Journal of the Economic Science Association (2024) 10:393–414
https://doi.org/10.1007/s40881-024-00163-9

ORIGINAL PAPER

Cognitive ability and the house money effect in public goods games

Julian Hackinger[1]

Received: 20 May 2021 / Revised: 8 January 2024 / Accepted: 19 January 2024 /
Published online: 1 April 2024

Abstract

I experimentally investigate the relation of endowment origin, cognitive abilities (as measured by the Cognitive Reflection Test, CRT), and co-operation in a one-shot linear public goods game. The results show that subjects' contributions depend on an interplay of cognitive abilities and endowment origin. A house money effect exists only for subjects with low CRT scores. They contribute more when income was allocated to them and less when income was obtained by effort. In contrast, subjects with high CRT scores contribute the same amount independent of income type. The findings have implications for redistribution, team production, and experimental designs.

Keywords Public goods game · Cognitive reflection · House money · Real effort · Income source

JEL Classification C91 · D03 · H41

1 Introduction and literature

Many individuals distinguish between money obtained from different sources, violating the assumption of fungibility of money.[1] This has important implications for various domains, for example, for the design and welfare effects of public policies. The observation that unearned income is treated differently than earned income is referred to as the house money effect. The common rationale is that windfall money evokes perceived property rights less strongly than earned money (see also Hoffman and Spitzer, 1985). As a result, in dictator games, subjects show less generosity when allocating earned income, both in the laboratory (Cherry,

[1] Corresponding behavioural operations are called mental accounting (Thaler, 1985, 1990, 1999).

✉ Julian Hackinger
julian.hackinger@tum.de

1 Technical University of Munich, Chair of Economics, Arcisstr. 21, 80333 Munich, Germany

2001; Cherry et al., 2002; Cherry & Shogren, 2008; Reinstein & Riener, 2012; Oxoby & Spraggon, 2008) and in the field (Carlsson et al., 2013).[2] Furthermore, Cárdenas et al. (2014) show that windfall money is spent more riskily. Houser and Xiao (2015) identify a house money effect in trust games. Transfers by investors and trustees are lower if they have to decide over earned money.

By contrast, in public goods games, corresponding evidence remains mixed, rather indicating no effect (Spraggon & Oxoby, 2009). Most studies find that contributions in public goods games are independent of endowment origin (Clark, 2002; Cherry et al., 2005; Antinyan et al., 2015; Bailey et al., 2022). Others indicate that contributions indeed depend on expended effort. Keeping subjects uninformed about the heterogeneity regarding the sources of endowment, Muehlbacher and Kirchler (2009) find that individuals who have to exert more effort contribute less. Conversely, Harrison (2007) re-analyses the data of Clark (2002) and reaches a different conclusion. He finds that individuals have a higher propensity to free-ride when playing with windfall money. On the intensive margin, there is no house money effect.

However, to date, the discussion has ignored a major factor that may be crucial for explaining the house money effect. The complexity of economic decisions requires analytical reasoning. Since humans vary in their cognitive abilities (Frederick, 2005), these reasoning skills are fundamental determinants of heterogeneous responses to economic problems. Dohmen et al. (2010) show that cognitive ability correlates with risk and time preferences. Furthermore, subjects appear to be less self-serving in dictator games when being under cognitive load (Schulz et al., 2014). By contrast, Chen et al. (2013) conclude that cognitive ability is positively correlated with generosity in dictator games. Recent research also suggests an effect of cognitive processes on co-operation. Interestingly, there exist both studies that find a positive (Clark, 1998; Lohse, 2016) and a negative link (Kanazawa & Fontaine, 2013; Nielsen et al., 2014) between cognitive skills and co-operation.

Regarding the house money effect, cognitive abilities appear to be equally relevant. Many researchers suggest that people apply strategies like choice bracketing and mental accounting to simplify economic decisions (Thaler, 1999; Read et al., 1999). Read et al. (1999, p. 187) argue that cognitive limitations are a key determinant for individuals to bracket narrowly, thereby facilitating decision-making. Empirically, cognitive ability decreases the use of heuristics (Frederick, 2005). Correspondingly, when spending income, individuals with low cognitive abilities are likely to be less capable to abstract from the source of income and to use the source for mental accounting. By contrast, subjects with higher cognitive capacity can be expected to be less reliant on applying heuristic simplification methods. Instead, they are more likely to keep track of the entire available budget when spending income. Thus, if the differentiation of income sources is the result of individuals having to simplify decision-making, this suggests that a

[2] Only Luccasen and Grossman (2017) obtain an opposing result. They find that warm-glow giving to charity or philanthropic institutions is higher for earned endowment. The authors hypothesise that subjects derive more utility from donating earned money than an equally sized windfall gain.

subject's cognitive capacity may also be associated with the extent to which she exhibits the house money effect. An experiment by Abeler and Marklein (2017) supports this prediction. They find that individuals with lower cognitive skills have a higher propensity to violate the assumption of the fungibility of money. Similarly, Antonides et al. (2011) find that higher education is negatively associated with mental budgeting.

In this paper, I examine whether individuals differ in the degree to which they exhibit the house money effect in a public goods game. I build on two documented facts: (1) unearned income generally appears to be donated or shared more easily (e.g., Cherry, 2001), and (2) violations of fungibility negatively correlate with cognitive skills (e.g., Abeler and Marklein, 2017).

On the basis of these observations, I hypothesise that in a public goods game, contributions by individuals with low cognitive skills should be smaller the higher the share of earned income. By contrast, I expect the origin of income to have less or no effect on contributions by individuals with high cognitive skills.

The results demonstrate that subjects' contributions depend on an interplay of cognitive abilities and endowment origin. While a house money effect exists for subjects with lower cognitive ability, there is no such effect for those with high cognitive ability. The former contribute more when income was allocated to them and less when income was obtained by effort. Contrarily, the latter contribute the same amount independent of income type. This finding may enable policy-makers to encourage subsidised households to act more co-operatively and spend their money more efficiently (e.g., purchase goods with positive household externalities rather than purely private goods).

