Gender matters - Performance spillover effects of symbolic recognition at school

Andrea Essl*, Frauke von Bieberstein†

Institute of Organization and Human Resource Management, University of Bern,

Engehaldenstrasse 4, 3012 Bern, Switzerland

Abstract

Symbolic recognition can be an efficient and powerful tool for motivating people. Despite much research on ex ante announced symbolic incentives, little is known about the spillover effect of different recognition types in a multitask setting. This paper is the first to examine not only positive but also negative symbolic recognition and its spillover effects. In a laboratory experiment, secondary school students had to work on two different tasks. In the experimental treatment, students received unannounced symbolic performance feedback for the first task. We find that the response to different symbolic recognition types is heterogeneous across gender. Female non-recipients as well as girls who received positive recognition significantly increased their performances in the second task. For girls receiving negative recognition and for boys independent of recognition type we do not find a spillover effect.

Keywords: symbolic recognition, performance spillover effects, gender differences, laboratory experiment

JEL codes: C93, D03, J16

*Corresponding author, phone: +41 31 631 3359, e-mail address: andrea.essl@iop.unibe.ch (A. Essl), Institute of Organization and Human Resource Management, University of Bern, Engehaldenstrasse 4, 3012 Bern, Switzerland

†E-mail address: frauke.vonbieberstein@iop.unibe.ch (F. von Bieberstein), Institute of Organization and Human Resource Management, University of Bern, Engehaldenstrasse 4, 3012 Bern, Switzerland

1 Introduction

According to the business book *1501 Ways to Reward Employees* by Nelson (2012), one of the strongest tools for increasing motivation is positive recognition. Many companies use symbolic non-monetary incentives to recognize their employees, such as Thinslices' and McDonald's "employee of the month" awards, J. C. Penney's "new managers ceremony" or Wells Fargo's "thank-you e-cards". Symbolic expressions of recognition are not only common in the business world, but also a popular part of school traditions. Schools make use of symbolic recognition, such as gold stars, certificates, and prizes for good performance to foster confidence and promote exemplary study habits.

At the same time, negative recognition for underachievement is often informal, e.g., when a boss criticizes individual or group performance. Further, negative recognition can be shown by taking away privileges like a private office or degrading someone in the hierarchy of the firm. In the school context, negative recognition can take forms like announced detention, sending a student out of the room, or announced underachievement in a test. In film and sports, formal negative recognition programs exist, like the "Golden Raspberry Awards" for the worst in film or the "Lanterne Rouge" being the competitor in last place in a cycling race such as the Tour de France. However, these negative awards are often extra-organizational, meaning they are not bestowed by the organization the recipient works for. Still, it has been argued that agents receive much more (informal) negative recognition as opposed to positive recognition (Nelson 2012).

In the current research, we examine symbolic recognition programs covering not only positive, but also negative recognition. In a multitask setting, we analyze the spillover effect of receiving symbolic positive or negative recognition in one task on performance in a second unrelated task. Recent research in economics has shown that non-monetary positive recognition programs can be an important and cost-efficient way of providing incentives in organizations. Some of these studies show that ex ante mentioned recognition programs increase individuals' performance (Frey and Neckermann 2008, Kosfeld and Neckermann 2011, Ashraf et al. 2014), whereas others find a positive ex post spillover effect of intra-organizational recognition programs (Neckermann et al. 2014, Bradler et al. 2016).[1] Bradler et al. (2016) provide evidence that unannounced recognition significantly improves subsequent performance in the same task, where the effect is mostly driven by performance increases of non-recipients. Most closely related to our study, Neckermann et al. (2014) find a positive ex post effect of awards on subsequent task performance in an unrelated task. Our contribution expands on this literature by investigating the ex post performance spillover effect of positive and negative recognition.

[1] The field study by Malmendier and Tate (2009) found that CEOs had a significant performance decline after receiving an award. However, the type of award they examined is extra-organizational.

To analyze performance spillover effects of different symbolic recognition types we conducted an experiment with 138 secondary school students. We believe that pupils are a suitable sample for examining the spillover effect of symbolic recognition on performance for two reasons. First, due to the high diversity of school subjects (e.g., math, languages, arts), students are used to working in a multitask environment. Second, recognition programs are common tools for rewarding academic achievements; thus, school students are also used to symbolic recognition. Students in our experiment were required to work on two different tasks. In the first task, they had to estimate the number of peas in a bowl and in the second task, they had to cut out advertising flyers for a university orchestra concert. Prior to the two main tasks, we introduced a stage to measure individual's baseline performance in the flyer-cutting task. In the experimental treatment, participants received unannounced symbolic performance feedback after the estimation task. The top third was rewarded with a smiley-sticker, the bottom third received a frowny-sticker, and the intermediate third did not receive symbolic recognition, but the information that they were ranked in the middle. Recognition remained private during the experiment. Students in the control treatment received no performance feedback after the estimation task. In both treatments, students received ex ante announced symbolic recognition for the main flyer-cutting task.

Our study shows that the response to different symbolic recognition types is heterogeneous across gender. We find that compared with female students in the control group, female non-recipients as well as females who received a smiley-sticker in the estimation task significantly improved in the subsequent flyer-cutting task. In contrast, for females who received a frowny-sticker we do not find a significant difference compared to the control group. Also, we do not find a spillover effect of the different recognition types on males' performance.

The effect of different symbolic recognition types on subsequent performance is not obvious. In particular, as is argued in more detail in Section 2, in our case the combination of two main mechanisms seems to best explain the pattern we find for females: First, reciprocity concerns entail an increase in subsequent performance by those who received a smiley-sticker, a decrease by those who received a frowny-sticker, and no reaction by non-recipients. Second, if positive recognition is the goal, goals as reference points can explain why those who are only slightly behind (in our case the non-recipients) increase their subsequent performance, while the effect for those who received a frowny-sticker can go in either direction depending on how far they consider themselves behind. At the same time, goals as reference point literature predicts no subsequent performance effect for those who already received positive recognition. Taken together, the combination of these two motives can explain the increase in performance of recipients of positive recognition and non-recipients as well as the lack of an effect for recipients of negative recognition.

There are several potential reasons why we find an effect for females but not for males. Previous research has shown that symbolic awards at school motivate girls but not boys (Jalava et al. 2015). In addition, compared to the case with only positive awards, our setting with positive and negative recognition provides more information on the relative ranking. Results from the literature on relative performance feedback show that gender differences often play a role in these settings (Barankay 2011, Kuhnen and Tymula 2012). Finally, recent research suggests that gender differences in competitive environments might be task-dependent (Günther et al. 2010, Grosse and Reiner 2010, Shurchkov 2012). Under competitive incentives, men often outperform women in stereotypically male tasks, whereas women frequently perform equally or better than men in stereotypically female tasks. As our task of flyer-cutting could be categorized as a female-oriented task (females performed better in the baseline stage), this might be another driver behind our results.

