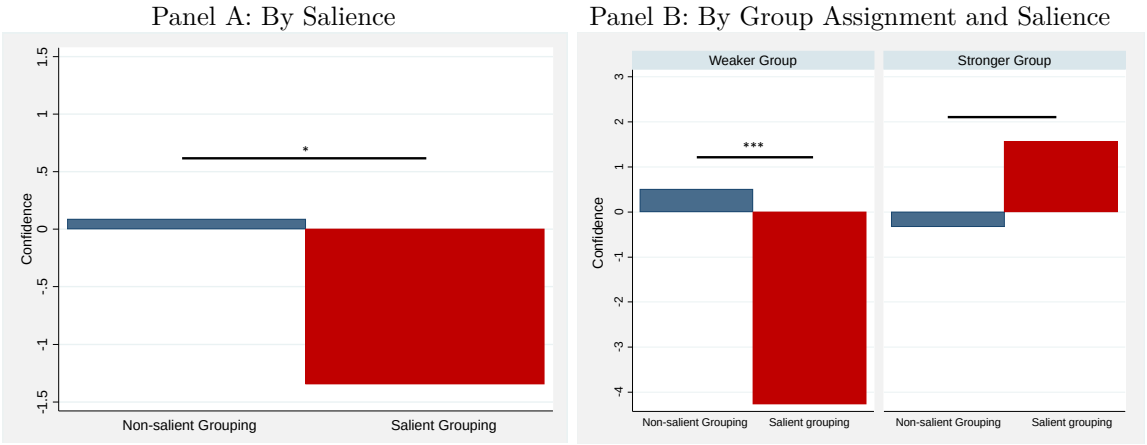
$$Confidence = Rank - RankBelief.$$

Recall from Section 3 that we elicited the $RankBelief$ by asking subjects to estimate their rank in the first test relative to all other participants in their session. Likewise, the $Rank$ measures a subject's actual performance in the first test relative to all other participants in the same session. Thus, our confidence measure is very intuitive as it captures the degree to which subjects overestimate or underestimate their performance relative to the other participants: If someone overestimates his performance relative to the other participants he will have $Confidence > 0$, while if he underestimates his performance relative to the other participants he will have $Confidence < 0$.

In the following, we will study the causal effects of salience of ability grouping as well as its interaction with assigned group on confidence. Then, we will study these

two effects, as well as the overall effect of group assignment, separately for higher and lower ability subjects. Note that while the causal effect of salience as well as its interaction with group assignment can be studied for the whole sample, the causal effect of group assignment by itself has to be studied separately for the higher and lower ability subjects as these two groups had different assignment probabilities.[5]

Figure 1: Effects of Salience of Ability Grouping and Group Assignment on Confidence



Note: Panel A shows the effect of salience of ability grouping on confidence. Panel B shows the interaction effect of salience of ability grouping and group assignment on confidence.

As can be seen in Panel A of Figure 1, confidence was higher when ability grouping was non-salient than when ability grouping was salient. Man-Whitney U (M-W U) tests[6] show that this difference is weakly significant. As can be seen in

[5]Higher ability subjects (who performed above the median in the first test) had a probability of being assigned to the stronger group that was twice as large as the probability of the lower ability subjects (who performed below the median in the first test). This means that, as intended, individuals in the weaker group had on average lower ability than individuals in the stronger group. Our confidence measure captures ability beliefs relative to true ability, so group differences in ability are controlled for in the graphs. However, as subjects had to state their beliefs in terms of ranks (#ranks = #subjects in session), the belief scale is restricted from above and from below, which means that higher ability subjects are more restricted in their possibility to report overconfidence than in their possibility to report underconfidence, while lower ability subjects are more restricted in their possibility to report underconfidence than in their possibility to report overconfidence. This may induce the overconfidence of higher ability subjects and the underconfidence of lower ability subjects to be underestimated. Within these two groups, the probability of being assigned to any of the two groups was perfectly random so that the restriction with respect to reporting overconfidence and underconfidence affected people assigned to the stronger group and the weaker group equally. Hence, by analyzing the effects of group assignment separately for higher and lower ability subjects, we can identify the causal effects of assignment to the weaker or stronger group on confidence.

[6]All tests in this paper are two-sided, unless stated otherwise.

Panel B of Figure 1 subjects who were assigned to the weaker group but did not know that their group was the weaker one were more confident than subjects who were assigned to the weaker group and knew that their group was the weaker one (M-W U test: p=0.00). On the contrary, when one was assigned to the stronger group, knowing whether one's group was the stronger one did not not significantly affect one's confidence (M-W U test: p=0.32).

Table 2: Effects of Salience of Ability Grouping and Group Assignment on Confidence

	(1)	(2)	(3)	(4)
Dependent Variable: Confidence	If Lower A.	If Higher A.	If Lower A.	If Higher A.
Non-salient Grouping	0.949	3.349**	3.952**	10.10***
	(0.55)	(2.54)	(2.07)	(5.02)
Stronger Group	3.282*	4.191***	6.839***	10.65***
	(1.78)	(3.25)	(2.71)	(6.36)
Non-salient Gr. × Stronger Group			-7.482*	-11.07***
			(-1.97)	(-4.33)
Observations	76	81	76	81
R^2	0.117	0.242	0.173	0.413

Note: This table presents the effect of non-salient versus salient ability grouping and assignment to a stronger versus a weaker group using a linear regression model including a constant, session fixed effects and robust standard errors. Dependent variable: confidence. Columns 1 and 3 (2 and 4) show results for lower (higher) ability subjects. t statistics are reported in parentheses * p<0.10, ** p<0.05, *** p<0.01.

In Table 2 we analyze, separately for higher and lower ability individuals, the effects of salience, of group assignment, as well as of the interaction between the two. The regressions are estimated by ordinary least squares and contain heteroscedasticity robust standard errors as well as session dummies and a constant, but no other control variables. Thus, all the coefficients show causal effects of our intervention. As can be seen in Columns 1 and 2 both lower and higher ability subjects were on average more confident (by 3.3 and 4.2 ranks, respectively) if they were assigned to the stronger group. These effects are marginally and highly significant, respectively. However, only higher ability subjects are affected by the salience of ability grouping overall. Thus, the difference in confidence shown in Panel A of Figure 1 are largely driven by the response of higher ability subjects. They were on average 3.3 ranks more confident when ability grouping was non-salient. Columns 3 and 4

present results for the interaction effects between group assignment and salience of the assignment mechanism. Qualitatively, lower and higher ability subjects respond similarly but the effects seem to be larger for higher ability subjects. When ability grouping is salient, both lower and higher ability subjects are more confident when they are assigned to the stronger group (by 6.8 and 10.7 ranks, respectively). Both effects are highly significant. Those who were assigned to the weaker group were more confident (by 4.0 and 10.1 rank, respectively) when they did not learn that their group was the weaker one. These effects are significant at the 5% and the 1% level, respectively. F-tests show that when ability grouping was non-salient, it did not matter for lower or higher ability subjects whether they were assigned to the weaker or the stronger group (for both p=0.81). Hence the differences presented in Panel B of Figure 1 are driven by both lower and higher ability subjects.[7]

4.2 Mechanisms

The above results show that when group assignment is salient, assignment to the weaker group causes individuals to be less confident than assignment to the stronger group. Furthermore, weaker group assignment causes subjects to be less confident when grouping is salient than when grouping is non-salient. The mechanisms underlying these observations can be further explored on three levels. First, we can investigate to what extend salient and non-salient ability grouping leads to a *de-calibration* of beliefs, i.e. to what extend they make people overconfident or underconfident.[8] Second, we can explore to what extend non-salient and salient ability

[7]Interestingly, higher but not lower ability subjects' beliefs in their intensity of effort (in terms of clicks), when ability grouping was salient, responds significantly to group assignment: When learning they were assigned to the weaker group, higher ability subjects believe to have exerted less effort than when learning they were assigned to the stronger group. This may indicate that higher ability subjects attribute weaker group assignment more strongly to effort (rather than to ability) than lower ability subjects.