I proceed as follows. In Sect. 2, I describe the experimental design. Section 3 presents results as well as further robustness checks. I conclude and discuss potential implications of the results in Sect. 4.

2 Experimental design

The experiment[3] consisted of four parts: (1) a real-effort task, (2) a three-person one-shot linear public goods game, (3) the cognitive reflection test (CRT) developed by Frederick (2005), and (4) a questionnaire on demographic information. It was conducted in four sessions with first-year business administration students using paper and pencil at the Technical University of Munich, Germany.[4] After undergraduate Economics tutorials, students were invited to remain seated to participate in the experiment where they could win a monetary reward. Subsequently, participating

[3] While I acknowledge the importance and value of pre-registration in scientific research, this experiment was not pre-registered. However, it was specifically designed to identify the interplay between cognitive skills and the house money effect rather than exploring multiple hypotheses. Thus, the experimental material was limited to elicit only data required to determine this interplay. While pre-registration is valuable for transparency, the focused nature of this study mitigates concerns regarding undisclosed analyses.

[4] Instructions translated to English are provided in Appendix A.

students were distanced from each other in a large classroom to prevent contact during the experiment.

Although there is evidence that business and economics students are different from the rest of the population (Meier & Frey, 2004; Kirchgässner, 2005; McCannon & Peterson, 2015; Bauman & Rose, 2011), the sample is well suited for the analysis. First, respective studies find level differences in social preferences between students of different subjects. Since the main focus of this study is to detect an interaction effect, this would only affect the analysis if the overall level of contributions was too low or too high to find an interaction. Second, students within one discipline are less heterogeneous. Hence, an interaction effect in this sample can be considered a lower bound of the interaction. Cognitive ability can be assumed to vary substantially more in the entire population. As a consequence, the house money effect is also more prevalent and the interaction between cognitive ability and the house money effect (i.e., the interaction effect) more strongly pronounced when considering more heterogeneous populations.

2.1 Real-effort task

At the beginning of the experiment, the participants were randomly divided into three treatment groups.[5] Each group had to colour in a different number of circles out of a total of 150.[6] This task was used to simulate a cognitively non-demanding real-effort task. Hereby, depending on the treatment, a fraction of the 150 circles had already been filled in on behalf of the participants, sparing them a part of the effort. More precisely, the subjects had to colour in either 15 circles with 135 circles already being filled in (*Low Effort*), 75 circles with the other half being filled in (*Medium Effort*[7]), or 135 circles with only 15 circles being filled in (*High Effort*). Hence, subjects in the *Low Effort* condition, for instance, only had to provide the effort of filling in 10% of the total 150 circles to earn the endowment. All participants were informed about these three different treatments.

In total, 161 students participated in the experiment. Seven participants did not finish the task and were excluded from the experiment. This resulted in 65, 38, and 51 subjects in the *Low*, *Medium*, and *High Effort* treatment, respectively.[8] For having coloured in all circles appropriately, all subjects received the same endowment of

[5] Assignment to tutorials was the students' choice and thus non-random.

[6] See Fig. 5 on the Task Sheet in Appendix A for the *High Effort* treatment which required participants to fill in 135 circles.

[7] The *Medium Effort* treatment was included as a manipulation check. As expected, contributions by subjects in the *Medium Effort* treatment are between those of the *Low* and the *High Effort* treatment, for the pooled sample as well as for the subjects with a low and a high CRT score separately.

[8] For the *Low* and *High Effort* treatments, the uneven number results from instructions being shuffled before they were handed out. As the *Medium Effort* treatment was primarily included to check for any non-monotone relationship between effort and contributions, and holds limited relevance for the majority of analyses (see Fig. 3, and Tables 1 and 2), a reduced number of corresponding instructions was included in the set of instructions to be distributed (the set consisted of 120, 110, and 120 instructions for the *Low*, *Medium*, and *High Effort* treatments, respectively). Two, one, and four subjects were dropped from the *Low*, *Medium*, and *High Effort* treatment, respectively.

100 tokens (10 tokens = 0.60 Euro). As the participants had to colour in different numbers of circles, this induced different proportions of earned and allocated income (i.e., either 10, 50, or 90% of the total income was earned by effort).

2.2 Public goods game

Following the task, individuals could decide on which proportion of their endowment to invest in a three-person one-shot linear public good with a marginal per capita return of 0.5. Therefore, the pay-off function π_i of player i with $i \in \{1, 2, 3\}$ who contributes $\theta_i \in \{0;100\}$ is given by

$$\pi_i = (100 - \theta_i) + 0.5 \cdot \sum_{j=1}^{3} \theta_j \tag{1}$$

with $\theta_j \in \{0;100\}$ being the contribution of player $j \in \{1, 2, 3\}$. Importantly, subjects did not know the required effort levels of the other two players in their group. However, they knew that all combinations of the three treatments were possible.[9]

2.3 Cognitive reflection test

Subsequently, the students had to perform the CRT to elicit cognitive ability. It contains three computationally easy questions that all have an intuitive, yet incorrect answer, and one correct answer that requires deliberation.[10] Despite the test's brevity, it significantly correlates with results from more sophisticated tests such as the *Wonderlic Personnel Test* or the *Wechsler Matrix Test* (Frederick, 2005; Toplak et al., 2011, 2014). It is also popular and frequently used in economics experiments, (e.g., Haita-Falah, 2017), including public goods experiments (Nielsen et al., 2014; Lohse, 2016). The test is particularly suitable in a setting with two different types of endowment. It aims at separating types that exercise deliberation before answering the questions from those who follow their first intuition or cannot solve the problem. Participants are making use of an endowment from two different sources, so it might be reasonable at first glance to use these sources differently as well. However, further cognitive reflection should make individuals realise that the two income types are perfect substitutes. For that reason, in the analyses, I do not further distinguish between the type of incorrect answers.

[9] Therefore, unlike experiments involving participant groups with heterogeneous sources of endowments (e.g., Oxoby and Spraggon, 2013), the groups in my study ex ante did not exhibit variation in the composition of their endowments. Oxoby and Spraggon (2013) find that contributions decrease when there is a strong minority in terms of endowment origin. However, all participants in my experiment faced the same group composition in expectation, thereby ensuring homogeneity across groups.

