

Grohmann, Antonia; Menkhoff, Lukas; Merkle, Christoph; Schmacker, Renke

Working Paper

Earn more tomorrow: Overconfidence, income expectations and consumer indebtedness

DIW Discussion Papers, No. 2065

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Grohmann, Antonia; Menkhoff, Lukas; Merkle, Christoph; Schmacker, Renke (2023) : Earn more tomorrow: Overconfidence, income expectations and consumer indebtedness, DIW Discussion Papers, No. 2065, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/283247>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

2065

Discussion Papers

Deutsches Institut für Wirtschaftsforschung

2023

Earn More Tomorrow: Overconfidence, Income Expectations and Consumer Indebtedness

Antonia Grohmann, Lukas Menkhoff, Christoph Merkle and Renke Schmacker

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

DIW Berlin, 2023

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<https://www.diw.de>

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<https://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<https://ideas.repec.org/s/diw/diwwpp.html>
<https://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

Earn more tomorrow: Overconfidence, income expectations and consumer indebtedness*

Antonia Grohmann,[†] Lukas Menkhoff,[‡] Christoph Merkle,[§] and Renke Schmacker[¶]

October 2023

Abstract

This paper examines whether biased income expectations due to overconfidence lead to higher levels of debt-taking. We show suggestive evidence for a link between overconfidence and borrowing behavior in a representative survey of German households (GSOEP-IS). This motivates a laboratory experiment to study causality behind these effects. In two experiments, participants can purchase goods by borrowing against their future income. We exogenously manipulate overconfidence about income expectations by letting income depend on relative performance in hard and easy quiz tasks. In the main experiment, we successfully generate biased income expectations and show that participants with higher income expectations initially borrow more. Overconfident participants scale back their consumption after income feedback. However, they remain in higher debt at the end of the experiment, which has real financial consequences. In a robustness experiment, we rule out that over-borrowing is driven by low prices of goods. Even though the expected income manipulation works less well in this experiment, debt-taking behavior is very similar and correlates with income expectations and overconfidence.

JEL classification: D14 (Personal Finance), D84 (Expectations), G40 (Behavioral Finance).

Keywords: Debt, Consumption, Borrowing, Overconfidence, Income Expectations.

*We thank for helpful comments Dirk Engelmann, Zack Enslin, Peter Haan, Rustamdjan Hakimov, Frank Heinemann, Olaf Hübler, Markku Kaustia, Melanie Koch, Dorothea Kübler, Camelia Kuhnen, Steffen Meyer, Luís Santos-Pinto, Roel van Veldhuizen, Georg Weizsäcker, participants of the Experimental Finance 2018, the ESA 2018 World Meeting, the Research in Behavioral Finance Conference 2018, the SAFE Household Finance Workshop 2018, the DGF Annual Meeting 2019, the Behavioral Finance Working Group Conference 2021, seminar participants in Aalborg, Berlin, and Dublin. We further thank the editor (Pok-Sang Lam) and two anonymous referees. Financial support by the German Research Foundation (DFG, grant CRC 190) and the Danish Finance Institute (DFI) is gratefully acknowledged.

[†]Aarhus University BSS, Fuglesangs Allé 4, 8210 Aarhus, Denmark, agrohmann@econ.au.dk.

[‡]DIW Berlin, Mohrenstrasse 58, 10117 Berlin, Germany, lmenkhoff@diw.de, and Humboldt-Universität zu Berlin and IfW Kiel

[§]Aarhus University BSS, Fuglesangs Allé 4, 8210 Aarhus, Denmark, cmerkle@econ.au.dk.

[¶]UNIL HEC Economics, Batiment Internef 531, 1015 Lausanne, Switzerland, Renke.Schmacker@unil.ch.

1. Introduction

Consumer indebtedness is a core economic issue in modern societies. There are millions of individuals in every major OECD country who are considered over-indebted. Their share is estimated to be between 5 and 10 percent in most developed countries; while in Germany 10% of the adult population is considered over-indebted (Creditreform, 2018), the US leads this statistic with shares of up to 20% (Fondeville, Özdemir and Ward, 2010; OECD, 2018).¹ An even higher fraction of households self-reports to experience problems with debt servicing and debt repayment (Lusardi and Tufano, 2015; Money Advice Service, 2017). Too high household indebtedness may not only reduce individual welfare, but can also have negative effects at a larger scale and may adversely affect the economy (Melzer, 2011; Livshits, Mac Gee and Tertilt, 2016; Sufi, Mian and Verner, 2017). Its role for growth, financial crises, and unemployment has, therefore, received increased policy attention (IMF, 2012; 2017).

However, the reasons why so many people incur more debt than they can repay are not well understood. It is argued that many phenomena related to over-borrowing are hard to reconcile with rational expectations and may reflect behavioral biases (Amar et al., 2011; Zinman, 2015; Alan et al., 2018; Beshears et al., 2018; Gathergood et al., 2019).

In this paper, we examine one possible cause of (over-)borrowing, namely the effect of overconfidence on borrowing behavior. If people are overconfident about their ability, they may overestimate their income potential and, therefore, form exaggerated expectations about future income (Smith and Powell, 1990; Reuben, Wiswall and Zafar, 2017; Buser et al., 2020).

¹International statistics often lack comparability, and the definition of over-indebtedness varies. For example, the OECD uses a debt-to-asset ratio above 75% and a debt-to-income ratio exceeding three as measures of over-indebtedness (Murtin and d’Ercole, 2015).

A classical assumption in household finance is that people maximize their inter-temporal utility and smooth consumption by borrowing and saving according to their expectations (e.g., Friedman, 1957). Assuming that households behave rationally, over-indebtedness should only occur in the face of negative shocks that exceed the shock absorbing capacity of the respective household. Such shocks often include unemployment, unexpected illness, or divorce (Statistisches Bundesamt, 2018). However, if people have too optimistic income expectations they may overconsume early in life and run into debt even without major negative shocks (Brunnermeier and Parker, 2005).

We first analyze evidence for the relationship between general overconfidence, debt taking and over-indebtedness in a representative sample of German households (GSOEP). The survey asks participants to judge their ability in two short tasks relative to a representative group of the German population. We show that those who overestimate their ability in these tasks are more likely to use overdrafts. We further observe that these overconfident individuals also expect to be able to make all repayments on time. This may be due to overconfident income expectations. However, the survey does not contain a viable measure of income expectations, which prevents us from studying the direct relation between income expectations and debt taking. In addition, data collected from household surveys only provides evidence of correlations.

In order to examine the causality of the relationship between overconfidence, income expectations and debt taking we conduct a novel laboratory experiment, in which we exogenously vary the confidence of participants about their future earnings. To create such variation we exploit the “reversed hard-easy effect” (Burson, Larrick and Klayman, 2006; Moore and Small, 2007), which refers to the tendency of people to overestimate their rela-

tive performance in easy tasks and to underestimate their relative performance in hard tasks. We prime participants with either hard or easy sample questions that are representative for general knowledge questions they answer to earn income. As participants are paid according to their relative ability, the actual income on group level is independent of task difficulty. However, in line with the literature, we find that those assigned to easy tasks form higher income expectations than those assigned to hard tasks or a random payment.

Income expectations are crucial to the experiment, as income determines the ability to consume (snacks, beverages, and other items), but is only gradually revealed and paid out over time. Participants need to form income expectations to optimally purchase goods in a sequence of markets. They know the price level of goods will increase over time, which provides an incentive to buy goods early. Early purchases, therefore, allow a higher consumption level at a given budget. To be able to buy early, participants can use interest-free debt. Earned income cannot be saved and paid out in cash, but can only be spent on goods. Thus, participants face a decision in which it is rational to take up as much debt as they expect to earn during the experiment. We repeat the quiz task (to earn income) and the market (to consume) after the initial run, which allows us to examine the dynamic interaction between overconfident income expectations and borrowing.

In the main experiment, results support the hypothesis that higher income expectations raise the level of borrowing, with the experimental design allowing for a causal identification. We first verify that the manipulation of income expectations is successful. Participants in the easy treatment expect significantly higher income than participants in the hard treatment. Moreover, we find that those in the easy treatment initially borrow more and have higher debt

levels throughout the experiment. When income is revealed, they reduce their consumption, but they are still more likely to remain in debt at the end of the experiment.

Participants with overconfident income expectations realize that their actual income does not match their consumption plans in later rounds. They more often leave the experiment with a negative balance that is deducted from their show-up fee. We interpret them as over-indebted, as most participants would prefer to avoid this cash penalty. This means that overconfident income expectations predict over-indebtedness in the experiment. In the second and third experimental market, the treatment effect of task difficulty becomes weaker, as participants adjust to the treatment condition. Instead, we find a stronger effect of overconfidence, defined by the difference between expected and actual income, on final debt. We establish the causality of these results using an instrumental variables strategy, in which we use treatment assignment as an instrument for overconfidence. We find that each Euro of overestimated income increases final debt by 0.30 Euros.

In a robustness experiment, we examine the possibility that excessive debt-taking is induced by the low prices of goods in the main experiment. We thus replace discounted prices with retail prices, while keeping all other aspects of the experiment the same, including the increasing price schedule, which now serves as the only motivation to borrow. In the robustness experiment, the manipulation of income expectations via the hard-easy effect produces insufficient variation across treatments, which means that we cannot exploit the exogenous shift in expectations. Importantly, this does not imply that income expectations do not affect debt taking, but that the robustness experiment does not allow us to investigate the causal effect. Instead, we have to rely on endogenous income expectations to provide correlational evidence. We find that high income expectations are associated with higher debt taking and

that overconfidence predicts who remains in debt at the end of the experiment. In terms of amounts borrowed, debt-taking behavior is very similar to that in the main experiment. There is no evidence that low prices of goods are responsible for excessive debt-taking. We conclude that even without a significant treatment effect, the robustness experiment broadly confirms our main results.

Several studies have examined biased expectations as a potential reason for high levels of borrowing by households. Biased expectations have further been associated with poor repayment rates and over-indebtedness. Overestimation of one's own self-control has so far gained most attention in the literature. Heidhues and Köszegi (2010) show in a theoretical model that consumers, who are unaware of their self-control problems, will put off repayment in back-loaded credit contracts. Lack of self-control has empirically been related to sub-optimal borrowing behavior and over-indebtedness (Gathergood, 2012). Relatedly, consumers underestimate their usage of late payments and overdrafts. As a consequence, they pay no attention to related fees when taking out a loan (Gabaix and Laibson, 2006). Another form of biased expectations relevant for debt decisions is the exponential growth bias. People underestimate the exponential growth of a loan amount due to compound interest, and more biased households tend to borrow more (Stango and Zinman, 2009).

Our results further contribute to the literature on the link between income expectations and debt-taking. Hyttinen and Putkuri (2018) analyze a survey in which participants are asked to predict how their financial situation will develop over the next year. They compare these predictions with the realizations one year later and find that those who make optimistic forecast errors have higher debt-to-income ratios. Similarly, Cocco, Gomes and Lopes (2019) find that households that experienced a decline in earnings and subsequently make

optimistic forecast errors are more likely to take out an additional mortgage on their house than those who do not make optimistic forecast errors. Souleles (2004) finds that households underestimate economic shocks and that their too positive sentiment is related to higher consumption. Our findings can help to explain the pattern they find: People with overconfident income expectations will be particularly affected by a negative shocks to the economy or their personal situation, as they already run overly high consumption levels.

Finally, we contribute to the literature on overconfidence in financial decision making. Mostly this literature is concerned with investment decisions (Odean, 1998, 1999; Barber and Odean, 2000). Overplacement relative to other people, which is the type of overconfidence we study, has been related to over-trading and greater risk taking (Glaser and Weber, 2007; Graham, Harvey and Huang, 2009; Merkle, 2017). It has further been shown that overconfident corporate managers use more debt, in particular long-term debt (Ben-David, Graham and Harvey, 2013). For household debt, however, overconfidence research is scarce. In contemporaneous work, Klühs, Koch and Stein (2019) administer a survey on income expectations and indebtedness in Thailand and run as an add-on a simplified version of our design as lab-in-the-field experiment. While their treatment has not shifted expectations sufficiently to study the causal effect of expectations on over-borrowing, they find an association between indebtedness in the experiment and “real life” debt indicators.