The paper proceeds as follows. Section 2 presents related literature including the different motives that could be at work. Section 3 explains the experimental design and procedure. Section 4 presents the results that are discussed in the concluding Section 5.

2 Related literature

Our paper builds on three strands of literature: symbolic recognition, relative performance feedback, and gender differences in competition. Regarding spillover effects of symbolic recognition, the papers by Neckermann et al. (2014) and Bradler et al. (2016) discuss different motives for why a spillover effect can occur. Reciprocity preferences can explain the actions of award winners, when they reciprocate the kind action of the employer by increasing their subsequent performance (Akerlof 1982, Fehr et al. 1993, Kube et al. 2012). Given negative recognition, the same argument results in negative reciprocity and would thus lead to a decline in subsequent performance of those people who received negative recognition. At the same time reciprocity predicts no effect for non-recipients.[2] In contrast, conformity preferences as in Bradler et al. (2016) would result in an opposite pattern: Recipients of negative recognition learn that their performance belongs to the bottom of the distribution and given a preference for conformity, they feel a need to increase subsequent performance. With the same reasoning, those who received positive recognition would reduce their performance for conformity reasons. Again, in our setup we would expect no effect for non-recipients given they are exactly in the middle of the performance distribution.

[2] A similar pattern can result from self-esteem concerns (Bénabou and Tirole 2006) and status concerns (Auriol and Renault 2008, Auriol et al. 2016). Positive recognition could raise employees' motivation and performance even in a subsequent unrelated task, because individuals want to live up to the experience of greater self-esteem and status. Conversely, negative recognition is often combined with a decline in self-esteem and status which may result in a performance decrease.

An explanation for increased effort provision of non-recipients could be that positive recognition serves as a goal and reference point (Heath et al. 1999). Berger and Pope (2011) show in different settings in the lab and in the field that being slightly behind the goal leads to an increase in motivation and in the subsequent probability of winning. The authors argue that if winning is the goal that acts as a reference point, being behind this goal is particularly painful due to loss aversion. In addition, due to diminishing sensitivity, people who are slightly below their goals should work harder than those for whom the goal is further away. In our setting, non-recipients who narrowly miss positive recognition in the first task thus increase their performance to achieve a positive award, that serves as a goal and reference point, in the following task. This motivational effect should be stronger for non-recipients than for recipients of negative recognition due to diminishing sensitivity. The goal as reference point literature predicts no effect for recipients of positive recognition.

Besides reflecting acknowledgment, recognition programs also provide feedback on relative performance.[3] Compared to the literature examining only positive recognition, our setting with positive and negative recognition provides participants with even more information regarding their relative performance, because they learn whether their performance was in the bottom, intermediate, or top third of the distribution. Charness et al. (2014) consider rank feedback with and without symbols to visually illustrate the relative performance in three-person groups and find that rank feedback increases effort as well as disreputable behavior. In addition, a growing economic literature has evolved studying the effects of giving relative performance information without handing out symbolic awards (in much of this research the feedback is private). The effect of feedback on performance is mixed: Whereas some studies find a positive effect of relative performance information on performance (Azmat and Iriberri 2010, Blanes i Vidal and Nossol 2011, Kuhnen and Tymula 2012, Tran and Zeckhauser 2012, Delfgaauw et al. 2013), others find no effect (Eriksson et al. 2009) or a negative effect (Barankay 2011, Bandiera et al. 2013[4]). Several of these studies have found strong gender effects, where often feedback seems to have stronger effects on men (, Barankay 2011, Kuhnen and Tymula 2012). In some cases this gender effect is more intricate, as in Delfgaauw et al. (2013) who find a positive effect of rank feedback for retail stores, but only if the store's manager and a sufficiently large fraction of the employees has the same gender. Jalava et al. (2015) consider different non-financial incentives in a school setting. They find that boys are only motivated by rank-based incentives, while girls are also motivated by receiving a symbolic reward.

[3] In their field experiment, Ashraf et al. (2014) unbundle the effect of disclosing rank information from the effect of employer recognition and social visibility in a field experiment conducted in Zambia. They find that employer recognition and social visibility increase performance, while social comparison reduces performance, especially for low-ability workers. In contrast, in a real-effort lab experiment Neckermann and Yang (2017) find that relative rank information is the driver behind positive ex post effects of unannounced financial or symbolic recognition. Within firms, symbolic recognition including relative rank information is often inferred from the position of an employee in the firm's hierarchy (Auriol et al., 2016).

[4] In Bandiera et al. (2013) the effect is driven by changes in team composition.

This important effect of gender is also documented in the literature examining preferences for competition.[5] In this literature, participants not only receive rank feedback, but in addition payments depend on rank. It has been shown that men outperform women in competitive environments (Gneezy et al. 2003, Gneezy and Rustichini 2004) and have a higher willingness to perform under a tournament compensation scheme (Niederle and Vesterlund 2007, Niederle and Vesterlund 2011, Datta Gupta et al. 2013, Sutter and Glätzle-Rützler 2015). Recent literate has suggested that there are gender differences in the reaction to winning and losing in a competition. While Gill and Prowse (2014) and Buser (2016) find that women, but not men, perform worse after losing a competition, Buser and Yuan (2016) show that women are less likely to compete again after losing a competition than are men. Furthermore, given recent evidence, gender differences in competitive settings might be task-dependent: Men outperform women in stereotypically male tasks, whereas women perform equally or better than men in stereotypically female tasks (Günther et al. 2010, Grosse and Reiner 2010, Shurchkov 2012).

3 Experimental design and procedure

3.1 Experimental design

The experiment consisted of two main tasks, both without any performance-dependent financial incentives. The first task was a simple estimation task (Falk and Zimmermann 2016). Students were shown a picture displaying a bowl filled with peas and were asked to estimate the number of peas inside it.[6] After all students had written down their estimates, in both treatments the experimenters announced the real number of peas in the bowl which was 3000. In the control treatment, students continued with the second task. In the experimental treatment, the experimenters announced that the top third would receive a smiley-sticker, the bottom third a frowny-sticker, and the intermediate third nothing other than a message saying that they were ranked in the middle. This recognition program is closely related to relative performance feedback, since individuals receive information about their relative performance and the incentives are not linked to monetary payments. The difference is that we provide relative performance information combined with tangible recognition stickers. After the stickers had been distributed, the experimenters asked participants to put the sticker on their sweatshirt. Given that students were seated separated from each other by cardboard blinds, recognition remained private during the experiment.[7]

