Effort and Social Comparison: Experimental Evidence from Uganda

Ben D'Exelle, Rik Habraken and Arjan Verschoor[1]

Abstract. Individual effort is key to generate income and escape from poverty. In small-scale societies in developing countries, where effort and resulting income are easily observable, social comparison can influence effort in both positive and negative ways. In addition, the effects of social comparison are less obvious where income depends on both the quality and quantity of production. To study the effects of social comparison, we use a lab experiment in which participants conduct a real-effort task characterized by a quantity-quality trade-off. We experimentally vary whether 1) feedback is provided about peers' performance and 2) one can reduce peers' earnings by so-called 'money-burning'. We find that feedback about peers' performance increases the earnings of low-performers as they shift effort from quantity to quality. High-performers only shift effort to quality when they are targeted by money-burning. Given the different response of low- and high-performers, the provision of feedback about peers' performance reduces inequality in earnings, but not when money-burning is an option. Anticipating these social comparison effects, can help policymakers who work with groups or communities to optimise policy design.

Keywords: Effort; social comparison; quantity-quality trade-off; money-burning.

JEL C91, D3, D8

---

[1] Acknowledgements: We thank The Field Lab Uganda for support with the fieldwork. We are grateful to Joshua Balungira, Floris Burgers, and the team of enumerators in Uganda for their assistance. The research documented in this article received ethical clearance from the ethics committee of the University of East Anglia. Rik Habraken acknowledges financial support from the University of East Anglia, where he was based when the work presented in the paper was carried out. Data are provided through Dataverse at https://doi.org/10.7910/DVN/ZHGUDU

1. Introduction

For designing policies that increase income and reduce poverty, it is crucial to know what drives individual effort. In small-scale societies in developing countries, such as rural or urban communities, groups or networks, effort and resulting income tend to be observable. This makes social comparison a potentially strong determinant of individual effort, which matters for the design of policy. For instance, education and training programmes frequently target groups of people who come together to improve their livelihoods (e.g., microfinance groups), with members originating from the same small-scale society. The effects of social comparisons within such groups may alter the influence of external interventions on the effort exerted by individuals. Anticipating this may help optimising the design of an intervention.

An extensive literature has documented the behavioural importance of social comparison. The inequality created by variation in effort can have a positive effect on individual effort as measured by, for example, worker productivity (Falk and Ichino, 2006; Eriksson et al., 2009; Blanes i Vidal and Nossol, 2011; Kuhnen and Tymula, 2012; Tafkov, 2013; Azmat and Iriberri, 2016; Gerhards and Siemer, 2016; Drouvelis and Paiardini, 2021) or school grades (Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012). Such an effect can be driven by so-called “behind-averse” or “ahead-seeking” motivations (Roels and Su, 2014; Gill et al., 2019). People try to minimize the differences with those who perform better (Agarwal et al., 2018), and once ahead, try to stay ahead of others (Bursztyn et al., 2018; Dohmen et al., 2011; Rustichini, 2008).

At the same time, growing evidence shows that social comparison can also make envious people reduce the earnings of those who are better off if they have the option to do so (Zizzo and Oswald, 2001; Zizzo, 2003; Dawes et al., 2007; Kebede and Zizzo, 2015; Fehr, 2018), and that the anticipation of such an envious reaction might reduce individual effort (Charness et al., 2014). Negative effects in developing countries have been documented mostly in the context of solidarity or informal insurance mechanisms (see e.g., Fafchamps, 1992). As argued by Platteau (2000), too large inequality among community members might be perceived as a threat, as the better-off might decide to break away from solidarity networks if they realize the required reciprocity fails to be balanced over time. Successful community members may therefore be looked at with suspicion.[2]

We see two important knowledge gaps that are relevant for small-scale societies in developing countries. First, while social comparison may have both positive and negative effects, little evidence exists on their separate and combined effect. Not only are we interested in effects on individual and total earnings, given the crucial importance of efficiency in communities characterized by high levels of poverty. We are also interested in effects on inequality. While small-scale societies are thought to have mechanisms in place that keep inequality within acceptable levels (see Platteau, 2000, above), few studies have looked at their effectiveness at doing so. Second, we do not know how social comparison influences effort and income where there is a quantity-quality trade-off in the production function, as is the case in many if not most tasks. Where income not only depends on the level of effort, but also on the allocation of effort between quantity and quality, the effects of social comparison are less obvious.

To fill these knowledge gaps, we use a lab experiment in Uganda, in which participants are organized in groups of four. Participants conduct an incentivized real-effort task, which

[2] Note that this mechanism is different from the pressure exerted by community members on successful members to share their profits, which is an essential element of solidarity mechanisms. Such income-sharing obligations can have a negative effect on productivity (Hadnes et al., 2013), liquid savings (Di Falco and Bulte, 2011), and information sharing about new productive activities (Di Falco et al., 2018), when community members try to avoid sharing obligations. D'Exelle and Verschoor (2015), in contrast, found that investment increased when an option was provided to share the investment output among friends.

involves counting the number of zeros in tables made up of ones and zeros. In this task, both quantity and quality are important, as participants only receive a payment for correctly solved tables. The same task is done twice. The first round creates inequality in output and earnings in each group. While this round is the same for all participants, we vary the second round, creating four between-subject treatments. Specifically, we use a factorial design that varies whether 1) feedback is provided about peers' performance in the first round and 2) individual earnings in the second round can be reduced by other group members through so-called 'money-burning' (Zizzo and Oswald, 2001). The comparison across these treatments allows us to test whether and how social comparison influences the allocation of effort to the quality and the quantity of production. We also investigate how the treatment effects interact with the participants' relative performance in the first round. This does not only provide insights in the heterogeneity of the effects of social comparison; it also helps us to understand potential changes in inequality at the group level.

We have three main results. First, giving feedback about individual performance increases the number of completed tables, but this translates nearly completely in an increase in incorrectly completed tables, so that it fails to increase earnings. It is only when we give feedback about the performance of other group members that participants manage to increase their earnings. They do so by shifting effort to the quality of the output. This effect is driven by respondents classified as 'low-performers', as measured by their performance in the first round. Second, it is only when the money-burning option is added that high-performers, who are the main target of money-burning, also shift effort to the quality of the production. Third, given the different response of low- and high-performers to relative performance feedback and the threat of money burning, inequality is reduced when feedback is provided about peers' performance, but not when it is combined with money-burning.

We make the following contributions to the literature. First, we contribute to the experimental literature that investigates the effect of social comparison on productivity in poor countries. Sseruyange and Bulte (2020) examine the effect of social comparison on effort in paid and unpaid tasks in rural Uganda. Bernard *et al.* (2014) in Ethiopia and Macours and Vakis (2014) in Nicaragua demonstrated that exposing the poorest in society to successful, similar people can increase the effort they invest in improving their livelihoods. Kebede and Zizzo (2015) investigate the negative effects of social comparison in Ethiopia, giving participants the option to reduce others' earning through so-called 'money-burning'. They find that money-burning levels correlate negatively with agricultural innovation at the village level. This suggests the existence of social preferences that are inimical to real-life agricultural innovations. We add to this literature an analysis that disentangles positive and negative effects of social comparison in a setting where both quantity and quality of production matter.

Second, a large literature has looked at the behavioural influence of social comparison using participant pools from the North (see the review above). Some studies have looked at quantity-quality trade-offs or tried to disentangle positive and negative effects of social comparison, but none have combined both. For example, Charness *et al.* (2014) disentangle positive and negative effects of social comparison in an effort task that uses a flat wage. They find that feedback about performance rank increases average effort in an effort task, while adding a money-burning option lowers effort. However, they do not distinguish quantity and quality of output. Other experimental papers have studied the allocation of effort where both quantity and quality of production matter (see e.g., Bracha and Fershtman, 2013; Rubin *et al.*, 2018), but none of them have looked at the positive and negative effects of social comparison.

A few papers come close to our design but differ in important ways. Eriksson *et al.* (2009) investigate the effect of feedback about peers' performance in a real-effort task. While they find some negative effects of relative performance feedback on the quality of the work of low performers, they ignore potential negative effects that work through anticipated envious reactions of peers. Carpenter *et al.* (2010) look at the effect of sabotage on the evaluation of the quality of production in a real-effort experiment. They find that effort increases in a tournament treatment, but not where others can sabotage evaluations. Their main contribution is an analysis of the effect of sabotage on the quantity and quality of performance, but they ignore the positive effects of performance feedback.

The rest of the paper is structured as follows. In Section 2, we present the experimental design, including the experimental task and treatments, the hypotheses, and a description of its implementation. Section 3 presents the results. Section 4 discusses the results and concludes.

2. Experimental design

2.1. Task and treatments

For the experiment, we matched each participant with three other randomly selected non-anonymous participants to mimic a small-scale society setting. Each participant was asked to individually count the number of zeros in a table of 110 (11x10) randomly ordered zeros and ones and record their answer in a designated space above each table (see Appendix F for an example).[3] No interaction with group members was allowed during the task. Participants had precisely 10 minutes to solve as many tables as possible out of 25 tables. Based on extensive testing, we confirmed that no prior knowledge was required to perform this task, performance was easy to measure, and the task was considered pointless, entailing a positive cost of effort for participants.

[3] For a similar task, see Abeler *et al.* (2011). As we expected that our participants had lower cognitive ability compared to the university students used in their study, we used a smaller grid (110 instead 150 cells) and added gridlines to the tables to lower the difficulty level of the task.

Earnings were individual and based on a piece-rate compensation scheme, which rewarded each correct answer. Each wrong answer incurred a small penalty, as we wanted to discourage subjects from ‘guessing’ the correct answer. If the number of zeros in each table was ‘guessed’ instead of counted, participants would not follow our instructions, according to which they were asked to ‘count’ the number of zeros. No feedback on performance was given *during* the 10 minutes of the task. Each participant did the task twice. The sets of tables of the first and second rounds were different, but the sequence in which we presented the tables was the same for all participants.

We use a between-subject treatment design with two rounds. In both rounds, all participants perform the same effort task. The differences between treatments arise from 1) whether relative performance feedback about round 1 is given at the start of round 2, and 2) whether participants can reduce the earnings other group members obtained in round 2. Table 1 presents the different treatments.

Table 1. Treatments

		IF vs RF	
Money-burning	No	IF	RF
	Yes	IF_MB	RF_MB

In a first treatment (IF), *individual feedback* is given at the end of round 1. Each participant receives information about the number of correct tables, the number of incorrect tables, and earnings obtained in round 1. We use this treatment to identify individual learning (as measured by the change between the two rounds), and as a control to compare the other treatments with.

In a second treatment (RF), *relative feedback* is provided at the start of round 2. In addition to feedback on how they performed individually (as provided in treatment IF), participants receive information on how each of the other three members in their group performed, in terms of the number of correct tables, the number of incorrect tables, and earnings. This information allows them to compare their performance with the other members in their group.

In a third treatment (RF_MB), participants are told at the beginning of round 2 that everyone will be offered the opportunity to lower the earnings other group members obtained in round 2, through so-called ‘money-burning’. Specifically, at the end of round 2, participants will receive information about the earnings each of the other group members obtained in round 2 and will be able to individually (and anonymously) reduce the earnings of specific group members.[4] The amount that they want to reduce can be different across group members. Participants can also decide not to reduce any earnings The cost of money-burning is 10% of the amount by which they want the earnings to be reduced. At the end of round 2, we randomly select one of the four group members’ money-burning decisions to be implemented.[5] To minimize the effect of anticipated post-experimental behaviour, they will not know whose money-burning decision is selected. All this information is given at the start of round 2.

In a fourth treatment (IF_MB), we keep the money-burning option but remove the relative performance feedback at the start of round 2 (but still provide individual performance feedback). A comparison with treatment IF allows us to test whether the threat of money-burning influences effort, without relative performance feedback. A comparison with RF_MB provides insight in a potential interaction between the effects of relative performance feedback and money-burning.

---

[4] As we expected that burning decisions would be driven primarily by earnings, information on the numbers of completed tables and correctly solved tables was not provided.

[5] As explained by Zizzo (2003), the advantage of selecting only one group member, is that a money-burning decision is independent of the expectations of the money-burning decisions of other group members.

After providing instructions at the start of round 1, four multiple-choice test questions are used to check whether subjects have correctly understood the main elements of the experiment. To keep the time between instructions and the actual task limited, the answers to the test questions were not assessed at individual level. Instead, the instructor provided the correct answers to the participants and explained each question. A similar procedure took place before the start of the second round.

2.2. Theory and hypotheses

2.2.1 Basic model

In this section, we show how effort allocation is optimized. In round 1, each participant decides on two-dimensional effort (e^1, e^2), with e^1 being the effort put in the number of tables solved and e^2 being the effort put in solving the tables *correctly*. The maximum amount of effort that they can allocate is E. The combination of both types of effort determines the expected number of *correctly solved* tables $Q = r(e^1).p(e^2)$, with production functions $r(.)$ and $p(.)$ determining the number of completed tables and the likelihood that the completed tables are solved correctly, respectively. Both functions are concave, such that $r'(.) > 0; r''(.) < 0$; $p'(.) > 0; p''(.) < 0$. In addition, each effort type generates a cost determined by the convex cost functions $c_1(e^1)$ and $c_2(e^2)$, with $c_1' > 0$, $c_1'' > 0$, $c_2' > 0$, $c_2'' > 0$. For each correctly completed table a non-zero pay-out w is received.[6]

Each participant chooses (e^1, e^2) such that it maximizes the expected profit $Q = w.r(e^1).p(e^2) - c_1(e^1) - c_2(e^2)$, s.t. $e^1 + e^2 \leq E$. Using the first-order conditions, and

[6] To simplify, we ignore the small penalty for each incorrectly solved table. The predictions would be the same if we included this penalty.

assuming that $e^1, e^2 > 0$ and the effort constraint is not binding, the effort allocation that maximizes the expected profit solves the following equation (see Appendix A.1 for details):[7]

$$w.r'(e^1).p(e^2) - w.r(e^1).p'(e^2) = c_1'(e^1) - c_2'(e^2) \qquad (1)$$

In other words, the optimal effort combination equalizes the differences in the marginal revenues and the marginal costs of both effort types. We next show how relative performance feedback and money-burning influence the optimal effort allocation.