2. Survey Evidence

We begin our analysis by examining the link between debt, expectations and overconfidence based on a representative sample of people living in Germany. The German Socio-Economic

Panel (GSOEP) is a representative longitudinal survey of German households aimed at examining their social and economic behavior. The variables of interest for this study are part of the Innovation Sample (GSOEP-IS), a supplementary stream of the GSOEP designed for more innovative research questions. We combine data from the 2016, 2017 and 2018 waves. Our sample comprises a total of 1,085 respondents. For a detailed description of the survey, the participant sample, and the used measures, see Online Appendix A.

2.1. Overconfidence, borrowing behavior and repayment

We use two measures of overconfidence available in the GSOEP-IS data. Participants are asked to compare their performance in two short tasks to a random sample of the German population. The first task asks participants to name as many numbers that are multiples of 9 (or 17) as they can in 20 seconds. The second task asks participants to name as many animals (or insects) as they can in 20 seconds. The questions ask for a judgment of relative performance and the tasks have an easy and a hard version, which is randomly assigned. To measure overconfidence, we compare respondents' belief about their relative ability to their actual position within the sample of participants who answered the same question (for details, see Online Appendix A).

We first link these overconfidence measures to self-reported borrowing behavior elicited in the survey. Overdraft use, expected repayment ability, and current debt payment problems are considered as indicators of potentially problematic borrowing behavior.² Participants state whether or not they currently use the overdraft facility of their checking account.

²The GSOEP-IS includes further measures of debt taking behavior, which we do not examine as their interpretation is more ambiguous (e.g., existence of mortgage debt). We select the measures that seem to be related to unintended debt, over-indebtedness, or repayment ability.

Overdrafts are a readily available but particularly expensive form of debt, for many banks this is the most expensive form of debt that they offer. The average interest rate on an overdraft in Germany is just under 10%³. Overdrafts have also been related to low self-control (Gathergood, 2012; Becker, Jaroszek and Weber, 2017). Continuous overdraft use might also signal that a household is living beyond its means. Overdraft use is conditional on participants stating that they have an account with overdraft facility (N=698).

The GSOEP-IS contains a question that directly aims at the expected ability to repay outstanding debt. Participants are asked for the probability that they will be able to make all scheduled repayments on time. This variable is presumably closest related to the experimental design, as in the experiment income expectations are crucial for the expected ability to repay ones debt. Overconfidence might lead participants to overestimate their ability to repay. As 92% of participants are certain to make all repayments on time in the GSOEP, we define repayment ability as a binary variable. Repayment ability is conditional on participants stating that they have debt outstanding (N=639). Participants also report whether they subjectively experience debt payments (installments and interest) as a problem. This feeling might be associated with over-indebtedness, as in this situation actual and psychological debt burden is particularly high (Keese, 2012). Debt problems is a binary variable conditional on participants stating that they are in debt (N=280).⁴

Table 1 shows the results of linear probability models with overdraft use, beliefs about repayment, and debt problems as dependent variables. The first two columns report results for the propensity to use overdraft depending on either of the two overconfidence measures.

³Stiftung Warentest: www.test.de/Girokonten-Dispozinsen-4586765-0 (accessed 05/29/2023).

⁴Sample sizes differ, as debt problems were elicited in the 2016 wave of the panel. The question also focuses mostly on consumer credit, which only about a third of participants indicate they have.

Both coefficients are positive, indicating that people that are more overconfident are more likely to use the overdraft facility of their checking account. The relationship is significant only for overconfidence in the numerical task, suggesting a higher importance of numeracy for financial decisions. A one-standard deviation increase in overconfidence corresponds to a 0.03 (3%-point) increase in overdraft use. This is a sizeable economic effect given the unconditional mean of 0.12.

In columns (3) and (4), we test for a relation between the expected repayment ability and the overconfidence measures. Coefficients are positive and significant for overconfidence in the numbers task with a similar economic magnitude as in the overdraft regression. However, the coefficient is insignificant and slightly negative for the other overconfidence measure. Finally, columns (5) and (6) of Table 1 display results regarding respondents experiencing debt problems. We do not find evidence that overconfident individuals are more likely to experience their debt payments problematic. While overconfidence positively affects expected repayment ability, it seems not to reduce current debt problems.

Overall, the survey results provide first evidence of a link between overconfidence and indebtedness. People who are more overconfident are more likely to engage in problematic borrowing behavior that may result in high debt levels. Income expectations which are closely related to repayment ability might be a channel through which overconfidence manifests in borrowing behavior. However, this is a conjecture, since we use overconfidence measures that are not directly related to income or debt, but are rather generic.

Therefore, in the experiment introduced in the next section, we manipulate income expectations directly to establish a causal relationship between income expectations, overconfidence and indebtedness.

3. Experimental Design

The purpose of the experiment is to study the effect of income expectations, in particular overconfident income expectations, on debt taking behavior.

The experimental design mimics actual consumption decisions, as participants earn income that they spend on real goods that are handed over to them at the end of the experiment. The possibility of buying on credit gives rise to borrowing decisions, which are the main interest of this study. We implement a between-subjects design with two treatment groups and a random group, which differ in the way income of participants is generated. The two main parts of the experiment are the selection of consumption goods to be bought (“market”) and the generation of income by solving questions in a general knowledge quiz (“income task”).

Figure 1 illustrates the general structure of the experiment. After the instructions, we conduct a test on the comprehension of the experiment. Participants failing this test are still allowed to participate but excluded from the analysis (see below). This is followed by a first elicitation of income expectations. Then the first market stage takes place, with further markets alternating with two rounds of the income task. Importantly, income is earned only after the first consumption decision is made. This introduces uncertainty about income and the opportunity to borrow against future income. The income task differs between the treatments as explained below. Income expectations are elicited a second time after the first income task. After the final market stage, participants complete a questionnaire on demographics and other control variables (including risk preferences and self-control).

The structure of the experiment and details on the different stages are explained to participants in the written instructions they receive before the start of the experiment. They complete a comprehension test on screen at the beginning of the experiment. The instructions and screenshots of the experiment can be found in Online Appendix B.

3.1. Income task and income expectations

Participants can earn income based on their performance in two rounds of a quiz task. Each of the two quizzes consists of ten general knowledge questions taken from a broad range of topics. Questions are asked in a multiple choice format with four answer alternatives (for the full set of questions see Online Appendix C). There is a time limit of five minutes for the completion of each quiz. Participants are randomly assigned to groups of eight participants and are paid depending on their relative rank within their group. Ranks are determined by the number of correct answers with completion time as a tie-breaker. Participants receive a maximum payment of €5 when finishing in first or second place, and the following pairs of ranks receive a reduced payment of €4, €2, and €1, respectively. Consequently, the range of total earnings in the two income tasks is between €2 and €10.

The treatment variation consists in the difficulty of the quizzes. While in one treatment the general knowledge questions are relatively easy (e.g., “*What does the chemical compound H_2O stand for?*”) in the other treatment the questions are considerably harder (e.g., “*What does the chemical compound NH_3 stand for?*”). We will refer to the respective treatments as the easy treatment and the hard treatment. Questions are matched in terms of topics across treatments to exclude any unintended effects of topic familiarity. To construct the quizzes, we have tested the difficulty of the questions in a pretest. Importantly, participants

are shown four sample questions representative for quiz difficulty at the beginning of the experiment.

Our treatment manipulation builds on the reversed hard-easy effect (Kruger, 1999; Burson, Larrick and Klayman, 2006; Moore and Small, 2007), which implies that people overestimate their relative position in easy tasks and underestimate it in hard tasks. In the psychological literature, differences in beliefs between groups solving easy or hard task are large and reliable.⁵ Based on this evidence it seems justified to rely on the hard-easy effect to produce different degrees of overplacement. Besides, several contemporaneous papers in economics have used a similar treatment manipulation (Dargnies, Hakimov and Kübler, 2019; Klühs, Koch and Stein, 2019; Colzani and Santos-Pinto, 2021; Barron and Gravert, 2022; Bruhin, Petros and Santos-Pinto, 2022). Participants in the easy treatment are thus predicted to expect a higher absolute income than participants in the hard treatment. Due to the identical payment scheme, the total actual income does not differ between treatments.

In the random group, income is determined by two independent random lotteries. To match the income distribution in the treatment groups, there is an equal chance to receive a payment of €1, €2, €4, and €5 in each lottery. To avoid any effects of a quicker sequence of markets or a shorter experiment duration, participants in the random group will nevertheless complete two quizzes (either hard or easy). It is made clear to them that the quiz performance is not payoff relevant. Participants in the random group are predicted to have income expectations close to mean lottery payouts.

⁵In Online Appendix D, we analyze results of five of the most cited articles on the hard-easy effect. In all studies, the effect is highly statistically significant ($p < 0.001$), and the average effect size measured by Cohen's d is 0.88.

After the instructions, but before the markets and income tasks, income expectations are elicited. At this stage, participants have been informed about the income tasks and have seen sample questions, but have not yet completed the quizzes. They are asked for their total expected income in € from the two quiz tasks. In addition, we follow Merkle and Weber (2011) and elicit probabilities for reaching each rank pair in a task. After an income task is completed, participants' true rank and payoff are revealed. After the first task, we ask for expected income in € for the remaining task. We do not incentivize the income expectations to avoid strategic behavior in the quiz tasks.⁶

3.2. Market for consumption goods

All three markets have the same structure. A total of ten goods are on display, of which multiple items can be purchased by participants. We select goods based on their assumed desirability for a student population (including chocolate, pens, and soft drinks), and goods remain constant across markets. The current price for each good is shown as well as the future prices in the remaining markets (see Online Appendix B for a screenshot). There is thus no uncertainty about prices or available quantity of goods. Prices for goods increase substantially over time. We sell products at a discount of 50% to retail price in the first market, at about retail price in the second market, and at a premium of 30% in the final market. This price structure is designed to induce borrowing, as usual borrowing motives are absent in the experiment. In particular, earlier purchases will not result in earlier consumption as all goods are handed over at the same time after the experiment.

⁶See the discussion by Schlag, Tremewan and Van der Weele (2015).

The maximum credit for purchases in the first market is €10 which corresponds to the maximum income in the income tasks. As the first market takes place before any income is earned, participants have to borrow against their expected future income. After the first income task, participants' payoff is added to their balance and they can use it in the second market in addition to a maximum credit of €5. After the second income task, participants' payoff is again added to their balance and they can spend any remaining positive balance in the third market. Importantly, the income from the quizzes can only be spent on goods, it will not be redeemed for cash at the end of the experiment. It is thus rational to spend any income from the income task on goods as long as the goods have positive utility. On the market screen, participants see their account balance and their debt level.

It follows from the design that participants can over-borrow. As all participants are allowed to take out a maximum of €10, but only few will actually earn €10, some might not be able to repay their debt. To make over-borrowing costly, these participants have to repay the debt out of their show-up fee. In contrast to the income from the income task, the show-up fee is paid out in cash. We assume and empirically corroborate in the final questionnaire that participants in general prefer cash to the offered goods and would like to avoid digging into their show-up fee (see also sections 4.4 and 5).

The experimental design provides relatively strong incentives to borrow. One may be concerned that subjects completely exhaust their credit independent of income expectations. However, this is neither what theory predicts nor what people do in practice when they observe a price discount. Theoretically, participants should spend their expected income (as they cannot keep this money), but beyond that only purchase further products if their marginal utility exceeds their price. Most people forgo most discounts in their daily grocery

shopping, and very few will exhaust their entire budget to buy as much as possible of an item that is on sale. On the contrary, it is rather difficult to find a small set of products that will appeal to a sufficient number of participants. A little thought experiment might illustrate this: Had we set up a store on campus trying to sell the exact same products at similar price discounts, most students would have ignored our efforts.⁷ We therefore predicted to see very little borrowing activity if we allowed for a cash payout. The role of the product prices is discussed and tested in section 5.