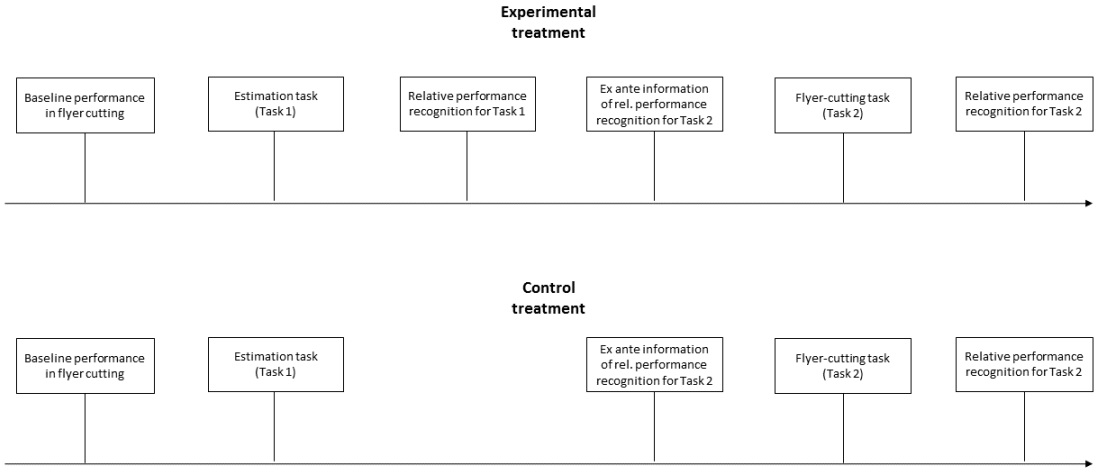
[5] Next to competition, it has also been found that women are less willing to take up challenging tasks in general compared to men (Niederle, M. and Yestrumskas, A. 2017, von Bieberstein and Jaussi 2017).

[6] Instructions, including the picture of the bowl, can be found in Appendix A.

[7] Of course, students could decide to keep the stickers on their sweatshirt after the experiment, if they wanted to.

In the second task, students were asked to cut out flyers from sheets. Each workstation was provided with scissors and a stack of sheets printed with the flyers.[8] On each sheet, four flyers with varying difficulty levels were printed in a random order. The task was to cut along a black line with a tolerance space of four millimeters shaded in gray. Due to differences in the difficulty level of the motifs, students were not allowed to start a new sheet before they had cut out all four flyers from the previous sheet. This mechanism prevented students from working just on the easy motifs. Note that all participants received the same motifs in the same order. The flyers promoted an upcoming concert by a university student orchestra. The participants knew that all the flyers that had been cut out correctly would be used for advertisement. The cutting task offered several advantages. Most importantly, compared with other real effort tasks (e.g., counting numbers, mathematical problems, transcribing), the flyer-cutting task introduced in this study was not useless, since flyers served as a real-world promotion campaign. Furthermore, the task required no special knowledge or cognitive abilities, it was easy to explain, and learning possibilities were trivial, as the school students at this age knew before the experiment how to cut out motifs with scissors.

FIGURE 1: Timeline of the experiment



[8] One example of a sheet is shown in Appendix A. The sheets were of size DIN A4 (similar to letter size).

Before participants started this second task, the experimenters in both treatments announced that the third of students with the largest number of correctly cut out flyers would receive a smiley-sticker, while the worst-performing third would receive a frowny-sticker, and the rest would just obtain a message notifying them that their performance was average. As this second recognition system was introduced in both the control and the experimental group, it allows us to isolate the ex post spillover effect of relative symbolic recognition from the ex ante incentive effect. Students had 15 minutes to work on the task. Nobody completed all the sheets provided within the 15 minutes. Participants were asked to put the cut out flyers into a non-transparent bag marked with their identification number. This procedure minimized peer effects as it prevented participants from comparing their work with each other. After students had completed the flyer-cutting task, they received the corresponding recognition incentive for the second task.

Prior to the two main tasks, we introduced a baseline stage, which was, apart from the time provided and the non-monetary incentive, identical to the cutting task. Students had five minutes to cut out motifs. This stage provided us with a baseline measure of students' ability in a non-competitive and non-incentivized setting. In addition, it allowed participants to become accustomed to the task. Students were also asked to fill in a questionnaire, which included basic demographic information like gender, age, and level of education as well as a question regarding whether they were right- or left-handed because the scissors provided were appropriate for right-handers. The timeline for both treatments is illustrated in Figure 1.

3.2 Procedure

In total, 138 students took part in the experiment. The participants were students of the seventh, eighth, and ninth grade from a secondary school in Switzerland. The participants' average age was 14 years, with a standard deviation of 1.2 years, and 47% of the students were female.[9] In the experimental group, 50 girls and 52 boys participated and the control group consisted of 15 girls and 20 boys. While 63 students attended the lower secondary school, 72 participants were enrolled in the upper secondary school.[10] Participation was voluntary.[11] The experiment was conducted in May 2014 and lasted for approximately 1 hour. Each student received a fixed payment of 8 CHF for participating.[12] After a short introduction, all participants were randomly assigned to

[9] One person did not reveal his or her sex.

[10] In Swiss secondary schools, students are separated according to their capabilities. Students with better grades are assigned to upper secondary level, whereas students who aspire to an apprenticeship are allocated to the so-called lower secondary school. Students from both types participated in this experiment. Three persons did not reveal their school type.

[11] All students in attendance took part in the experiment. This is not surprising because the experiment took place during school time and in the students' classrooms.

[12] The payment of 8 CHF corresponds to the average daily pocket money of a Swiss student at that age.

treatments and classrooms. In both treatments, participants received written instructions for each task separately and were also asked to answer control questions for the cutting task.

A total of 103 students participated in the experimental treatment and 35 students took part in the control treatment. This ensured that for the estimation task about one fourth of the participants received a smiley-sticker and another fourth a frowny-sticker, while another fourth received a message saying that their performance was average. The remaining fourth of the participants took part in the control treatment in which neither a recognition incentive nor feedback was provided for the estimation task. Because of space constraints, the experimental group was divided into three sub-groups (three classrooms, n=34, 34, and 35). In the experimental group, in each classroom, the top third received positive symbolic recognition, the bottom third negative symbolic recognition, and the intermediate third only the information that their performance had been average. In each classroom each student was seated at a workstation separated by cardboard blinds. These laboratory conditions were installed prior to the experiment to ensure anonymity and a standard procedure. The experiment was conducted simultaneously in four classrooms. This allowed us to rule out the possibility that students would hear about the treatments before they actually took part in the experiment. Since all four experimenters followed a strict protocol, the procedure for the experimental sub-groups was exactly the same.