[8]In the benchmark treatment, in which people were not assigned to different groups and received feedback relative to the whole session, people's mean confidence was 0.31, which is not significantly different from 0 (t-test: p=0.69). Hence, without ability grouping, subjects were on average well calibrated.

grouping lead people to state "*irrational*"[9] beliefs, i.e. rank beliefs that could not possibly be correct given the feedback someone received. Third, we can shed light on how non-salient and salient ability grouping affect the *distributions* of beliefs. This may help us to better understand the average treatment effects as well as the decalibration of beliefs.

Overconfidence and underconfidence When grouping is salient, confidence could be lower with weaker group assignment than with stronger group assignment due to (1) weaker group assignment making people underconfident and/or (2) stronger group assignment making them overconfident. Panel B of Figure 1 suggests that the effect is driven mostly by salient grouping making people assigned to the weaker group underconfident, while they seem well calibrated when grouping is non-salient. Furthermore, when ability grouping is salient, people tend to be on average less confident than when ability grouping is non-salient. This could be due to (1) non-salient grouping making people overconfident and/or (2) salient grouping making people underconfident. Panel A of Figure 1 suggests that while with non-salient grouping people have on average quite correct beliefs, they seem to be very underconfident on average with salient ability grouping.

Using one-sided t-tests of the means of the four groups (stronger group–non-salient / weaker group–non-salient / stronger group–salient / weaker group–salient) in Panel B of Figure 1 against the null hypothesis that people have correct beliefs (Confidence=0) reveals that when grouping was non-salient, subjects were neither significantly overconfident when assigned to the weaker group (p=0.30) nor significantly underconfident when assigned to the stronger group (p=0.36). However, if grouping was salient, subjects who were assigned to the weaker group were significantly underconfident (p=0.00) and subjects who were assigned to the stronger

[9]The feedback given to each person implied that there were certain ranks they were definitely not occupying. As people were paid for correct rank estimates, it was never optimal for one's monetary payoff to report rank beliefs that are definitely false. However, one could think of a model where an agent benefits from incorrect beliefs, e.g. with respect to his self-image or his motivation. In this case, false beliefs could possibly be optimal. In our setting, we will abstract from this possibility and will call beliefs "irrational" if they indicate a rank that was impossible for a person given the information they had received.

group were weakly significantly overconfident (p=0.08). Furthermore, a M-W U test reveals that if grouping was salient, people who were assigned to the weaker group were significantly more underconfident than people who were assigned to the stronger group were overconfident (p=0.03). This shows that people assigned a larger weight to the ability signal from group assignment when it was negative than if it was positive.

Overall, people become more decalibrated by salient than by non-salient ability grouping. When ability grouping is salient, they are more decalibrated if they are assigned to the weaker group than if they are assigned to the stronger group.

"Irrational" beliefs In the following, we will address the question to what extend the stronger decalibration from salient than from non-salient grouping is "irrational" given the feedback information subjects received. The feedback given to each person, while imprecise about their relative position, ruled out certain ranks for them.Thus some rank beliefs were "irrational" for them to hold. We will also shed light on the mechanisms that may explain why, when ability grouping is salient, weaker group assignment leads people to become more decalibrated than stronger group assignment.

Note that the feedback types explained in Section 3 are not equal to the four groups analyzed above (stronger group–non-salient / weaker group–non-salient / stronger group–salient / weaker group–salient). This is because in the non-salient grouping treatment (stronger group–non-salient / weaker group–non-salient) subjects did not learn their group assignment but only which half they occupied within their group. Hence, the two feedback types with non-salient grouping are "upper half within group" and "lower half within group", which we will call "Non-salient Grouping – 1" and "Non-salient Grouping – 2", respectively. By contrast, in the salient grouping treatment, people learned both whether their group was the weaker or the stronger one as well as their half within their group. Thus, with salient grouping, we have four feedback types: "upper half in stronger group", "lower half in stronger group", "upper half in weaker group" and "lower half in weaker group",

which we will call "Salient Grouping – 1", "Salient Grouping – 2", "Salient Grouping – 3", and "Salient Grouping – 4", respectively. Furthermore, subjects knew that their group assignment depended imperfectly on their ability. Hence, they knew that stronger group assignment did not necessarily imply that one's performance was above average, while weaker group assignment did not necessarily imply that one's performance was below average.

Figure 2: Information Content of Feedback and Distribution of Beliefs

Treatment	Feedback		Information/Belief Distribution				
	Group	Half within Group	1st Quart. (0.00-0.25)	2nd Quart. (0.26-0.50)	3rd Quart. (0.51.-0.75)	4th Quart. (0.76-1.00)	
Non-salient Grouping – 1	?	Upper	60.47	37.20	0	2.33	A
Non-salient Grouping – 2	?	Lower	2.78	2.78	38.88	61.11	
Salient Grouping – 1	Stronger	Upper	72.22	27.78	0	0	B
Salient Grouping – 2	Stronger	Lower	0	52.38	14.29	33.33	C
Salient Grouping – 3	Weaker	Upper	5.56	0	77.77	16.67	
Salient Grouping – 4	Weaker	Lower	0	9.52	9.52	80.95	

Likely Possible Impossible

Note: This table indicates the likelihood, conditional on feedback, of being ranked in a given quartile (dark gray: likely, light gray: possible, white: impossible). The numbers indicate the percentage of people believing to be ranked in a given quartile, conditional on feedback.

Figure 2 shows the six different types of feedback that were given during the experiment. For example, if someone was in the non-salient ability grouping treatment he was either told that he was in the upper half within his group or that he was in the lower half within this group. If he was in the upper half within his group (feedback type "Non-salient Grouping – 1"), and the ability distributions of the two groups were not too different, he was likely in the upper half with respect to all the participants in the session. However, it was theoretically possible that his group was much worse than the other group. In this case, being in the upper half within this group could possibly entail being only in the 3^{rd} quartile with respect to all participants. However, even if his group was so bad compared to the other group that the two groups' ability rank distributions did not overlap, given that he was

told he was in the upper half within his group, it was impossible that he occupied an ability rank in the 4^{th} quartile (percentile ranks >= 0.75) with respect to all people in the session. Applying the same reasoning to the other five types of feedback as well produces the different zones (likely range, possible (less likely) range, impossible range) that are indicated by the different shadings for the four quartiles. The numbers in Table 2 indicate the percentage of people, in a given feedback category, who reported a rank belief in the respective quartile. To give an applied example, consider subjects who were in salient grouping and were told that they occupied a rank in the upper half of the stronger group (feedback type "Salient Grouping – 1). Among them 72.22 percent indicated a rank belief in the first quartile (for them, the likely range), while 27.78 percent reported a rank belief in the second quartile (for them, the possible range). None of these people reported a rank in the 3^{rd} or 4^{th} quartile. We can conclude that none of the people who received this type of feedback reported an "irrational" belief.

With non-salient grouping, the two groups ("upper half within group" and "lower half within group") have similar belief distributions over the likely, possible and impossible range. However, with salient grouping, the picture is different. Here, of those who were assigned to the weaker group 16.7 and 9.5 percent, respectively, report beliefs in the impossible range while none of those assigned to the stronger group do so. Furthermore, those who were assigned to the weaker group seem to state fewer beliefs in the possible range than those assigned to the stronger group. Among those of the weaker group, the proportion of people stating a belief in the likely range seems to be larger (at 77.77 and 80.95 percent, respectively) than among those of the stronger group (52.38 and 72.22 percent, respectively). In the following, we will study how similar, overall and within the four quartiles, the belief distributions of people who received the different types of feedback are.