2.2.2 Relative performance feedback and money-burning

There is abundant evidence that people not only care about their own earnings, but also how their earnings compare with others' earnings. Applied to our setting, individual effort allocation is not only influenced by expected individual earnings, but also by expected inequality between one's individual earnings and the earnings of each of the other group members. In Appendix A.2 we show that we can model this by multiplying the left-hand side of equation (1) by the term $(1 + \sigma)$, giving the following equation:

$$(1 + \sigma).\left[\left(w.r'(e^1).p(e^2)\right) - \left(w.r(e^1).p'(e^2)\right)\right] = c_1'(e^1) - c_2'(e^2) \qquad (2)$$

Parameter $\sigma > 0$ weighs the marginal utility (disutility) derived from being ahead (behind) of other group members. We assume that with feedback about the earnings of other group members after the first round, participants are made aware of the inequality with other group

[7] We ignore the possibility of corner solutions, as setting e^1 or e^2 equal to zero leads to zero profits. Given the small amounts of effort needed to solve one puzzle correctly and the substantial remuneration, one would be better off by solving at least one table correctly than setting e^1 or e^2 equal to zero. This is confirmed by the analysis, where we find that all participants solve at least one table correctly. We also assume that the effort constraint is not binding. See Appendix A.1 for the case where the effort constraint is binding. The predictions do not depend on this assumption.

members and become sensitive to the inequality in round 2. As a result, the left-hand side of equation (1) is multiplied with a factor larger than one. If the difference in the marginal revenues is different from zero, the optimal effort allocation changes. Specifically, to restore the equality of this equation, we should increase e^1 and/or decrease e^2 if the difference in marginal revenues is positive, and decrease e^1 and/or increase e^2 if the difference in marginal revenues is negative. This implies that more effort is allocated to the effort type that has the highest marginal revenue. As a result, the provision of relative performance feedback increases the number of correct tables and the earnings. This leads to our first hypothesis.[8]

Hypothesis H1. *Treatment RF leads to a larger increase in the number of correct tables and earnings relative to treatment IF.*

In treatment RF_MB, we add a money-burning option. We expect that money-burning is used to reduce inequality, by targeting group members with higher earnings, as demonstrated by Zizzo (2003) and Zizzo and Oswald (2001). We therefore hypothesise that money-burning will weaken the effect of relative performance feedback. The reasoning is that the larger the number of correct tables one produces the more likely it is that one will obtain higher earnings than the other group members, and hence be targeted by money-burning (for details see Appendix A.2). As a result, we expect $\triangle Q_{RF\_MB} - \triangle Q_{IF}$ to be smaller than $\triangle Q_{RF} - \triangle Q_{IF}$, or put differently, $\triangle Q_{RF\_MB} < \triangle Q_{RF}$. We test these predictions as a second hypothesis.

[8] Note that we do not elaborate hypotheses about the effort allocated to quantity and quality, as this depends on the specific functional form of $r(.)$ and $p(.)$. Also, whether one of the effort types needs to be decreased depends on whether the effort constraint is binding. If it is binding, the increase of one type requires the decrease of the other one. If it is not binding, it depends on the specific functional forms of the production and cost functions.

Hypothesis H2. *a) Money-burning targets high-performers; b) Treatment RF_MB leads to a smaller increase in the number of correct tables and earnings relative to treatment RF.*

Finally, in treatment IF_MB, we do not provide relative performance feedback, but keep the money-burning option. This allows us to investigate the interaction between relative performance feedback and money-burning. Without relative performance feedback, the participants would not know their relative position at the start of round 2. If money-burning mainly targets the high-performers (see hypothesis H2), they are less able to anticipate that they might be targeted by money-burning. This could reduce the effect of money-burning (for details see Appendix A.2). Put differently, we would only expect an effect of money-burning if it is combined with relative performance feedback. Therefore, we expect no difference between treatments IF and IF_MB. This leads to our third hypothesis.

Hypothesis H3. *The change in the number of correct tables and earnings is not different between treatments IF_MB and IF.*

2.3. Sampling and implementation

To implement the experiment, we used a sample of final-year students from a random selection of secondary schools in the urban Mbale district and rural Sironko district in eastern Uganda. The use of students in this area avoids illiteracy and cognitive ability issues that could affect the understanding of the experiment. To recruit participants, we randomly selected a set of classes, from which we randomly invited around 60% of the students in each class to participate in the experiment. A week before the experimental session, participants filled out a questionnaire at the school location, capturing important socioeconomic characteristics. Thereafter, students received an invitation to take part in the experiment.

As is the case with many secondary school students in Uganda, many of them are in boarding arrangement and schools are often far away from where parents or guardians are located. Therefore, asking their consent for students below the age of 18 would have been logistically challenging.[9] The invite that students received was accompanied by a statement indicating that participation should be entirely voluntary. At the start of the experiment the instructor would explicitly repeat that statement and students were entirely free to decide to not participate at any time throughout the experiment. Each student was then asked to give their written consent if they were at least 18 years old or assent if below 18 years old. The consent/assent form also stated clearly that students could withdraw at any time during the experiment without given reason. During the instructions and during the effort tasks, no student decided to leave.

To select classes, we used an exhaustive list of schools and classes provided by local officials. To ensure that each treatment is implemented at least once in each class, so that treatment assignment is orthogonal to class, a minimum class size of 30 students was set as a precondition for the inclusion of classes in the sampling frame.[10] Classes with more than 125 students were also excluded from the sampling frame to avoid the need to organize multiple sessions with participants of the same class, which might contaminate decisions across sessions. In addition, schools were only included if they offered mixed-gender education to allow the inclusion of boys and girls. Altogether, the selection criteria led to a sampling frame of 32 classes, from which we randomly selected 10.[11] To avoid contamination across

[9] In each school, consent to approach students and organise the effort experiment at the school location was asked from the school director and the dean of school.

[10] With 60% of the students in each class being recruited for the experiment, a class of 30 students gives a sample of 18 participants. As we need at least 16 students from each class to be able to do the four treatments in the same session/class, this gives us two spare students, which can be included if some of them do not show up.

[11] The sampling frame consists of 24 senior six classes in Mbale, and 8 Senior six classes in Sironko. Most secondary schools in rural Sironko have a limited number of students enrolled in Senior six or do not offer Senior

experimental sessions, classes belonging to the same school participated in the same experimental session. In total, we conducted eight experimental sessions. Sessions were organised at the school location and took place on Saturday morning to avoid interference with school activities.

In each experimental session, we implemented all four treatments. Students were randomly assigned to subgroups and treatments, which was organised in a reception room before the start of the experiment. Students were randomly organised into subgroups of four participants, after which subgroups were randomly allocated to treatments. To ensure that treatment assignment was orthogonal to class, each of the four treatments required a minimum of one subgroup of four students, which meant that a minimum of 16 students per class was needed in each session. To keep the sessions manageable, we did not allow more than four subgroups per treatment. In other words, we did not have more than 16 subgroups and 64 students in one experimental session.

To avoid contamination across treatments, each treatment was assigned to a separate classroom where participants received treatment-specific instructions. Members of the same group were seated next to each other in the classroom, but seats were sufficiently far apart to avoid that they could see what other group members were doing. In each classroom, we had two research assistants who explained the instructions. To minimize experimenter bias, we ensured that treatment assignment was orthogonal to research assistants by randomly assigning research assistants to one of the treatments at the start of each experimental session.

The experiment was run with pen and paper. After the instructions, but before the start of the experiment, participants received a pen and paper (calculators were not allowed). To

six education, which explains the lower number of classes in Sironko. From this sampling frame, we randomly selected five classes in Mbale from five different schools and five classes in Sironko from three different schools.

organize the performance feedback, we used information slips. Participants in treatments RF and RF_MB were shown the name of the other group members on the feedback information slip. For the money-burning in treatments RF_MB and IF_MB, we used decision cards that indicated the names of each of the other group members, so that money-burning could be targeted to specific group members. See Appendix E for the feedback slip and the money-burning decision card. For the experimental instructions, see Appendix D.[12]

At the end of the experimental session, we randomly selected one of the two rounds to determine the individual pay-outs. We informed participants at the start of the experiment that they would play the same game twice, but that only one round would be randomly selected for pay-out. At that time, they did not know whether and how the second round would differ from the first round.

Participants received 500 UGX (Ugandan Shillings) per correct answer and received a show-up fee of 2,000 UGX.[13] Per round, a maximum of 25 tables could be completed. To discourage guessing, a penalty of 100 UGX for each incorrect answer was applied, but earnings could never go below 0 UGX and penalties were not deducted from the show-up fee. On average, participants earned 6,676.09 UGX, including the show-up fee, which is slightly less than 2 US dollars. Since the daily wage for semi-skilled labourers is 4,000-5,000 UGX in this part of Uganda, this represents a substantial incentive for the students in our sample. The experimental sessions took approximately 1.5 to 2 hours in total, depending on the number of participants per session.

---

[12] The participants know with whom they are matched. As they are in the same class, they also know each other.

[13] At the time of the fieldwork, 1 USD was equal to 3,600 UGX.

2.4. The sample

We invited 332 students, of which 276 participated in the experiment (83%). We noticed that a small number of participants did not follow the rules of the experiment, as they tried to 'guess' the number of zeros in the tables rather than 'count' them, despite the penalty we applied for each incorrect answer. We classified 14 participants as 'guessers' and removed them from the analysis, leading to a final sample size of 262 participants.[14]

As participants were assigned to treatments randomly, differences in participants' characteristics between treatments should be due to chance. Table 2 shows that we cannot reject the null hypothesis that the treatments are balanced on a wide range of socioeconomic characteristics. We also find that effort levels in round 1, as measured by the number of completed tables and number of correctly completed tables, are balanced across treatments.

Table 2. Descriptive statistics of participants

	IF	RF	RF_MB	IF_MB	p-value
Female	33%	48%	36%	46%	0.218
Household wealth	0.20 (1.99)	-0.16 (1.84)	0.03 (1.74)	-0.04 (1.95)	0.743
Bagisu tribe	76%	65%	61%	68%	0.310
Number of siblings	6.84 (4.14)	5.81 (3.03)	5.86 (3.17)	5.91 (3.26)	0.243
Father higher education	47%	51%	53%	49%	0.917
Mother higher education	43%	46%	39%	34%	0.530
Round 1 completed tables [a]	13	13	12	14	0.172
Round 1 correct tables [a]	11	10	9	10	0.161
Number of subgroups	18	17	17	17	69
Number of participants	70	63	64	65	262

Notes. Two-sided p-values reported of an ANOVA test (for continuous variables) and a chi-square test (for categorical variables). For the wealth index we used the first factor of a principal component analysis, which uses the following items: independent house, owned house, burnt bricks with cement walls, cement floor, electricity, number of bikes, number of fridges, number of beds, number of radios, number of televisions, and number of mobile phones owned. [a] To minimize the influence of outliers, we use a non-parametric equality-of-medians test. As some participants were removed from the analysis, the number of participants per treatment is not necessarily a multiple of four.

[14] See Appendix C for details on how we identified them and some robustness tests.

2.5. Empirical strategy

To investigate the treatment differences, we compare the change in output (between the two rounds) between the control treatment IF and each of the other treatments.[15] To do so, we use the following regression specification:

$$\Delta Y_i = \alpha + \beta_1.FB_i + \beta_2.FB\_MB_i + \beta_3.IF\_MB_i + \varepsilon_i \qquad (1)$$

where ΔY_i is the change in output between both rounds generated by individual i. To measure output, we use the total number of completed tables, the number of correct tables and the number of incorrect tables. In addition, we look at the treatment differences in accuracy and individual earnings. The former is measured as the proportion of completed tables that is correctly completed. The earnings also include the penalty incurred for incorrect tables, but not the show-up fee. We distinguish the earnings before and after money-burning costs are deducted.

We include binary variables equal to 1 for treatments RF, RF_MB, and IF_MB, with IF as reference category. They capture the additional change in output between round 1 and round 2 in treatments RF, RF_MB, and IF_MB, relative to the control treatment IF. The constant term α captures the change in output in treatment IF. ε_i is the error term.

As experimental sessions took place at school level, and classes belonging to the same school participated in the same experimental session, we adjust the standard errors for non-independencies within schools. To do so, we estimate standard errors with bootstrapping, which provides a better method than clustered standard errors when the number of clusters is low (Cameron, Gelbach, & Miller, 2008).

---

[15] See Appendix B for descriptive statistics (Tables B.1 and B.2) and the distributions of the dependent variables by treatment (Figures B.1 – B.4).

3. Results

We start this section with a presentation of the treatment effects. Thereafter, we look at money-burning decisions, heterogeneity in the treatment effects and inequality at the group level.

3.1. Treatment differences

Table 3 compares the treatments. The constant term is statistically significant in Columns 1, 2 and 4. The size and sign of the coefficients indicate that in treatment IF the number of completed tables increases by an average of 2.3 in round 2 relative to round 1, while the number of incorrect tables increases by 1.8 tables. These results indicate that the increase in the number of completed tables translates nearly completely in an increase in incorrectly completed tables. This results in a decrease in accuracy of 7 percentage points.

The coefficients of RF estimate the additional change in treatment RF relative to treatment IF. Their statistical significance in Columns 3, 4, 5 and 6 confirm that the increase in the number of correctly completed tables and earnings is significantly larger, and the increase in incorrectly completed tables (marginally) significantly smaller in treatment RF relative to treatment IF.

Looking at the coefficients of RF_MB, we observe they are statistically significant in Columns 1 and 4. The sign of the coefficients indicates that the increase in the total number of completed tables and incorrectly completed tables is significantly lower in treatment RF_MB relative to treatment IF. Focusing on the size of the negative coefficient in Column 4, we see that in absolute terms it comes close to the positive coefficient of the constant term, which suggests that the increase in the number of incorrect tables in treatment IF disappears in treatment RF_MB. Finally, the results in the fourth row show that the change in the performance indicators is not statistically different between treatments IF_MB and IF.[16]

Table 3. Average treatment effects

	Completed tables	Accuracy [a]	Correct tables	Incorrect tables	Earnings [b]	Earnings [c]
	(1)	(2)	(3)	(4)	(5)	(6)
RF	0.133	0.038	0.838**	-0.705*	489.524**	489.524**
	(0.433)	(0.027)	(0.346)	(0.397)	(189.562)	(189.562)
RF_MB	-0.827**	0.077	0.676	-1.503**	488.259	282.009
	(0.381)	(0.059)	(0.778)	(0.730)	(453.967)	(417.396)
IF_MB	-0.266	0.086	1.073	-1.338	670.110	385.341
	(0.596)	(0.063)	(0.697)	(0.987)	(431.860)	(367.860)
Constant	2.343***	-0.071**	0.543	1.800***	91.429	91.429
	(0.379)	(0.029)	(0.406)	(0.411)	(228.737)	(228.737)
R-squared	0.021	0.031	0.023	0.034	0.026	0.014
RF = RF_MB [d]	0.009	0.311	0.793	0.076	0.997	0.514
RF = IF_MB [d]	0.093	0.347	0.684	0.372	0.612	0.717
RF_MB = IF_MB [d]	0.258	0.802	0.184	0.801	0.363	0.546

Notes. N = 262. OLS regressions. Difference between both rounds of the measure presented in first row used as dependent variable. Bootstrap clustered standard errors (in parentheses) at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. [a] Defined as number of correct tables divided by number of completed tables. [b] Earnings before money-burning costs are included. [c] Earnings after money-burning costs are included. The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in this variable, while the guessers' earnings after money-burning are not included in this variable. [d] Two-sided p-value of a chi-square test.