4 Results

Recognition groups are classified according to the recognition type received in the estimation task. Importantly, performance in the estimation tasks is not correlated with performance either in the baseline or in the main flyer-cutting task (correlation coefficients corr=-0.08, p=0.351 or, respectively, corr=-0.05, p=0.523). While in the control treatment (n=35) nobody received performance feedback in the estimation task, in the experimental treatment the top third was awarded a smiley-sticker (n=37), the bottom third received a frowny-sticker (n=32), and the intermediate third received only the information that their performance was average (n=34). If two or more students had the same threshold rank, they were assigned to the higher recognition incentive.[13] Since it was explicitly mentioned that only correctly cut out flyers can be used for the promotion campaign and the ex ante announced recognition in the flyer-cutting stage was linked to the number of correctly cut out flyers, the analysis focuses on this measure. The number of correctly cut out flyers is represented by the number of motifs that were cut out within the gray shaded area. As the baseline stage and the main flyer-cutting stage differed in length, the performance was measured in terms of productivity. Therefore, the performance measure is the number of correctly cut out flyers per minute. The

[13]The situation that two students had the same rank happened twice.

descriptive statistics for the control and the experimental groups are shown in Table 1. Consistent with previous literature, we find that in the control as well as in the experimental group, ex ante announced symbolic recognition has a highly significant positive effect on performance (for both groups $p<0.001$; Wilcoxon signed-rank test).[14] In the control group, performance increases from 0.39 to 0.86 correct flyers and in the experimental group from 0.56 to 0.95 correct flyers. However, as we are not able to separate the sole incentive effect from a possible learning effect, we suggest that this performance increase may only partly be driven by ex ante mentioned symbolic recognition.

Next, we analyze whether different recognition types received in the estimation task influence subsequent performance in the main flyer-cutting task. Descriptive statistics for the different subgroups of the experimental group are also presented in Table 1. Mann-Whitney tests reveal that individuals who received no recognition in the estimation task cut out subsequently significantly more flyers correctly per minute than students from the control group and those from the negative recognition group ($p=0.045$ or, respectively, $p=0.042$). The differences in performance between the control group and either the positive recognition group or the negative recognition group are not significant ($p=0.159$ or, respectively, $p=0.920$; Mann-Whitney test). In addition, we do not find significant performance differences either between the positive recognition group and the negative recognition group or between the positive recognition group and the no recognition group ($p=0.126$ or, respectively, $p=0.522$; Mann-Whitney test).

[14] All statistical tests are two-sided.

TABLE 1: Descriptive statistics of correctly cut out flyers by recognition type

	Control	Experimental	Neg. recognition	No recognition	Pos. recognition
All					
N	35	103	32	34	37
Correct flyers baseline	0.39	0.56**	0.54*	0.52	0.61**
	(0.30)	(0.34)	(0.31)	(0.37)	(0.34)
Correct flyers main task	0.86	0.95	0.84	1.03**	0.99
	(0.37)	(0.39)	(0.39)	(0.46)	(0.34)
Female					
N	15	50	17	18	15
Correct flyers baseline	0.51	0.64	0.56	0.59	0.77*
	(0.35)	(0.36)	(0.34)	(0.34)	(0.38)
Correct flyers main task	0.87	1.07**	0.93	1.17***	1.11**
	(0.40)	(0.37)	(0.30)	(0.47)	(0.27)
Male					
N	20	52	14	16	22
Correct flyers baseline	0.31	0.49**	0.53**	0.45	0.49**
	(0.24)	(0.31)	(0.29)	(0.40)	(0.25)
Correct flyers main task	0.84	0.85	0.75	0.88	0.90
	(0.36)	(0.37)	(0.33)	(0.40)	(0.37)

The table reports the means for each recognition group. The standard deviations are displayed in parentheses. The Mann-Whitney test is used for comparison. The significance levels are based on a comparison with the control group. One student did not report his or her sex.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Since important gender effects are documented in the literature examining reactions to relative performance feedback and competitive incentives, we further analyze whether gender matters in our setting as well. Importantly, a Mann-Whitney test shows that the guesses in the estimation task did not differ among genders (p=0.197), whereas flyer-cutting can be classified as a female-oriented task, because boys cut out fewer flyers correctly than girls, not only in the main stage but also in the baseline stage (p=0.003 or, respectively, p=0.098; Mann-Whitney test). Furthermore, our results show that in the control and the experimental group, females and males cut out significantly more flyers correctly in the main stage than in the baseline stage (in the control group for girls p=0.007 and for boys p=0.001; in the experimental group for girls and boys p<0.001; Wilcoxon signed-rank test). Interestingly, we find that in the control treatment the positive reaction of male students to pre-announced symbolic recognition was stronger than that of females (p=0.092; Mann-Whitney test). This finding is in line with previous literature that has revealed that men react stronger to competitive incentives than women (e.g., Gneezy et al. 2003, Gneezy and Rustichini 2004).

Next, we consider the effects of different recognition types by gender (see also Table 1). Mann-Whitney tests show that female non-recipients and females who received positive recognition in the estimation task cut out significantly more flyers correctly than female students in the control group

(p=0.010 or, respectively, p=0.046). Girls who received no recognition in the estimation task also performed significantly better than female students in the negative recognition group (p=0.041; Mann-Whitney test). However, the comparisons of the performance of females in the other groups are not significant (for smiley vs. frowny p=0.191, for no recognition vs. smiley p=0.425, for frowny vs. control p=0.382; Mann-Whitney test). Turning to the boys, we find that compared with the control group, male students did not significantly change their performance in response to previous recognition (for frowny p=0.409, for no recognition p=0.873, for smiley p=0.869; Mann-Whitney test). Furthermore, there are no significant performance differences between boys in the different recognition groups (for smiley vs. frowny p=0.368, for smiley vs. no recognition p=0.847, for no recognition vs. frowny p=0.478; Mann-Whitney test).

As recognition groups are not balanced with respect to baseline performance, a simple comparison of correctly cut out flyers in the second task is not sufficient.[15] Thus, to provide meaningful estimates of the spillover effect of different recognition types from one task to another unrelated task, we applied a regression model in which we control for baseline performance. We run regressions of the following default form:

$$y_{i,t=2} = \beta_0 + \beta_1 PosR_i + \beta_2 NoR_i + \beta_3 NegR_i + \beta_4 y_{i,t=1} + \beta_5 G_i + \beta_6 \mathbf{X}_i + \epsilon_i \tag{1}$$

where $y_{i,t=2}$ is the number of correctly cut out flyers per minute in the second (main) task and $y_{i,t=1}$ represents baseline performance. We consider three different rank dummies in the experimental group: the top third ($PosR_i$), who received a smiley-sticker, the intermediate third (NoR_i), who had no symbolic recognition but the information that their performance was average, and the bottom third ($NegR_i$), who received a frowny-sticker. In each case the omitted category is the control treatment. Furthermore, we include a gender dummy variable equal to one for males and zero for female (G_i). The vector $\mathbf{X}_i$ represents other control variables such as age, school type, and whether the students are right- or left-handed. The question of how the spillover effect of different recognition types varies by gender is examined by including interactions between the recognition type dummies and the gender dummy in the regression function. In all specifications, robust standard errors are clustered on individual levels.

[15] Students in the negative and in the positive recognition group cut out significantly more flyers correctly in the baseline stage than individuals in the control group (p=0.077 or, respectively, p=0.013; Mann-Whitney test). A manipulation check also shows that the recognition groups and the control group can be considered as balanced regarding gender, age, school type, right- or left-hander, and the performance in the main flyer-cutting task. Table 3 in Appendix B reports the descriptive statistics by treatment and recognition group. There are statistically significant differences regarding the estimation deviation between the negative recognition group and the control group (p<0.001; Mann-Whitney test) as well as between the positive recognition group and the control group (p< 0.001; Mann-Whitney test). However, these differences in the accuracy of estimation are caused by the experimental manipulation itself, and when comparing the average estimation deviation between the complete experimental group and the control group, the differences are not significant.