[10] The questions can be found in Sect. *Additional Questions* in Appendix A. Overall, 67% of all questions were answered correctly, 20% were answered with the intuitive yet incorrect answer, and only 13% were answered with an incorrect answer unequal to the intuitive one.

It is feasible that both the real-effort task and the deliberation when choosing the contribution to the public good impact the performance in the CRT. Johnson et al. (2014) show that CRT scores decrease with cognitive load. However, a real-effort task was chosen that is cognitively non-demanding. Moreover, participants did not have to memorise anything after making the decision in the public goods game, the standard procedure to induce cognitive load. Finally, the research question dictates the real-effort task to precede the decision in the public goods game, and the CRT, since it can be considered a real-effort task, to follow it.

Finally, participants had to complete a questionnaire on demographic information.

2.4 Payment

After all experimental sessions had been completed, a total of 18 participants[11] were randomly chosen using a random number generator and assigned into groups of three to receive the resulting pay-off. Nevertheless, the participants remained completely anonymous. The composition of groups was only announced using participation numbers. Also, final payments were subsequently done in private while ensuring that subjects could not see which participants were drawn or formed a group. Both was communicated at the beginning of each session. I also informed the participants prior to the experiment that the chances of being drawn were at least ten percent. Following Dohmen et al. (2010, p.1245) this ensures incentive compatibility although ultimately not everyone is being paid. On average, the selected students earned 7.68 Euro.

3 Results

On average, subjects contribute around 51% of their initial endowment. In the CRT, 40% answered all three questions correctly, followed by 32% with two, 16% with one, and 12% with zero correct answers.[12]

3.1 Contributions by endowment origin

In line with the existing literature, contributions in the public goods game do not differ significantly across treatments, i.e., effort levels. Of their equally high endowment, subjects contribute 55% in the *Low* and 48% in the *High Effort* treatment ($p = .210$, independent t test[13]; $p = .230$, Mann–Whitney U test, $n = 116$; Fig. 1).

[11] Six participants were chosen from both Session 1 with a total of 58 participants, and Session 4 (66). Additionally, three participants each were selected from Session 2 (24) and Session 3 (13).

[12] As the sample consists of first-year students who have not yet had opportunities to participate in experiments, I am confident that they did not know the test and preclude concerns of familiarity raised by recent studies (e.g., Toplak et al., 2014).

[13] Unless otherwise stated, every time an independent t test is conducted, I consider a two-tailed t test. With reported t tests, corresponding Shapiro–Wilk tests of normality and two-sample variance-comparison tests do not reject normality or equal variances.

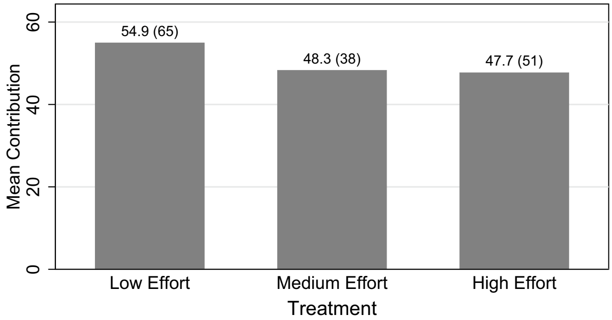

Fig. 1 Mean contributions by treatment group (sample size in parentheses)

3.2 Contributions by cognitive ability

For the analysis of contributions by cognitive ability, I divide the sample into individuals with high (40% of the sample) and low cognitive skills (60%). *High CRT* (*Low CRT*) individuals answered all three (two or less) questions correctly. Across treatments, contributions of individuals with a high CRT score are not significantly different to those of their low CRT counterparts, both contributing around 51% (p = .885, t test with unequal variances; p = .826, Mann–Whitney U test, n = 154; Fig. 2). This suggests that subjects with low and high cognitive ability do not differ in how much they contribute to a public good when the dominant strategy is to give nothing.

3.3 Contributions by endowment origin and cognitive ability

A breakdown of contributions by treatment and CRT results confirms the hypothesis derived in Sect. 1. As depicted in Fig. 3, I find that subjects' contributions depend on the interaction of their cognitive abilities and their endowment source. On the one hand, individuals with a low CRT score contribute 62.9% in the *Low Effort* treatment and 43.9% in the *High Effort* treatment. An independent t test shows that this difference is significant (p = .008, t test; p = .016, Mann–Whitney U test, n = 63) and demonstrates that a low CRT score is associated with behaviour exhibiting the house money effect. On the other hand, contributions by individuals with a high CRT score do not differ significantly by endowment source, 48% in the *Low* versus 55% in the *High Effort* treatment (p = .496, t test; p = .578, Mann–Whitney U test, n = 53). This suggests that subjects with higher cognitive ability are less likely to exhibit the house money effect.

The endowment source determines whether subjects with a low CRT score appear to be more or less co-operative than subjects with a high CRT score. In the *Low Effort* treatment, contributions by subjects with low CRT scores are 14.8 percentage

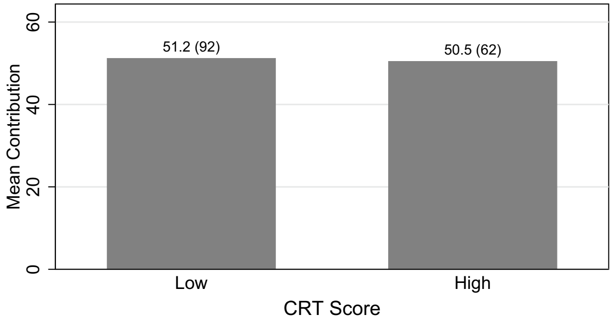

Fig. 2 Mean contributions by CRT scores (sample size in parentheses)