The lower panel of the table compares treatments RF, RF_MB and IF_MB. We observe that the decrease in the number of completed tables is significantly larger in RF_MB compared to RF and marginally significantly larger in IF_MB compared to RF. In addition, the decrease

[16] As described in footnote 12, all analyses exclude 'guessers'. However, as they might also influence the behaviour of non-guessers in the same group, it is important to test whether the results are robust after excluding *all* participants in subgroups with at least one guesser. Table B.3 in Appendix B shows that the treatment effects are robust.

in the number of incorrectly completed tables is marginally significantly larger in RF_MB than in RF. We summarize these findings in a first result:

Result 1. *The increase in earnings and the number of correct tables between both rounds is larger in RF than in IF. The decrease in the number of completed tables is larger in treatment RF_MB relative to both IF and RF. The decrease in the number of incorrect tables is larger in treatment RF_MB relative to RF.*

3.2. Money-burning

In this section, we analyse the money-burning decisions. Specifically, we test whether: 1) high-performers are more targeted by money-burning than low-performers, and 2) money-burning is equally common in the two treatments with money-burning.[17] The first analysis tests an important assumption that we used in our theoretical model, and on which the predicted effect of money-burning on effort is based. The second analysis can provide further insights in the interaction between the effects of relative performance feedback and money-burning on effort.

In treatment RF_MB, money-burning occurred in 21% of the decisions. On average, targeted participants' earnings were reduced by 1260 UGX per burning decision. In treatment IF_MB, money-burning was slightly less common. Participants decided to reduce the earnings of others in 19% of the decisions, and an average of 1231 UGX per decision was burned. However, a test of proportions reveals that there are no statistically significant differences in the decision to use money-burning between treatments RF_MB and IF_MB (two-sided p-value

[17] An analysis of the participants' beliefs of the likelihood that they would be affected by money-burning could provide important insights, as it is the 'anticipated' money-burning that influences effort. However, we did not elicit these beliefs, as doing so might influence effort if these beliefs are elicited before the start of round 2 or before they take a money-burning decision (through experimenter demand effects), or they might be biased by post-rationalization if they are elicited after the money-burning decisions are taken.

= 0.710, N = 408) while a t-test reveals no statistically significant differences in the amount of money burned (two-sided p-value = 0.895, N = 81).

Table 4. Inequality in earnings in round 2 as determinant of money-burning

	Burning decision			Amount burned (in UGX)		
	RF_MB	IF_MB	RF_MB + IF_MB	RF_MB	IF_MB	RF_MB + IF_MB
	(1)	(2)	(3)	(4)	(5)	(6)
Positive distance	-0.092***	-0.039**	-0.039**	-91.563**	-69.065**	-69.065**
	(0.024)	(0.017)	(0.017)	(42.769)	(30.024)	(30.174)
Negative distance	0.050**	0.002	0.002	141.682**	5.343	5.343
	(0.021)	(0.015)	(0.015)	(55.819)	(22.941)	(23.137)
RF_MB * Pos. distance			-0.053*			-22.498
			(0.029)			(52.219)
RF_MB * Neg. distance			0.047*			136.339**
			(0.027)			(58.725)
Constant	0.256***	0.236***	0.246***	200.301**	313.197***	256.749***
	(0.063)	(0.054)	(0.042)	(93.574)	(104.496)	(70.140)
Observations	204	204	408	204	204	408
R-squared within	0.261	0.072	0.191	0.212	0.101	0.182

Notes. OLS regressions with individual fixed effects. Bootstrap clustered standard errors (in parentheses) at school level, using 2000 iterations. *** p < 0.01, ** p < 0.05, * p < 0.10. To calculate the positive (negative) distances in earnings in round 2, we set negative (positive) differences equal to zero. The distance variables measure the absolute differences between the decision-maker and the paired person, divided by 1000. Decisions that involve 'guessers' (i.e., their money-burning decisions and money-burning targeted at them) are included.

Next, we test whether money-burning is more likely used by low-performers to target the high-performers than by high-performers to stay ahead of others. To test this hypothesis, we regress (a) the decision to reduce the earnings of another group member (yes/no), and (b) the amount of earnings burned, on the absolute positive and negative distances in earnings in round 2 between the decision-maker and the other group member. The positive distance in earnings indicates the extent to which the decision-maker's earnings are higher than the other group member. If the latter's earnings are higher, the value is set to 0. Similarly, the negative distance in earnings indicates the extent to which the earnings of the decision-maker are lower than the earnings of the other group member, with a value of 0 if the earnings of the decision-maker are higher than the other group member. The results are presented in Table 4.

In the first row, we observe that with each 1000 UGX that the decision-maker earned more than the other group member, the likelihood that money-burning is used decreases with 9% and 4% in RF_MB and IF_MB, respectively. A similar difference is observed in the amount that is burned (Columns 4-5). In addition, the negative distance has a positive effect on the likelihood that money-burning is used and on the amount of earnings reduced in RF_MB, as evidenced by two statistically significant coefficients in the second row of Columns 1 and 4. These findings confirm our hypothesis that money-burning is more likely used to reduce the inequality with group members who obtain higher earnings.

The effects are stronger in RF_MB than in IF_MB, as shown by the interaction terms in Columns 3 and 6. We therefore conclude that relative feedback on previous performance increases the likelihood that money-burning is used as well as the amount of money-burning. This suggests that the ineffectiveness of money-burning at influencing effort when it is not preceded by relative performance feedback (which we observed in Table 3) could not only be due to participants being unsure about the chance they end up highly ranked in round 2, and hence whether they could become a target of money-burning. It could also be driven by an anticipation that money-burners would be more forgiving if high-performers did not know they were highly ranked in round 1, and thus their high rank in the second round is less the result of their deliberate choice. We summarize these new observations in a second result.

Result 2. *A group member is less (more) inclined to reduce the earnings of another group member, the larger the positive (negative) difference in earnings in round 2 with that person. Money-burning is more common when it is preceded by relative performance feedback.*

3.3. Heterogeneity: high-performers versus low-performers

The average treatment effects might conceal important heterogeneity. An important source of heterogeneity relates to the performance in the first round. We see two reasons why the treatment effects might vary along this dimension. First, the effect of the provision of feedback about the performance of peers might differ between high- and low-performers, if ahead-seeking behaviour is weaker than behind-averse behaviour. Second, given that we found that money-burning tends to target peers who perform better, we expect that high-performers are more influenced by money-burning.[18]

To define high- and low-performers, we use the participants' ranks in their group based on their earnings in round 1. We define high-performers as participants who ranked first or second in their group, while low-performers ranked third and fourth. High-performers have average earnings in round 1 that are 2,389 UGX higher than low-performers. This difference is statistically significant (two-sided p-value of a t-test = 0.000). We also find that high-performers completed on average 4.30 more tables correctly (two-sided p-value of a t-test = 0.000) than low-performers. In addition, they completed 1.89 more tables (two-sided p-value of a t-test = 0.000) and made 2.41 fewer mistakes in round 1 (two-sided p-value of a t-test = 0.000). They also worked more accurately (mean difference in accuracy = 0.22, two-sided p-value of a t-test = 0.000).

To identify differences in treatment effects between high- and low-performers, we use the same regression as we used before, and we add an interaction between the set of explanatory variables and a dummy variable equal to one for high-performers, zero otherwise. The results are presented in Table 5.

Looking at control treatment IF (last row in both panels), we observe that high- and low-performers experience a statistically significant increase in the number of completed tables between both rounds (Column 1). On average, low-performers complete 2.9 tables more in

---

[18] This might also explain why some of the coefficients in Table 3 that aimed to identify average treatment effects had large standard errors, especially in treatments with money-burning.

round 2 than in round 1, while high-performers complete 1.8 tables more. This difference between high- and low-performers is statistically significant, as indicated by the significant coefficient of ‘High performer’ in panel a. This indicates that in round 2 the low-performers reduce the difference with high-performers in terms of the number of completed tables. Remember that on average they completed 4.30 tables less in round 1.[19] Both high- and low-performers also significantly increase the number of incorrectly completed tables. We also observe that high-performers experience a significant decrease in accuracy and low-performers obtain a marginally significant increase in the number of correct tables. However, the effects on these three performance indicators are not significantly different between high- and low-performers.

We now look at the changes in treatments RF, RF_MB and IF_MB relative to treatment IF. Starting with the high-performers (Panel b), we observe that the coefficient of ‘RF + RF x High performer’ is statistically significant. The sign and size of this coefficient indicates that the increase in completed tables is 1.1 tables larger in treatment RF than in treatment IF. We also find that the increase in completed tables is significantly larger in RF compared to RF_MB (chi2 = 3.92, p-value = 0.048) and IF_MB (chi2 = 17.42, p-value = 0.000).

Looking at the other performance indicators, none of the coefficients are statistically significant, which confirms that the change in these indicators is not significantly different from treatment IF. However, we do find significant differences among treatments RF, RF_MB and IF_MB. The change in incorrectly solved tables is statistically different between RF and RF_MB (chi2 = 6.78, p-value = 0.009) and between RF and IF_MB (chi2 = 5.58, p-value =

---

[19] To the extent that at least part of the performance is random, this could be influenced by “regression towards the mean”. While we acknowledge the latter possibility, we expect that this affects all treatments equally. As such, it does not confound the treatment effects or the heterogeneity in treatment effects.

0.018). The difference in the change in accuracy between RF and RF_MB is also statistically significant (chi2 = 4.27, p-value = 0.039). The sign of these differences indicates that the increase in incorrect tables is significantly lower in treatments RF_MB and IF_MB relative to treatment RF, and the reduction in accuracy is significantly lower in treatment RF_MB relative to treatment RF.

Table 5. Treatment effects by rank in earnings in round 1

	Completed tables	Accuracy [a]	Correct tables	Incorrect tables	Earnings [b]	Earnings [c]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel a						
RF	-1.015*	0.130**	1.305*	-2.320***	884.532**	884.532**
	(0.563)	(0.052)	(0.674)	(0.627)	(379.724)	(379.724)
RF_MB	-1.651*	0.132	0.856	-2.507**	678.558	562.429
	(0.965)	(0.103)	(1.356)	(1.188)	(768.057)	(741.736)
IF_MB	-0.586	0.129	1.211	-1.797	785.009	510.816
	(0.863)	(0.101)	(1.181)	(1.512)	(719.868)	(679.323)
High performer	-1.163**	-0.041	-1.232	0.069	-622.876	-622.876
	(0.525)	(0.063)	(0.860)	(0.836)	(500.058)	(500.058)
RF x High performer	2.126***	-0.157**	-0.694	2.820***	-628.976	-628.976
	(0.635)	(0.064)	(1.024)	(0.794)	(576.349)	(576.349)
RF_MB x High performer	1.600	-0.106	-0.346	1.946*	-367.447	-542.227
	(1.208)	(0.088)	(1.222)	(0.995)	(658.987)	(664.300)
IF_MB x High performer	0.632	-0.081	-0.243	0.875	-209.192	-229.411
	(0.818)	(0.093)	(1.268)	(1.449)	(758.317)	(764.614)
Constant	2.941***	-0.050	1.176*	1.765**	411.765	411.765
	(0.595)	(0.058)	(0.715)	(0.744)	(410.927)	(410.927)
Panel b (treatment effects on high-performers) [d]						
RF	1.111**	-0.027	0.611	0.500	255.556	255.556
	(0.485)	(0.030)	(0.549)	(0.489)	(305.160)	(305.160)
RF_MB	-0.051	0.026	0.510	-0.561	311.111	20.202
	(0.329)	(0.021)	(0.391)	(0.420)	(225.690)	(176.650)
IF_MB	0.046	0.047	0.967	-0.922	575.817	281.405
	(0.545)	(0.045)	(0.613)	(0.864)	(377.954)	(319.855)
Predicted within-change in control treatment (IF)	1.778***	-0.091***	-0.056	1.833***	-211.111	-211.111
	(0.271)	(0.019)	(0.435)	(0.382)	(248.717)	(248.717)
R-squared	0.049	0.150	0.109	0.109	0.117	0.118

Notes. N = 262. OLS regressions. Difference between both rounds of the measure presented in first row used as dependent variable. Bootstrap clustered standard errors (in parentheses) at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Rank is based on the earnings of round 1 within each subgroup. [a] Defined as number of correct tables divided by number of completed tables. [b] Earnings before money-burning costs. [c] Earnings after money-burning costs. The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in the calculation of this variable. [d] Predictions of treatment effects for high-performers, using the pooled sample.

Continuing with the low-performers (Panel a), we find that low-performers experience a larger increase in the number of correct tables in treatment RF than in treatment IF, which is marginally significant. This, together with a significantly lower increase in the number of incorrectly solved tables and a significantly larger increase in accuracy, translates in a significantly larger increase in earnings. In treatment RF_MB – similar to treatment RF – they experience a (marginally) significantly lower increase in completed tables and a significantly lower increase in incorrectly solved tables. We do not find any significant differences between treatments IF_MB and IF. Comparing the effects among treatments RF, RF_MB and IF_MB, we find no statistically significant differences.

Comparing high and low-performers (Panel a), we observe that the coefficient of 'RF x High performer' is statistically significant in Columns 1, 2 and 4. The sign and size of this coefficient in Column 1 indicates that the positive effect of relative performance feedback on the number of completed tables is higher among high-performers than among low-performers, while its positive effect on accuracy and negative effect on the number of incorrect tables is stronger among low-performers compared to high-performers. When money-burning is added to relative performance feedback (treatment RF_MB) the differences in the change in performance between high- and low-performers become smaller (incorrect tables) or disappear (completed tables, and accuracy). When money-burning is used without relative performance feedback (treatment IF_MB) no differences in the change in performance are identified between high- and low-performers.[20] We summarize the main findings in a third result.

[20] To test whether the results depend on how we define high- and low-performers, Tables B.4 and B.5 in Appendix B run the same regressions with alternative definitions. The effects in treatment FB among low-performers are stronger when we define high- and low-performers by the quantity produced in the first round, instead of earnings (Table B.4). The difference between high-performers and low-performers on the number of completed tables in

Result 3. *High-performers experience a larger increase in the number of completed tables in RF relative to IF, while low-performers experience a larger increase in earnings, which is the result of a larger increase in accuracy. The differences in the performance change between high- and low-performers become smaller or disappear when money-burning is used (RF_MB).*

3.4. Treatment differences in inequality

The positive and negative effects of social comparison might not only influence individual earnings. It is also important to look at their effects at the society level. Specifically, the treatment differences in earnings that we identified, which vary between high- and low-performers, may translate in treatment differences in intra-group inequality in earnings. To investigate this, we calculate the absolute distance between individual earnings in round 2 and the group mean, and we compare this across treatments. Table 6 presents the results.

We observe that inequality is significantly lower in treatment RF compared to each of the other treatments. This suggests that relative performance feedback without money-burning has an inequality-reducing influence. This finding is in line with the previous observation that low-performers manage to increase their earnings in treatment RF, while high-performers are unable to do so. Inequality in RF is also lower than in RF_MB. This indicates that the inequality reducing effect of relative performance feedback disappears once it is combined with money burning. This is consistent with the previous finding that differences in performance between

---

this treatment disappears, while the difference in the number of correct tables and earnings is now statistically significant. Defining high- and low-performers by their accuracy in the first round (Table B.5 in Appendix B), in contrast, weakens the differences between high- and low-performers.

high- and low-performers become smaller or disappear when money-burning is used (RF_MB). We summarize these new findings in a fourth result.