TABLE 2: Effect of different recognition types on correctly cut out flyers

	1	2	3	4	5	6
Negative recognition	-0.020	-0.075	-0.036	0.058	0.045	0.053
	(0.085)	(0.078)	(0.080)	(0.126)	(0.124)	(0.121)
No recognition	0.176*	0.126	0.147	0.299*	0.280*	0.300**
	(0.101)	(0.098)	(0.094)	(0.152)	(0.152)	(0.140)
Positive recognition	0.132	0.051	0.090	0.240*	0.179	0.216**
	(0.085)	(0.081)	(0.080)	(0.124)	(0.111)	(0.107)
Baseline performance		0.384***	0.304***		0.230	0.179
		(0.100)	(0.104)		(0.153)	(0.159)
Male			-0.139**	-0.028	-0.061	-0.108
			(0.064)	(0.131)	(0.177)	(0.174)
Negative recognition x Male				-0.154	-0.247	-0.185
				(0.173)	(0.159)	(0.167)
No recognition x Male				-0.268	-0.316	-0.314*
				(0.199)	(0.194)	(0.189)
Positive recognition x Male				-0.180	-0.207	-0.227
				(0.168)	(0.157)	(0.154)
Baseline performance x Male					0.254	0.296
					(0.196)	(0.198)
Secondary school			-0.006			-0.019
			(0.063)			(0.065)
Right-handed			0.010			0.010
			(0.100)			(0.099)
Age			0.060**			0.064**
			(0.028)			(0.029)
Constant	0.855***	0.704***	-0.030	0.871***	0.755***	-0.104
	(0.063)	(0.081)	(0.350)	(0.104)	(0.154)	(0.331)
N	138	138	135	137	137	135
R^2	0.047	0.156	0.216	0.114	0.203	0.247

The table shows the OLS estimates. The robust standard errors clustered on individual levels are in parentheses. The sample sizes differ because three students did not reveal the school type and one of them further did not report his or her sex and age. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2 presents the results of the OLS regression analysis, where the number of correctly cut out flyers in the main task is regressed on the different recognition types received in the estimation task. Specifications 4-6 display the OLS estimation when including the interactions between the recognition type dummies and gender. In specification 2 and 5, we additionally control for the baseline performance, and in specification 3 and 6, we further add gender, age, school type, and whether the students are right- or left-handed as control variables.[16] The results of specification 1

[16]Also with respect to gender, we find that there are statistically significant differences between the recognition groups and the control group in some of the covariates. The share of female upper secondary school students was

confirm the findings of the descriptive analysis and show that the non-recipients in the experimental group performed significantly better than the individuals assigned to the control treatment. However, when controlling for baseline performance, this effect diminishes and is no longer significant. Not surprisingly, the results of specifications 2 and 3 show that the baseline performance in flyer-cutting significantly influences the productivity in the main flyer-cutting task. Specification 3 also indicates that productivity significantly increases with age. In addition, the regression analysis reveals that males cut out significantly fewer flyers correctly than females.

The regression results regarding gender differences are in line with the descriptive analysis. Taking specification 6 as a benchmark, female students who received no recognition and females who were awarded a smiley-sticker in the estimation task cut out significantly more flyers correctly than female pupils in the control treatment. Girls in the no recognition group increased their performance by 0.30 and females in the positive recognition group by 0.22 correctly cut out flyers per minute. In contrast, the positive effect of a smiley-sticker as well as of no recognition is offset by negative interaction coefficients when we compare the performance of boys in the experimental group with those in the control group. In addition, the significant interaction between no recognition and being male indicates that boys cut out 0.31 fewer flyers than girls when they received only the information that their performance was average. Furthermore, specification 6 shows that performance slightly but significantly increased with age.

5 Discussion and conclusion

In an experiment with secondary school students we examine the ex post performance spillover effects of unannounced positive and negative recognition. We find that female non-recipients and female students who received a positive recognition (a smiley-sticker) in an estimation task increased their performance in the subsequent flyer-cutting task. These results remain statistically and economically significant when we control for baseline performance and demographics. In contrast, we do not find a significant performance effect for females who received negative recognition (a frowny-sticker). For male students we do not find a spillover effect.

A combination of reciprocity preferences and positive recognition serving as a goal and reference point is best able to explain our findings for females. First, reciprocity preferences have also been

slightly higher in the positive recognition group than in the control group. In addition, male students in the negative recognition group were on average younger than those in the control group. Furthermore, there are significant differences regarding the estimation deviation between boys in the control group and those in the experimental group. More critically, the baseline performance of boys was unbalanced across treatments. In the Appendix B, Table 4 shows descriptive statistics by treatment and recognition type for girls and boys separately. Due to the experimental manipulation, there are significant differences between the control group and the recognition groups regarding estimation deviations.

identified in Bradler et al. (2016) as a main driver for ex post performance spillover effects. Positive reciprocity can explain the increase in performance after receiving the smiley-sticker. The same argument predicts no effect for non-recipients and a decrease in performance for recipients of the frowny-sticker due to negative reciprocity. A stronger effect of positive reciprocity compared to negative reciprocity for females is in line with findings in Dohmen et al. (2009), who show that both types of reciprocity seem to be fundamentally different traits that are only weakly correlated. Furthermore, Buchan et al. (2008) and Dohmen et al. (2009) show that women have a higher tendency for positive reciprocity and lower tendency for negative reciprocity compared to men. Second, the smiley-sticker serving as a goal and reference point as in Berger and Pope (2011) predicts a positive effect for non-recipients and recipients of negative recognition due to loss aversion. Because of diminishing sensitivity, the positive effect should be stronger for non-recipients compared to recipients of the frowny-sticker (Heath et al. 1999, Berger and Pope 2011).

Besides reflecting acknowledgment, recognition programs also provide feedback on relative performance and entail a competitive element. Literature on both of these topics regularly finds strong gender effects with respect to performance (e.g., Gneezy et al. 2003, Günther et al. 2010, Barankay 2011, Delfgaauw et al. 2013, Gill and Prowse 2014, Jalava et al. 2015, Buser 2016, Buser and Yuan 2016). These differences are often driven by the specific institutional setting and the nature of the task. Closely related to our school setting is Jalava et al. (2015), who also find a positive effect of symbolic recognition for girls at school but not for boys. In addition, our main task of flyer-cutting can be considered as decorative work that is more female-oriented and girls were significantly better at the task than boys. Thus, the nature of the task could also be a driver of our results.