Belief distributions by feedback types As as shown in Table 2, people who received the different feedback types had different ranges of likely, possible, and impossible beliefs. To illustrate this, Figure 14 in Appendix A.3 shows the ex-

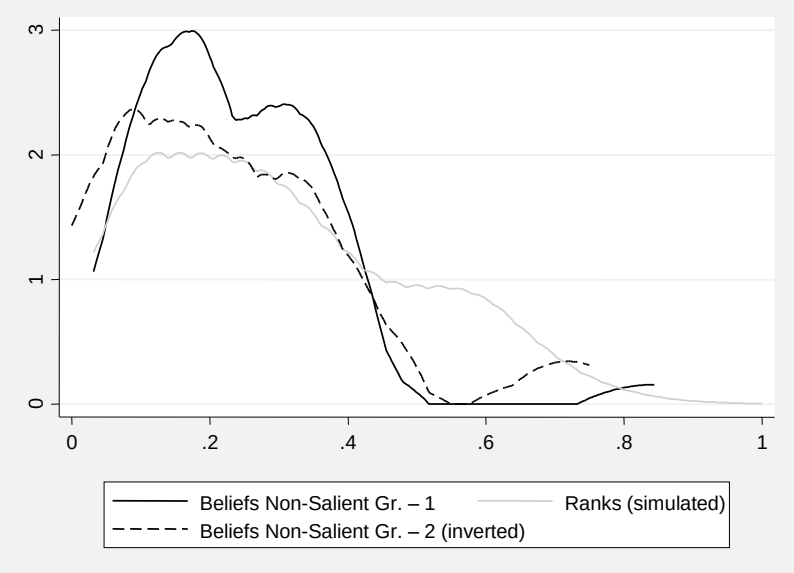
pected ability rank distributions by feedback type resulting from our assignment mechanism. We can see that the expected rank distributions for subjects who received feedback types "Non-salient Grouping – 1" and "Non-salient Grouping – 2", and likewise for "Salient Grouping – 1" and "Salient Grouping – 4" as well as for "Salient Grouping – 2" and "Salient Grouping – 3" are mirror images of each other. Hence, within these pairs of feedback types the rank distributions that subjects had to match with their beliefs were the same except for being inverted. Thus, a straightforward way for testing whether the belief distributions differed from each other, conditional on feedback, within each of the three pairs is to invert the elicited rank beliefs of one of the groups within each of the pairs. Next, we can run statistical tests for the equality of distributions.

Figures 4.3, 4.4, and 4.5 show the inverted belief distributions from the feedback types whose expected rank distributions are shown on the right hand side of Figure 14 in Appendix A.2 mapped onto the belief distributions from the feedback types whose expected rank distributions are shown on the left hand side. Furthermore, they are depicting the expected rank distributions that are shown in Figure 14, which are identical within each pair after the right hand side distributions have been inverted. As can be seen in Figure 4.3, which corresponds to comparison "A" in Figure 2 and Panel A in Figure 14, with non-salient grouping, when people receive positive feedback (Non-salient Grouping – 1 (NSG–1)), they have a very similar belief distribution, conditional on feedback, as people who receive negative feedback (Non-salient Grouping – 2 (NSG–2)). Subjects in NSG–1 seem largely not to take into consideration that they could occupy a rank in the lower half while subjects in NSG–2 seem to largely ignore their rank could be in the upper half with respect to the whole session. A Kolmogorov–Smirnov test shows that the two distributions are not significantly different overall (p=0.58)[10]. Testing the distributions in the four quartiles separately reveals that the 1^{st} quartile of NSG–1 is not significantly different from the 4^{th} quartile of NSG–4 and the 2^{nd} quartile of

[10]For all Kolmogorov–Smirnov tests in the paper exact p-values from combined (two-sided) tests are reported

NSG–1 is not significantly different from the 3^{rd} quartile of NSG–2. However, while NSG–1 has no observations in the 3^{rd} and the 4^{th} quartile NSG–2 has observations both in the 2^{nd} and in the 1^{st} quartile (as can also be seen in Figure 2) .

Figure 3: Comparison of Beliefs in Non-Transparent grouping (A)

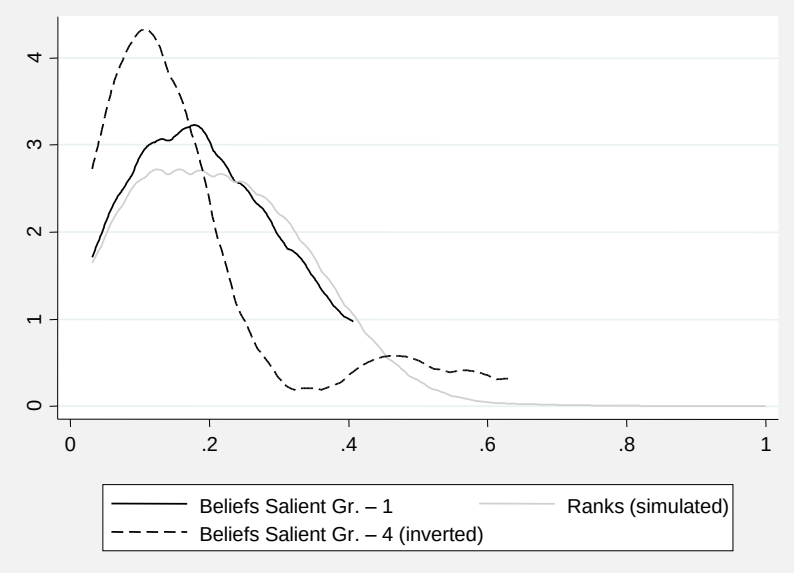


Note: This graph shows the distributions of the rank beliefs of subjects who received feedback type "Non-Salient Grouping – 1", the inverted rank beliefs of subjects who received feedback type "Non-Salient Grouping – 2", and the (inverted) expected ability rank distribution for subjects who received feedback type "non-Salient Grouping – 1 " ("Non-Salient Grouping – 2 ").

As can be seen in Figure 4.4, which corresponds to comparison "B" in Figure 2 and Panel B in Figure 14, with salient grouping, when people get extreme feedback ("upper half in stronger group" or "lower half in weaker group") and it is negative (Salient Grouping – 4 (SG–4)), "lower half in weaker group"), they tend to interpret it more extremely than when they get positive feedback (Salient Grouping – 1 (SG–1), "upper half in stronger group"). However, among those who get negative feedback some take into account the possibility that they might have been in the upper half overall (inverted percentile rank >0.5). When people get extreme positive feedback they seem to have surprisingly correct beliefs overall. However, they seem to ignore the possibility that they might have performed in the lower half (percentile rank >0.5). Kolmogorov–Smirnov tests show that the two distributions

are not significantly different overall (p=0.25). Testing the four quartiles separately reveals that the 2^{nd} quartile in SG – 1 and the 3^{rd} quartile in SG – 4 ((inverted) percentile rank >0.25 and <0.5) are weakly significantly different from each other (p=0.09). Furthermore, the distributions are different in the 3^{rd} quartile in SG – 1 and the 2^{nd} quartile in SG – 4 ((inverted) percentile rank >0.50 and <0.75), as SG–1 does not have any observations in the 3^{rd} quartile while SG–4 has observations in the 2^{nd} quartile (as can also be seen in Figure 2).