Result 4. *Intra-group inequality in earnings is lower in treatment RF compared to each of the other treatments.*

Table 6. Inequality within groups

	Earnings before money-burning		Earnings after money-burning [a]	
RF – IF	-311.164**	(156.128)	-311.164**	(156.128)
RF_MB – IF	106.726	(175.063)	48.653	(173.351)
IF_MB – IF	243.810	(177.791)	234.604	(177.843)
RF – RF_MB	-417.890***	(154.182)	-359.817**	(152.015)
RF – IF_MB	-554.974***	(158.119)	-545.768***	(158.184)
RF_MB – IF_MB	-137.083	(178.267)	-185.951	(76.512)

Notes. Pairwise treatment comparison of the average absolute distance between individual earnings and the group mean earnings in round 2. Standard errors (in parentheses) reported of a t-test. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Number of observations per treatment: IF: N = 70, RF: N = 63, RF_MB: N = 64, IF_MB: N = 65. [a] The money-burning decisions of 'guessers', and the cost of money-burning of non-guessers targeting guessers are included in this variable. The guessers' earnings are not included in this variable.

4. Discussion and conclusion

In this section, we discuss the results and conclude. We first assess whether the empirical evidence supports our hypotheses, after which we discuss some findings for which we did not have any hypotheses. In the theoretical section, we presented three hypotheses. First, under hypothesis H1 we predicted that relative performance feedback would increase the number of correct tables and earnings. This hypothesis is supported by the evidence from a comparison between treatments RF and IF. Second, the observation that the addition of a money-burning option in treatment RF_MB weakens the decrease in incorrect tables observed in RF provides some support for hypothesis H2. Third, the observation that money-burning without relative performance feedback (treatment IF_MB) does not increase the number of correct tables and earnings relative to control treatment IF is in line with hypothesis H3. Our results contribute to the existing body of literature on the effects of social comparison (see the references listed in the introduction), by disentangling positive and negative effects of social comparison in a setting where both quantity and quality of production matter.

We also identified interesting findings for which we did not develop any hypotheses. First, we found heterogeneity in the treatment effects, along the individual performance in the first round. Specifically, the larger increase in earnings in treatment RF relative to treatment IF is driven by low-performers. This allows low-performers to catch up with others in their group. The few papers that investigated how the effects of social comparison vary between high and low-performers obtained mixed results. For example, Gill *et al.* (2019) find in a real-effort experiment that the positive effect of the provision of rank-order relative performance feedback is highest among the lowest and highest-ranked participants in a group. Eriksson *et al.* (2009), described in the introduction, in contrast, found that feedback had a negative effect on the quality of production among low-performers. They attribute this to increased stress among low-performers when they realise that they are lagging behind, which then increases the number of mistakes they make. Interestingly, we find the opposite effect: low-performers increase the quality of output when they receive feedback about their peers' performance.

Second, comparing the treatment effects on quantity and quality of output, for the entire sample and separately for low- and high-performers, we found opposing effects on quantity and quality of production. While in treatment IF, the quantity of output increases at the expense of quality between both rounds, in treatment RF the quality increases at the expense of quantity. Importantly, it is only in the RF treatment that respondents manage to increase their earnings. Our theoretical model cannot explain this change in effort allocation in treatment IF and the differences between treatments IF and RF.

A potential mechanism that could explain these findings is incomplete information about the quantity-quality trade-off. As the task is new for the participants, they need to learn by their own experience (treatment IF) or the experience of others (treatment RF) how the number of correct tables and consequent earnings can be increased. Specifically, they need to obtain an accurate estimate of the marginal revenues of both effort types. While individual feedback (treatment IF) turns out to be insufficient to increase their earnings, it is through social learning in treatment RF that participants obtain a more accurate view of the marginal revenues of the effort put in quality and quantity of production. They realize that they underestimated the marginal revenue of the effort put in quality versus the effort put in quantity, which makes them shift more effort to quality in treatment RF.

Interestingly, this is only observed among low-performers. While both low- and high-performers tend to focus on quantity in treatment IF, it is only low-performers who shift effort to the quality of production in treatment RF, while high-performers increase the effort in quantity even further. This leads to an increase in earnings among low-performers, while high-performers fail to increase their earnings. By comparing their performance in round 1 with the other group members, low-performers realize that effort in quality has the highest marginal revenue, which makes them allocate more effort to the quality of the output. High-performers do not have such learning opportunity, as their earnings are higher in the first round. Relative performance feedback in treatment RF pushes them to increase the number of completed tables even further, while the addition of a money-burning option in treatment RF_MB weakens this effect. Specifically, money-burning leads to a lower decrease in accuracy among high-performers, who are the main target of money-burning.

There are only a handful of studies that have looked at effort provision where there is a quantity-quality trade-off (see e.g., Bracha and Fershtman, 2013; Rubin *et al.*, 2018). We have added the insight that the provision of relative performance feedback increases earnings where there is a quantity-quality trade-off. The size of this effect and how it changes the allocation of both types of effort crucially depends on the difference in marginal revenues of the effort types. Where people have incomplete information about these marginal revenues, individual and social learning become important. As this was not a mechanism we had anticipated with our design and theoretical model, further research that analyses belief updating about the marginal revenues of different effort types could provide support for the interpretation of our findings.

Third, our results also contribute new insights to the literature on inequality mitigating mechanisms in small-scale societies (on this see e.g., Platteau (2000) in the introduction) in the following two ways. First, our finding that money-burning is more prevalent if high-performers knew they were highly ranked in round 1 (treatments RF_MB versus MB), indicates that it is not aversion to inequality per se that triggers a negative response. It suggests a fairness norm that prescribes that inequality should be challenged if it is the result of one's deliberate action to get ahead of others. Second, we found that money-burning is little effective at reducing inequality. Social comparison without money-burning is very effective at reducing inequality, but its effect disappears once it is combined with money-burning. This suggests that the negative response driven by norms around inequality that are thought to limit inequality miss their goal and are thus less effective at limiting inequality than suggested by previous literature.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101: 470-92.

Agarwal, Sumit, Vyacheslav Mikhed, and Barry Scholnick. 2018. "Does the Relative Income of Peers Cause Financial Distress? Evidence from Lottery Winners and Neighboring Bankruptcies." Working paper 18-16: Federal Reserve Bank of Philadelphia.

Azmat, Ghazala., and Nagore Iriberri. 2010. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students." *Journal of Public Economics* 94(7-8): 435-52.

Azmat, Ghazala, and Nagoe Iriberri. 2016. "The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction." *Journal of Economics & Management Strategy* 25(1): 77-110.

Bandiera, Oriana, and Imran Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique." *Economic Journal* 116(514): 869-902.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse. 2014. "The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia." CSAE Working Paper Series 2014-16. Oxford: Centre for the Study of African Economies, University of Oxford.

Blanes i Vidal, Jordi, and Mareike Nossol. 2011. "Tournaments without Prizes: Evidence from Personnel Records." *Management Science* 57(10): 1721-36.

Bracha, Anat, and Chaim Fershtman. 2013. "Competitive Incentives: Working Harder or Working Smarter?" *Management Science* 59(4): 771-81.

Bursztyn, Leonardo, Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao. 2018. "Status Goods: Experimental Evidence from Platinum Credit Cards." *The Quarterly Journal of Economics* 133(3): 1561-95.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90(3): 414-27.

Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm. 2010. "Tournaments and Office Politics: Evidence from a Real Effort Experiment." *American Economic Review* 100(1): 504-17.

Charness, Gary, David Masclet, and Marie Claire Villeval. 2014. "The Dark Side of Competition for Status." *Management Science* 60(1): 38-55.

Conley Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review* 100(1): 1-28.

Dawes, Christoper D., James H. Fowler, Tim Johnson, Richard McElreath, Oleg Smirnov. 2007. "Egalitarian Motives in Humans." *Nature* 446(7137): 794-96

D'Exelle, Ben, and Arjan Verschoor. 2015. "Investment Behaviour, Risk Sharing and Social Distance." *The Economic Journal* 125(584): 777-802.

Di Falco, Salvatore, and Erwin Bulte. 2011. "A Dark Side of Social Capital? Kinship, Consumption, and Savings." *Journal of Development Studies* 47(8): 1128-51.

Di Falco, Salvatore, Francesco Feri, Paolo Pin, and Xavier Vollenweider. 2018. "Ties that Bind: Network Redistributive Pressure and Economic Decisions in Village Economies." *Journal of Development Economics* 131: 123-131.

Dohmen, Thomas, Armin Falk, Klaus Fliessbach, Uwe Sunde, and Bernd Weber. 2011. "Relative versus Absolute Income, Joy of Winning, and Gender: Brain imaging evidence." *Journal of Public Economics* 95(3-4): 279-85.

Drouvelis, Michalis, and Paola Paiardini. 2021. "Feedback Quality and Performance in Organisations." *Leadership Quarterly*, 101534.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval. 2009. "Feedback and Incentives: Experimental Evidence." *Labour Economics* 16(6): 679-88.

Fafchamps, Marcel. 1992. "Solidarity Networks in Preindustrial Societies: Rational Peasants with a Moral Economy." *Economic Development and Cultural Change* 41(1): 147-74.

Falk, Armin, and Andrea Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics* 24: 39-57.

Fehr, Dietmar. 2018. "Is Increasing Inequality Harmful? Experimental Evidence." *Games and Economic Behavior* 107: 123-34.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3): 817-68.

Foster, Andrew D., and Mark R. Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *Journal of Political Economy*, 103(6): 1176-1209.

Gerhards, Leonie, and Neele Siemer. 2016. "The Impact of Private and Public Feedback on Worker Performance: Evidence from the Lab." *Economic Enquiry* 54: 1188-1201.

Gill, David, Zdenka Kissova, Jaesun Lee, and Victoria Prowse. 2019. "First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision." *Management Science* 65(2): 494-507.

Hadnes, Myriam, Björn Vollan, and Michael Kosfeld. 2013. "The Dark Side of Solidarity." Washington, DC: World Bank.

Kebede, Bereket, and Daniel J. Zizzo. 2015. "Social Preferences and Agricultural Innovation: An Experimental Case Study from Ethiopia." *World Development* 67: 267-280.

Kuhnen, Camelia M., and Agnieszka Tymula. 2012. "Feedback, Self-Esteem, and Performance in Organizations." *Management Science* 58(1): 94-113.

Macours, Karen, and Renos Vakis. 2014. "Changing Households' Investment Behaviour through Social Interactions with Local Leaders: Evidence from a Randomised Transfer Programme." *The Economic Journal* 124(576): 607-33.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60(3): 531-42.

Munshi, Kaivan. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics* 73(1): 185-213.

Platteau, Jean-Philippe. 2000. *Institutions, Social Norms, and Economic Development*. Routledge.

Roels, Guillaume, Xuanming Su. 2014. "Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points." *Management Science* 60(3): 606-27.

Rubin, Jared, Anya Samek, and Roman M. Sheremeta. 2018. "Loss Aversion and the Quantity–Quality Tradeoff." *Experimental Economics* 21(2): 292-315.

Rustichini, Aldo. 2008. "Dominance and Competition." *Journal of the European Economic Association* 6(2-3): 647-56.

Sseruyange, John, and Erwin Bulte. 2020. "Wage Differentials and Workers' Effort: Experimental Evidence from Uganda." *Oxford Bulletin of Economics and Statistics* 82(3): 647-68.

Tafkov, Ivo D. 2013. "Private and Public Relative Performance Information under Different Compensation Contracts." *The Accounting Review* 88: 327-50.

Tran, Anh, and Richard Zeckhauser. 2012. "Rank as an Inherent Incentive: Evidence from a Field Experiment." *Journal of Public Economics* 96(9-10): 645-50.

Zizzo, Daniel J., and Andrew J. Oswald. 2001. "Are People Willing to Pay to Reduce Others' Incomes?" *Annales d'Economie et de Statistique* 63/64: 39-65.

Zizzo, Daniel J. 2003. "Money Burning and Rank Egalitarianism with Random Dictators." *Economics Letters* 81: 263-66.

For Online Publication

Appendix A. Theory

A.1. Optimization of basic model

Each individual maximizes the expected profit $Q = w.r(e^1).p(e^2) - c_1(e^1) - c_2(e^2)$ subject to $e^1 + e^2 \leq E$. Using the first-order conditions, and assuming $e^1, e^2 > 0$ and $e^1 + e^2 < E$, we take the first derivate to each of the two effort levels:

$$\frac{\partial Q}{\partial e^1} = w.r'(e^1).p(e^2) - c_1'(e^1) = 0$$

$$\frac{\partial Q}{\partial e^2} = w.r(e^1).p'(e^2) - c_2'(e^2) = 0$$

Merging both equations, we obtain that:

$$w.r'(e^1).p(e^2) - w.r(e^1).p'(e^2) = c_1'(e^1) - c_2'(e^2) \quad (1)$$

If the effort constraint is binding, we substitute $e^2 = E - e^1$, so that this equation becomes:

$$\left(w.r'(e^1).p(E-e^1)\right) - \left(w.r(e^1).p'(E-e^1)\right) = c_1'(e^1) - c_2'(E-e^1) \quad (2)$$

A.2. The effect of relative performance feedback and money-burning

We develop predictions on the effect of relative performance feedback on individual effort allocation. We assume that the effort constraint is not binding. However, all predictions remain the same without this assumption. We first demonstrate that relative performance feedback increases the number of correct tables. Thereafter, we show that this increase is weakened when a money-burning option is added.