Prior research has focused mostly on the effects of positive recognition programs, while informal negative recognition is quite common in practice (Nelson 2012). However, formal negative recognition is less often observed. One reason might be the fear of retaliation produced by negative reciprocity. At least for our setting, we do not find such an adverse effect of negative recognition, potentially because it is offset by other motives such as goals serving as a reference point. In the paper closely related to our study, Neckermann et al. (2014) find a positive performance spillover effect for employees who received positive recognition in an unrelated task. We also find this effect for female students who received a smiley-sticker. In addition, we find a positive spillover effect for non-recipients. This additional effect could be due to the existence of negative recognition in our setting that was absent in the study of Neckermann et al. (2014). In our case, non-recipients know that they are not at the bottom of the distribution but in the middle and thus narrowly missed positive recognition. Due to diminishing sensitivity, this could strengthen the motivational effect of goals as reference points.

Next to symbolic recognition, our paper also contributes to the literature on gender differences in a competitive setting among adolescents (Sutter and Glätzle-Rützler 2015). As Dreber et al. (2014) mention, finding suitable incentive schemes and environments for this age group is especially important, since adolescents have to take decisions with long-term consequences. The right incentive system might contribute to reducing the still existing gender wage gap and occupational segregation.

Being among the first papers to examine the effects of both positive and negative recognition, there are several interesting avenues for future research. First, in order to get a better understanding of the drivers behind the gender effect, it would be interesting to consider a non-female oriented second task. This would allow to examine whether females respond strongly because of the existence of negative recognition or because of the female-oriented task. Also, to stay closer to the practice in firms, it is of interest to study the effect of informal public negative recognition. In addition, while we focus on ex post performance spillover effects, it would be interesting to consider the effect of ex ante announced negative recognition. Finally, although school students are an important and worthwhile sample to study, it would be interesting to test how adults react to different recognition types.

Acknowledgments

We are grateful to Roja Abbassi who conducted the experiment as part of her master thesis and the teachers and students of the school Möriken-Wildegg. Lucian Hunger, Katja Rubin and Flavio von Rickenbach provided excellent research assistance. We gratefully acknowledge financial support from Förderverein IOP. We would like to thank Florian Englmaier, Susanne Neckermann, Stefanie Jaussi, and participants at NIBS (University of Nottingham) for helpful comments. All errors are ours.

References

Akerlof, G. A. (1982). Labor contracts as a partial gift exchange. *Quarterly Journal of Economics*, 97(4):543–569.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization*, 100:44–63.

Auriol, E., Friebel, G., and von Bieberstein, F. (2016). The firm as the locus of social comparisons: Standard promotion practices versus up-or-out. *Journal of Economic Behavior & Organization*, 121:41–59.

Auriol, E. and Renault, R. (2008). Status and incentives. *RAND Journal of Economics*, 39(1):305–326.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7):435–452.

Bandiera, O., Barankay, I., and Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5):1079–1114.

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. *University of Pennsylvania*, unpublished manuscript.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Berger, J. and Pope, D. (2011). Can losing lead to winning? *Management Science*, 57(5):817–827.

Blanes i Vidal, J. and Nossol, M. (2011). Tournaments without prizes: Evidence from personnel records. *Management Science*, 57(10):1721–1736.

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, 62(11):3085–3099.

Buchan, N. R., Croson, Rachel T. A., and Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the investment game. *Journal of Economic Behavior & Organization*, 68(3–4):466–476.

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

Buser, T. and Yuan, H. (2016). Do women give up competing more easily? Evidence from the lab and the dutch math olympiad. *Tinbergen Institute Discussion Paper 16-096I*.

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

Datta Gupta, N., Poulsen, A., and Villeval, M. C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1):816–835.

Delfgaauw, J., Dur, R., Sol, J., and Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, 31(2):305–326.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2009). Homo reciprocans: Survey evidence on behavioural outcomes. *Economic Journal*, 119(536):592–612.

Dreber, A., Essen, E. v., and Ranehill, E. (2014). Gender and competition in adolescence: Task matters. *Experimental Economics*, 17(1):154–172.

Eriksson, T., Poulsen, A., and Villeval, M. C. (2009). Feedback and incentives: Experimental evidence. *Labour Economics*, 16(6):679–688.

Falk, A. and Zimmermann, F. (2016). Consistency as a signal of skills. *Management Science*, forthcoming.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics*, 108(2):437–459.

Frey, B. S. and Neckermann, S. (2008). Awards. *Journal of Psychology*, 216(4):198–208.

Gill, D. and Prowse, V. (2014). Gender differences and dynamics in competition: The role of luck. *Quantitative Economics*, 5(2):351–376.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, 118(3):1049–1074.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, 94(2):377–381.

Grosse, N. D. and Reiner, G. (2010). Explaining gender differences in competitiveness: Gender-task stereotypes. *Jena Economic Research Papers 2010-017*.

Günther, C., Ekinci, N. A., Schwieren, C., and Strobel, M. (2010). Women can't jump? - An experiment on competitive attitudes and stereotype threat. *Journal of Economic Behavior & Organization*, 75(3):395–401.

Heath, C., Larrick, R. P., and Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1):79–109.

Jalava, N., Joensen, J. S., and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior & Organization*, 115:161–196.

Kosfeld, M. and Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3):86–99.

Kube, S., Maréchal, M. A., and Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, 102(4):1644–1662.

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

Malmendier, U. and Tate, G. (2009). Superstar CEO's. *Quarterly Journal of Economics*, 124(4):1593–1638.

Neckermann, S., Cueni, R., and Frey, B. S. (2014). Awards at work. *Labour Economics*, 31:205–217.

Neckermann, S. and Yang, X. (2017). Understanding the (unexpected) consequences of unexpected recognition. *Journal of Economic Behavior & Organization*, 135:131–142.

Nelson, B. (2012). *1501 Ways to Reward Employees*. Workman Publishing Company, New York, 3rd edition.

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

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3:601–630.

Niederle, M. and Yestrumskas, A. (2017). Gender differences in challenge seeking: The role of institutionsl. *NBER Working Paper 13922*.

Shurchkov, O. (2012). Under pressure: Gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association*, 10(5):1189–1213.

Sutter, M. and Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, 61(10):2339–23354.

Tran, A. and Zeckhauser, R. (2012). Rank as an incentive: Evidence from a field experiment. *Journal of Public Economics*, 96(9-10):645–650.

von Bieberstein, F. and Jaussi, S. (2017). Perceived opportunities in the workplace: A new explanation fort he gender wage gap. *Working paper*.

6 Appendix A: Experimental instructions

(Original instructions are in German)

General instructions

Welcome to this experiment and thank you very much for your participation. Please read these general instructions carefully:

- The experiment consists of three parts and a short questionnaire.
- The single tasks are explained thoroughly with simple instructions. If you have any questions please raise your hand clearly. We will then come to your place and answer your questions.
- During the experiment you may not use any other devices than those that are mentioned in the following instructions. Please consider that you are not allowed to speak during the experiment. The use of mobile phones, smartphones, tablet-PCs, and so on is forbidden. Interferences lead to exclusion from the experiment.
- As a matter of course, all information is evaluated absolutely confidentially and anonymously.
- Provided that you do not breach these rules, you will receive compensation of 8 CHF.