3.3. Control variables

After the main experiment, a questionnaire asks for demographics of participants (including gender, age, and education). We ask whether they would prefer a cash payout instead of goods. We further test for financial literacy using six standard questions similar to those used by van Rooij, Lusardi and Alessie (2011). We measure risk aversion using a self-assessment on a scale between 0 and 10 (Dohmen et al., 2011) and using the staircase method developed by Falk et al. (2016). Finally, we measure self-control using the 13-item scale developed by Tangney, Baumeister and Boone (2004).

3.4. Procedures and participants

After the questionnaire, the experiment ends and participants receive the goods they purchased privately in a separate room. The total show-up fee amounts to €13 and thus sufficiently covers any possible amount of credit.⁸ Participants leaving the experiment without

⁷In fact after the experiment, we tried to sell off leftover products at one of the authors' institution. Even though they were easily accessible at low prices, it took substantial time to get rid off even a portion of the products.

⁸Laboratory rules require a minimum show-up fee of €5 to be paid in cash. The additional €8 correspond to the maximum shortfall from credit (maximum credit – minimum income = €10 – €2 = €8). To make

debt receive the full €13 in addition to the products they purchased. Participants leaving the experiment with debt receive the €13 minus their debt outstanding in addition to the products they purchased.

The experiment has been registered in the RCT registry of the American Economic Association under the identifier AEARCTR-0002634. We filed our main hypotheses in the registration, a description of the three treatment designs, as well as the planned number of participants ($n = 288$). The registration was filed on December 12, 2017, prior to the first experimental session. The experiment was programmed using the experimental software z-tree (Fischbacher, 2007) and was conducted in the experimental laboratory of Technical University Berlin, Germany, in December 2017. Participants were invited using the recruiting software ORSEE (Greiner, 2015).

A total of 285 participants completed the experiment in sessions of 24.⁹ In the main analysis, we exclude 33 participants from the analysis who answer less than four of the five comprehension questions correctly.¹⁰ As the exclusion criterion is based on comprehension of the general instructions, it is random across treatments. Table 2 shows demographic information for the final sample of 252 participants. We obtain an almost equal proportion of female and male participants. Average age of participants is 23 and most of them are studying for a Bachelor’s degree. About a third of the students work and their monthly

this clear to participants, the two parts of the show-up fee are designated as “show-up fee” and “participation fee.”

⁹The easy and hard treatments require a group of eight participants to enable relative comparisons. An equally large random group was targeted. Sessions were slightly overbooked, but due to no-shows in two sessions did not reach the targeted number of participants. In this case, we used a smaller random group.

¹⁰As an experimenter approaches and talks to participants with incorrect responses, we discover that insufficient comprehension of the German language is responsible for the errors in many cases. The laboratory indicates the language of experiments in the invitation (English or German), but this might be overlooked. To avoid disruptions, participants are allowed to regularly continue with the experiment. We provide additional analysis on excluded participants in Online Appendix E.1.

income by this or other means is on average about €700. They show high financial literacy but moderate risk tolerance and self-control.

Table 2 also provides means by treatment group as a balance test of the randomization. Differences between groups are small for most demographic variables. We find higher average income among participants in the easy treatment and also slightly higher risk tolerance. While we believe these differences are due to chance, we nevertheless control for these variables in the regressions.

4. Results

4.1. Income expectations

We first examine whether the treatment variation leads to differences in expected income during the experiment. Participants in the two treatment groups were exposed to different sample questions (hard or easy). Figure 2 shows average expected income at the beginning of the experiment, separately for the two treatments and the random group. The range of possible income in the two quiz tasks is between €2 and €10. Participants in the hard treatment on average expect to earn €5.61, whereas participants in the easy treatment expect to earn €6.67. The difference of more than 1 Euro is highly significant ($p < .001$). Participants in the random group know that they are paid according to the outcome of two lotteries. They expect to earn €6.04, which is very close to the expected value of the lotteries (€6). Unsurprisingly, the variance of income expectations is lowest in this group.

Figure 3 shows income expectations after the first income task and before the second market stage. Participants are asked to provide their expected income for the remaining

income task; the range of possible values is thus reduced to between €1 and €5. They respond after receiving feedback on their income in the first quiz task. The average income expectation of participants in the hard treatment is €2.77, while those in the easy treatment expect to earn €3.16. Participants in the easy treatment still expect higher income, but the difference is smaller and statistically significant at a level of 10% ($p = 0.07$). The difference is slightly smaller than a proportional decrease would suggest, which probably results from updating after receiving feedback. However, the feedback is not sufficient to close the gap between the two treatments. Feedback in the experiment is timely and unambiguous. Overconfidence might be even more persistent in real world settings, where feedback is delayed or performance is affected by chance. The expected income in the random group is again almost exactly in line with the expected value of the lottery.

Overall, we confirm that the manipulation of income expectations is successful. Participants in the easy treatment expect significantly higher income at the beginning of the experiment than participants in the hard treatment. The difference decreases but persists throughout the experiment.

4.2. Actual income and overconfidence

The hard treatment proves to be harder in terms of quiz difficulty, as participants on average answer 5.9 of 20 questions correctly, while in the easy treatment they answer 14.7 questions correctly. In this respect, participants expectations about quiz difficulty induced by the sample questions are confirmed in the income tasks. However, as incentives are based on relative performance, the average actual income does not differ between the treatment groups. It amounts to €6 in all treatment groups. As a consequence, the income expectations in the

easy treatment group are on average too high and in the hard treatment group too low. This has been explained by egocentrism in comparative judgments (Kruger, 1999): people tend to think more about their own performance than about how difficult the task will be for other participants. A simple measure for individual overconfidence is the difference between income expectations and income realizations. As we elicit income expectations twice, we can likewise calculate two overconfidence variables. We label these variables *overconfidence* and *overconfidence 2*.

Average overconfidence in the full sample is 0.03 and not significantly different from zero (overconfidence 2 = -0.13 , see Table 3). Thus, there is no general tendency to overestimate one's income in the experiment. However, there are large treatment differences resulting from the differences in income expectations. Panel A of Table 3 shows the results by treatment group and the difference between the easy treatment and the hard treatment. Participants in the easy treatment are on average overconfident about their income, while participants in the hard treatment are underconfident. The difference amounts to 1.19 and is statistically significant ($p < 0.01$).

After receiving feedback, income expectations become more realistic and overconfidence goes down (Panel B). There still remains a difference between treatment groups, which is no longer statistically significant ($p = 0.29$). While the differences in overconfidence on group level are induced exogenously by the experimental design, we also observe considerable heterogeneity within treatment. This supports the view of overconfidence as an individual trait. However, we do not find strong associations between overconfidence and other observables like age or gender (see Online Appendix Table E.6). We find Pearson correlations between

overconfidence and overconfidence 2 of 0.55. In subsection 4.4, we will further investigate the consequences of overconfidence beyond the treatment effect.

4.3. Consumption and debt taking

We next examine the consumption decisions in the market stage. Participants are active in the markets and spend on average €3.94 in the first market, €1.40 in the second market, and €0.97 in the final market. Only 20% of participants spend less than the minimum income of €2 in round 1. This means that the offered products reasonably appeal to participants and the market rules are understood. The expenditures are highest in the first market, presumably because prices are lowest and participants are still unrestricted by their actual income. Their debt limit corresponds to the maximum possible income.

Importantly, the consumption in the first market is identical to the debt taken out, as participants have not yet earned any income. Panel A of Table 4 shows the debt level of participants after each of the three markets. The average debt level decreases over time as participants repay their debt from the earned income they receive before markets two and three. There is some new borrowing in market two as participants who have not maxed out their credit line can take out additional debt (no new debt is possible in market three). The debt level after market three corresponds to the final debt that participants are unable to repay from their income. We find that 38% of participants have final debt, which is deducted from their show-up fee.¹¹ The table further shows that participants on average do not spend

¹¹The counterpart to leaving the experiment with debt is leaving it with unspent income, which is forfeited after the last market stage ends. However, as the cheapest product in the final market costs €1.20, we consider it only unreasonable if participants leave more than €1.20 on the table (under the assumption that products have positive utility). Such high unspent income is observed for only 3% of participants. As our focus is on debt, we will not discuss this issue further.

their entire expected income. At least part of this can be explained by a cautionary motive, as income is uncertain. Such caution is in line with our premise that participants are unwilling to dig into their show-up fee.

Table 4 shows in Panels B-D the borrowing behavior in the different treatment groups. In line with their higher income expectations, participants borrow most in the easy treatment and least in the hard treatment. The differences decrease over time, as the experimental design allows participants to adjust their spending to their actual income, but remain visible until the end of the experiment. This is a first indication of the treatment effect on borrowing behavior.

We study the effect now more formally in a regression framework. Table 5 shows the results of debt variables regressed on treatment indicators. In column (1), the dependent variable is the initial debt from consumption in market one. Coefficients have a natural interpretation in terms of Euro. Participants in the easy treatment spend 73 cents more than participants in the hard treatment (omitted category), while participants in the random group spend 59 cents more. Given a baseline consumption of €3.50, the treatment effects are economically and statistically significant. The difference between the hard treatment and the random group, however, is only statistically significant at the 10%-level.

Columns (2) and (3) show how the treatment effect evolves throughout the experiment. As already evident from the descriptive statistics, the effect decreases, but it remains at least marginally significant until the end of the experiment. This decline is expected as participants receive feedback about their income, and the erroneous income expectations induced by the treatments are gradually corrected. They can also reduce consumption in later markets. In fact, they can stop consumption altogether, an option that real debtors

usually do not have. Nevertheless, participants in the easy treatment are consistently more indebted than those in the hard treatment. This means that they are not able to make up for their initial overspending completely and remain in debt.

It is worth noting that the difference in initial borrowing is smaller than the difference in income expectations (see Figure 2). Participants with high income expectations spend proportionally less of their expected income. One reason might be risk aversion, as the risk to fall short of their expected income is higher for these participants. Another reason might be decreasing marginal utility of consumption. Interestingly, this finding reverses after market two: the between-treatment differences in debt levels are now larger than the differences in income expectations for the second income task. This suggests that income expectations adjust more quickly than debt levels can, a problem that seems relevant for real life debt as well.

Column (4) of Table 5 reports results of a linear probability model with a binary variable whether participants have debt at the end of the experiment ($\text{final debt} > 0$) as the dependent variable. The results suggest that participants in the easy treatment are about 11% more likely to have debt at the end of the experiment. This difference, however, is not statistically significant. We conclude that the treatment has a significant and persistent causal effect on the borrowing behavior of participants in the experiment. Participants in the easy treatment borrow more initially and have higher debt levels throughout the experiment. We find that the treatment effect operates more strongly on the intensive margin than on the extensive margin of debt taking. These results show the causal effect of our experimental intervention and the coefficients should be interpreted as intention-to-treat (ITT) estimates. As all other factors are held constant in the lab setting, we can attribute the difference in debt levels to

the treatment. As we are ultimately interested in the effect of overconfidence on borrowing, we will use an IV strategy to estimate the local average treatment effect (LATE) in the next section.

4.4. Mechanism

In this subsection, we examine the mechanisms that drive the treatment effect on indebtedness. The main variables of interest are income expectations and overconfidence, as these were intended to be manipulated by the treatment. First, we conduct analyses of their total effect using OLS regressions. Here, we regress the debt level on income expectations and overconfidence, acknowledging that these variables are not exogenous. They in part represent individual differences that create variation also within treatment. While endogenous, it is still interesting to explore, whether the within treatment variation aligns with the treatment effect.