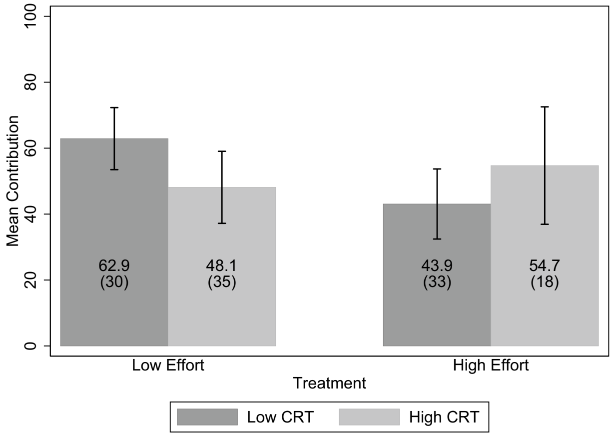

Fig. 3 Mean contributions by CRT score and treatment group. Error bars indicate 95% confidence intervals (sample size in parentheses)

points higher than contributions by subjects with high scores, 62.9 versus 48.1% ($p = .044$, t test; $p = .053$, Mann–Whitney U test, $n = 65$). Thus, subjects with low CRT scores behave relatively more co-operatively when they had to exert less effort to obtain their endowment. Contrarily, in the *High Effort* treatment, their contributions are 10.8 percentage points lower than those of subjects with high scores (43.9 versus 54.7%). However, this difference is not statistically significant, possibly due to a too small sample size ($p = .250$, t test; $p = .295$, Mann–Whitney U test, $n = 51$).[14] Hence, while being more co-operative than subjects with high CRT scores in the *Low Effort* treatment, subjects with lower CRT scores behave similarly to the high

[14] The observed substantial difference between contributions of subjects with low CRT scores compared to those with high CRT scores, along with the moderate p-value, prompts the need for a power analysis to determine the probability of detecting a difference if it exists. Based on this analysis ($\alpha = .1$ and $n = 51$), I found that the statistical power of the t test is estimated to be.295. This result suggest that with a feasibly higher sample size, a difference would be detectable. The sample size that would be needed to identify a significant difference with the given means and standard deviation is 262 observations.

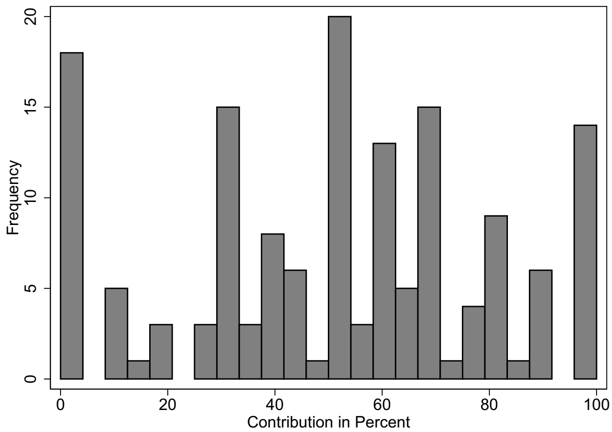

Fig. 4 Histogram of contributions

CRT scorers and possibly less co-operative when using earned income. An independent t test shows that this interaction of treatment and CRT scores is statistically significant ($p = .029$).

3.4 Multivariate analysis

The interaction effect is also robust to controlling for gender, age, and session effects. Similar to Harrison (2007), I resort to the Hurdle model (2) that accommodates the extensive and intensive margin of decisions in public goods games. As depicted by Fig. 4, a large fraction of subjects (11.0%) decides to contribute zero to the public good.

The Hurdle model therefore has two components. First, it estimates whether a subject contributes anything at all (3). Second, it fits a linear outcome model (4). It can be characterised by the following set of relations (see Botelho et al., 2009, for a discussion of the Hurdle model)

$$y_i = s_i h_i^*, \tag{2}$$

$$\text{with } s_i = \begin{cases} 1 & \text{if } \mathbf{x}_i\gamma + \epsilon_i > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

$$\text{and } h_i^* = \mathbf{x}_i\beta + v_i, \tag{4}$$

where y_i is the estimated contribution. s_i is the selection variable which is 1 if a subject is estimated to contribute a positive amount and 0 otherwise. h_i^*, the latent variable, is a subject's expected contribution, conditional on the contribution being positive. $\mathbf{x}_i$ is the vector of explanatory variables. I include the same variables in both the selection and the outcome model. Thus, γ and β are the corresponding vectors of coefficients, and ϵ_i and v_i are error terms.

Table 1 Maximum-likelihood estimates of the Hurdle model of contributions using the binary treatment variable (*High Effort*) and the binary CRT variable *High CRT*

	Contribution	Selection (Probit)
High Effort	− 21.474*	− 0.620
	(8.605)	(0.629)
High CRT	− 11.746	− 0.836
	(7.145)	(0.573)
High Effort × High CRT	30.617**	0.670
	(11.565)	(0.763)
Age	1.603	− 0.101
	(1.316)	(0.074)
Male	7.077	− 0.863
	(5.977)	(0.505)
Constant	30.724	4.859**
	(27.811)	(1.754)
Session Effects	Yes	Yes
N	116	

Notes: The variable *High Effort* is a binary variable, which takes a value of one if the subject is assigned to the *High Effort* treatment and zero for the *Low Effort* treatment. Therefore, the regression is run on 116 observations of the *Low* and *High Effort* treatment, excluding those of the *Medium Effort* treatment. *High CRT* is equal to 1 if the subject has answered all CRT questions correctly and 0 otherwise. Male is equal to one for males and zero otherwise

Standard errors in parentheses; $*$ $p < 0.05$, $**$ $p < 0.01$, $***$ $p < 0.001$

Table 1 reports the estimation results using the dummy for CRT score. While none of the control variables has a significant effect on the contribution to the public good, the required effort to earn the original endowment significantly reduces contributions by subjects with low cognitive skills by −21.5, given they decide to contribute. Due to the interaction effect, High Effort × High CRT of 30.6, this effect vanishes for subjects with high CRT scores. Surprisingly, none of the variables has a significant effect on the extensive margin. I am grouping the intensive and extensive margin in Appendix B, obtaining a qualitatively identical result.