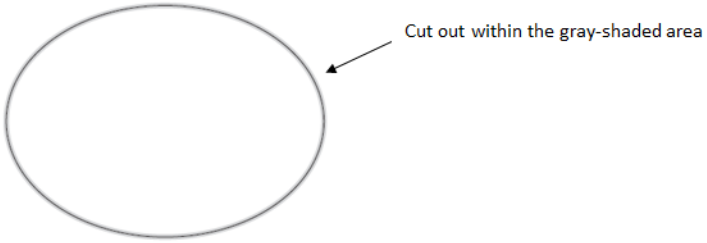
Thank you very much for your participation in the experiment.

Instructions Part 1 ***(baseline stage)***
For the first task you have scissors and a stack of A4 sheets with flyers on the desk. The Aarau Students' Orchestra (ASTOR) will play a concert in the church of Buchs next Sunday, on Mother's Day. To have some additional promotion, we kindly ask you to cut out the flyers according to the motifs. This evening, the members of the orchestra will distribute these flyers to the public.

We kindly ask you to cut out as many motifs as possible following the pattern during **5 minutes**.

- On the back of the A4-sheet you will see different motifs.
- Cut out the motifs within the marked frame (see the example picture below). Please cut out each motif individually: good quality can only be achieved like this. As we need the same number of each motif, it is especially important that you always cut out all four motifs on an A4-sheet and only then start the next A4-sheet.

- Place the finished motifs on the plastic plate on your desk. Please consider that only those motifs that have been cut out within the given frame will be counted and distributed in the end.



Please answer the following control questions by ticking the right answers:

1. Where do the finished motifs need to be placed?
O On the surface of the table O Into the plastic plate O Into the envelope

2. Which motifs will be counted at the end and distributed by the orchestra?
O All the motifs that have been cut out outside the frame
O All the motifs that have been cut out within the frame
O All the motifs that have been cut out

If you have any questions by now, please raise your hand. The experimenter will come to your place. Otherwise please wait until the experimenter calls to start cutting out.

Instructions Part 2 ***(estimation task)***

Please estimate the number of peas in the pictured bowl. You should estimate the amount of peas as exactly as possible. Enter your answer into the provided gap.

What do you estimate? How many peas are in the pictured bowl? ..... peas

In the case that you have any questions concerning this task, please raise your hand. The experimenter will come to your place. After you have made your estimation, please turn over the sheet. The experimenter will collect it afterwards.

[Note: After the estimation task, the experimenters in the experimental groups made the following announcement: *Great, you have successfully finished this task of the experiment. For your estimation we now bestow awards. The top third of the students in this room receives a smiley-sticker (show example) while the bottom third of all the students in this room receives a frowny-sticker (show example). The intermediate third of all the students in this room receives no sticker but a message that they have performed averagely. Your answers are now being evaluated so that we can afterwards assign the smiley- and frowny-stickers. As soon as the stickers have been distributed, we kindly ask you to stick them to your t-shirt for the rest of the experiment.*]

Instructions Part 3 ***(Flyer-cutting task, second stage)***

This is now the last part of the experiment. Again, you can find scissors and a stack of A4-sheets on the desk. As before, the flyers are for the Aarau Students' Orchestra (ASTOR). This evening they will be distributed by members of the orchestra. After reading the instructions carefully, please repeat the task of the first part during the next **15 minutes**.

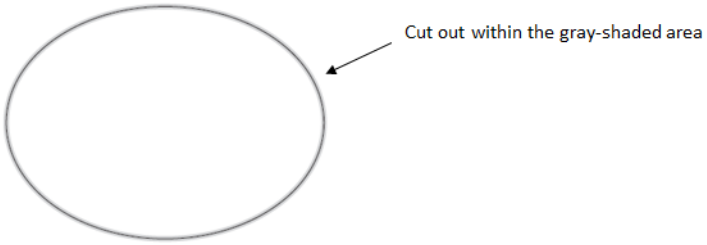
Award

Please consider that you can win an award after completing this task. The top third, meaning the students who cut out the most motifs within the given frame, will receive a smiley-sticker. The bottom third, meaning the students who cut out the fewest motifs within the given frame, will receive a frowny-sticker. The intermediate third will receive no recognition but a message that they have performed averagely.

Task

We kindly ask you to cut out as many motifs as possible following the pattern during the next 15 minutes (in exactly the same way as in part 1):

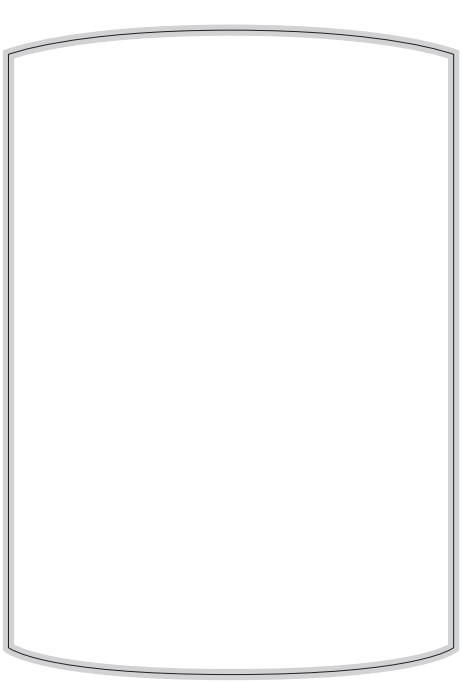
- On the back of the A4-sheet you will see different motifs.
- Cut out the motifs within the marked frame (see the example picture below). Please cut out each motif individually: good quality can only be achieved like this. As we need the same number of each motif, it is especially important that you always cut out all four motifs on an A4-sheet and only then start the next A4-sheet.
- Place the finished motifs on the plastic plate on your desk. Please consider that only those motifs that have been cut out within the given frame will be counted and distributed in the end.