Second, we employ an IV approach, in which we instrument overconfidence with the treatment assignment. Here, we study the causal effect by exploiting only experimentally induced variation in overconfidence.

4.4.1. OLS Regressions

Table 6 shows OLS regressions with debt levels after the first market, second market, and at the end of the experiment as dependent variables. We also examine the propensity to remain in debt at the end of the experiment as additional dependent variable. The dependent variables are regressed on income expectations and overconfidence as defined above. We include a range of control variables to account for observable differences between participants.

Column (1) shows that income expectations are a strong predictor for debt levels after the first market. For each additional expected Euro of income, participants take out 25 cents more debt. While the expectations coefficient is similarly strong after the second market, the impact of expectations on final debt reduces to 11 cents, as can be seen in columns (3) and (5). For overconfidence, we observe the opposite pattern. While the impact of overconfidence on initial debt levels is insignificant (see column (2)), it strongly predicts debt levels and remaining in debt later in the experiment (see columns (4), (6) and (8)). We can use the OLS results for comparison with the GSOEP results as the regression specifications are similar. In particular, “has debt” is also a binary variable as are the outcome variables in the GSOEP. A one-standard deviation increase in overconfidence corresponds to a 0.23 (23%-point) increase in the probability to end the experiment in debt. Unsurprisingly, the effect size in the experiment is much larger than in the survey. One reason is that overconfident income expectations have direct consequences for spending and debt-taking behavior in the experiment, while GSOEP questions elicit generic overconfidence that is inconsequential for income.

The results suggest that individuals with high income expectations initially borrow more, but that for some of them high expectations are backed by actual income. Overconfidence zooms in on those participants who have unrealistically high income expectations. They are the ones who have a high risk to stay in debt and to become over-indebted. Table E.7 in the Online Appendix shows that this is particularly true for participants who remain overconfident regarding the second income task despite receiving feedback (Overconfidence 2).

4.4.2. IV Strategy and Results

We now use an instrumental variable strategy with the random treatment assignment as instrument. We follow previous research that has used instrumental variables created in laboratory environments to study the causal effect of beliefs on actions (e.g., Costa-Gomes, Huck and Weizsäcker, 2014).

The estimation strategy is based on a two-stage least squares estimation, in which we use the treatment indicators as excluded instrument. In the first stage, we estimate:

$$\text{overconfidence}_i = \alpha_0 + \alpha_1 T_i + \alpha_2 X_i + \varepsilon_i \quad (1)$$

where overconfidence_i is the difference between income expectations before round 1 and total payout (as defined before). T_i is the treatment indicator and X_i is a set of control variables. In the second stage, we estimate:

$$\text{debt}_i = \beta_0 + \beta_1 \widehat{\text{overconfidence}}_i + \beta_2 X_i + \varepsilon_i \quad (2)$$

where debt_i are the debt levels after round 1, 2, and at the end of the experiment, respectively, and $\widehat{\text{overconfidence}}_i$ are predicted values of overconfidence from the first stage. Hence, in the IV estimation we only use exogenous variation in overconfidence. The variation in overconfidence is induced in the treatments by shifting income expectation since the average payout does not vary across treatments.

We restrict the analysis to participants in the easy and hard treatment to adhere to the monotonicity assumption of IV.¹² The exclusion restriction requires the assumption that the treatment affects the consumption decision only through income expectations. As the treatment is assigned randomly and everything else is held constant as part of the experiment, we consider it credible that the exclusion restriction holds. Using the language of heterogeneous treatment effects, we can interpret the IV estimates as local average treatment effect (LATE) for the compliers (Imbens and Angrist, 1994; Angrist and Pischke, 2009). Compliers are those participants whose confidence is shifted due to the treatment variation, where more weight is given to participants whose confidence responds the most to the treatment variation.

In Table 7, we show results of IV regressions with debt levels after market 1, market 2, and at the end of the experiment, as well as the propensity to remain in debt, as dependent variables. The table shows the impact of instrumented overconfidence on these variables with and without control variables. Column (2) shows that for each Euro of overestimated income, participants borrow 47 cents more in market 1. After receiving feedback over the course of the experiment, they scale back on borrowing. However, the effect of overconfidence on final debt is still sizeable at 30 cents per Euro of overestimated income and statistically significant (column (6)).

Moreover, we find that overconfidence has a positive effect on the propensity to remain in debt at the end of the experiment. Overconfident income expectations increase the likelihood that participant are unable to repay their debt by 10% (per €) in the IV results. We

¹²The monotonicity assumption requires that the treatment moves income expectations unequivocally upwards or downwards (Angrist and Pischke, 2009). While this assumption is likely fulfilled for the easy compared to the hard treatment, it is less clear for the random group compared to the easy and hard quiz treatments. In Table E.1, we reproduce Table 5 including only the easy and hard treatment (with and without controls). This specification constitutes the reduced-form of the IV specification.

thus observe a positive effect of overconfidence on the intensive and extensive margin of over-indebtedness (the effect on debt levels also holds when restricting to participants with positive debt, see Online Appendix E).

The results have to be viewed in light of a relatively weak first stage with an F-statistic below 10. We follow recommendations of the literature on weak IVs and report Anderson-Rubin (AR) p-values for the null that the overconfidence coefficient equals zero (Andrews, Stock and Sun, 2019).¹³ However, inference based on the Anderson-Rubin p-values leads to qualitatively similar results.

We observe that the IV estimates are larger than the OLS estimates reported in Table 6. Interpreting the IV estimates as local average treatment effects, the results suggest that the nexus between overconfidence and overborrowing is stronger for those individuals whose income expectations respond more to the treatment variation.¹⁴ Figure D.1 in the Online Appendix shows that the easy treatment induces an almost uniform first-order stochastic shift of the cumulative distribution function (CDF) of overconfidence compared to the hard treatment. This suggests that the results are not driven by a complier sample with specific manifestations of overconfidence.

As the IV estimations only use exogenous variation in overconfidence, the results differ from Table 6 that includes within-treatment variation in income expectations. However, in

¹³Moreira (2009) has shown that in the single-instrument case the AR test is uniformly most accurate unbiased. In the single-instrument case, the AR t-statistic coincides with the associated t-statistic of the instrument in the reduced form regression (Angrist and Kolesár, forthcoming). This can be easily verified by comparing the p-value in square brackets with the statistical significance of the reduced-form coefficients in Table E.1.

¹⁴An alternative interpretation for the difference in the IV and the OLS results is that income expectations are measured with error and that this measurement error leads to attenuation bias in the OLS estimates.

both cases we find a statistically significant effect of overconfidence on debt taking at the end of the experiment.

We consider final debt the most relevant outcome variable since we assume that subjects plan to repay their debt over the course of the experiment and that staying in debt at the end of the experiment is unintended (resembling over-indebtedness). Several observations support this assumption: first, we ask people in the post-experimental questionnaire whether they would prefer to receive additional income in cash or in products at prices of the first market. Only 15% respond that they would prefer the products, which suggests that the vast majority intends to consume only the income that is not convertible to cash.¹⁵ Secondly, we rarely see participants spend more than their expected income (19% in the first market). This strongly suggests that most participants do not intend to spend more than they earn in the income tasks. The reported results are robust to the exclusion of either group.

5. Robustness

A concern with the registered experiment is that the artificially low prices may provoke borrowing behavior that otherwise would be absent. First, people with low income expectations may borrow a lot in the first market, simply because products are cheap and they do not mind having their cash reward reduced in exchange for the cheap products. Second, at low prices, participants are able to buy large quantities of the offered products. Participants with high income expectations may therefore reach the point of satiation before they have spent

¹⁵In addition, the lab is an experimental economics lab in which cash incentives are the norm. People, who are responsive to cash incentives, self-select into the subject pool. In Online Appendix E, we show results excluding the group of participants who prefer the products over the cash payout. In section 5, we discuss a robustness test with prices at or above retail prices.

as much as they expect to earn. Both effects would reduce the sensitivity of consumption behavior to income expectations and thus work against our results. However, they should not introduce systematic variation between treatments.

Nevertheless, we conduct a replication of the experiment in which we eliminate the price discounts. In this version, prices in the first market stage correspond to the retail prices of the products. Prices in the second market stage are 40% higher and in the final market stage 80% higher than the retail price. We keep the ascending price profile to provide an incentive for early borrowing. Otherwise the optimal course of action would simply be to wait with consumption until realized income is revealed. The robustness experiment was conducted in July 2018 in the experimental laboratory of the TU Berlin, excluding subjects who had participated in the main experiment. We run ten sessions with a total of 219 participants; descriptive statistics on the participants can be found in Online Appendix F. As in the main experiment, we exclude participants with insufficient understanding of the experimental instructions and remain with a final sample of 193 participants.

Panel A of Table 8 shows income expectations and overconfidence of participants regressed on treatment indicators. We find that the manipulation of income expectations by the treatment works less well in this sample compared to the sample in the main experiment. Participants in the easy treatment expect to earn 57 cents more, but the gap between the easy and the hard treatment is much smaller than in the main experiment and only marginally significant. While standard deviations remain unchanged, Cohen's d decreases from 0.59 to 0.31, which is 53% of the original effect size. Treatment differences in overconfidence do not obtain significance. As in the main experiment, differences are further reduced after feedback on income is obtained.

We next examine whether debt-taking behavior changes in the high price condition. Participants borrow €3.86 on average in the first market, have debt levels of €2.16 in the second market, and €0.72 in the final market. This is very similar compared to the main experiment (all $p > 0.30$), but as prices are considerably higher, participants purchase far fewer products in this condition. Participants again spend less than their expected income, as it is particularly undesirable to pay for the expensive products out of the cash reward. Only 5% state in the high price condition that they would prefer the goods over cash. Nevertheless, 31% remain in debt at the end of the experiment compared to 36% in the main experiment.

Given the small treatment differences in expectations, it is unlikely to find a strong treatment effect on borrowing behavior. Panel B of Table 8 confirms this by showing results for debt levels of participants in analogy to Table 5. We observe positive effects of the easy treatment on indebtedness, but coefficients are smaller than in the main experiment and not statistically significant. The effect size of the treatment effect on debt after market 1 ($d=0.15$) is 50% of the effect size in the main experiment ($d=0.30$). The reduction in effect size is proportional to the reduction in effect size in the income expectation manipulation. This is exactly what one would expect if there was a true link between income expectations and debt taking. We further find a negative effect for the random group. It is possible that participants in the random group do not want to consume at high prices given the risk of the income lottery.

To determine, whether direct effects of income expectations and overconfidence hold despite the weak treatment manipulation, we analyze the impact of expectations and overconfidence on debt levels throughout the experiment (see Table 9). As there is no significant effect of the manipulation on overconfidence, we are unable to use an IV strategy in these

regressions and use the OLS specification instead. We find a similar relationship between income expectations and initial debt taking as in the main experiment (column (1) and (3)). As before, overconfidence becomes more important in the later stages of the experiment, in particular for final debt (columns (4) and (6)), and the likelihood to remain in debt (column (8)). The regression coefficients for overconfidence are larger than in the main experiment and strongly significant. Thus, based on correlational evidence, we can confirm income expectations and overconfidence as predictors of borrowing behavior.

Since our treatment manipulation is not as powerful in the second experiment, we cannot fully replicate the results of the first experiment. We find that effect sizes in contemporaneous economic studies that build on the hard-easy effect are in general less strong than in the original psychology experiments (see Online Appendix D). While this can be partly attributed to publication bias, the file drawer problem, or a regression effect (Camerer, Dreber and Johannesson, 2019; Fiedler and Prager, 2018), other factors such as subtle differences in design may also contribute. We elicit beliefs about ranks and income expectations based on sample questions before participants take the actual quizzes. The timing was intentional as it more accurately reflects expectations of future income, but it might increase uncertainty about performance and lead to a less distinct treatment effect. Recent studies that also elicit beliefs before the task tend to have weaker effect sizes (Klühs, Koch and Stein, 2019; Bruhin, Petros and Santos-Pinto, 2022).