Computing the predicted values based on the coefficients of both stages of the Hurdle model confirms the interaction effect and the differences within treatments observed in the non-parametric analysis (see Table 2). Individuals with low CRT scores contribute 22.1 percentage points less in the *High* than in the *Low Effort* treatment (combined marginal effect of the variable *High Effort* for *Low CRT* subjects, $p = .006$). For subjects with high CRT scores, the difference is positive (they contribute more), but not significant (7.9, combined marginal effect of the variable *High Effort* for *High CRT* subjects, $p = .443$). The difference between individuals with low and high CRT scores is 16.8 (combined marginal effect of the variable *High CRT*, $p = .023$) in the *Low Effort* treatment and −13.1 (combined marginal effect of the variable *High CRT*, $p = .144$) in the *High Effort* treatment.

Table 2 Contributions as estimated by the Hurdle model

	Low effort	High effort	*p*-value
Low CRT	64.6	42.6	0.006
High CRT	47.7	55.6	0.443
p-Value	0.023	0.144	

To conclude, subjects with low CRT scores exhibit the house money effect, giving more in the *Low* than in the *High Effort* condition. By contrast, contributions by subjects with high CRT scores are statistically indistinguishable. Moreover, subjects with low CRT scores give more than their high CRT counterparts in the *Low Effort* treatment, but less in the *High Effort* treatment. As these differences partly cancel out, it seems that cognitive ability does not affect co-operative behaviour when pooling treatments.[15]

3.5 Discussion

The results raise the question whether studies that find no significant effect of effort in public goods games but do not account for cognitive ability should at least find suggestive evidence for a house money effect.

In fact, Antinyan et al. (2015) obtain a negative but insignificant effect of effort on contributions in their *No Punishment* treatment as well as in their *Punishment* treatment without controlling for an interaction of time and effort. Further, effort reduces the proportion of endowment contributed to the public good in Cherry et al. (2005). Finally, Clark (2002) finds suggestive but insignificant evidence for a house money effect in the initial round of a repeated public goods game. Similarly, with an experimental design that addresses issues with the design of Clark (2002), Bailey et al. (2022) find no significant house money effect, but slightly higher contributions by subjects who received the endowment on the day of the experiment (their house money treatment) versus 3 weeks prior to the experiment (which subjects mostly viewed to be their own money by the time the public goods game was played).

Hence, the interaction between cognitive ability and the house money effect in my experiment, combined with previous studies showing similar but insignificant effects in the same direction, suggests that their insignificance results may be the consequence of homogenous subject pools (university students in all mentioned

[15] I run several robustness checks to further confirm these results. Collapsing the extensive and intensive margin as well including CRT scores and Effort as continuous variables as well as controlling for risk preferences does not change the conclusion drawn from the main analysis (see Appendix B). Using final math grades instead of CRT results provides suggestive yet non-significant evidence of an interaction effect.

studies). A house money effect might be more prevalent in a cross-section of society. Also, my experiment was not designed to investigate the persistence of a house money effect. However, the revealed interplay between cognitive skills and endowment origin asks for further experiments investigating the house money effect over time across distinct cognitive types.

Despite the evidence of a house money effect among subjects with lower cognitive skills and the broad existence of a house money effect in dictator games (Cherry, 2001; Cherry et al., 2002; Cherry & Shogren, 2008; Reinstein & Riener, 2012; Oxoby & Spraggon, 2008; Carlsson et al., 2013), it remains unclear whether the origin of the endowment indeed affects social preferences. Alternatively, unearned income or income that required little effort might just be more loosely handled without any pro-social motive. Identifying the exact psychological mechanism should be the purpose of future research.

4 Conclusion

Experimental studies have frequently used public goods games to shed light on the dynamics behind co-operative behaviour (see Zelmer, 2003, for an overview on public goods games). This is the first study to show that contributions to a public good depend on an interplay of income origin and cognitive skills. This result has implications for the interpretation of past and future experiments in which subjects are either given or earn their endowments. For instance, studies that find a relationship between factors such as cognitive ability and generosity or willingness to co-operate[16] may only hold true for the specific experimental design used, for example, if income is given by the experimenter, but the relationship may vanish or even be reversed if the design is changed, for example, if income has to be earned. Therefore, the generalisability of these findings should be considered with caution, and the specific experimental conditions under which the relationship was observed should be taken into account. In the experiment, subjects with a high CRT score behave more co-operatively in the *High Effort* treatment but less so in the *Low Effort* treatment. Thus, the external validity of respective results might particularly be questioned, considering populations that are more cognitively heterogeneous than students. In economics, the participant pool often consists solely of university students who are likely to be rather homogenous (Frederick, 2005) and to have above average cognitive abilities. Nevertheless, I am able to identify an interaction effect even within this group.

[16] This applies to studies that find subjects with higher cognitive ability are more (Chen et al., 2013) or less (Schulz et al., 2014) generous or more (Clark, 1998; Lohse, 2016) or less (Kanazawa & Fontaine, 2013; Nielsen et al., 2014) co-operative.

Furthermore, contributions in a public goods game can *inter alia* be regarded as subjects' preferences for redistribution. Thus, the results are relevant for studying redistributive taxation and tax compliance (e.g., Bühren and Kundt, 2014). My results provide evidence that people with low cognitive skills may exhibit different spending patterns for earned income than for unearned income, which could have implications for the design of public policies that aim to support low-income individuals or reduce inequality. Assuming some degree of correlation between cognitive skills and income, the fact that non-earned income is spent more co-operatively can be considered as an argument in favour of existing and additional income subsidies, allowing these families to exploit positive externalities within families.

This study also shows that people with low cognitive skills are less willing to give away or invest earned income, suggesting that they may perceive it as more valuable for their own consumption. This could have implications for compensation practices and the allocation of performance-based and flat pay. Employers may need to consider how they structure pay and benefits packages to ensure that employees with different cognitive skills feel valued and motivated.

Finally, team production can also be compared to a public goods game (Alchian & Demsetz, 1972). Contributions by one team member benefit all other team members. For this reason, team production is equally exposed to a free-rider problem. Hence, a variation in how agents receive the resources they can use as production inputs (e.g., information) can alter team member's propensity to co-operate.

Appendix A: Instructions

Dear participant,

To begin with, I would like to thank you for partaking in this experiment.