Figure 4: Comparison of Beliefs in Transparent grouping (Extreme Feedback) (B)

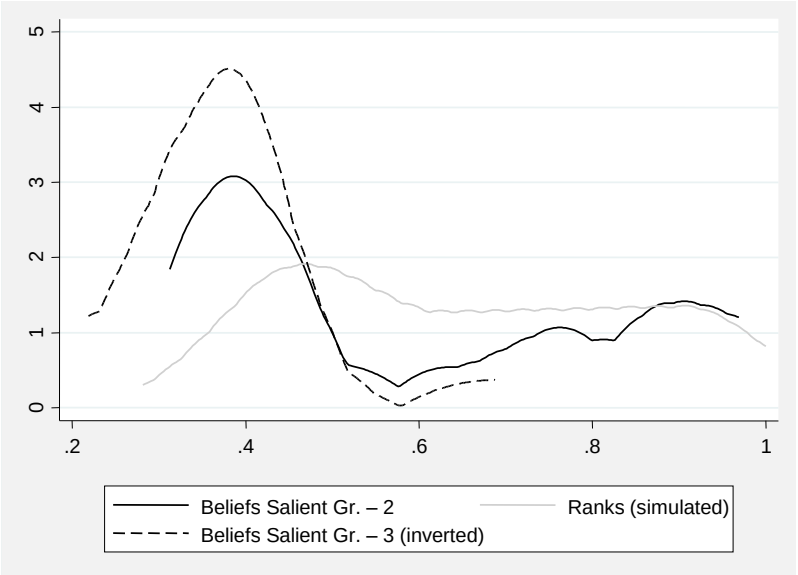


Note: This graph shows the distributions of the rank beliefs of subjects who received feedback type "Salient Grouping – 1", the inverted rank beliefs of subjects who received feedback type "Salient Grouping – 4", and the (inverted) expected ability rank distribution for subjects who received feedback type "Salient Grouping – 1" ("Salient Grouping – 4").

As can be seen in Figure 4.5, which corresponds to comparison "C" in Figure 2 and Panel C in Figure 14, with salient grouping, when people get positive feedback about their group but negative feedback about their standing within their group (Salient-grouping – 2 (SG–2), "lower half in stronger group"), many of them correctly take into account that they might in fact have performed in the lower half relative to the whole session. However, when people get negative feedback about their group but positive feedback about their standing within their group (Salient-

grouping – 3 (SG–3), "upper half in weaker group"), they largely ignore the possibility that they might have performed in the upper half overall. Kolmogorov–Smirnov tests show that the belief distributions with these two ambivalent feedback types are significantly different overall (p=0.039). Testing the four quartiles separately reveals that the 2^{nd} quartile of SG–2 and the 3^{rd} quartile of SG–3 as well as the 3^{rd} quartile of SG–2 and the 2^{nd} quartile of SG–3 are not significantly different from each other. The 1^{st} quartile of SG–2 has no observations while the 4^{th} quartile of SG–3 does. Furthermore, the 4^{th} quartile of SG–2 does have observations while the 1^{st} quartile of SG–3 does not (as can also be seen in Figure 2). Those who received "lower half in stronger group" feedback seem to correctly take into account that the partial randomness of our group assignment mechanism implies that one may have below average performance in spite of being assigned to the stronger group. On the contrary, those who received "upper half in weaker group" feedback seem to ignore the partial randomness of our group assignment mechanism and that they may well have above average performance in spite of being assigned to the weaker group. Note that the group who seems to ignore the partial randomness of assignment has on average higher performance in the first test than the group who takes it into account (M-W U: p=0.062). Thus, the resulting more decalibrated beliefs among those receiving bad between-group and good within-group information than those receiving good between-group and bad within-group information cannot be explained by the former having lower ability as measured by the test (which may be correlated with the ability to understand the feedback). Rather, negatively surprising group assignment seems to lead to a larger decalibration of beliefs than positively surprising group assignment.

Figure 5: Comparison of Beliefs in Transparent grouping (Ambivalent Feedback) (C)



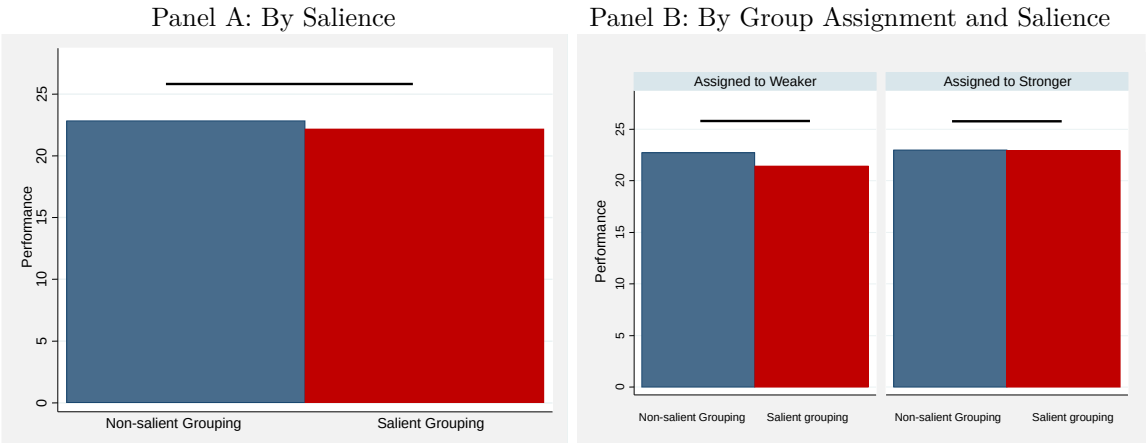
Note: This graph shows the distributions of the rank beliefs of subjects who received feedback type "Salient Grouping – 2", the inverted rank beliefs of subjects who received feedback type "Salient Grouping – 3", and the (inverted) expected ability rank distribution for subjects who received feedback type "Salient Grouping – 2" ("Salient Grouping – 2").

Implications of these findings will be discussed in Section 5 together with the results for performance.

4.3 Effects of Salience of Ability Grouping and Group Assignment on Performance

We will now analyze whether ability grouping affects participants' outcomes in the second test. First, we will compare the test score averages between people in the non-salient and the salient grouping treatment. Then we will look at the interaction effects between the assigned group and salience of group assignment on average test scores. Note that the bar graphs in Figure 6 are showing raw scores from the second test. As can be seen in Figure 6, there is neither a significant overall effect of salience of ability grouping nor an interaction effect of salience of ability grouping with group assignment on performance.

Figure 6: Effects of Salience of Ability Grouping and Group Assignment on Performance



Note: Panel A shows the effect of salience of ability grouping on test scores. Panel B shows the interaction effect of salience of ability grouping and group assignment on test scores.

In Table 3 the treatment effects of ability grouping on performance are analyzed separately for lower and higher ability subjects (who had below and above median performance, respectively, in the first test). Interestingly, we find opposite and significant effects for the two groups that are disguised when looking at the average over both groups as in Figure 6. As can be seen in Columns 1 and 2, while lower ability subjects perform significantly worse (by 3.1 points), higher ability subjects perform significantly better (by 2.7 points) with non-salient ability grouping than with salient ability grouping. Columns 3 and 4 show that when assigned to the weaker group, lower ability subjects benefit from learning that their group is the weaker one (by 4.1 points), while higher ability subjects suffer from learning that their group is the weaker one (by 6.6 points).[11]

Hence, we find that salient ability grouping has a positive effect on the performance of lower ability individuals while it has a negative effect on the performance of higher ability individuals. This is driven by opposite effects for these groups

[11] We do not find that people's effort, in terms of revealing information by clicking on city names in the learning phase, which was meant to measure the intensity of their learning, responded to our treatments (see Table 7 in Appendix A.3). We infer that subjects rather responded to the intervention by adjusting their mental efforts and that it may be better use of the revealed information that improves test outcomes.