An additional complication arises from translating beliefs about relative performance into income expectations. Relative ranks in the quiz task directly determine income, but the correlation between elicited rank beliefs and income expectations is only 0.7. At the same time, the treatment effect is stronger for rank beliefs than for income expectations. Apparently,

the hard-easy manipulation loses strength when participants have to take an additional cognitive step to arrive at income expectations (even though we test in the comprehension check whether they understand the relation between rank and income).

While the above considerations are relevant for both experiments and do not specifically explain the weaker effect in the robustness experiment, a lower initial effect size and significance will typically also reduce replicability (Camerer, Dreber and Johannesson, 2019; Davis et al., 2023). We perform a post-hoc power analysis using data from the main experiment that reveals 94% power to find a hard-easy effect at 5% significance in the robustness experiment (actual $p=.06$). We interpret the result as a false negative, as the hard-easy effect has been replicated many times and usually produces strong results. The confidence interval for the effect size in the robustness experiment is $[-0.02, 0.64]$, which includes zero but also the effect size in the main experiment ($d=0.59$) and similar recent economic experiments (see Online Appendix Table D.2).

The change in the price structure cannot be responsible for the weaker hard-easy effect, as income expectations are submitted before participants learn about prices in the market stage. The instructions, the income task, and the treatments remained the same. Variation in the subject pool across the semester may contribute to the difference in results given that the robustness experiment was conducted at the very end of the semester (Ebersole et al., 2016), but we do not find evidence for lower attention or motivation.¹⁶ We finally consider a contamination of the subject pool, as unbeknownst to us Dargnies, Hakimov and Kübler

¹⁶The exclusion rate for failing the comprehension test is 12% in both experiments, the quiz performance is very similar, and participants translate rank beliefs into income expectations equally well.

(2019) and Barron and Gravert (2022) use a hard-easy design in the same laboratory and prior to the robustness experiment, although the overlap in participants is unknown.

We conclude that the low realization of the effect size for the hard-easy manipulation in the robustness experiment is most likely a chance result. This in itself is unproblematic, as the aim of our experiment was not to make a claim about the existence of the hard-easy effect. However, it is unfortunate for our ability to derive a treatment effect on borrowing. Stroebe (2019) points out that replications can technically fail if researchers auxiliary hypotheses are not met, which does not invalidate the underlying theory of interest. We thus emphasize the very stable correlational results across both experiments and the similar debt-taking behavior.

By running a robustness experiment, we contribute to the literature that evaluates the replicability of experimental findings. Camerer et al. (2016) find that a considerable fraction of results in experimental economics do not replicate. Internal replication efforts within original studies can be a remedy. They can also pinpoint reasons for limited replication success more clearly than large scale replication efforts (Chen, Chen and Riyanto, 2021). A downside to internal replications could be an author bias to confirm initial findings, as chances for publication are reduced if a replication is unsuccessful (Christensen and Miguel, 2018). In our case, a failure to replicate the main effect, would have undermined the relevance of overconfident income expectations for debt taking. The failure to replicate the belief manipulation is less severe, but still does not deliver the intended robustness. On the other hand, external replications battle the opposite problem of an “overturn bias,” as replications that confirm initial results are less publishable (Galiani, Gertler and Romero, 2017).

6. Conclusion

Household indebtedness is a problem in many countries and the fraction of households that have to be considered over-indebted rises. Economic shocks and structural changes in lending technology (Livshits, Mac Gee and Tertilt, 2016) as classic explanations for over-indebtedness can only partly explain the rising trend, as the overall economic situation has been favorable in the 2010s. Little is known about behavioral biases that may play a role in whether and how much debt people take.

In this paper, we examine a particular behavioral bias, overconfident income expectations, and analyze how it influences borrowing behavior. We first provide evidence for this using representative household survey data from Germany. In the survey, overconfidence is measured in domains which are independent from income or debt levels, but still show positive correlations with overdraft use, subjective debt burden, and expected repayment ability.

In a laboratory experiment with real consumption, we exogenously manipulate income expectations of participants. In two treatments, we induce either overconfident or underconfident income expectations for income earned in two quiz tasks. We find that participants with higher income expectations consume more, take out more debt and keep higher debt levels throughout the experiment. In particular, overconfident participants run the risk to stay in debt at the end of the experiment. Based on IV regressions, we claim that these links are causal.

Overconfident income expectations might interact with other behavioral issues in debt taking such as limited self-control and exponential growth bias. It might serve as an early

warning indicator, as overconfident expectations manifest itself long before indebtedness builds up. The findings are important for financial advice and debt counselling, which may help consumers to obtain a more realistic picture of their debt and repayment capacity.

References

- Alan, Sule, Mehmet Cemalcilar, Dean Karlan, and Jonathan Zinman. 2018. “Unshrouding: Evidence from Bank Overdrafts in Turkey.” *The Journal of Finance*, 73(2): 481–522.
- Amar, Moty, Dan Ariely, Shahar Ayal, Cynthia E. Cryder, and Scott I. Rick. 2011. “Winning the Battle but Losing the War: The Psychology of Debt Management.” *Journal of Marketing Research*, 48(SPL): S38–S50.
- Andrews, Isaiah, James Stock, and Lu Sun. 2019. “Weak Instruments in IV Regression: Theory and Practice.” *Annual Review of Economics*, 11: 727–753.
- Angrist, Joshua, and Michal Kolesár. forthcoming. “One instrument to rule them all: The bias and coverage of just-ID IV.” *Journal of Econometrics*.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, New Jersey: Princeton University Press.
- Barber, Brad M., and Terrance Odean. 2000. “Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors.” *The Journal of Finance*, 55(2): 773–806.
- Barron, Kai, and Christina Gravert. 2022. “Confidence and career choices: an experiment.” *The Scandinavian Journal of Economics*, 124(1): 35–68.
- Becker, David, Lena Jaroszek, and Martin Weber. 2017. “Borrower Heterogeneity and the (Ir)Rational Demand for Short-Term Credit.” Working Paper.
- Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. 2013. “Managerial Miscalibration.” *The Quarterly Journal of Economics*, 128(4): 1547–1584.
- Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. 2018. “Behavioral Household Finance.” In *Handbook of Behavioral Economics: Foundations and Applications 1*, ed. B. Douglas Bernheim, Stefano DellaVigna and David Laibson, 177–276. Amsterdam: Elsevier.
- Bruhin, Adrian, Fidel Petros, and Luis Santos-Pinto. 2022. “The Role of Self-Confidence in Teamwork: Experimental Evidence.” Working Paper.
- Brunnermeier, Markus K., and Jonathan A. Parker. 2005. “Optimal Expectations.” *American Economic Review*, 95(4): 1092–1118.
- Burson, Katherine A., Richard P. Larrick, and Joshua Klayman. 2006. “Skilled or Unskilled, but Still Unaware of it: How Perceptions of Difficulty Drive Miscalibration in Relative Comparisons.” *Journal of Personality and Social Psychology*, 90(1): 60–77.
- Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Jol van der Weele. 2020. “Overconfidence and Gender Gaps in Redistributive Preferences: Cross-Country Experimental Evidence.” *Journal of Economic Behavior & Organization*, 178: 267–286.
- Camerer, Colin F., Anna Dreber, and Magnus Johannesson. 2019. “Replication and other practices for improving scientific quality in experimental economics.” *Handbook of Research Methods and Applications in Experimental Economics*, Chapter 5, 83–102. Cheltenham, UK:Edward Elgar Publishing.

- Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu. 2016. “Evaluating Replicability of Laboratory Experiments in Economics.” *Science*, 351(6280): 1433–1436.
- Chen, Roy, Yan Chen, and Yohanes E Riyanto. 2021. “Best Practices in Replication: A Case Study of Common Information in Coordination Games.” *Experimental Economics*, 24: 2–30.
- Christensen, Garret, and Edward Miguel. 2018. “Transparency, Reproducibility, and the Credibility of Economics Research.” *Journal of Economic Literature*, 56(3): 920–980.
- Cocco, João F., Francisco Gomes, and Paula Lopes. 2019. “Evidence on Expectations of Household Fiances.” Working Paper.
- Colzani, Paola, and Luis Santos-Pinto. 2021. “Does Overconfidence Lead to Bargaining Failures?” Working Paper.
- Costa-Gomes, Miguel A., Steffen Huck, and Georg Weizsäcker. 2014. “Beliefs and Actions in the Trust Game: Creating Instrumental Variables to Estimate the Causal Effect.” *Games and Economic Behavior*, 88: 298–309.
- Creditreform. 2018. “SchuldnerAtlas Deutschland: Überschuldung von Verbrauchern – Jahr 2018.” published online: www.creditreform.de/nc/aktuelles/news-list/details/news-detail/schuldneratlas-deutschland-2018.html (retrieved 01/30/2019).
- Dargnies, Marie-Pierre, Rustamdjan Hakimov, and Dorothea Kübler. 2019. “Self-Confidence and Unraveling in Matching Markets.” *Management Science*, 65(12): 5603–5618.
- Davis, Andrew M, Blair Flicker, Kyle B Hyndman, Elena Katok, Samantha Keppler, Stephen Leider, Xiaoyang Long, and Jordan Tong. 2023. “A Replication Study of Operations Management Experiments in Management Science.” *Management Science*, forthcoming.
- Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jrgen Schupp, and Gert G. Wagner. 2011. “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences.” *Journal of the European Economic Association*, 9(3): 522–550.
- Ebersole, Charles R., Olivia E. Atherton, Aimee L. Belanger, Hayley M. Skulborstad, Jill M. Allen, Jonathan B. Banks, Erica Baranski, Michael J. Bernstein, Diane B.V. Bonfiglio, Leanne Boucher, Elizabeth R. Brown, Nancy I. Budiman, Athena H. Cairo, Colin A. Capaldi, Christopher R. Chartier, Joanne M. Chung, David C. Cicero, Jennifer A. Coleman, John G. Conway, William E. Davis, Thierry Devos, Melody M. Fletcher, Komi German, Jon E. Grahe, Anthony D. Hermann, Joshua A. Hicks, Nathan Honeycutt, Brandon Humphrey, Matthew Janus, David J. Johnson, Jennifer A. Joy-Gaba, Hannah Juzeler, Ashley Keres, Diana Kinney, Jacqueline Kirshenbaum, Richard A. Klein, Richard E. Lucas, Christopher J.N. Lustgraaf, Daniel Martin, Madhavi Menon, Mitchell Metzger, Jaclyn M. Moloney, Patrick J. Morse, Radmila Prislín, Timothy Razza, Daniel E. Re, Nicholas O. Rule, Donald F. Sacco, Kyle Sauerberger, Emily Shrider, Megan Shultz, Courtney Siensen, Karin Sobocko, R. Weylin Sternglanz, Amy Summerville, Konstantin O. Tskhay, Zack van Allen, Leigh Ann Vaughn, Ryan J. Walker, Ashley Weinberg, John Paul Wilson, James H. Wirth, Jessica Wortman, and Brian A. Nosek. 2016. “Many Labs

- 3: Evaluating participant pool quality across the academic semester via replication.” *Journal of Experimental Social Psychology*, 67: 68–82. Special Issue: Confirmatory.
- Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2016. “The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences.” IZA Discussion Paper Series No. 9674.
- Fiedler, Klaus, and Johannes Prager. 2018. “The Regression Trap and Other Pitfalls of Replication Science Illustrated by the Report of the Open Science Collaboration.” *Basic and Applied Social Psychology*, 40(3): 115–124.
- Fischbacher, Urs. 2007. “z-tree - Zurich Toolbox for Readymade Economic Experiments.” *Experimental Economics*, 10: 171–178.
- Fondeville, Nicole, Erhan Özdemir, and Terry Ward. 2010. “Over-Indebtedness: New Evidence from the EU-SILC Special Module.” European Commission Research Note 4/2010.
- Friedman, Milton. 1957. *Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.
- Gabaix, Xavier, and David Laibson. 2006. “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets.” *The Quarterly Journal of Economics*, 121(2): 505–540.
- Galiani, Sebastian, Paul Gertler, and Mauricio Romero. 2017. “Incentives for Replication in Economics.” National Bureau of Economic Research Working Paper 23576.
- Gathergood, John. 2012. “Self-control, Financial Literacy and Consumer Over-Indebtedness.” *Journal of Economic Psychology*, 33(3): 590–602.
- Gathergood, John, Neale Mahoney, Neil Stewart, and Jörg Weber. 2019. “How Do Individuals Repay Their Debt? The Balance-Matching Heuristic.” *American Economic Review*, 109(3): 844–75.
- Glaser, Markus, and Martin Weber. 2007. “Overconfidence and Trading Volume.” *The GENEVA Risk and Insurance Review*, 32(1): 1–36.
- Graham, John R., Campbell R. Harvey, and Hai Huang. 2009. “Investor Competence, Trading Frequency, and Home Bias.” *Management Science*, 55(7): 1094–1106.
- Greiner, Ben. 2015. “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE.” *Journal of the Economic Science Association*, 1(1): 114–125.
- Heidhues, Paul, and Botond Köszegi. 2010. “Exploiting Naviets about Self-Control in the Credit Market.” *American Economic Review*, 100(5): 2279–2303.
- Hyttinen, Ari, and Hanna Putkuri. 2018. “Household Optimism and Overborrowing.” *Journal of Money, Credit and Banking*, 50(1): 55–76.
- Imbens, Guido, and Joshua Angrist. 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica*, 62: 467–476.
- International Monetary Fund (IMF). 2012. “Dealing with Household Debt.” In *World Economic Outlook: Growth Resuming, Dangers Remain*, Chapter 3, 89–124. IMF.