1. Relative performance feedback

The optimization problem outlined in our theoretical model is to maximize:

$$w.r(e^1).p(e^2) - c_1(e^1) - c_2(e^2)$$

To model the influence of relative performance feedback, we add a utility term that depends on the inequality in the number of correct tables with the other group members, such that the utility of individual *i* becomes:

$$U_i = Q_i - c_1(e^1) - c_2(e^2) - \sum_{j=1}^{3} f\left(Q_i, Q_j\right) + \sum_{j=3}^{3} g\left(Q_i, Q_j\right) \quad (3)$$

with $Q_i = w.r(e_i^1).p(e_i^2)$, $Q_j = w.r(e_j^1).p(e_j^2)$, and $f(Q_i, Q_j)$ and $g(Q_i, Q_j)$ defined as follows:

$$f(Q_i, Q_j) = \begin{cases} \alpha(Q_j - Q_i) & if \;\; Q_i \leq Q_j \\ 0 & if \;\; Q_i \geq Q_j \end{cases}$$

$$g(Q_i, Q_j) = \begin{cases} 0 & if \;\; Q_i \leq Q_j \\ \beta(Q_i - Q_j) & if \;\; Q_i \geq Q_j \end{cases}$$

The parameters α, β are such that $0 < \alpha, \beta < 1$. They capture “behind-averse” and “ahead-seeking” motivations, respectively. $f(Q_i, Q_j)$ represents the utility loss due to being behind others, while $g(Q_i, Q_j)$ captures the utility increase due to being ahead of others.[21] We demonstrate that for each combination of effort (e_j^1, e_j^2) of the other group members, the optimal effort levels of e_i^1 and e_i^2 are such that more (or at least as many) number of correct tables are produced where there is social comparison than where there is no social comparison. As this applies to all participants irrespective of the effort allocations of the other group members, we have that in equilibrium the number of correct tables will be at least as high as without social comparison. To show this formally, we maximize the following function:

$$U_i = w.r(e^1).p(e^2) - c_1(e^1) - c_2(e^2) - \sum_{j=1}^{3} f\left(Q_i, Q_j\right) + \sum_{j=3}^{3} g\left(Q_i, Q_j\right)$$

using the first-order conditions of utility maximization

$$\frac{\partial U_i}{\partial e^1} = w.r'(e^1).p(e^2) - c_1'(e^1) - \sum_{j=1}^{3} \frac{\partial f(Q_i, Q_j)}{\partial e_i^1} + \sum_{j=3}^{3} \frac{\partial g(Q_i, Q_j)}{\partial e_i^1} = 0$$

$$\frac{\partial U_i}{\partial e^2} = w.r(e^1).p'(e^2) - c_2'(e^2) - \sum_{j=1}^{3} \frac{\partial f(Q_i, Q_j)}{\partial e_i^2} + \sum_{j=3}^{3} \frac{\partial g(Q_i, Q_j)}{\partial e_i^2} = 0$$

with

$$\frac{\partial f(Q_i, Q_j)}{\partial e_i^1} = \begin{cases} -\alpha\left(w.r'(e^1).p(e^2)\right) & if \;\; Q_i < Q_j \\ 0 & if \;\; Q_i > Q_j \end{cases}$$

$$\frac{\partial g(Q_i, Q_j)}{\partial e_i^1} = \begin{cases} 0 & if \; Q_i < Q_j \\ \beta\left(w.r'(e^1).p(e^2)\right) & if \;\; Q_i > Q_j \end{cases}$$

[21] The utility reducing effect of being ‘behind’ is similar to the effect of envy as modelled by the utility function of Fehr & Schmidt (1999). In line with Roels & Su (2014) and supported by empirical evidence of Gill et al. (2019) we assume that being ‘ahead’ does not decrease utility caused by ‘guilt’ as modeled by Fehr & Schmidt (1999) but actually increases uility.

$$\frac{\partial f(Q_i,Q_j)}{\partial e_i^2} = \begin{cases} -\alpha\big(w.r(e^1).p'(e^2)\big) & if\ \ Q_i < Q_j \\ 0 & if\ \ Q_i > Q_j \end{cases}$$

$$\frac{\partial g(Q_i,Q_j)}{\partial e_i^2} = \begin{cases} 0 & if\ Q_i < Q_j \\ \beta\big(w.r(e^1).p'(e^2)\big) & if\ \ Q_i > Q_j \end{cases}$$

<u>Case 1</u>: $Q_i \neq Q_j$, for all j.

In this case, we have that:

$$(1+\sigma).\big[\big(w.r'(e^1).p(e^2)\big) - \big(w.r(e^1).p'(e^2)\big)\big] = c_1'(e^1) - c_2'(e^2) \qquad (4)$$

with $\sigma = n\alpha + (3-n)\beta$, n and $(3-n)$ being the number of dyads where $Q_i < Q_j$ and $Q_i > Q_j$, respectively. If we set $\sigma = 0$, the first order conditions are the same as with the basic model. With $\sigma > 0$, however, the LHS is either more positive or more negative. To maintain the equality, and depending on the sign of the differences, we should either increase e^1 and decrease e^2 (if the LHS is positive), or decrease e^1 and increase e^2 (if the LHS is negative). In both cases, the number of correct tables will be larger compared to the case where $\sigma = 0$.

<u>Case 2</u>: $Q_i = Q_j$, for some j.

Note that the partial derivatives of $f(Q_i,Q_j)$ and $g(Q_i,Q_j)$ to e_i^1 and e_i^2 are undefined when $Q_i = Q_j$. For example, if $Q_i = Q_j$, the partial derivative of $f(Q_i,Q_j)$ can either be $-\alpha\big(w.r'(e^1).p(e^2) + w.r(e^1).p'(E-e^1)\big)$ or 0. In the case where the four group members put the same effort in round 1, the predicted effort levels are such that either effort levels are the same as predicted in Case 1, or they are the same as with $\sigma = 0$. If at least one individual *j* has a different output, the provision of feedback will increase the number of correct tables.

2. Money-burning

We show that the effect of relative performance feedback is weakened when money-burning is added. To implement money-burning, a second stage is added in which each participant can reduce the earnings of fellow group members at the end of round 2. In other words, after each player simultaneously decides on their effort combination (e_i^1, e_i^2) at the start of round 2, player i decides on money-burning b_{ij} targeted at player j at the end of round 2, with $b_{ij} \leq b_{ij}^{max} = Q_j - Q_i$ and $b_{ji} \leq b_{ji}^{max} = Q_i - Q_j$.

We assume that the amount of money-burning increases with the inequality in earnings between individuals i and j, and that player i only uses money-burning against player j if the latter performs relatively better than player i. We will test these assumptions in the empirical section. Using these assumptions, we distinguish the following cases.

Case 1: money-burning decisions of the highest-performer are implemented

As money-burning would only target high-performers, no money-burning would be used. Using backward induction, the effort allocation of the high-performer in the first stage would be same as in treatment FB.

Case 2: money-burning decisions of the lowest-performer are implemented

In this case, the earnings of the other three group members would be reduced with certainty. As a result, the expected inequality with money-burning is reduced. Using backward induction, this would weaken the effect of social comparison for all group members, including the lowest-performer, such that the number of correct tables and earnings would be lower with money-burning.

Case 3*: money-burning decisions of the second or third best performer are implemented*

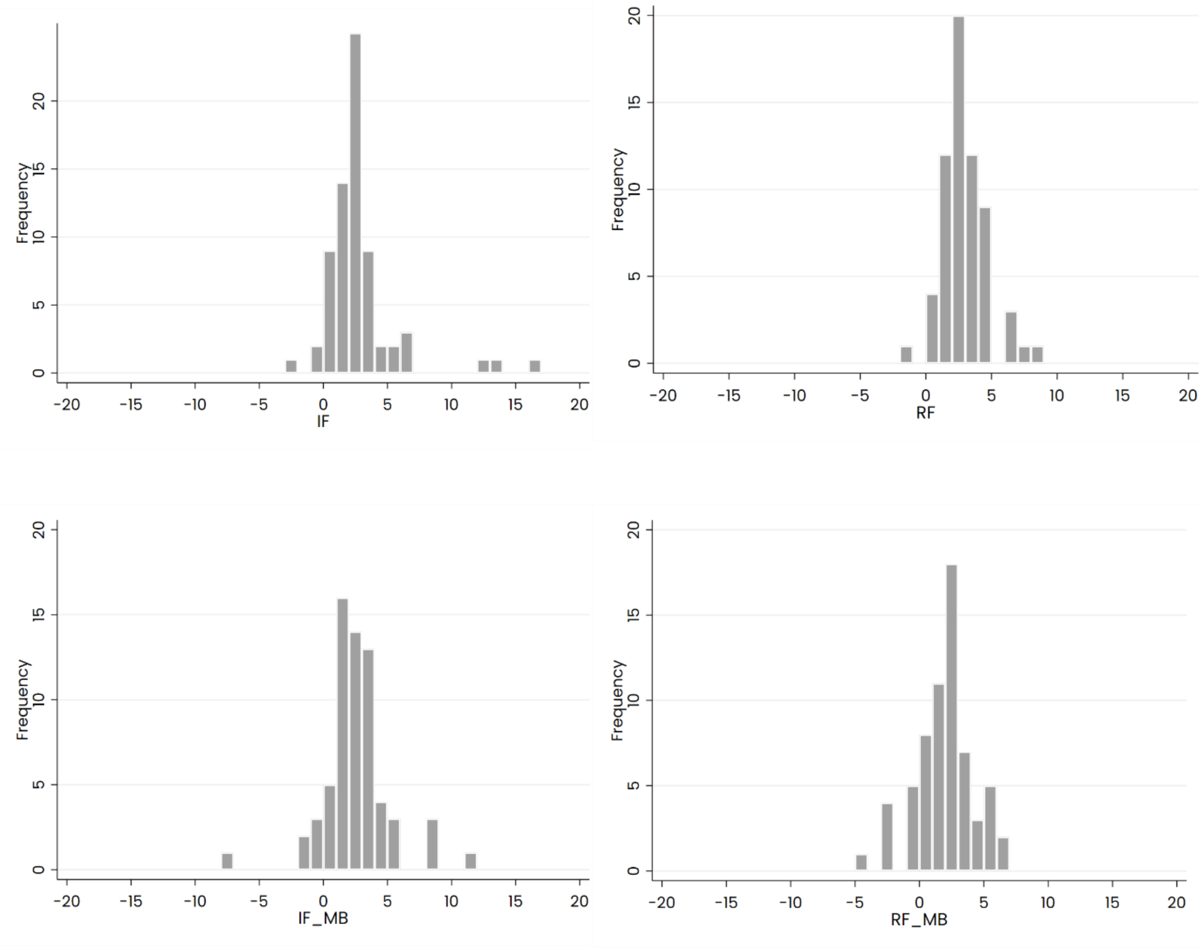
If the money-burning decisions of the second-best performer is implemented, then only the earnings of the best performer would be reduced. If the money-burning decisions of the third best performer is implemented, then the earnings of the best performer and the second-best performer would be reduced. Using backward induction, in both cases the positive effect of social comparison would be weakened for all group members, such that the number of correct tables and earnings would be lower with money-burning.

As the selection of the group member whose money-burning decisions are implemented is random, the effort decisions in the first stage cannot influence whose money-burning decisions are implemented. Also, money-burning decisions will be independent of the expectations of the money-burning decisions of other group members. It does obviously influence the likelihood that one is targeted by money-burning. This implies that each player will be able to estimate the expected money-burning for each effort level, conditional on the beliefs about the effort provision of the other group members. Using backward reduction, they optimize their effort levels in the first stage.

Two additional notes are needed. First, as the optimal effort level is conditional on the effort levels of other group members it is important to consider the beliefs one has about the effort levels of other groups, and how it creates the possibility of multiple equilibria. However, it does not change the predicted effect of money-burning. A group member with higher-order beliefs might anticipate that other group members also reduce their effort to minimize the risk of being targeted by money-burning. This implies that they will have to reduce their effort even further to avoid money-burning. This dynamic will reinforce the effect of money-burning. Second, beliefs about money-burning will be more accurate if money-burning is combined with relative performance feedback. This might increase the effect on the effort allocation of high-performers, who are most targeted by other group members.