For this experiment, we do not use Euro as our currency, but **ECU (Experimental Currency Units)** instead. Upon completion of the experiment, the **ECUs** you have earned will be converted to **Euro**. The exchange rate equals **10 ECU = 0.60 EURO**. After the experiment ends, randomly selected students will receive the payoff they have obtained.

This experiment consists of two parts: the first requires you to fulfil a task; in the second you will be asked to invest ECUs.

Part 1: Task

To complete part one of the experiment, 10 rows of circles must be filled in while either 1, 5, or 9 rows have already been filled in. For completing this task, you receive an initial endowment of **100 ECU**. To participate in the draw, determining which students receive monetary payoffs, all rows must be filled in.

Part 2: Investment

In part two, you anonymously play an economic game with two other participants. The amount of rows these participants had to fill in was randomly determined.

This game provides you with the option to invest a share of your initial endowment. The investment of all three group members is added up, then **multiplied by 1.5**, and subsequently **split evenly among all three group members**.

The share of your initial endowment you chose not to invest goes directly towards your balance at the end of a round

$$\text{Payoff} = (\text{Initial Endowment} - \text{Investment}) + 1/3 \cdot (1.5 \cdot \text{Sum of Investments}).$$

Example:

Of her 100 ECU initial endowment, a participant (group member 1) decides to keep 20 ECU and invest 80 ECU. The two other group members decide to invest 40 ECU (group member 2) and 60 ECU (group member 3), respectively. In total, 80 ECU + 40 ECU + 60 ECU = 180 ECU were invested. Multiplied by 1.5, this amounts to 270 ECU, which is then divided evenly among all group members (90 ECU per person). As a result, the individual group members receive the following payoffs:

- Group member 1 keeps the 20 ECU she did not invest and receives an additional 90 ECU from the investment, a total of 110 ECU.
- Group member 2: 60 ECU (= 100 ECU − 40 ECU) + 90 ECU (Investment) = 150 ECU.
- Group member 3: 40 ECU (= 100 ECU − 60 ECU) + 90 ECU (Investment) = 130 ECU.

Payoff

Following this experiment, all task and decision sheets will be collected. For this reason, please detach this sheet from the second one. After the collection of the sheets, the winners will immediately and anonymously be determined. These individuals' responses will be used to calculate their respective payoffs. In the case of an incomplete response sheet, the draw will be repeated. The winners will be able to receive their payoffs at my office (2423) after presenting their title sheet and subject id/participant number, which can be found at the end of all sheets.

Task sheet *(separate page)*

Please fill in all empty circles with a ballpoint pen.

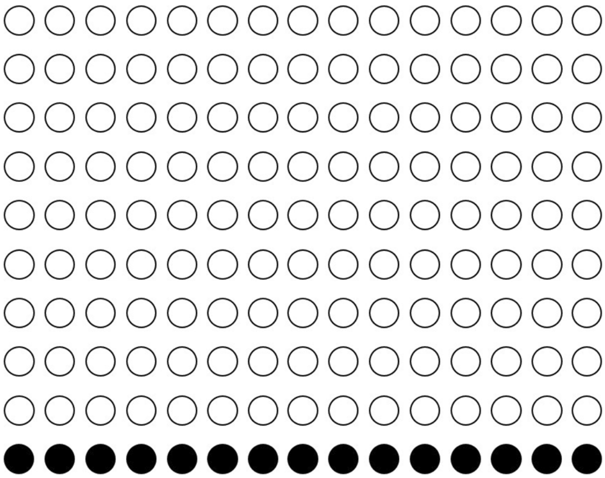

Fig. 5 Circles to be filled in by participants

Decision sheet *(separate sheet)*

Investment decision

What amount would you like to invest?

Please choose a number between 0 and 100. Note that any amount of this endowment, which you choose not to invest is counted directly towards your payoff.

Additional questions

1. A bat and a ball together cost 110 cents. The bat costs 100 cents more than the ball. How much does the ball cost? (in Euro)

Euro

2. If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? (in minutes)

Min

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (in days)

Days

Risk preferences

Assess yourself: Are you more of a risk-taking person or do you think of yourself as a risk-avoider? Please tick a box on the scale below, 0 indicating "no tolerance for risk" and 10 indicating "very risk-seeking". The values in between can help you more finely represent your image of yourself.

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7	8	9	10

Demographic information

In closing, I would like to ask you to give some information on yourself. It is important for analysing the data created in this experiment and will be treated strictly confidentially.

Your gender: ☐ Female

☐ Male

Your age: ☐

Your final math grade: ☐

Appendix B: Robustness checks

I first check if I obtain the same results without separating the extensive and intensive margin [Specifications (1) and (2) in Table 3]. To test whether the results are driven by including the CRT scores as a binary variable, I run an OLS regression replacing the dummy with a continuous variable (*Correct*) that indicates the number of correctly answered CRT questions [Specification (4) in Table 3]. Finally, while I have excluded the *Medium Effort* treatment so far, I treat effort as a continuous variable and therefore include corresponding observations in Specification (6). Across specifications, significance levels and directions are similar for the treatment variable (*High Effort*), for the measures of cognitive ability (the dummy *High CRT* indicating that a subject has answered all CRT questions correctly and *Correct* as the total number of correct answers), and the interaction effect (*High Effort* × *High CRT* and *High Effort* × *Correct*).

Although the payoffs in a one-shot linear public good game are completely deterministic and do not entail risk, there is evidence of a correlation between risk preferences and contributions to a public good (Charness & Villeval, 2009). Therefore, Specifications (3), (5), and (7) control for participants' risk attitudes. Including risk preferences does not impact the existence of an interplay between cognitive skills and the house money effect. Even with Specification (3), where the interaction effect itself is just not significant at the 10% level anymore ($p = .101$), computing average marginal effects shows that participants with low CRT scores contribute significantly less in the *High Effort* treatment than in the *Low Effort* treatment ($p = 0.007$), while participants with high CRT scores don't ($p = 0.819$).