- International Monetary Fund (IMF). 2017. "Household Debt and Financial Stability." In *Global Financial Stability Report: Is Growth at Risk?* Chapter 2, 53–89. IMF.
- Keese, Matthias. 2012. "Who Feels Constrained by High Debt Burdens? Subjective vs. Objective Measures of Household Debt." *Journal of Economic Psychology*, 33(1): 125–141.
- Klühs, Theres, Melanie Koch, and Wiebke Stein. 2019. "Don't Expect too Much: The Effect of Biased Expectations on (Over)-Indebtedness." Working Paper.
- Kruger, Justin. 1999. "Lake Wobegon Be Gone! The "Below Average Effect" and the Egocentric Nature of Comparative Ability Judgements." *Journal of Personality and Social Psychology*, 77(2): 221–232.
- Lakens, Daniel. 2013. "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs." *Frontiers in Psychology*, 4.
- Livshits, Igor, James C. Mac Gee, and Michle Tertilt. 2016. "The Democratization of Credit and the Rise in Consumer Bankruptcies." *The Review of Economic Studies*, 83(4): 1673–1710.
- Lusardi, Annamaria, and Peter Tufano. 2015. "Debt Literacy, Financial Experiences, and Overindebtedness." *Journal of Pension Economics and Finance*, 14(4): 332–368.
- Melzer, Brian T. 2011. "The Real Costs of Credit Access: Evidence from the Payday Lending Market." *The Quarterly Journal of Economics*, 126(1): 517–555.
- Merkle, Christoph. 2017. "Financial Overconfidence over Time: Foresight, Hindsight, and Insight of Investors." *Journal of Banking and Finance*, 84: 68–87.
- Merkle, Christoph, and Martin Weber. 2011. "True Overconfidence: The Inability of Rational Information Processing to Account for Apparent overconfidence." *Organizational Behavior and Human Decision Processes*, 116(2): 262–271.
- Money Advice Service. 2017. "Levels of Over-Indebtedness in the UK: Technical Report." published online: www.moneyadviceservice.org.uk/en/corporate/a-picture-of-over-indebtedness-in-the-uk (retrieved 01/30/2019).
- Moore, Don A., and Deborah A. Small. 2007. "Error and Bias in Comparative Judgment: on Being Both better and Worse than We Think We Are." *Journal of Personality and Social Psychology*, 92(6): 972–989.
- Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review*, 115(2): 502–517.
- Moore, Don A., and Tai Gyu Kim. 2003. "Myopic social prediction and the solo comparison effect." *Journal of Personality and Social Psychology*, 85(6): 1121–1135.
- Moreira, Marcelo J. 2009. "Tests with correct size when instruments can be arbitrarily weak." *Journal of Econometrics*, 152(2): 131–140.
- Murtin, Fabrice, and Marco Mira d'Ercole. 2015. "Household Wealth Inequality Across OECD Countries: New OECD Evidence." OECD Statistics in brief, June 2015, No. 21.

- Odean, Terrance. 1998. "Are Investors Reluctant to Realize their Losses?" *The Journal of Finance*, 53(5): 1775–1789.
- Odean, Terrance. 1999. "Do Investors Trade Too Much?" *American Economic Review*, 89(5): 1279–1298.
- OECD. 2018. "Wealth: OECD.Stat." data available online: stats.oecd.org/index.aspx?datasetcode=Wealth (retrieved 01/30/2019).
- Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. 2017. "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *The Economic Journal*, 127(604): 2153–2186.
- Schlag, Karl H., James Tremewan, and Joël J. Van der Weele. 2015. "A Penny for Your Thoughts: A Survey of Methods for Eliciting Beliefs." *Experimental Economics*, 18(3): 457–490.
- Smith, Herbert L., and Brian Powell. 1990. "Great Expectations: Variations in Income Expectations Among College Seniors." *Sociology of Education*, 63(3): 194–207.
- Souleles, Nicholas S. 2004. "Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys." *Journal of Money, Credit and Banking*, 36(1): 39–72.
- Stango, Victor, and Jonathan Zinman. 2009. "Exponential Growth Bias and Household Finance." *The Journal of Finance*, 64(6): 2807–2849.
- Statistisches Bundesamt. 2018. "Statistik zur Überschuldung privater Personen 2017." Fachserie 15 Reihe 5.
- Stroebe, Wolfgang. 2019. "What Can We Learn from Many Labs Replications?" *Basic and Applied Social Psychology*, 41(2): 91–103.
- Sufi, Amir, Atif Mian, and Emil Verner. 2017. "Household Debt and Business Cycles Worldwide." *The Quarterly Journal of Economics*, 132(4): 1755–1817.
- Tangney, June P., Roy F. Baumeister, and Angie L. Boone. 2004. "High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success." *Journal of Personality*, 72(2): 271–324.
- van Rooij, Marten, Annamaria Lusardi, and Rob Alessie. 2011. "Financial Literacy and Stock Market Participation." *Journal of Financial Economics*, 101(2): 449–472.
- Zinman, Jonathan. 2015. "Household Debt: Facts, Puzzles, Theories, and Policies." *Annual Review of Economics*, 7(1): 251–276.

TABLE 1

Overconfidence and borrowing behavior in the GSOEP data

	Overdraft use		Belief to repay		Debt problems	
	(1)	(2)	(3)	(4)	(5)	(6)
Overconfidence (numbers)	0.089** (0.045)		0.089** (0.039)		-0.002 (0.056)	
Overconfidence (animals)		0.058 (0.037)		-0.021 (0.032)		-0.058 (0.046)
Gender	0.022 (0.026)	0.024 (0.025)	0.004 (0.021)	-0.002 (0.021)	-0.027 (0.036)	-0.033 (0.035)
Age	-0.001 (0.001)	-0.001* (0.001)	0.001** (0.001)	0.001** (0.001)	-0.003** (0.002)	-0.003** (0.002)
Education	0.001 (0.008)	0.001 (0.008)	-0.013 (0.008)	-0.010 (0.007)	0.008 (0.012)	0.010 (0.012)
Works	0.020 (0.031)	0.025 (0.031)	-0.045** (0.023)	-0.045** (0.022)	-0.110** (0.048)	-0.118** (0.048)
Log of income	0.006 (0.025)	0.002 (0.025)	0.042* (0.024)	0.036 (0.024)	-0.016 (0.046)	-0.014 (0.046)
Financial literacy	-0.012 (0.010)	-0.013 (0.010)	0.000 (0.008)	0.000 (0.008)	-0.051*** (0.017)	-0.051*** (0.017)
Risk tolerance (self assessed)	0.010 (0.006)	0.010 (0.006)	-0.000 (0.005)	-0.001 (0.005)	0.001 (0.007)	0.002 (0.008)
Risk tolerance (choices)	0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	-0.000 (0.002)	0.003 (0.002)	0.003 (0.002)
Self-control	-0.012 (0.017)	-0.012 (0.017)	0.014 (0.014)	0.014 (0.014)	0.030 (0.024)	0.030 (0.024)
Constant	0.131 (0.218)	0.166 (0.221)	0.544*** (0.193)	0.613*** (0.189)	0.668* (0.398)	0.668* (0.394)
Unconditional mean	0.12	0.12	0.92	0.92	0.10	0.10
R ²	0.022	0.019	0.035	0.026	0.099	0.104
Observations	698	698	639	639	280	280

Notes: The table shows results of linear probability models (OLS) with overdraft use, belief in ones ability to repay, and debt problems as dependent variables. Overdraft use is an indicator variable whether a household currently uses the overdraft facility of their checking account, repayment ability is an indicator whether or not a household is certain to repay its debt, and debt problems is an indicator whether a household experiences problems with debt payments. Overconfidence (numbers) and overconfidence (animals) are the difference between the believed relative performance and the actual performance in the respective domain. Control variables are defined in the Online Appendix, Table A.1. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

TABLE 2
Demographic and control variables

	<i>n</i>	Mean	Mean by treatment			F-test
			Hard	Easy	Control	<i>p</i> -value
Gender (female=1)	252	0.49	0.49	0.45	0.50	0.66
Age	252	22.64	22.17	23.20	23.05	0.02
Bachelor degree	252	0.18	0.18	0.23	0.14	0.32
Masters degree	252	0.06	0.02	0.06	0.10	0.12
Works	252	0.31	0.29	0.32	0.34	0.77
Income	248	701	635	834	627	0.01
Financial literacy	252	5.26	5.10	5.42	5.29	0.37
Risk tolerance (choices)	248	10.1	9.64	11.11	9.35	0.05
Risk tolerance (self-assessed)	252	4.75	4.55	5.22	4.45	0.04
Self-control	252	−0.03	−0.18	0.04	0.04	0.30

Notes: The table shows means of demographic variables and controls for the full sample of participants and by treatment group. Gender is an indicator variable taking a value of one if female. Age is reported in years. Bachelor degree and Masters degree are indicator variables taking a value of one if a participant completed the respective degree. Works is an indicator variable whether a participant works. Income is the monthly income in € irrespective of source. Financial literacy is the number of correct answers in a financial literacy test (six questions taken from the German SOEP-IS 2016 survey). Risk tolerance (choices) is based on lottery choices using the staircase method of Falk et al. (2016) with values from 1 (least risk tolerant) to 32 (most risk tolerant). Self-assessed risk tolerance is measured on a scale from 0=“completely unwilling to take risk” to 10=“very willing to take risk.” Self control is a factor score based on Tangney, Baumeister and Boone (2004) with values from −2.90 (lowest self-control) to 2.44 (highest self-control). Different numbers of observation reflect non-responses. The *p*-values of one-way ANOVA F-tests for between group differences are reported.