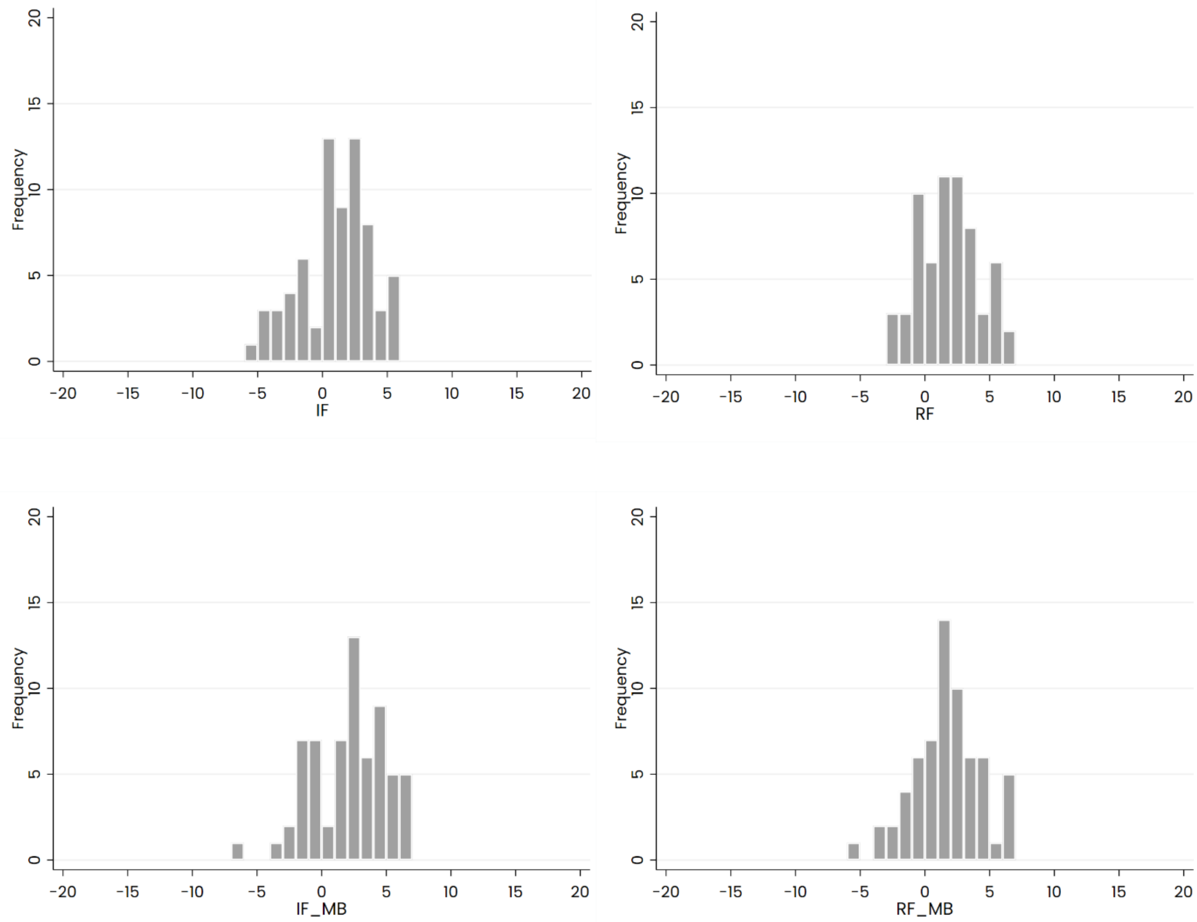
Appendix B. Additional results

Figure B.1. Within-change in number of completed tables, by treatment



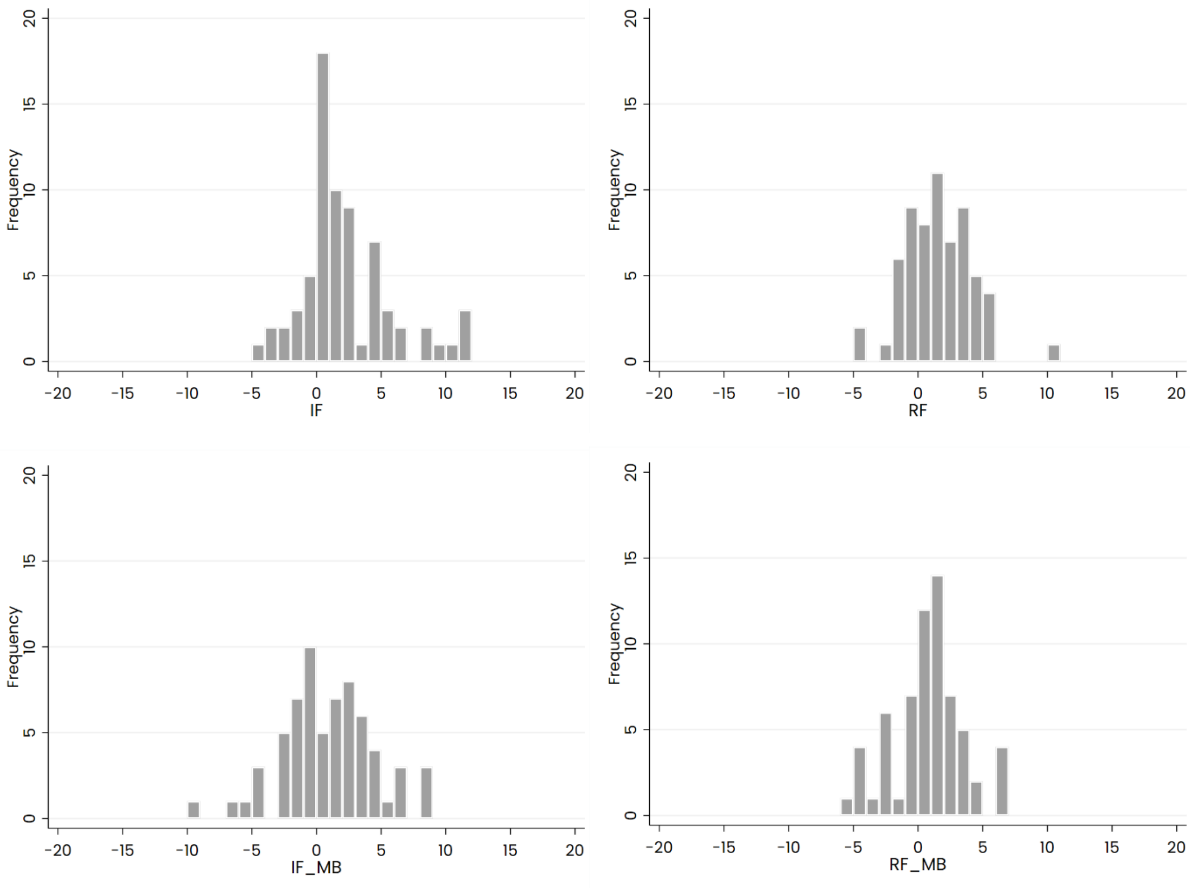
Notes. Change between rounds 1 and 2 in number of completed tables

Figure B.2. Within-change in number of correct tables, by treatment



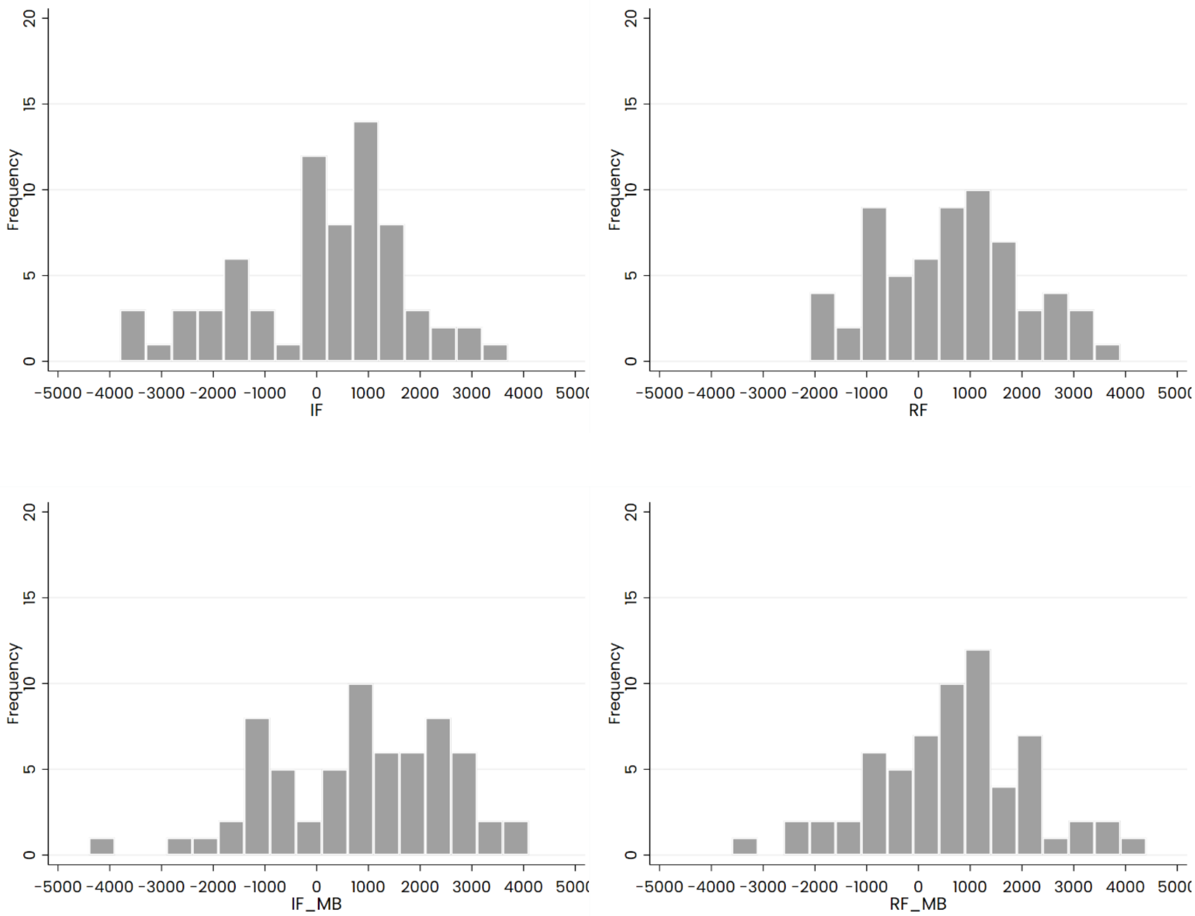
Notes. Change between rounds 1 and 2 in number of correct tables.

Figure B.3. Within-change in number of incorrect tables, by treatment



Notes. Change between rounds 1 and 2 in number of incorrect tables

Figure B.4. Within-change in earnings, by treatment



Notes. Change between rounds 1 and 2 in earnings (ignoring money-burning costs)

Table B.1. Descriptive statistics per round per treatment

	Completed tables		Accuracy (correct / total)		Correct tables		Incorrect tables		Earnings before money-burning		Earnings after money-burning [a]	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Round 1												
IF	13.21	13	0.78	0.83	10.33	11	2.89	2	4875.71	5000	4875.71	5000
	(2.22)		(0.19)		(2.85)		(2.45)		(1598.68)		(1598.68)	
RF	12.97	13	0.79	0.82	10.21	10	2.76	2	4826.98	4900	4826.98	4900
	(2.70)		(0.18)		(2.81)		(3.00)		(1596.08)		(1596.08)	
RF_MB	12.70	12	0.72	0.76	9.13	9	3.58	3	4204.69	4300	4204.69	4300
	(3.05)		(0.19)		(3.35)		(2.56)		(1818.11)		(1818.11)	
IF_MB	13.95	14	0.69	0.72	9.60	10	4.35	4	4364.62	4500	4364.62	4500
	(3.91)		(0.23)		(3.86)		(3.67)		(2126.06)		(2126.06)	
Round 2												
IF	15.56	15	0.71	0.74	10.87	11	4.69	4	4967.14	5000	4967.14	5000
	(3.25)		(0.21)		(3.23)		(3.87)		(1873.44)		(1873.44)	
RF	15.44	15	0.76	0.79	11.59	12	3.86	3	5407.94	5400	5407.94	5400
	(3.03)		(0.16)		(2.70)		(3.23)		(1536.47)		(1536.47)	
RF_MB	14.22	14	0.72	0.75	10.34	10	3.88	3	4784.38	4800	4578.13	4550
	(3.24)		(0.18)		(3.67)		(2.79)		(1995.81)		(2005.43)	
IF_MB	16.03	16	0.70	0.75	11.22	12	4.82	4	5126.15	5600	4841.38	5000
	(3.51)		(0.22)		(3.86)		(3.60)		(2151.62)		(2220.18)	

Notes. N = 262. Standard deviations in parentheses. [a] The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in this variable, while the guessers' earnings after money-burning are not included in this variable.

Table B.2. Descriptive statistics per treatment

	Change in completed tables	Change in accuracy [a]	Change in correct tables	Change in incorrect tables	Change in earnings [b]	Change in earnings [c]
IF	2.34	-0.07	0.54	1.80	91.43	91.43
	(2.98)	(0.20)	(2.74)	(3.61)	(1607.75)	(1607.75)
RF	2.48	-0.03	1.38	1.10	580.95	580.95
	(1.79)	(0.17)	(2.34)	(2.66)	(1379.09)	(1379.09)
RF_MB	1.52	0.01	1.22	0.30	579.69	326.56
	(2.31)	(0.19)	(2.63)	(2.78)	(1506.20)	(1578.05)
IF_MB	2.08	0.01	1.62	0.46	761.54	476.77
	(2.68)	(0.22)	(2.86)	(3.67)	(1703.92)	(1598.18)

Notes. N = 262. Average within-subject change reported. Standard deviations in parentheses. [a] Defined as number of correct tables divided by number of tables completed. [b] Earnings before money-burning costs are deducted. [c] Earnings after money-burning costs are deducted. The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in this variable.

Table B.3. Average treatment effects (excluding subgroups with at least one guesser)

	Completed tables	Accuracy [a]	Correct tables	Incorrect tables	Earnings [b]	Earnings [c]
	(1)	(2)	(3)	(4)	(5)	(6)
RF	0.107	0.032	0.746***	-0.638	436.607***	436.607***
	(0.432)	(0.027)	(0.215)	(0.413)	(121.630)	(121.630)
RF_MB	-0.798**	0.064	0.454	-1.252*	352.404	119.712
	(0.361)	(0.064)	(0.845)	(0.757)	(492.061)	(460.288)
IF_MB	-0.071	0.080	1.103*	-1.174	668.750	352.500
	(0.558)	(0.061)	(0.642)	(1.000)	(409.862)	(330.694)
Constant	2.375***	-0.063**	0.719*	1.656***	193.750	193.750
	(0.386)	(0.026)	(0.397)	(0.392)	(221.670)	(221.670)
R-squared	0.018	0.025	0.024	0.025	0.025	0.013

Notes. N = 228. OLS regressions. Difference between both rounds of the measure presented in first row used as dependent variable. Bootstrap clustered standard errors (in parentheses) at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Subgroups that contain one or more guessers are removed from the regressions. [a] Defined as number of correct tables divided by number of completed tables. [b] Earnings before money-burning costs are included. [c] Earnings after money-burning costs are included.

Table B.4. Treatment effects, by rank in tables completed in round 1

	Completed tables	Accuracy [a]	Correct tables	Incorrect tables	Earnings [b]	Earnings [c]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel a						
RF	-0.137	0.143***	2.089***	-2.226***	1267.262***	1267.262***
	(0.688)	(0.035)	(0.495)	(0.462)	(250.405)	(250.405)
RF_MB	-1.162	0.140*	1.131	-2.293***	794.762	641.429
	(1.017)	(0.077)	(1.135)	(0.890)	(618.282)	(607.176)
IF_MB	0.143	0.166	2.071*	-1.929	1228.571*	842.857
	(0.859)	(0.102)	(1.075)	(1.329)	(644.883)	(604.074)
High performer	-0.976	0.064**	0.012	-0.988*	104.762	104.762
	(0.796)	(0.029)	(0.470)	(0.554)	(230.130)	(230.130)
RF x High performer	0.467	-0.172***	-2.022***	2.488***	-1259.570***	-1259.570***
	(0.671)	(0.050)	(0.560)	(0.679)	(314.827)	(314.827)
RF_MB x High performer	0.504	-0.110*	-0.855	1.359**	-563.389	-662.997
	(1.118)	(0.060)	(1.102)	(0.563)	(566.679)	(567.077)
IF_MB x High performer	-0.771	-0.138	-1.754	0.983	-975.418	-798.082
	(0.732)	(0.093)	(1.105)	(1.189)	(651.597)	(683.822)
Constant	2.929***	-0.110***	0.536	2.393***	28.571	28.571
	(0.810)	(0.035)	(0.585)	(0.551)	(296.518)	(296.518)
Panel b (treatment effects on high performers) [d]						
RF	0.330	-0.028	0.068	0.262	7.692	7.692
	(0.389)	(0.035)	(0.384)	(0.508)	(228.264)	(228.264)
RF_MB	-0.658***	0.030	0.276	-0.934	231.373	-21.569
	(0.208)	(0.052)	(0.774)	(0.650)	(450.238)	(387.827)
IF + MB	-0.628	0.028	0.317	-0.945	253.153	44.775
	(0.577)	(0.047)	(0.671)	(0.960)	(416.919)	(386.481)
Predicted within-change in control treatment (IF)	1.952***	-0.046*	0.548	1.405***	133.333	133.333
	(0.228)	(0.027)	(0.370)	(0.422)	(221.642)	(221.642)
R-squared	0.065	0.066	0.088	0.053	0.082	0.067

Notes. N = 262. Difference between both rounds of the measure presented in first row used as dependent variable. Bootstrap clustered standard errors (in parentheses) at the school level. *** p < 0.01, ** p < 0.05, * p < 0.10. Rank is based on the earnings of round 1 within each subgroup. [a] Defined as number of correct tables divided by number of completed tables. [b] Earnings before money-burning costs. [c] Earnings after money-burning costs. The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in the calculation of this variable. [d] Predictions of treatment effects for high-performers, using the pooled sample.