when they are assigned to the weaker group. While the performance of lower ability individuals increases when they learn that they were assigned to the weaker group, the performance of higher ability individuals decreases when they learn that they were assigned to the weaker group. This suggests that, in our setting, higher confidence in ability as measured by the learning test does not clearly result in better test outcomes. In fact, only for higher ability subjects confidence and subsequent performance are positively correlated, while they are negatively correlated for lower ability subjects. For the whole sample, confidence predicts subsequent outcomes negatively (p=0.027, see Table 8 in Appendix A.3). Although we intended our feedback about performance in the "learning test" to influence subjects' beliefs about their marginal productivity of effort, which we expected to be positively related to effort, our feedback possibly (also) influenced a different type of belief.[12]

Table 3: Effects of Salience of Ability Grouping and Group Assignment on Performance

	(1)	(2)	(3)	(4)
Dependent Variable: Test Score	If Lower A.	If Higher A.	If Lower A.	If Higher A.
Non-salient Grouping	-3.053**	2.743*	-4.067**	6.581**
	(-2.14)	(1.74)	(-2.22)	(2.59)
Stronger Group	-1.783	0.406	-2.984	4.080
	(-1.13)	(0.26)	(-1.33)	(1.53)
Non-salient Grouping × Stronger Group			2.528	-6.293*
			(0.83)	(-1.82)
Observations	76	81	76	81
R^2	0.144	0.099	0.153	0.142

Note: This table presents the effect of non-salient versus salient ability grouping and assignment to a stronger versus a weaker group using a linear regression model including a constant, session fixed effects and robust standard errors. Dependent variable: test score. Columns 1 and 3 (2 and 4) show results for lower (higher) ability subjects. t statistics are reported in parentheses * p<0.10, ** p<0.05, *** p<0.01.

5 Discussion

We studied the causal effects of assignment to a weaker or a stronger group as well as its interaction with salience of the assignment mechanism on confidence in

[12]The belief we manipulated does not seem to be (only) a person's baseline belief in receiving a good outcome, which Fischer and Sliwka (2017) show may be negatively related to subsequent performance, because we find the inverse relationship for higher and lower ability subjects compared to what they find.

learning ability and outcomes in a subsequent learning test. To do so, we designed a feedback intervention that gave people imprecise feedback about either (1) their standing within their group (whether they performed in the upper or the lower half relative to their group) or (2) their standing within their group plus their group's standing relative to another group (whether their group was stronger or weaker than the other group). We expected, based on empirical research that finds that students become more confident in their academic abilities when they have worse classmates (Murphy and Weinhardt, 2014; Elsner and Isphording, 2017), that when only learning about their standing within their group, subjects would become more confident when they were assigned to the weaker group. Furthermore, when learning about both their standing within their group and their group's standing relative to another group, this effect should be expected to disappear if subjects assign correct weights to within-group and between-group ability signals, as in this case between-group information would counterbalance within-group information. However, if subjects overweigh between-group information, the effect of weaker group assignment would be negative, while if they overweigh within-group information the effect of weaker group assignment would still be positive.

Our results show that, in the setting we studied, when the group assignment mechanism was non-salient, it did not matter for subjects confidence whether they were assigned to the weaker or the stronger group. The signs of the effects suggest that in this case subjects were slightly more confident when assigned to the weaker group, however the effect sizes are so small that it would need a much larger sample size to possibly find a significant effect. Furthermore, we find that if the group assignment mechanism was salient, weaker group assignment made people less confident. This effect is highly significant and much larger than the positive effect of weaker group assignment when the assignment mechanism was non-salient. We find this effect both for lower and higher ability individuals, although it seems to be even stronger for the latter. We also find that subjects are on average less confident when the group assignment mechanism is salient than when it is non-salient.

This is found to be the case due to salient grouping causing subjects' beliefs to become decalibrated, especially when learning they were assigned to the weaker group. When grouping was non-salient, subjects on average gave quite correct estimates of their ability rank. However, when grouping was salient, subjects who were assigned to the stronger group were significantly overconfident while subjects who were assigned to the weaker group were significantly underconfident, indicating that people *overweighed* ability signals coming from *between-group* information.

When ability grouping was salient, subjects assigned to the weaker group were more underconfident than subjects assigned to the stronger group were overconfident, indicating that people *overweighed negative* information as compared to positive information. Some of those who are told they are in the weaker group report "irrational" rank beliefs (i.e. beliefs that must be false given the subject's information), while none of those who are told they were in the stronger group do so. When comparing people who received extreme feedback ("upper half in stronger group" and "lower half in weaker group") we find that, conditional on feedback, these groups did not have significantly different belief distributions although they represented the two extremes of the ability distribution. However, when it comes to ambivalent feedback, we find more decalibrated beliefs among those receiving bad between-group and good within-group information ("upper half in weaker group") than those receiving good between-group and bad within-group information ("lower half in stronger group"), which cannot be explained by lower abilities of the former group as compared to the latter. Thus, group assignment information seems to lead to stronger decalibration of beliefs if it is negatively surprising than if it is positively surprising. This is in line with the finding that people's beliefs respond more strongly to negative information (Ertac, 2011) but contradicts the possibly more common finding that people incorporate positive information into their beliefs more strongly than negative information (Eil and Rao, 2011; Mobius et al., 2011; Grossman and Owens, 2012; Wiswall and Zafar, 2015).

With respect to test outcomes, we find that salient ability grouping has a positive

effect on the performance of lower ability individuals while it has a negative effect on the performance of higher ability individuals. This is driven by opposite effects for these groups when they are saliently assigned to the weaker group. While the performance of lower ability individuals increases when they learn that they were assigned to the weaker group, the performance of higher ability individuals decreases when they learn that they were assigned to the weaker group. Past research has also variously found that performance increases (Kuhnen and Tymula, 2012; Azmat et al., 2016; Fischer and Wagner, 2017) or decreases (Buser, 2016) in response to negative performance information. Our findings suggest that in our setting, higher confidence in learning ability as measured by the test does not have clear benefits for people in terms of improving their test outcomes. In fact, confidence overall predicts subsequent test outcomes negatively.

6 Conclusion

To our knowledge, this is the first study to investigate the causal effects of within-group and between-group information on people's ability beliefs and performance. Overall, our results suggest that ability grouping may have negative effects on people's confidence in their ability, especially for those who are assigned to a weaker group. Being part of a weaker peer group should not generally be expected to make people more confident. Our results imply that the positive effect of weaker peers on confidence if relative ability between groups is non-salient may be greatly outweighed by the negative effect of having weaker peers when people know that their peers are relatively weaker compared to another group. In line with past findings (Coffman, 2014), negative information about one's group may lead people to self-stereotype, i.e. to believe that one has worse characteristics than one actually does. Our results also suggest that, in settings where ability grouping is done visibly, forming ability groups may risk harming those people who are negatively surprised by weaker group assignment more than it may benefit those who are positively surprised by stronger group assignment.

The results of this study demonstrate that the effects of one's group's abilities on beliefs in own ability and subsequent performance are sensitive to information about the group assignment process. Because of this, one should be careful when interpreting effects of peer group ability on performance from field experiments where the group assignment mechanism is non-salient (as e.g. in Duflo et al., 2011; Carrell et al., 2013; Booij et al., 2017) as other effects may prevail once people find out that groups of different abilities were deliberately formed.

Overall, our findings suggest that the relationship between ability beliefs and motivation are complex and should be further investigated in future research. Our study may help to understand the effects of ability grouping in the field by isolating the effects it may have on ability beliefs. However, we caution that our results are based on a laboratory experiment that studies the effects in an abstract setting and further research needs to be done to confirm that our findings hold in educational or workplace settings.