TABLE 3
Overconfident income expectations

Panel A: Overconfidence	<i>n</i>	Mean	Std.dev.	5p	95p	<i>p</i> -value
All participants	252	0.03	3.05	−5	5	0.88
Easy treatment	88	0.67	3.02	−4	6	0.04
Hard treatment	84	−0.52	2.74	−5	3	0.08
Random group	80	−0.10	3.27	−5	5	0.78
Easy − hard		1.19				<0.01
Panel B: Overconfidence 2	<i>n</i>	Mean	Std.dev.	5p	95p	<i>p</i> -value
All participants	218	−0.13	1.95	−3	3	0.33
Easy treatment	76	0.05	1.79	−3	4	0.80
Hard treatment	73	−0.26	1.82	−3	3	0.23
Random group	69	−0.18	2.24	−4	3	0.49
Easy − hard		0.31				0.29

Notes: The table shows summary statistics of overconfidence variables for the full sample of participants and by treatment group. Overconfidence is the difference between the income expectations at the beginning of the experiment and the actual income from both income tasks (Panel A). Overconfidence 2 is the difference between the income expectations before market stage 2 and the actual income from the second income task (Panel B). The table reports the number of observations, the mean, standard deviation, the 5th percentile, and the 95th percentile for both overconfidence variables. Differences in number of observations are due to the exclusion restriction and non-responses. Easy − hard is the difference between overconfidence in the easy treatment and overconfidence in the hard treatment. The *p*-values of a two-sided *t*-test are reported, testing for a zero mean or a zero between-group difference, respectively.

TABLE 4
Borrowing behavior

Panel A: All participants	<i>n</i>	Mean	Std.dev.	5p	95p
Debt after market 1	252	3.94	2.30	0	8.53
Debt after market 2	252	2.44	1.97	0	5.99
Debt after market 3	252	0.67	1.30	0	3.96
New debt market 2	252	1.39	1.62	0	4.51
Unspent income after market 3	252	0.41	0.93	0	1.12
Expected income – debt	252	2.17	2.60	−1.98	6.44
Panel B: Easy treatment	<i>n</i>	Mean	Std.dev.	5p	95p
Debt after market 1	88	4.23	2.42	0	8.54
Debt after market 2	88	2.72	2.00	0	5.99
Debt after market 3	88	0.85	1.41	0	3.98
Panel C: Hard treatment	<i>n</i>	Mean	Std.dev.	5p	95p
Debt after market 1	84	3.50	2.38	0	7.93
Debt after market 2	84	2.12	1.90	0	5.78
Debt after market 3	84	0.49	1.09	0	2.97
Panel D: Random group	<i>n</i>	Mean	Std.dev.	5p	95p
Debt after market 1	80	4.10	2.03	0.93	8.09
Debt after market 2	80	2.44	2.00	0	6.40
Debt after market 3	80	0.65	1.37	0	4.36

Notes: The table shows in Panel A summary statistics for the debt level of participants after each of the three market stages. Debt after market 1 is the amount in Euro that participants spend of their maximum initial credit of €10. Debt after market 2 is the debt amount in Euro after the first income is added and the second round of spending subtracted (in case of a positive account balance, debt is zero). Debt after market 3 is the debt amount in Euro after the second income is added and the final round of spending subtracted. This is also referred to as “final debt” (in case of a positive account balance, debt is zero). Panel A further shows new debt taken out in the second market, which is additional credit taken out that was not used so far. Unspent income after market 3 is any positive account balance after the final round of spending (in case of debt, unspent income is zero). Expected income – debt is the difference between income expectations and debt after market 1. Panels B-D show the debt levels separately for the two treatment groups and the random group. The table reports the number of observations, the mean, standard deviation, the 5th percentile, and the 95th percentile.

TABLE 5
Debt taking and treatment effect

	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	0.728** (0.366)	0.600** (0.297)	0.361* (0.192)	0.111 (0.073)
Random group	0.592* (0.345)	0.313 (0.305)	0.169 (0.194)	0.028 (0.074)
Constant	3.501*** (0.259)	2.128*** (0.208)	0.489*** (0.119)	0.310*** (0.051)
R ²	0.019	0.016	0.013	0.010
Observations	252	252	252	252

Notes: The table shows results of regressions of debt taking variables on treatment indicators. Columns (1)-(3) show results of OLS regressions with the debt level after each market stage as the dependent variable. Debt after market 1 is the amount in Euro that participants spend of their maximum initial credit of €10. Debt after market 2 is the debt amount in Euro after the first income is added and the second round of spending subtracted. Final debt is the debt amount in Euro after the second income is added and the final round of spending subtracted. Column (4) shows a linear probability model with a binary variable whether a participant has debt at the end of the experiment (final debt > 0) as the dependent variable. Independent variables are treatment indicators with the hard treatment as the omitted category. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

TABLE 6
Indebtedness over the course of the experiment (OLS)

	Debt after market 1		Debt after market 2		Final debt		Has debt	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expectations	0.246*** (0.094)		0.279*** (0.082)		0.114** (0.052)		0.029 (0.019)	
Overconfidence		0.078 (0.048)		0.177*** (0.040)		0.187*** (0.029)		0.076*** (0.009)
Female	-0.848*** (0.317)	-0.949*** (0.316)	-0.408 (0.250)	-0.523** (0.246)	-0.301* (0.179)	-0.348** (0.160)	-0.090 (0.069)	-0.102* (0.059)
Age	-0.012 (0.049)	-0.005 (0.048)	0.014 (0.043)	0.018 (0.041)	0.008 (0.026)	0.004 (0.024)	-0.005 (0.010)	-0.008 (0.009)
Bachelor	0.012 (0.388)	0.058 (0.388)	-0.300 (0.311)	-0.233 (0.306)	-0.179 (0.203)	-0.133 (0.185)	-0.049 (0.088)	-0.033 (0.078)
Masters	0.727 (0.993)	0.674 (1.000)	0.648 (0.764)	0.621 (0.738)	0.504 (0.479)	0.537 (0.436)	0.214 (0.176)	0.233 (0.159)
Works	0.058 (0.332)	0.149 (0.345)	-0.088 (0.272)	-0.019 (0.272)	0.074 (0.185)	0.057 (0.163)	-0.024 (0.072)	-0.040 (0.063)
Log of income	-0.076 (0.229)	-0.061 (0.239)	-0.215 (0.187)	-0.204 (0.196)	-0.152 (0.132)	-0.155 (0.120)	0.000 (0.045)	-0.002 (0.041)
Financial literacy	0.156 (0.115)	0.149 (0.117)	0.225** (0.087)	0.249*** (0.089)	0.045 (0.064)	0.096 (0.064)	-0.001 (0.023)	0.022 (0.021)
Risk tolerance (choices)	0.011 (0.031)	0.023 (0.031)	0.013 (0.029)	0.022 (0.027)	0.033 (0.025)	0.032 (0.021)	0.001 (0.007)	-0.000 (0.006)
Risk tolerance (self-assessed)	0.079 (0.079)	0.108 (0.078)	0.100 (0.061)	0.129** (0.061)	-0.034 (0.039)	-0.028 (0.035)	-0.000 (0.016)	-0.000 (0.014)
Self-control	0.116 (0.144)	0.103 (0.144)	-0.005 (0.114)	-0.010 (0.113)	-0.077 (0.068)	-0.067 (0.061)	-0.016 (0.029)	-0.010 (0.027)
Constant	2.239 (1.638)	3.294** (1.667)	0.199 (1.304)	1.426 (1.333)	0.442 (0.769)	0.983 (0.768)	0.337 (0.332)	0.485 (0.302)
R ²	0.13	0.11	0.16	0.18	0.09	0.26	0.04	0.24
Observations	241	241	241	241	241	241	241	241

Notes: The table shows results of OLS regressions of the debt level after market 1, 2, and at the end of the experiment on income expectations and overconfidence. Debt is the amount in Euro that participants spend of their maximum initial credit of € 10. Income expectations is the expected income in Euro from both income tasks. Overconfidence is the difference between the expected income and the actual income. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

TABLE 7
Indebtedness over the course of the experiment (IV)

Panel A: Second stage								
	Debt after market 1		Debt after market 2		Final debt		Has debt	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Overconfidence	0.609*	0.470	0.502*	0.373	0.303*	0.297**	0.093	0.101*
	(0.366)	(0.323)	(0.282)	(0.241)	(0.157)	(0.141)	(0.057)	(0.055)
Female		-1.051**		-0.480		-0.232		-0.084
		(0.444)		(0.320)		(0.206)		(0.073)
Age		-0.020		-0.032		-0.020		-0.016
		(0.077)		(0.051)		(0.031)		(0.013)
Bachelor		0.241		-0.024		-0.024		-0.002
		(0.633)		(0.427)		(0.244)		(0.103)
Masters		0.985		1.142		0.394		0.256
		(1.691)		(0.725)		(0.449)		(0.242)
Works		0.646		0.488		0.262		0.028
		(0.520)		(0.363)		(0.208)		(0.080)
Log of income		-0.296		-0.387*		-0.235*		-0.061
		(0.284)		(0.225)		(0.134)		(0.051)
Financial literacy		0.293		0.332**		0.112		0.032
		(0.227)		(0.152)		(0.103)		(0.037)
Risk tolerance (choices)		0.016		0.014		0.045**		-0.001
		(0.038)		(0.027)		(0.020)		(0.006)
Risk tolerance (self-assessed)		0.022		0.088		-0.033		0.002
		(0.119)		(0.086)		(0.044)		(0.018)
Self control		0.119		0.071		0.034		-0.004
		(0.200)		(0.141)		(0.072)		(0.034)
Constant	3.821***	4.598**	2.391***	3.292**	0.647***	1.677*	0.358***	0.954**
	(0.218)	(2.345)	(0.168)	(1.638)	(0.093)	(0.899)	(0.035)	(0.380)
AR p-value	0.048	0.102	0.045	0.113	0.060	0.049	0.131	0.106
Observations	172	165	172	165	172	165	172	165
Panel B: First stage								
Easy treatment	1.194***	1.247***	1.194***	1.247***	1.194***	1.247***	1.194***	1.247***
	(0.440)	(0.469)	(0.440)	(0.469)	(0.440)	(0.469)	(0.440)	(0.469)
Control variables	no	yes	no	yes	no	yes	no	yes
1st stage F	7.37	7.08	7.37	7.08	7.37	7.08	7.37	7.08
Observations	172	165	172	165	172	165	172	165

Notes: The table shows results of IV regressions of the debt level after market 1, 2, and at the end of the experiment on overconfidence. Debt is the amount in Euro that participants use of their maximum initial credit of €10. Overconfidence is the difference between the expected income and the actual income. Panel A shows results for the second stage of the IV-regressions. Panel B shows results for the first stage. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$. Robust standard errors are shown in parentheses. AR p-values indicate Anderson-Rubin p-values shown at the bottom of Panel A. 1st stage F is the F-statistic of the first stage.