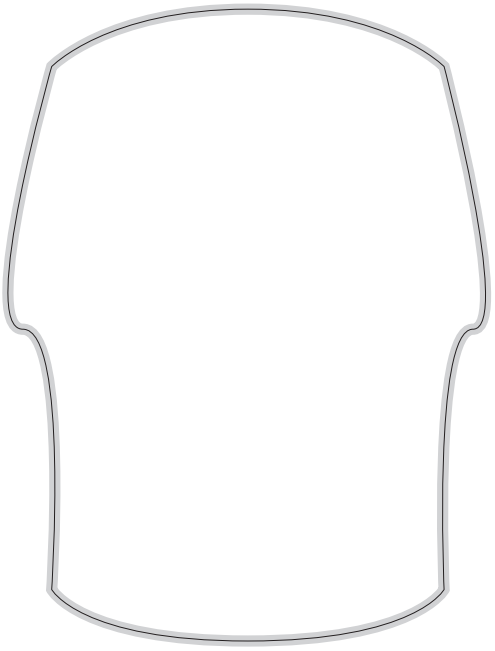


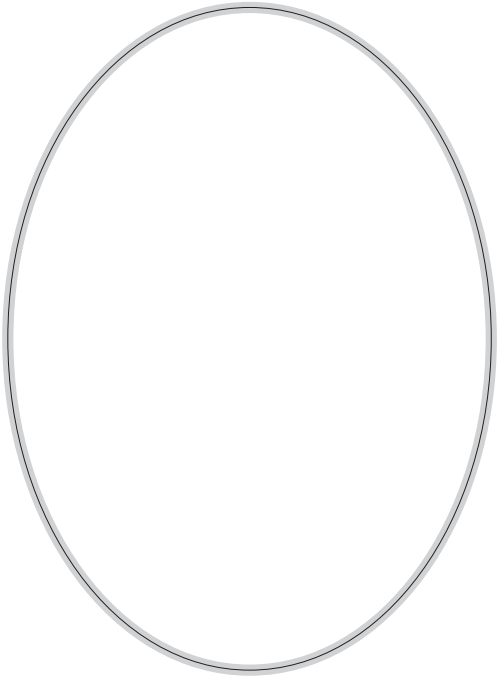
If you have any questions, please raise your hand. The experimenter will come to your place. Otherwise please wait until the experimenter calls to start cutting out.

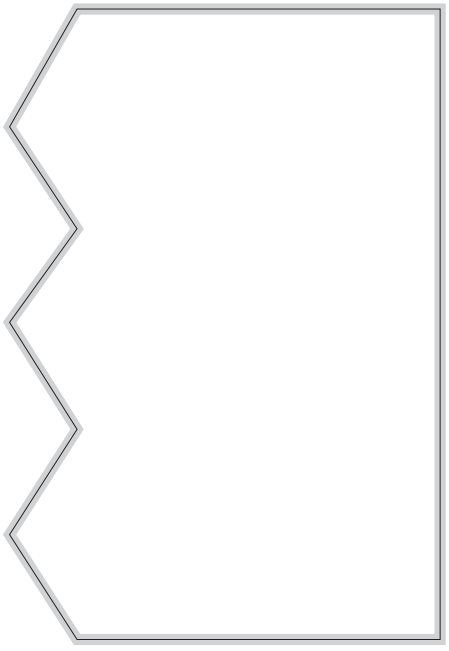
Front face of the flyers

Example of sheets with motifs









7 Appendix B: Further results

TABLE 3: Summary statistics by recognition type

	Control	Experimental	Neg. recognition	No recognition	Pos. recognition
N	35	103	32	34	37
Female	0.43	0.49	0.55	0.53	0.41
	(0.50)	(0.50)	(0.51)	(0.51)	(0.50)
Age	13.89	13.68	13.52	13.71	13.78
	(1.18)	(1.25)	(1.15)	(1.22)	(1.38)
Upper secondary school	0.53	0.53	0.58	0.38	0.64
	(0.51)	(0.50)	(0.50)	(0.49)	(0.49)
Right-hander	0.89	0.86	0.90	0.79	0.89
	(0.32)	(0.35)	(0.30)	(0.41)	(0.31)
Estimation deviation	2183.26	4.761.27	10759.03***	2433.82	1712.76***
	(650.96)	(22984.6)	(40967.68)	(145.89)	(481.65)
Flyers baseline	0.77	0.77	0.74	0.81	0.77
	(0.31)	(0.28)	(0.25)	(0.29)	(0.29)
Correct flyers baseline	0.39	0.56**	0.54*	0.52	0.61**
	(0.30)	(0.34)	(0.31)	(0.37)	(0.34)
Flyers main task	1.06	1.09	0.99	1.17	1.11
	(0.43)	(0.38)	(0.29)	(0.44)	(0.36)
Correct flyers main task	0.86	0.95	0.84	1.03**	0.99
	(0.37)	(0.39)	(0.39)	(0.46)	(0.34)

The table reports the means for each recognition group. The standard deviations are displayed in parentheses. The Mann-Whitney test is used for numerical data and the Chi-squared test for categorical data. The significance levels are based on a comparison with the control group. Three students did not reveal the school type and one of them further did not report his or her sex and age.
Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE 4: Summary statistics by recognition type and gender

	Control	Experimental	Neg. recognition	No recognition	Pos. recognition
Female					
N	15	50	17	18	15
Age	13.80	13.62	13.76	13.56	13.53
	(1.42)	(1.42)	(1.30)	(1.38)	(1.30)
Upper secondary school	0.40	0.50	0.47	0.33	0.73*
	(0.51)	(0.51)	(0.51)	(0.49)	(0.46)
Right-hander	0.93	0.84	0.82	0.78	0.93
	(0.26)	(0.37)	(0.39)	(0.43)	(0.26)
Estimation deviation	2473.93	7073.90	16,861.94**	2,411.28**	1,575.93***
	(491.40)	(32913.93)	(56238.84)	(157.64)	(550.75)
Flyers baseline	0.84	0.84	0.78	0.89	0.87
	(0.34)	(0.34)	(0.26)	(0.31)	(0.34)
Correct flyers baseline	0.51	0.64	0.56	0.59	0.77*
	(0.35)	(0.36)	(0.34)	(0.34)	(0.38)
Flyers main task	1.03	1.17*	1.06	1.25**	1.20
	(0.53)	(0.41)	(0.33)	(0.53)	(0.30)
Correct flyers main task	0.87	1.07**	0.93	1.17***	1.11**
	(0.40)	(0.37)	(0.30)	(0.47)	(0.27)
Male					
N	20	52	14	16	22
Age	13.95	13.73	13.21**	13.88	13.95
	(1.00)	(1.21)	(0.89)	(1.02)	(1.43)
Upper secondary school	0.63	0.57	0.71	0.44	0.57
	(0.50)	(0.50)	(0.47)	(0.51)	(0.51)
Right-hander	0.85	0.88	1.00	0.81	0.86
	(0.37)	(0.32)	(0.00)	(0.40)	(0.35)
Estimation deviation	1,965.25	2,573.38*	3,909.71***	2,459.19***	1,806.05
	(680.90)	(1,497.00)	(2,357.23)	(131.78)	(415.91)
Flyers baseline	0.72	0.71	0.71	0.71	0.71
	(0.29)	(0.23)	(0.22)	(0.25)	(0.23)
Correct flyers baseline	0.31	0.49**	0.53**	0.45	0.49**
	(0.24)	(0.31)	(0.29)	(0.40)	(0.25)
Flyers main task	1.08	1.02	0.92	1.08	1.05
	(0.35)	(0.33)	(0.22)	(0.30)	(0.40)
Correct flyers main task	0.84	0.85	0.75	0.88	0.90
	(0.36)	(0.37)	(0.33)	(0.40)	(0.37)

The table reports the means for each recognition group. The standard deviations are displayed in parentheses. The Mann-Whitney test is used for numerical data and the Chi-squared test for categorical data. The significance levels are based on a comparison with the control group. Three students did not reveal the school type and one of them further did not report his or her sex and age. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.