Table B.5. Treatment effects, by rank in accuracy in round 1

	Completed tables	Accuracy [a]	Correct tables	Incorrect tables	Earnings [b]	Earnings [c]
	(1)	(2)	(3)	(4)	(5)	(6)
Panel a						
RF	-0.044	0.064	0.901	-0.945	545.168	545.168
	(0.422)	(0.051)	(0.586)	(0.839)	(369.107)	(369.107)
RF_MB	-0.994*	0.111	0.961	-1.955**	675.882	555.882
	(0.599)	(0.090)	(1.388)	(0.962)	(784.575)	(761.345)
IF_MB	-0.227	0.111	1.327	-1.555	819.216	535.882
	(0.484)	(0.091)	(1.169)	(1.358)	(713.322)	(657.645)
High performer	0.095	-0.118**	-1.289	1.384*	-783.007	-783.007
	(0.774)	(0.050)	(0.865)	(0.744)	(477.124)	(477.124)
RF x High performer	0.312	-0.037	-0.018	0.330	-41.993	-41.993
	(0.892)	(0.077)	(1.069)	(1.356)	(642.805)	(642.805)
RF_MB x High performer	0.311	-0.059	-0.495	0.806	-328.170	-490.523
	(0.794)	(0.076)	(1.262)	(0.825)	(697.834)	(700.319)
IF_MB x High performer	-0.076	-0.042	-0.416	0.340	-241.755	-244.422
	(0.923)	(0.081)	(1.234)	(1.307)	(718.896)	(727.398)
Constant	2.294***	-0.011	1.206*	1.088*	494.118	494.118
	(0.323)	(0.048)	(0.694)	(0.585)	(399.833)	(399.833)
Panel b (treatment effects on high performers) [d]						
RF	0.268	0.027	0.883	-0.615	503.175	503.175
	(0.768)	(0.042)	(0.670)	(0.724)	(369.677)	(369.677)
RF_MB	-0.683	0.051	0.466	-1.149	347.712	65.359
	(0.534)	(0.042)	(0.359)	(0.704)	(232.706)	(171.031)
IF_MB	-0.303	0.069	0.912	-1.215	577.460*	291.460
	(0.943)	(0.055)	(0.586)	(1.009)	(345.651)	(316.220)
Predicted within-change in control treatment (IF)	2.389***	-0.129***	-0.083	2.472***	-288.889	-288.889
	(0.689)	(0.024)	(0.470)	(0.521)	(242.458)	(242.458)
R-squared	0.025	0.180	0.106	0.107	0.117	0.117

Notes. N = 262. OLS regressions. Difference between both rounds of the measure presented in first row used as dependent variable. Bootstrap clustered standard errors (in parentheses) at the school level. *** p < 0.01, ** p < 0.05, * p < 0.10. Rank is based on the earnings of round 1 within each subgroup. [a] Defined as number of correct tables divided by number of completed tables. [b] Earnings before money-burning costs. [c] Earnings after money-burning costs. The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in the calculation of this variable. [d] Predictions of treatment effects for high-performers, using the pooled sample.

Appendix C. The identification of ‘guessers’

In the instructions, it was clearly explained that we wanted participants to count rather than guess. It could therefore be concluded that the subjects who guessed the number of zeros did not follow our instructions. The inclusion of these subjects would increase noise in the data, which would complicate the analysis. For these reasons, we removed these subjects from the analyses.[22]

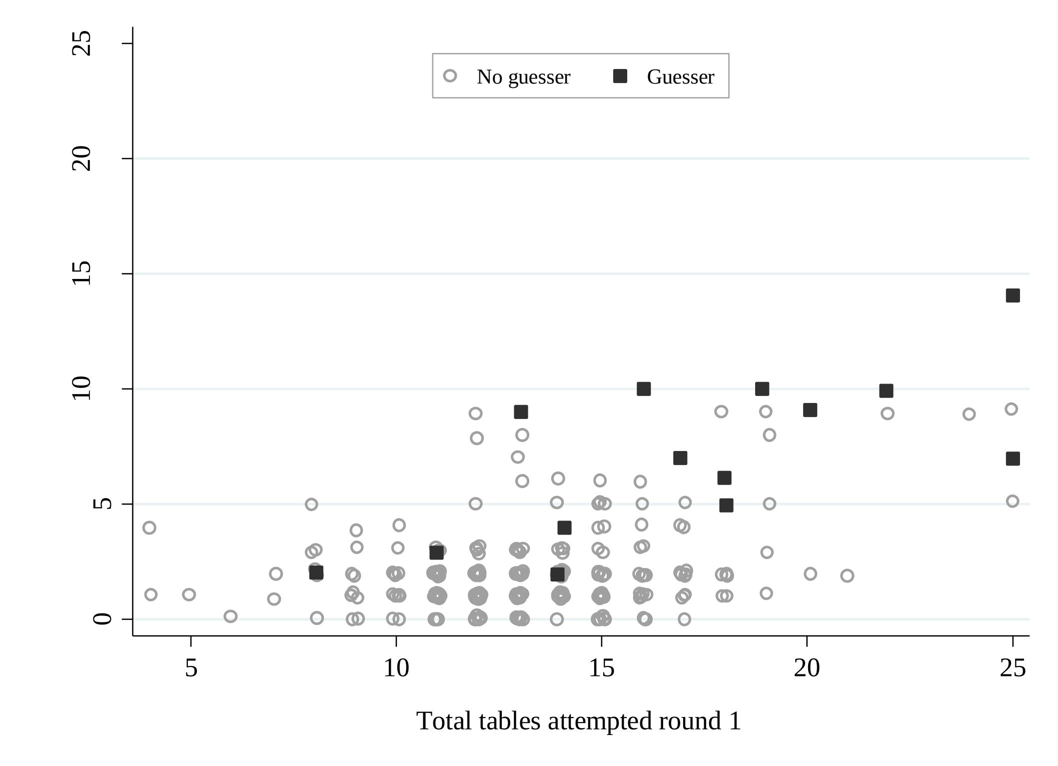
We tried to identify them in the following way. We looked at the number of consecutive errors and considered participants to engage in ‘guessing’ behaviour if *their largest number of consecutive errors in one of the two rounds exceeded 3 times the interquartile range* away from the first quartile or the third quartile of the sample. Following this procedure, we classified 14 participants as ‘guessers’.[23] The pilot confirmed that it is extremely difficult to do all 25 tables in the given time. Many participants who we identified as ‘guessers’ reached this maximum, as demonstrated by the scatterplots in Figure C.1. Plotting for each participant the largest number of consecutive errors by the number of completed tables, for each round, we observe that ‘guessers’ not only have many consecutive errors. They also solve many tables. This confirms that they are correctly identified as ‘guessers’ and their large number of errors is not the result of low skills.

While removing them from the analysis reduces noise in the data, we do not expect this to bias the treatment comparisons, as the proportion of guessers is equally balanced across the treatments (ANOVA: F= 0.55, p = 0.645), and are not statistically different from ‘non-guessers’ on observable socio-economic characteristics, as shown in Table C.1. We also observe that the treatment effects are weaker when the guessers are included (Table C.2). This is not surprising as these participants are less likely to respond to the different treatment conditions.

[22] A guessing strategy did not pay off as the probability of guessing the right answer was very low, and the effect of making mistakes led to negative earnings due to the penalty of 100 UGX. Guessers earned on average 2,357.14 UGX in round 1, while non-guessers earned 4,573.28 UGX, which is more than guessers in round 1 (difference = 2,216.14 UGX, two-sided p-value of a t-test = 0.000). In round 2, guessers earned 1,242.85 UGX, while non-guessers earned 5,067.94 UGX, and the difference increased to 3,825.08 UGX (two-sided p-value of a t-test = 0.000).

[23] In both rounds, the cut-off point is 9 consecutive errors, so that participants who made 10 consecutive errors were excluded from the sample. Following this procedure, we identified 4 ‘guessers’ in round 1 and 12 ‘guessers’ in round 2. 2 participants were identified as ‘guessers’ in both rounds.

Figure C.1. Scatter plots identifying students who structurally 'guessed' answers



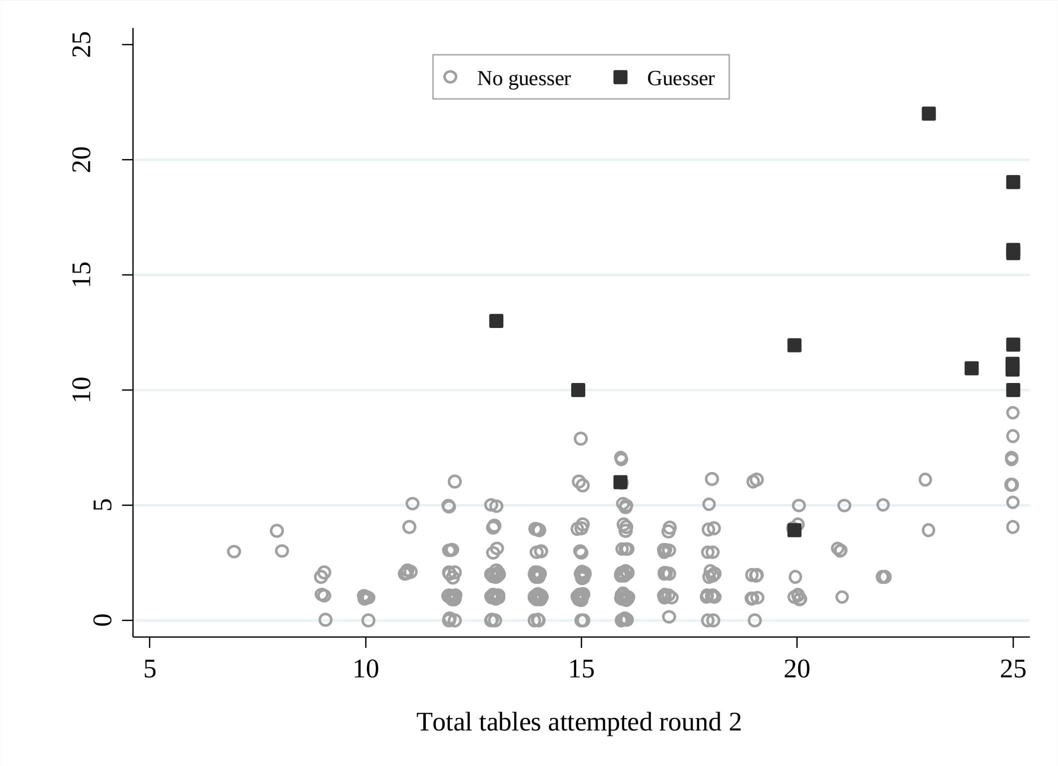


Table C.1. Guessing vs. non-guessing students

	Guessing students	Non-guessing students	Mean difference	p-value
Mbale (urban)	58%	79%	20%	0.134
Female	40%	57%	17%	0.217
Household wealth	0.01	-0.22	0.23	0.660
Bagisu tribe	68%	71%	4%	0.763
Number of siblings	6.12	6.29	0.16	0.862
Firstborn	20%	14%	6%	0.609
Father higher education	50%	57%	7%	0.603
Mother higher education	40%	36%	4%	0.724
Observations	14	262	276	

Notes. *** p < 0.01, ** p < 0.05, * p < 0.10. To test for statistical differences, we use a two-sample proportion test for binary variables. For continuous variables we use a two-sided t-test.

Table C.2. Average treatment effects (with guessers included)

	Completed tables	Accuracy [a]	Correct tables	Incorrect tables	Earnings [b]	Earnings [c]
	(1)	(2)	(3)	(4)	(5)	(6)
RF	0.706	0.027	0.749*	-0.043	378.922*	378.922*
	(0.451)	(0.028)	(0.382)	(0.420)	(210.391)	(210.391)
RF_MB	-0.456	0.066	0.587	-1.043	398.039	203.922
	(0.353)	(0.061)	(0.793)	(0.866)	(477.119)	(443.052)
IF_MB	-0.294	0.084	0.999	-1.293	628.922	355.539
	(0.632)	(0.063)	(0.721)	(1.017)	(445.158)	(383.622)
Constant	2.250***	-0.072**	0.486	1.764***	66.667	66.667
	(0.393)	(0.028)	(0.411)	(0.400)	(229.333)	(229.333)
R-squared	0.024	0.028	0.019	0.025	0.020	0.009
RF = RF_MB [d]	0.009	0.277	0.759	0.134	0.952	0.549
RF = IF_MB [d]	0.003	0.232	0.645	0.105	0.467	0.934
RF_MB = IF_MB [d]	0.789	0.642	0.161	0.765	0.300	0.471

Notes. N = 276. OLS regressions. Difference between both rounds of the measure presented in first row used as dependent variable. Bootstrap clustered standard errors (in parentheses) at the school level. *** p < 0.01, ** p < 0.05, * p < 0.10. [a] Defined as number of correct tables divided by number of completed tables. [b] Earnings before money-burning costs are included. [c] Earnings after money-burning costs are included. The money-burning decisions from 'guessers', and the cost of money-burning of non-guessers targeting guessers is included in this variable, while the guessers' earnings after money-burning are not included in this variable. [d] Two-sided p-value of a chi-square test.

Appendix D. Experimental Instructions

Registration (all treatments)

- [Welcome the participants].
- [When participants enter the meeting room, they are asked for their name. We have a list of invited candidates. Their name is marked and they are given a sticker with an identity number, which we ask them to stick on their shirt. It is explained that this identity number is unique and allows us to identify them during the exercise while guaranteeing complete confidentiality. This is important, as they are able to earn real money in the exercise. All students belonging to class [A] are asked to stand together in one area, and students belonging to class [B] are asked to group together in another area of the meeting room].
- [Make sure that the number of students per class is dividable by 4 (so 4, 8, 12, 16, 20, 24, etc.). If not, give the students that came in last the show up fee of 2.000 UGX, thank them and explain to them that we always invite more people than needed; this will be done at the start of the session, rightly after welcoming the participants and registering their presence].

Plenary Welcome (all treatments)

- *Welcome. Thank you for taking the time to come today. [Introduce Experimenters and Assistants] You can ask any of us questions during today's program.*
- *We have invited you here, today, because we want to learn about how people in this area take decisions. Your decisions are about money, and the money that results from this will be yours to keep.*
- *What you need to do will be explained fully in a few minutes. First, we want to make a couple of things clear.*
    - *First of all, this is not our money. We belong to a university in the UK, and this money has been given to us for research.*
    - *Participation is voluntary. You may still choose not to participate in the exercise.*
    - *We also have to make clear that this is research about your decisions. Therefore, you cannot talk with others. This is very important. I am afraid that if we find you talking with others, we will have to ask you to leave the room, and you will not be able to earn any money here today. Of course, if you have questions, you can ask one of us by simply raising your hand.*
    - *We would also like to ask you to switch off your mobile phones. Please do that now.*
- *Make sure that you listen carefully to us. You will be able to make a good amount of money here today, and it is important that you follow our instructions.*
- *During today's program, you will be asked to put as much effort in a task as possible, which will be explained to you very clearly. The amount of effort you put in determines the money you will be paid. Any money you earn will be paid out to you privately and confidentially after all parts of the exercise are complete.*

- *All of you will receive a 2.000 UGX show up fee, no matter what kind of decisions you take or how you perform in today's program.*
- *First, however, we are going to determine with whom you are going to form a group, and to which room you will be assigned to.*

Assigning subgroups and treatments on the day of the experiment (all treatments)

- [A bag and cards depicting numbers (maximum 16 different numbers, 4 cards per number) will be used to assign students to groups. There are 4 cards of each number. The students (maximum 64; 4x4x4) draw cards without putting them back, determining to which group they belong. After all cards have been drawn, students are asked to group together with others that have drawn the same number. In order to determine to which group they go, 2 separate bags will be used). The first bag contains the group numbers; the second bag contains the letters A – B – C – D representing the 4 different rooms. First a group number card from the first bag is drawn, followed by the drawing of a letter card from the second bag. These cards will not be put back in the bag. After doing this 4 times (4 different rooms), new treatment cards will be put in the bag, and the procedure is repeated until all groups are assigned to a room].
- [The numbering of the subgroups (as used in the registration Excel sheet) per treatment is irrelevant for the experiment, but group identifiers need to be registered in order to know to which groups participants belong].