A Appendix

A.1 Details on the Experimental Procedure

Test and Feedback Screens

Figure 7: Test 1 (Test Phase)

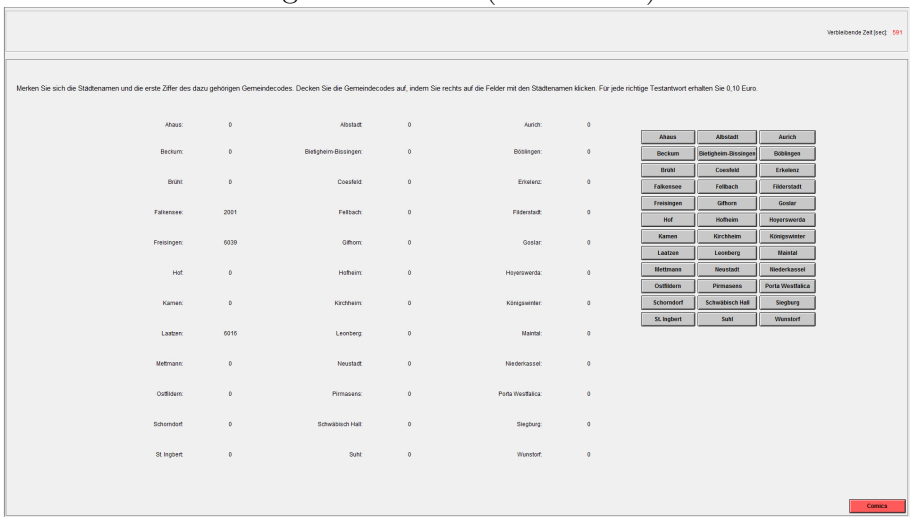


Figure 8: Test 1 (Learning Phase)

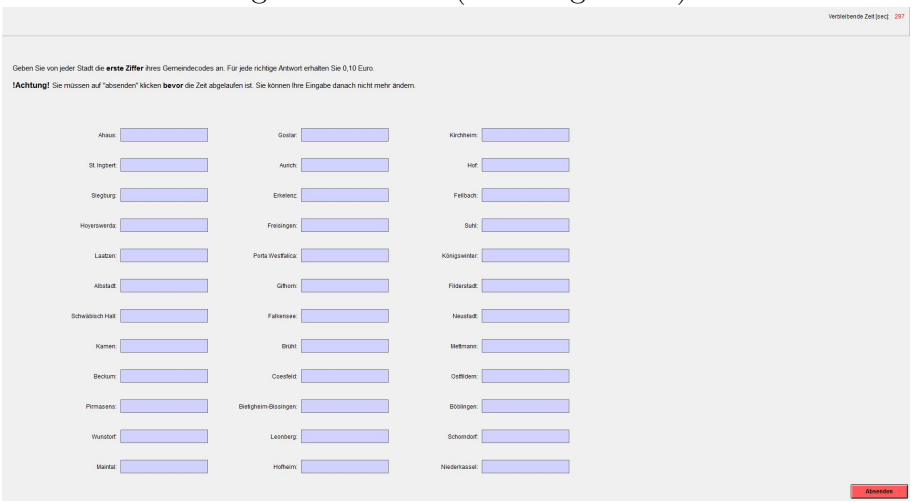


Figure 9: Sample Feedback: Non-salient Grouping

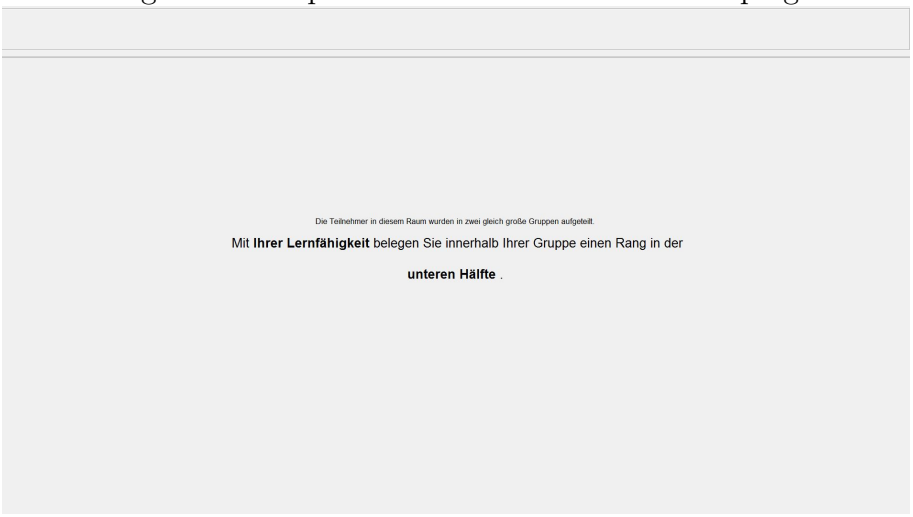


Figure 10: Sample Feedback: Salient Grouping

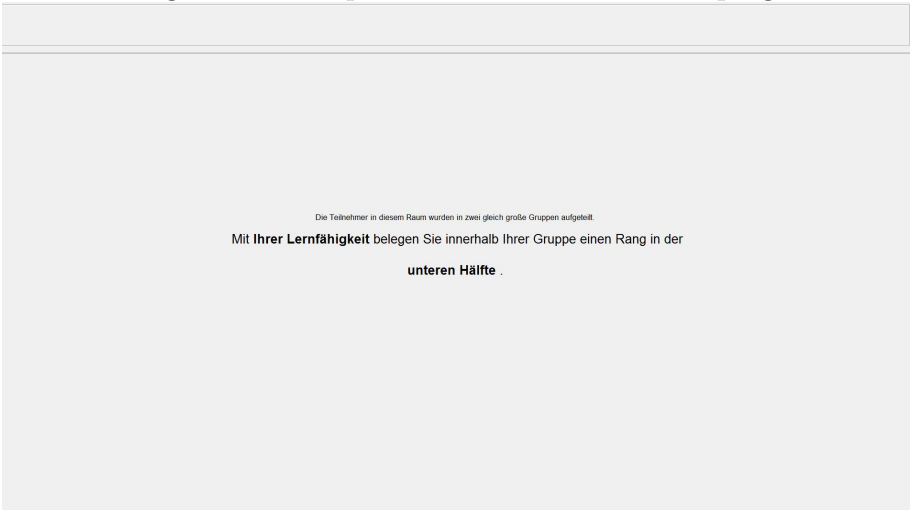


Figure 11: Sample Feedback: No Grouping

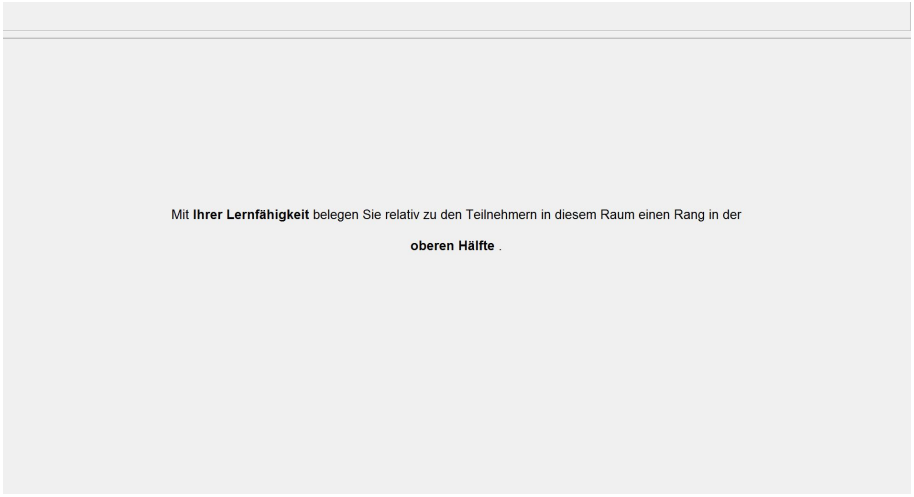


Figure 12: Test 2 (Test Phase)