In addition to the CRT as a measure for cognitive ability or ability to overcome one's faulty intuition, I test whether the interaction effect exists when taking the

Table 3 OLS regression results for *Contribution* using binary (*High CRT*, (1)–(3)) and continuous (*Correct*, (4)–(7)) measurements of cognitive ability, binary (*High Effort*, (1) - (5)) and continuous (*Effort Level*, (6) and (7))) treatment variables, and including *Risk Attitude* as control variable (3), (5), and (7))

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High Effort	− 19.021**	− 22.155**	− 19.851**	− 35.144**	− 31.944**		
	(6.891)	(8.171)	(7.255)	(11.446)	(10.049)		
Effort Level						− 0.438**	− 0.397**
						(0.143)	(0.126)
High CRT	− 14.786*	− 15.839*	− 8.689				
	(7.092)	(7.407)	(6.521)				
Correct				− 8.090*	− 5.546	− 8.649*	− 5.962
				(3.678)	(2.908)	(3.951)	(3.387)
High Effort × High CRT	25.629*	29.033*	17.850				
	(12.097)	(12.559)	(10.801)				
High Effort × Correct				12.271*	9.501*		
				(5.365)	(4.515)		
Effort Level × Correct						0.144*	0.113*
						(0.066)	(0.056)
Age		0.349	0.277	0.213	0.210	0.519	0.842
		(1.282)	(1.103)	(1.356)	(1.096)	(1.054)	(0.911)
Male		− 0.112	− 2.115	1.516	− 1.129	0.274	− 3.095
		(5.910)	(5.270)	(5.956)	(5.184)	(4.758)	(4.337)
Risk Attitude			7.406***		7.507***		5.794***
			(1.227)		(1.209)		(1.105)
Constant	62.900***	60.299*	19.021	71.385*	27.042	65.988**	23.961
	(4.604)	(27.453)	(23.223)	(30.305)	(23.600)	(23.905)	(20.881)
Session Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
N	116	116	115	116	115	154	153

Notes: The variable *High Effort* is a binary variable, which takes a value of one if the subject is assigned to the *High Effort* treatment and zero for the *Low Effort* treatment. Therefore, the regression is run on 116 (115 if *Risk Attitude* is included) observations of the *Low* and *High Effort* treatment in Specifications (1) - (5), excluding those of the *Medium Effort* treatment. Specifications (6) and (7) include the 38 observations from the *Medium Effort* treatment. The variable *Effort Level* is the relative amount in percentages of circles that the participant had to fill in. *High CRT* is equal to 1 if the subject has answered all CRT questions correctly and 0 otherwise. The variable *Correct* is the number of correctly answered questions in the CRT. Male is equal to 1 for males and 0 otherwise. Risk Attitude ranges from low (0) to high (10)

Heteroskedasticity robust standard errors in parenthesis

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

students' final math grade as an explanatory variable as proposed by Abeler and Marklein (2017). The results also point towards an interaction effect. However, the coefficients are not statistically significant (see Table 4 and Fig. 6 for the estimated coefficients and the correspondingly predicted contributions in a Hurdle model.)

Table 4 Maximum-likelihood estimates of the Hurdle model of contributions using the continuous treatment variable (*Effort Level*) and the final math grade of the students as a measure for cognitive ability

	Contribution	Selection (Probit)
Effort level	− 0.561*	0.004
	(0.264)	(0.016)
Final math grade	− 2.513	0.138
	(1.324)	(0.095)
Effort level × Final math grade	0.036	− 0.000
	(0.023)	(0.001)
Age	1.398	0.009
	(1.107)	(0.083)
Male	8.148	− 1.209*
	(4.942)	(0.539)
Constant	57.160*	0.689
	(27.585)	(1.959)
Session effects	Yes	Yes
N	147	

Notes: The variable *Effort Level* is the relative amount in percentages of circles that the participant had to fill in. The final math grade ranges from 0 (worst grade, no observations) to 15 (best grade). A minimum of five points is required to pass. The variable has a mean of 10.97 and is only available for students with grades conforming with the German system (N = 147). Male is equal to one for males and zero otherwise

Heteroskedasticity robust standard errors in parenthesis

$* \ p < 0.05$, $** \ p < 0.01$, $*** \ p < 0.001$

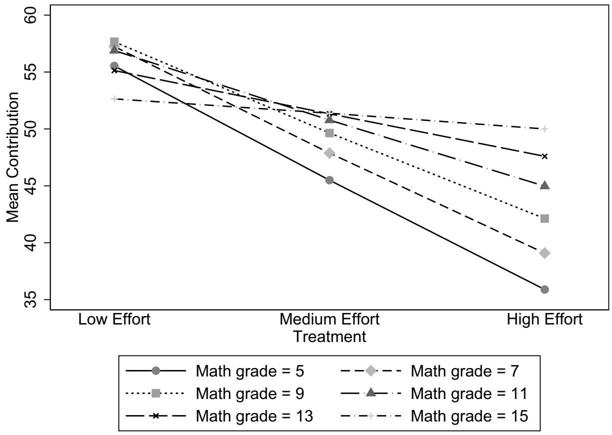

Fig. 6 Mean contributions by treatment group and math grade as estimated by the Hurdle model in Table 4. Notes: The final math grade ranges from 0 (worst grade, no observations) to 15 (best grade). A minimum of five points is required to pass. The variable has a mean of 10.97 and is only available for students with grades conforming with the German system

Acknowledgements First of all, I would like to thank Robert K. von Weizsäcker for his invaluable advice and support. Furthermore, I acknowledge helpful comments and suggestions from Miriam Leidinger, Christoph March, Michael Kurschilgen, Rachel Croson, Simon Gächter, Glenn W. Harrison, and Arnd Klein, and participants of the SMYE in Lisbon, ABBESS in Athens, M-BEES in Maastricht, ASFEE in Cergy (Paris), BMD in Augsburg, SABE/IAREP in Wageningen, BEWIP in Munich, and of the research seminar at the TUM School of Management.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The replication material for the study is available at https://doi.org/10.17605/OSF.IO/ZH6QS.

References

Abeler, J., & Marklein, F. (2017). Fungibility, Labels, and Consumption. *Journal of the European Economic Association, 15*(1), 99–127.

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *American Economic Review, 62*(5), 777–795.