Experimental script

Preparations in classroom (all treatments)

- [Set up benches to facilitate 1 to 4 groups (depending on number of participants) with each 4 participants].
- [Make sure you have all the necessary documents:
  - 2 piles of sets of tables for round 1 and round 2. Make sure you keep sets together (they are stapled), so participants do not get the same tables in their sets. Check whether each set has the proper front page with test questions.
  - Tablet with the Excel template prepared for registering responses. Fill out the date, the name of the school, the class number, the name of the treatment, and the names of the enumerators. Save the document as *[yymmdd – Name school –treatment & number]. For e*xample: treatment 1 conducted on the 1st of July 2016 at School 1 will be saved as: *160701 – School 1 – Treatment 1*. Have this prepared before starting the experiment.
  - The answer sheet in hard copy. This sheet is also available in Excel (but protected, preventing accidental changes).
  - Scrap paper.

- Pens.
- Ethical consent forms.
- Pay slips.
- Envelopes to hide earnings.
- Reducing scenario sheets.[24]

- [Put 1 pen and 1 piece of scrap paper on each table].

Registration and seating (all treatments)

- [Participants are asked to take a seat in the classroom. There are four groups of two rows of benches, placed orthogonal to the instruction table. Ask students to take their seats on the benches. Both rows of participants are seated back-to-back. The benches/chairs should be arranged so that no participant can see what another participant is looking at. No swapping or trading places is allowed, even within groups; once seated students remain seated].
- [Take the laptop, open the registration document, go to each group and register their name & ID (see sticker). Make sure you SAVE the Excel document].
- [Ask participants in each group to sit in a circle, facing each other, when you are giving instructions. This should enhance the group feeling].

General instructions (all treatments)

- *Before I will tell you what we are going to do, we will give you an attendance fee of 2.000 UGX.*
- *Today you can earn money by solving tables. Each of you gets a set of tables with zeros and ones, and we would like to ask you to count the number of <u>ZEROS</u>. [Show the tables]. Above the table, we would like to ask you to write down your name and answer. All answers NOT written in the box will automatically be considered as incorrect, so you need to write your answer IN that designated box. If you count 70 zeros in the table, you write '70' ONLY here [indicate the box]. Writing the number is enough, so do NOT write the word 'zeros' or make sentences like 'there are 70 zeros in this table'.*
- *We would like to ask you to solve as many tables as possible, and you have 10 minutes exactly. You will be asked to do this 2 times, so we will have 2 rounds of table solving of 10 minutes each. For each correct answer, you will earn 500 UGX. For example, if you provide correct answers to 15 tables, you earn 7.500 UGX. Be aware that when you provide a wrong answer, 100 UGX will be subtracted from your total earnings so far. For example: by providing 10 correct you earn 5.000 UGX. If you have also given 5 incorrect answers it means 500 UGX will be subtracted from the 5.000 UGX, leaving you with 4.500 UGX.*

[24] While in the paper we call them money-burning decision cards, during the actual experiment we wanted to use a more neutral term.

- *You will all get 1 piece of scrap paper and 1 pencil, but you are also allowed to make handwritten notes on the tables. As mentioned earlier, we will do 2 rounds of table solving. After each round, we will check your answers, and after checking your answers in the second round, we will calculate your total earnings and ask each of you to individually step forward to receive your pay-out. After collection, we would like to ask you to leave this room, so we can pay out everyone here in a discrete and organized way. We would also like to emphasize that all the information gathered today will be FULLY confidential, only accessible for the research team, and used for research purposes only.*
- *When you receive the set of tables, we would like to ask you not to open the document, ensuring that every one of you is given an equal start. You can start solving the tables once we say the word 'START', and please turn over the set of tables again once we say the word 'STOP' [show by using one of the sets of tables]. Altogether, you have 10 minutes to solve as many tables as possible. Please make sure you write your name and group number in the designated boxes on top of the tables. Without name on the table, we do not know to whom it belongs, and unfortunately, we cannot credit you for any correct answers.*
- *We will do this exercise two times in total, but at the end, we will select ONLY ONE of the rounds for pay-out. So, if you earned 5.000 UGX in the first round and 4.000 UGX in the second round, and the second round is selected, we will pay you 4.000 UGX, on top of the 2.000 UGX show-up fee you receive. It can also be the other way around of course.*
- *We would like to start asking you to fill out a consent form, indicating that you agree to participate in our program, and return it to us. Please read this form carefully and tick the boxes. After that, we start handing out the sets of tables plus a cover page with questions about the exercise. These questions are meant to indicate whether everyone has understood the instructions well.*

- [Hand out the ethical consent form. Ask them to read the statements, tick the boxes, and sign the document. Give them 5 minutes to do so and then collect all forms].
- [Hand out the sets of tables with the test questions on top so the participants cannot see any exercises, and also give them 1 piece of scrap paper and 1 pen per participant. Emphasize again they are not allowed to look at the tables].
- [After handing out all the sets, ask the participants to fill out the 4 questions].

Round 1 (all treatments)

- [Reset stopwatch and set alarm on 10 minutes (BE UTTERLY PRECISE WITH THIS AS TIMING IS ESSENTIAL). Mention again very clearly participants can only start when you say the word 'START'. They need to stop when you say 'STOP'. Mention that during the exercise 2 indications about remaining time will be given ('5 minutes left' & '1 minute left')].
- [Tell them not to forget to write their name, and group number on top of EACH table they answer, after you say the word 'START'].
- [START the experiment].
- [Walk around and keep an eye on the stopwatch. Also, start preparing the small forms for performance feedback].
- [After the alarm of the stopwatch goes off say 'STOP' loudly, meaning participants immediately need to put their pencils down, and turn their tables over so the cover page with the test questions is up].
- [Collect the tables as quickly as possible, preventing students to solve more tables].

Checking and registering answers round 1 (all treatments)

- [Emphasise that while you check the answers, talking is not allowed. Mention that, altogether, this might take 10 minutes].
- [After collecting the forms, start checking the answers by making use of the answer sheet in Excel or hard copy. Each table has a unique id, which should help finding the corresponding answer].
- [Register in Excel which tables are answered correctly, and which are answered wrongly by the participant. Insert a small 'c' when answers are correct, a small 'f' when answers are incorrect. Leave open those boxes that refer to tables that are not answered at all. Make sure you SAVE the Excel document].

- *We will now start the second round.*

Round 2

Individual performance feedback (for treatments IF and MB only):

- [Prepare the individual performance feedback forms depicting how each individual has performed].
- *You will now receive feedback on how you have performed in the first round. We will inform you on the number of correct answers, number of incorrect answers, and the money you earned in that round. We would like to ask you not to show this to other participants here today.*
- [Each participant receives a small piece of paper, showing (in handwriting) how many correct answers and incorrect answers they had and the money they earned. Emphasise that participants are NOT allowed to show this to other participants].

- [After handing out the last one, participants have 3 minutes to absorb the information, then start collecting the pieces of paper].

Relative performance feedback (for treatments RF and RF_MB only):

- [Prepare the performance feedback forms depicting how all group members have performed].

- *First, you will receive feedback on how you and your group members have performed in the first round. We will inform you on the number of correct answers, number of incorrect answers, and the money you earned in that round. We would like to ask you not to show this to any members of the other groups.*
- [Each participant receives a small piece of paper with 4 rows, showing (in handwriting) how many correct answers and incorrect answers they had and the money they earned, together with the performance of all the other individual group members. Emphasise that participants are NOT allowed to show this to other groups].
- [After handing out the last one, participants have 3 minutes to absorb the information, then start collecting the pieces of paper].

Announcing money-burning (for treatments RF_MB and MB only):

- [Explain the possibility to reduce group members earnings];

In this second round, there will be an extra element we have not told you about yet. After this round of counting zeros in tables, all of you will be given the opportunity to reduce the earnings of your three fellow group members. This opportunity is, however, LIMITED to the money that is earned in round 2 ONLY.

You cannot reduce your own money earned in round 2, and you have to pay a price for reducing the money of others; the price of reducing is one tenth of the amount you would like to reduce. So, for example, if you want to reduce 1.000 UGX of all your group members in total, you pay 100 UGX, if you want to reduce 10.000 UGX in total, you pay 1.000 UGX. To reduce the earnings of others, you need to have earned at least one tenth of the amount you want to reduce. If you did not earn anything, you cannot reduce anything. Please be aware you are NOT obliged to reduce anything; the decision is totally up to you. After the second round of table solving, you receive a piece of paper which shows your earnings and the individual earnings of each of your group members in the second round. You have 5 minutes to absorb that information and think about which reducing decisions you would like to take. Fill out on the form how much you would like to reduce from each group member, and fold the paper when you are done. After that, we will collect all pieces of paper.

We would strongly like to emphasize that this part of today's game is about your decisions, so there are NO right or wrong answers. You should make your decisions as you wish and on your own; you can reduce the money of others in any way you want or perhaps you decide to not reduce at all. Your decisions will remain FULLY confidential, so no one will ever find out about your decisions].

Bear in mind that others can also reduce your earnings. *After recording all your decisions, we select 1 'scenario' per group and apply it to the earnings of those within that particular group. Every one of you has 25% chance of having his or her scenario selected. It is possible your scenario is selected, which could reduce your earnings with the price you are willing to pay for reducing the earnings of the others,* ***but it is more likely that one of the scenarios of your group members is selected, which possibly reduces YOUR earnings****. We will not tell you which scenario is selected, and your reducing scenario will stay ANONYMOUS at all times; no one will ever know how much you wanted to reduce from others. Think seriously about your choices because they will affect how much money you can take home.*

First, we start again with questions that will indicate whether everyone in this group has understood the instructions.

- [Hand out the new sets of tables and emphasize participants are not allowed looking at the tables. Ask them to fill out first the questions on the first page].
- [Reset stopwatch and set alarm on 10 minutes. Indicate we will do the same as in round 1. Mention again very clearly participants can only start when you say the word 'START'. They need to stop when you say 'STOP'. Mention that during the exercise 2 indications about remaining time will be given ('5 minutes left' & '1 minute left')].
- [Walk around and keep an eye on the stopwatch. Also, start preparing the money reducing scenarios and payslips by already inserting the names].
- [After the alarm of the stopwatch goes off say 'STOP' loudly, meaning participants immediately need to put their pencils down, and turn their sets of tables over so the cover page with the test questions is up].
- [Collect their sheets as fast as possible, possibly preventing students to solve more tables than allowed].

Checking and registering answers round 2 (all treatments)

- [After collecting the forms, start checking the answers by making use of the answer sheet in Excel. Each table has a unique id].
- [Register in Excel which tables are answered correctly and which are answered wrongly by the participant. Leave open those boxes that refer to tables that are not answered at all].
- [The pay-out of round 2 per individual will be automatically calculated in the template in Excel].

Money-burning (for treatments RF_MB and MB only):

- [On the ‘scenario notes’ write for all participants in a specific group the amount that he or she earned in round 2. The earnings in round 1 are not mentioned as we don’t want the earnings of round 1 to interfere with the reduction of the earnings of round 2. Make sure you SAVE the Excel document].
- [Hand out the reducing scenario forms with feedback on how they and their fellow group members have performed in round 2].
- *You now have 5 minutes to absorb this information, and think seriously about which reducing decisions you would like to take. Again, you need to pay a price for reducing the earnings of the others, which is one tenth of the total amount you would like to reduce. The first column on this overview provides information how you and your group members performed, but in the second column, you can write the amount that you would like to reduce from the others. After that, fold your piece of paper, ensuring that your decisions remain fully confidential.*
- [Give them 5 minutes]
- [Collect the forms]
- *We will now start registering all your decisions, please give us a couple of minutes for this. After that, we randomly select one scenario and apply that to your earnings. Again, which scenario we select remains a secret.*
- [Register their decisions directly in the Excel sheet before calling over the next participant].
- [After registering all scenarios of ALL groups put per group all the forms in a bag, and draw 1 scenario per group. Repeat this procedure until you have sampled a scenario for all groups. Make sure that participants do not get to know which scenario is selected].
- [Register in the Excel sheet which scenario is sampled. Apply the scenario and the corrected earnings for each participant in each group are automatically calculated. Make sure you SAVE the Excel document].

Pay-out (all treatments)

- *We will now select only ONE of the rounds for pay-out. So, there is 50% chance you will receive the money you have earned in round 1, and there is a 50% chance you will receive the money from round 2.*
- [Put 2 counters with different colours in a bag, each representing a different round. Draw 1 counter, show it to the entire group, and mention which round will be paid out. In the Excel sheet insert the selected round and the final earnings will be automatically calculated].
- [Write the names of the participants and the corresponding earnings on the ‘pay slips’ for pay out].

- [Tell the participants they now will be paid their earnings of the selected round. Explain that if round 2 is selected, their earnings might be less than anticipated, as some of it might be reduced by others. They are still not allowed to talk and they will be called over individually].
- [Ask the first participant to come over, hand over the template ('payslip') with the name of the participant, and write down the amount the participant has earned. Then hand over the actual amount in an envelope and ask the participant to sign for reception (on the template). Make sure you keep the pay slips and do not give it to the participant].
- [Before asking the next participant to come over, prepare the exact amount and 'payslip' and repeat this procedure until all participants have been paid].
- [Make sure you collect and document all the payslips, feedback notes, and individual scenarios. This will be used to check the Excel file later. Also, check whether the Excel file is filled out correctly, before saving a final version the document].

Appendix E. Additional materials used in the experiment

Table E.1. Example of a table used for the real-effort task

NAME PARTICIPANT:									
NUMBER OF ZEROS:									
1	1	1	1	1	0	0	1	1	0
0	1	0	1	1	1	0	0	0	0
0	1	1	0	0	1	1	1	0	1
0	0	1	0	1	1	0	0	1	0
1	0	0	0	1	0	0	0	1	0
0	1	1	1	0	1	1	0	1	0
0	0	1	0	1	1	1	0	1	1
0	1	0	0	0	0	0	1	1	0
1	0	1	1	1	1	1	0	1	1
0	1	0	0	1	0	0	0	1	0
0	0	0	1	0	0	1	1	1	1

Table E.2. Feedback slip

Name	Correct answers	Incorrect answers	Earnings round 1

Table E.3. Money-burning decision card

Name:		Earnings Round 2:	
#	**Name other group members**	**Earnings Round 2**	**Amount you want to reduce**