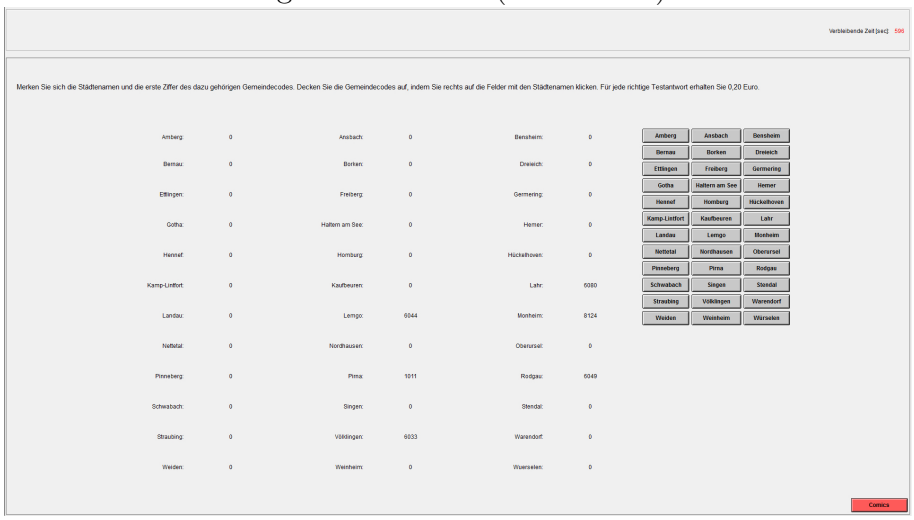
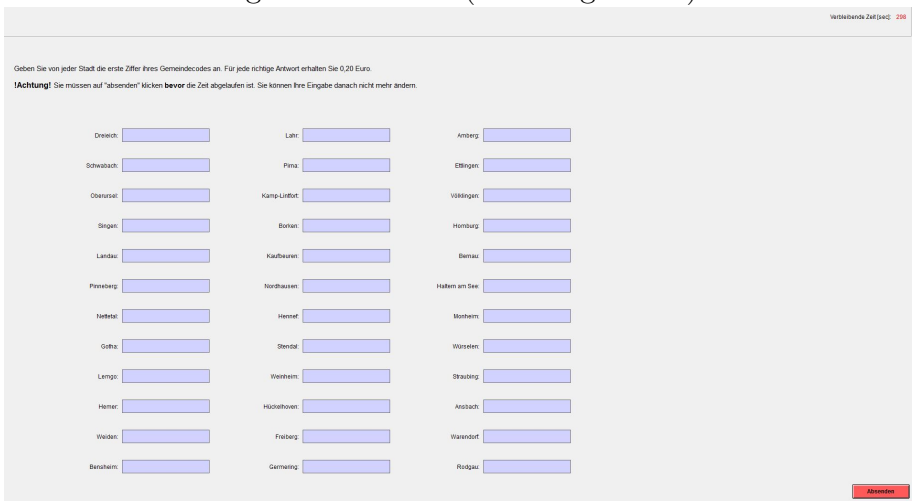


Figure 13: Test 2 (Learning Phase)



Treatment Messages

Table 4: Message by Treatment

Treatment:	**Non-salient grouping**	**Salient grouping**
Message:	"The participants in this room were divided into two equally sized groups. With your learning ability you occupy a ranking in the upper [lower] half within your group."	"The participants in this room were divided into to equally sized groups: The stronger learners and the weaker learners. There, a better result makes it much more likely to be assigned to the stronger learners. You were assigned to the stronger (weaker) learners. With your learning ability you occupy a rank in the upper [lower] half among the stronger (weaker) learners."

A.2 Summary Statistics and Balance Checks

Table 5: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Points Test 1	19.881	7.676	3	36	193
Points Test 2	22.668	7.888	0	36	193
Better Half	0.508	0.501	0	1	193
Confidence	-0.451	6.406	-20	19	193
Decalibration	4.793	4.261	0	20	193
Effort 1	239.539	117.775	51	898	193
Effort 2	242.518	120.898	56	672	193
Non-salient Tracking	0.409	0.493	0	1	193
Salient Tracking	0.404	0.492	0	1	193
Stronger Group	0.497	0.502	0	1	157
Better Half in Group	0.409	0.493	0	1	193
Extreme Feedback	1.538	0.505	1	2	39
Ambivalent Feedback	1.462	0.505	1	2	39
Female	0.492	0.501	0	1	193
Semester	5.611	3.483	1	15	193
School GPA	2.574	6.424	0	90	193
Profit	11.41	2.295	5.8	17.6	193
Session 1	0.145	0.353	0	1	193
Session 2	0.135	0.342	0	1	193
Session 3	0.155	0.363	0	1	193
Session 4	0.135	0.342	0	1	193
Session 5	0.15	0.358	0	1	193
Session 6	0.119	0.325	0	1	193
Session 7	0.161	0.368	0	1	193
Humanities	0.098	0.299	0	1	193
Social Science	0.078	0.268	0	1	193
Law	0.109	0.312	0	1	193
Busines Administration	0.295	0.457	0	1	193
Economics	0.161	0.368	0	1	193
Medicine	0.062	0.242	0	1	193
Natural Sciences	0.078	0.268	0	1	193
Other Fields	0.119	0.325	0	1	193