Antinyan, A., Corazzini, L., & Neururer, D. (2015). Public good provision, punishment, and the endowment origin: Experimental evidence. *Journal of Behavioral and Experimental Economics, 56*, 72–77.

Antonides, G., De Groot, I. M., & Van Raaij, W. F. (2011). Mental budgeting and the management of household finance. *Journal of Economic Psychology, 32*(4), 546–555.

Bailey, N. T., Ramalingam, A., & Stoddard, B. V. (2022). Experimental (re-) analysis of the house-money effect in a public goods game. *Journal of the Economic Science Association*, (pp. 1–14).

Bauman, Y., & Rose, E. (2011). Selection or indoctrination: Why do economics students donate less than the rest? *Journal of Economic Behavior & Organization, 79*(3), 318–327.

Botelho, A., Harrison, G. W., Pinto, L. M. C., & Rutström, E. E. (2009). Testing static game theory with dynamic experiments: A case study of public goods. *Games and Economic Behavior, 67*(1), 253-265.e3.

Bühren, C., & Kundt, T. C. (2014). Does the level of work effort influence tax evasion: Experimental evidence. *Review of Economics, 65*(2), 137–158.

Cárdenas, J. C., de Roux, N., Jaramillo, C. R., & Martinez, L. R. (2014). Is it my money or not? An experiment on risk aversion and the house-money effect. *Experimental Economics, 17*(1), 47–60.

Carlsson, F., He, H., & Martinsson, P. (2013). Easy come, easy go: The role of windfall money in lab and field experiments. *Experimental Economics, 16*(2), 190–207.

Charness, G., & Villeval, M.-C. (2009). Cooperation and competition in intergenerational experiments in the field and the laboratory. *American Economic Review, 99*(3), 956–978.

Chen, C.-C., Chiu, I.-M., Smith, J., & Yamada, T. (2013). Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency. *Journal of Economic Behavior & Organization, 90*, 112–122.

Cherry, T. L. (2001). Mental accounting and other-regarding behavior: Evidence from the lab. *Journal of Economic Psychology, 22*(5), 605–615.

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review, 92*(4), 1218–1221.

Cherry, T. L., Kroll, S., & Shogren, J. F. (2005). The impact of endowment heterogeneity and origin on public good contributions: Evidence from the lab. *Journal of Economic Behavior & Organization, 57*(3), 357–365.

Cherry, T. L., & Shogren, J. F. (2008). Self-interest, sympathy and the origin of endowments. *Economics Letters, 101*(1), 69–72.

Clark, J. (1998). Fairness preferences and optimization skills: Are they substitutes? *Journal of Economic Behavior & Organization, 34*(4), 541–557.

Clark, J. (2002). House money effects in public good experiments. *Experimental Economics, 5*(3), 223–231.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review, 100*(3), 1238–1260.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.

Haita-Falah, C. (2017). Sunk-cost fallacy and cognitive ability in individual decision-making. *Journal of Economic Psychology, 58*, 44–59.

Harrison, G. W. (2007). House money effects in public good experiments: Comment. *Experimental Economics, 10*(4), 429–437.

Hoffman, E., & Spitzer, M. L. (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *The Journal of Legal Studies, 14*(2), 259–297.

Houser, D., & Xiao, E. (2015). House money effects on trust and reciprocity. *Public Choice, 163*(1–2), 187–199.

Johnson, E. D., Tubau, E., & De Neys, W. (2014). The unbearable burden of executive load on cognitive reflection: A validation of dual process theory. In *Proceedings of the annual meeting of the Cognitive Science Society*, volume 36.

Kanazawa, S., & Fontaine, L. (2013). Intelligent people defect more in a one-shot prisoner's dilemma game. *Journal of Neuroscience, Psychology, and Economics, 6*(3), 201–213.

Kirchgässner, G. (2005). (Why) are economists different? *European Journal of Political Economy, 21*(3), 543–562.

Lohse, J. (2016). Smart or selfish—When smart guys finish nice. *Journal of Behavioral and Experimental Economics, 64*, 28–40.

Luccasen, A., & Grossman, P. J. (2017). Warm-glow giving: Earned money and the option to take. *Economic Inquiry, 55*(2), 996–1006.

McCannon, B. C., & Peterson, J. (2015). Born for finance? Experimental evidence of the impact of finance education. *Journal of Behavioral Finance, 16*(3), 199–205.

Meier, S., & Frey, B. S. (2004). Do business students make good citizens? *International Journal of the Economics of Business, 11*(2), 141–163.

Muehlbacher, S., & Kirchler, E. (2009). Origin of endowments in public good games: The impact of effort on contributions. *Journal of Neuroscience, Psychology, and Economics, 2*(1), 59–67.

Nielsen, U. H., Tyran, J.-R., & Wengström, E. (2014). Second thoughts on free riding. *Economics Letters, 122*(2), 136–139.

Oxoby, R. J., & Spraggon, J. (2008). Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization, 65*(3–4), 703–713.

Oxoby, R. J., & Spraggon, J. (2013). A clear and present minority: Heterogeneity in the source of endowments and the provision of public goods. *Economic Inquiry, 51*(4), 2071–2082.

Read, D., Loewenstein, G., & Rabin, M. (1999). Choice Bracketing. *Journal of Risk and Uncertainty, 19*(1/3), 171–197.

Reinstein, D., & Riener, G. (2012). Decomposing desert and tangibility effects in a charitable giving experiment. *Experimental Economics, 15*(1), 229–240.

Schulz, J. F., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology, 41*, 77–87.

Spraggon, J., & Oxoby, R. J. (2009). An experimental investigation of endowment source heterogeneity in two-person public good games. *Economics Letters, 104*(2), 102–105.

Thaler, R. H. (1985). Mental Accounting and Consumer Choice. *Marketing Science, 4*(3), 199–214.

Thaler, R. H. (1990). Anomalies: Saving, Fungibility, and Mental Accounts. *Journal of Economic Perspectives, 4*(1), 193–205.

Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral decision making, 12*(3), 183.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & cognition, 39*(7), 1275–1289.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning, 20*(2), 147–168.
Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics, 6*(3), 299–310.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.