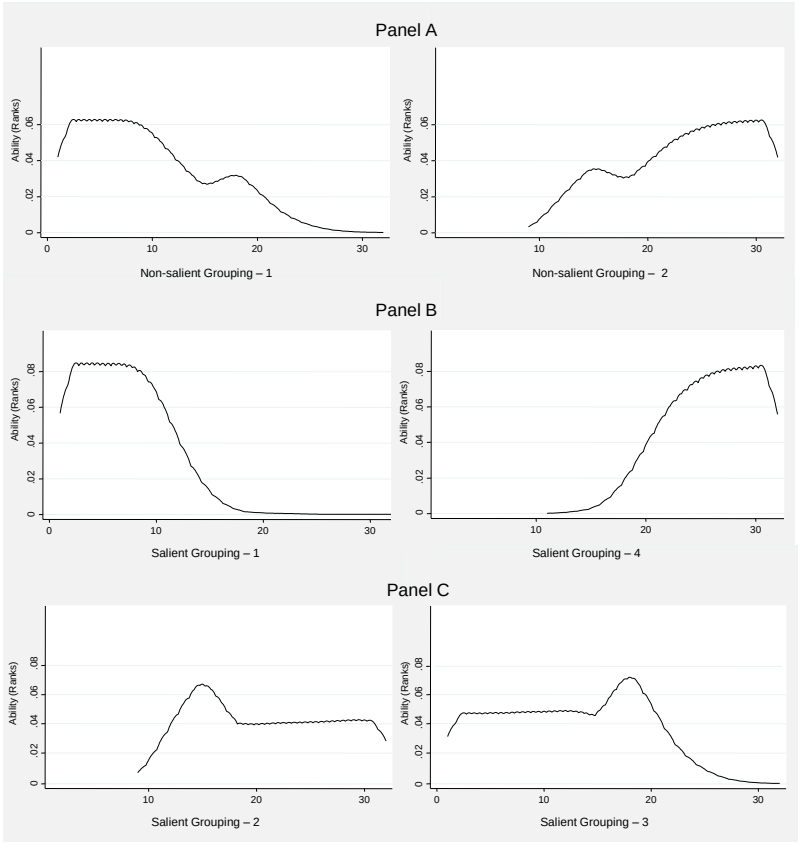
Table 6: Balance Check

	(1) Non-Transparent Tracking	(2) Transparent Tracking	(3) No Tracking	(4) Overall	(1) vs. (2), p-value	(1) vs. (3), p-value	(2) vs. (3), p-value
Female	0.481	0.513	0.472	0.492	0.692	0.931	0.690
	(0.057)	(0.057)	(0.084)	(0.036)			
Points Test 1	20.418	19.628	19.250	19.881	0.513	0.478	0.801
	(0.907)	(0.791)	(1.391)	(0.553)			
School GPA	2.101	1.982	2.092	2.051	0.239	0.936	0.399
	(0.068)	(0.075)	(0.100)	(0.045)			
Semester	5.532	6.000	4.944	5.611	0.408	0.362	0.156
	(0.358)	(0.438)	(0.534)	(0.251)			
Field of Study	4.304	4.782	4.611	4.554	0.201	0.479	0.710
	(0.252)	(0.274)	(0.322)	(0.163)			
Session No.	4.177	3.872	3.944	4.010	0.357	0.566	0.856
	(0.235)	(0.233)	(0.303)	(0.146)			
N	79	78	36	193			
Proportion	0.409	0.404	0.187	1.000			

Standard errors in parentheses.

A.3 Simulations and Further Results

Figure 14: Expected Ranks by Feedback Type



Note: This figure shows the distributions of the expected ability ranks by feedback type. The graphs are based on simulations applying our ability group assignment mechanism to 64,000 observations.

Table 7: Effort Intensity

Dependent Variable: Effort	(1) If Lower Ability	(2) If Higher Ability	(3) If Lower Ability	(4) If Higher Ability
Non-salient Grouping	29.90	-13.18	34.34	17.08
	(1.17)	(-0.46)	(0.97)	(0.38)
Stronger Group	-43.74	-14.50	-38.48	14.46
	(-1.56)	(-0.52)	(-1.06)	(0.32)
Non-salient Gr. × Stronger Group			-11.06	-49.62
			(-0.20)	(-0.76)
Observations	76	81	76	81
R^2	0.181	0.128	0.181	0.135

Note: This table presents the effect of non-salient versus salient ability grouping and assignment to a stronger versus a weaker group using a linear regression model including a constant, session fixed effects and robust standard errors. Dependent variable: effort in terms of clicks. Columns 1 and 3 (2 and 4) show results for lower (higher) ability subjects. t statistics are reported in parentheses * p<0.10, ** p<0.05, *** p<0.01.

Table 8: Correlation between Confidence and Subsequent Performance

Dependent Variable: Test Score	(1) All	(2) If Lower Ability	(3) If Higher Ability
Confidence	-0.205**	-0.148	0.0312
	(-2.32)	(-1.27)	(0.24)
Observations	157	76	81
R^2	0.081	0.096	0.064

Note: This table presents the correlation between confidence and subsequent performance using a linear regression model including a constant, session fixed effects and robust standard errors. Dependent variable: test score. Column 1 shows results for all subjects, and columns 2 and 3 show results for lower and higher ability subjects, respectively. t statistics are reported in parentheses * p<0.10, ** p<0.05, *** p<0.01.

References

Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature*, 40(4):1167–1201.

Albrecht, K., Parys, J., Szech, N., and von Essen, E. (2013). Updating, self-confidence and discrimination. *European Economic Review*, pages 144–169.

Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A., editors, *Discrimination in Labor Markets*. Princeton University Press, Princeton, N.J.

Azmat, G., Bagues, M., Cabrales, A., and Iriberri, N. (2016). What you don't know... Can't hurt you? A field experiment on relative performance feedback in higher education. Discussion Paper DP11201, Centre for Economic Policy Research.

Benabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Benabou, R. and Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, 70(3):489–520.

Benabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164.

Booij, A. S., Leuven, E., and Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, 84(2):547–578.

Buser, T. (2016). The impact of losing in a competition on the willingness to seek further challenges. *Management Science*, 62(12):3439–3449.

Butler, J. V. (2016). Inequality and relative ability beliefs. *The Economic Journal*, 126(593):907–948.

Caliendo, M., Cobb-Clark, D. A., and Uhlendorff, A. (2015). Locus of control and job search strategies. *Review of Economics and Statistics*, 97(1):88–103.

Camerer, C. and Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1):306–318.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3):855–882.

Cebi, M. (2007). Locus of control and human capital investment revisited. *The Journal of Human Resources*, 42(4):919–932.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4):1625–1660.

Dee, T. S. (2014). Stereotype threat and the student-athlete. *Economic Inquiry*, 52(1):173–182.

Dohmen, T. and Falk, A. (2010). You get what you pay for: Incentives and selection in the education system. *The Economic Journal*, 120(546):F256–F271.

Dohmen, T. and Falk, A. (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2):556–90.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy*, 19(3):733–769.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–138.

Elsner, B. and Isphording, I. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3).

Ertac, S. (2011). Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532 – 545.

Fang, H. and Moro, A. (2011). Theories of statistical discrimination and affirmative action: A survey. In Benhabib, J., Jackson, M., and Bisin, A., editors, *Handbook of Social Economics*, volume 1A, pages 133–200. North Holland, The Netherlands.

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

Filippin, A. and Paccagnella, M. (2012). Family background, self-confidence and economic outcomes. *Economics of Education Review*, 31(5):824–834.

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

Fischer, M. and Sliwka, D. (2017). Confidence in knowledge or confidence in the ability to learn: An experiment on the causal effects of beliefs on motivation. *mimeo*.

Fischer, M. and Wagner, V. (2017). Effects of timing and reference frame of feedback: Evidence from a field experiment in secondary schools. *unpublished*.

Fryer, R. G., Levitt, S. D., and List, J. A. (2008). Exploring the impact of financial incentives on stereotype threat: Evidence from a pilot study. *American Economic Review*, 98(2):370–75.

Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen*, pages 79–93. Ges. fuer Wiss. Datenverarbeitung, Goettingen.

Grossman, Z. and Owens, D. (2012). An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, 84(2):510–524.

Guyon, N., Maurin, E., and McNally, S. (2012). The effect of tracking students by ability into different schools: A natural experiment. *Journal of Human Resources*, 47(3):684–721.

Hanushek, E. A. and Wößmann, L. (2006). Does educational tracking affect performance and inequality? differences- in-differences evidence across countries*. *The Economic Journal*, 116(510):C63–C76.

Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Heineck, G. and Anger, S. (2010). The returns to cognitive abilities and personality traits in germany. *Labour Economics*, 17(3):535–546.

Kerr, S. P., Pekkarinen, T., and Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, 31(3):577–602.

Kluger, A. N. and DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, 7(3):67–72.

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

Malamud, O. and Pop-Eleches, C. (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, 95(11):1538 – 1549. Special Issue: International Seminar for Public Economics on Normative Tax Theory.

Malmendier, U. and Tate, G. (2005). Ceo overconfidence and corporate investment. *Journal of Finance*, 60(6):2661–2700.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79(3):280–295.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2011). Managing self-confidence: Theory and experimental evidence. *NBER Working Papers*, 17014.

Murphy, R. and Weinhardt, F. (2014). Top of the Class: The Importance of Ordinal Rank. CESifo Working Paper Series 4815, CESifo Group Munich.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4):659–661.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, pages n/a–n/a.

Santos-Pinto, L. (2008). Positive self-image and incentives in organisations. *The Economic Journal*, 118(531):1315–1332.

Schneider, D. J. (2004). *The Psychology of Stereotyping*. Guilford Press, New York.

Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association*, 13(1):130–167.

Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: how the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology*, 90(2):334–349.

Wiswall, M. and Zafar, B. (2015). How do college students respond to public information about earnings? *Journal of Human Capital*, 9(2):117–169.