TABLE 8
Robustness test of treatment effects

Panel A: Expectations	Income Expectations (1)	Income Expectations 2 (2)	Overconfidence (3)	Overconfidence 2 (4)
Easy treatment	0.569* (0.305)	0.167 (0.219)	0.113 (0.083)	−0.066 (0.081)
Random group	0.064 (0.309)	−0.222 (0.220)	−0.054 (0.089)	−0.007 (0.091)
Constant	5.732*** (0.237)	2.915*** (0.149)	0.380*** (0.058)	0.394*** (0.058)
R ²	0.02	0.01	0.02	0.00
Observations	193	193	193	193
Panel B: Debt taking	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has Debt (4)
Easy treatment	0.367 (0.406)	0.473 (0.373)	0.046 (0.289)	0.019 (0.079)
Random group	−0.482 (0.400)	−0.467 (0.348)	−0.464** (0.235)	−0.079 (0.083)
Constant	3.842*** (0.283)	2.102*** (0.275)	0.818*** (0.201)	0.324*** (0.056)
R ²	0.02	0.03	0.02	0.01
Observations	193	193	193	193

Notes: The table shows in Panel A, OLS regression results for income expectations and overconfidence. Income expectations are the income expectations in Euro at the beginning of the experiment for both income tasks, and income expectations 2 are the income expectations in Euro for the second income task. Overconfidence is the difference between the income expectations at the beginning of the experiment and the actual income from both income tasks, and overconfidence 2 is the difference between income expectations 2 and the actual income from the second income task. Independent variables are treatment dummies with the hard treatment as the omitted category. The table shows in Panel B results of regressions of debt variables on treatment dummies. Columns (1)-(3) show results of OLS regressions with the debt level after each market as the dependent variable. Debt after market 1 is the amount in Euro that participants spend of their maximum initial credit of €10. Debt after market 2 is the debt amount in Euro after the first income is added and the second round of spending subtracted. Final debt is the debt amount in Euro after the second income is added and the final round of spending subtracted. Column (4) shows marginal effects of a probit regression with a binary variable whether a participant has debt at the end of the experiment (final debt > 0) as the dependent variable. Independent variables are treatment dummies with the hard treatment as the omitted category. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

TABLE 9
Robustness test of overconfidence and borrowing behavior

	Debt after market 1		Debt after market 2		Final debt		Has debt	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expectations	0.258** (0.112)		0.273*** (0.098)		0.125 (0.081)		0.009 (0.019)	
Overconfidence		0.088 (0.071)		0.245*** (0.062)		0.272*** (0.051)		0.077*** (0.010)
Constant	2.322*** (0.660)	3.866*** (0.169)	0.533 (0.588)	2.183*** (0.144)	-0.027 (0.450)	0.740*** (0.101)	0.258** (0.118)	0.317*** (0.030)
R ²	0.04	0.01	0.05	0.11	0.02	0.24	0.00	0.22
Observations	193	193	193	193	193	193	193	193

Notes: The table shows results of OLS regressions with debt levels after market stage 1, 2 and at the end of the experiment as the dependent variable. Income expectations is the expected income in Euro for the two income tasks. Overconfidence is the difference between the expected income and the actual income. Easy treatment and control group are indicator variables for the respective treatment (with the hard treatment as the omitted category). Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

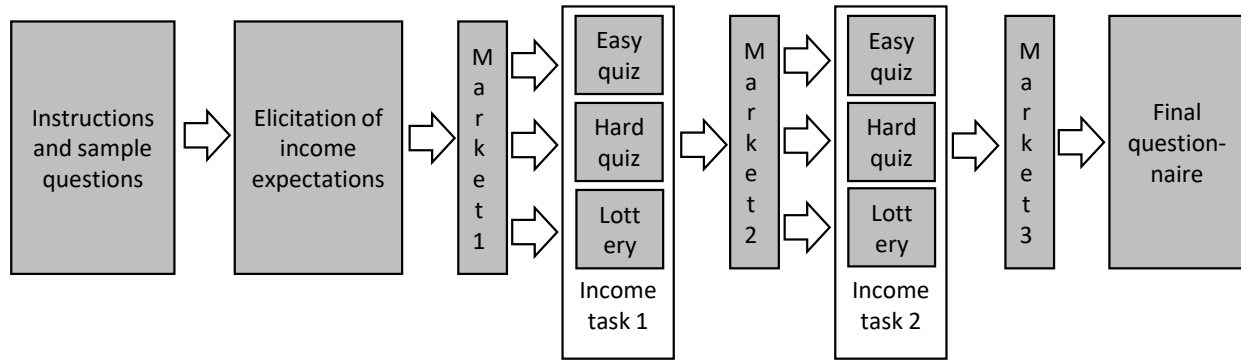


Fig 1. Sequence of the experiment.

Notes: The figure shows the sequential structure of the experiment. Arrows indicate how participants progress from one stage to another. Multiple arrows indicate instances in which different treatment groups enter different tasks. Randomization in treatment groups is done at the beginning of the experiment, as already sample questions are different across groups.

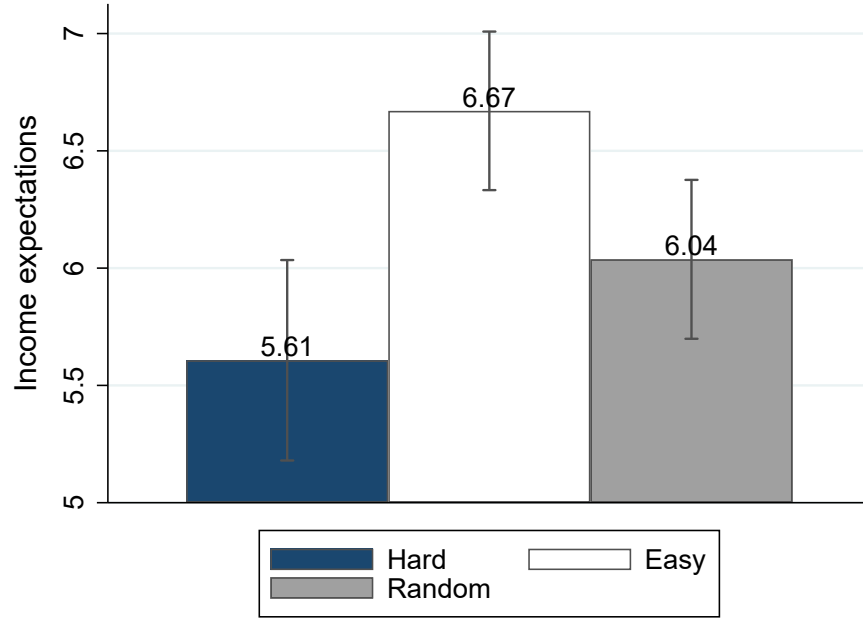


Fig 2. Income expectations by treatment before market 1.

Notes: Average income expectations in Euro for both income tasks by treatment (possible range is between 2 and 10 Euro). The 95-percent confidence interval is indicated.

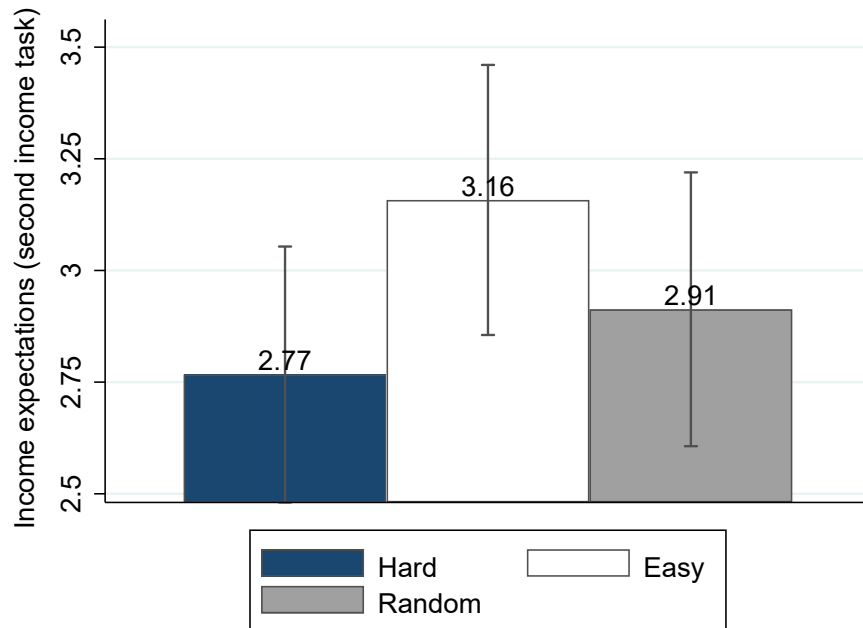


Fig 3. Income Expectations by Treatment before Market 2.

Notes: Average income expectations in Euro for the second income task by treatment group (possible range is between 1 and 5 Euro). The 95-percent confidence interval is indicated.

Earn more tomorrow: Overconfidence, income expectations and consumer indebtedness

Online Appendix

This online appendix contains supplementary material to the paper “Earn more tomorrow: Overconfident income expectations and consumer indebtedness.” Section A provides details on the survey evidence from the German SOEP panel, section B contains instructions and screenshots from the experiment, section C provides the questions used in the quiz tasks, section D examines evidence on the hard-easy effect, section E reports additional tables and robustness tests for the main experiment, and section F presents further results of the robustness experiment.

Online Appendix A. The German Socio-Economic Panel

A.1. GSOEP supplementary data description

The German Socio-Economic Panel Innovation Sample (GSOEP-IS) is a supplementary stream of the much larger GSOEP household survey (Richter and Schupp, 2012). The GSOEP-IS is designed to field more innovative questions to a representative sample of German households. We use data from the panel that started in 2016 and continued in 2017. This sample answers some of the questions that are also covered in the main GSOEP, however, households mostly respond to the GSOEP-IS questionnaire, which covers different questions. In 2016, 1,000 households were randomly sampled throughout Germany. The field work includes several attempts to contact selected households. Non-responding households are replaced by similar households to preserve representativeness. In each household all adults complete the survey, which results in 1,556 individuals taking part in 2016. Sample size is reduced to 1,085 participants in 2017 and to 902 participants in 2018 due to attrition.

We combine responses from the surveys in 2016, 2017 and 2018, as they contain different questions. The 2017 survey includes a number of detailed questions on borrowing behavior by the household and household members (for the code book see Horneber, 2017). We use this sample for our analysis, and, therefore, cite $N=1,085$ as the sample size in the paper. Table A.1 shows descriptive statistics about these participants. Just over half of participants are women, the average age is almost 55, and 26% of all participants have a lower secondary degree, whereas 18% have a university degree. Only about half of the respondents is employed. Average net monthly household income is €3040.

Some of the control variables in the experiment were modelled after the questions included in the GSOEP. The values for financial literacy, self-control, and risk tolerance can thus be directly compared between the two samples. Unsurprisingly, financial literacy is worse for the general population (4.3 correct responses) than for a student sample (5.1). Self-assessed risk tolerance is lower with 4.0 compared to 4.8 in the student sample. In the choice method, risk tolerance is more similar at 9.7 for the representative sample compared to 10.3 for students. As self-control is the only variable taken from the survey wave in 2018, we have fewer observations for this variable. We thus impute missing values of self-control in the regression analysis.

Table A.1. GSOEP data descriptive statistics

	<i>n</i>	Mean	Std.dev.	5p	95p
Gender	1085	0.52	0.50	0.00	1.00
Age	1084	54.68	18.08	23.00	81.00
Education	1085	1.38	1.51	0.00	4.00
Works	1085	0.46	0.50	0.00	1.00
Household income (net)	1032	3040	1737	950	6000
Financial literacy	1039	4.31	1.61	1.00	6.00
Risk tolerance (self-assessed)	1037	4.00	2.26	0.00	8.00
Risk tolerance (choices)	1014	9.06	7.48	1.00	24.00
Self control	835	−0.02	0.87	−1.45	1.39

Notes: The table shows means, standard deviations, the minimum, and the maximum of demographic variables and controls for the participants in the GSOEP-IS in 2017. Gender is an indicator variable taking a value of one if female. Age is reported in years. Education is an ordinal variable containing the highest achieved educational degree from 0 (no educational degree) to 4 (university degree). Works is an indicator variable taking the a value of one if a participant works. Household income (net) is the monthly net income of the household from all sources. Financial literacy is the number of correct answers in a financial literacy test (six questions taken from the German SOEP-IS 2016 survey). Self-assessed risk tolerance is measured on a scale from 0=“completely unwilling to take risk” to 10=“very willing to take risk.” Risk tolerance (choices) is based on lottery choices using the staircase method of Falk et al. (2016) with values from 1 (least risk tolerant) to 32 (most risk tolerant). Self control is a factor based on 13 questions using the German translation of Tangney, Baumeister and Boone (2004) included in the GSOEP-IS in 2018.

A.2. Measures of overconfidence

To measure overconfidence, we use two short tasks which are part of the questionnaire. The first task asks participants to name as many numbers that are multiples of 9 (or 17) as they can in 20 seconds. The second task asks participants to name as many animals (or insects) as they can in 20 seconds. The version of the task (9 or 17, animals or insects) is randomly assigned and is intended to reflect different levels of difficulty. Each respondent answers only one question in each domain. Participants manage on average to name 10.7 correct multiples of 9 and 14.1 animals, while they name only 5.8 correct multiples of 17 and 8.2 insects.

After they have completed these tasks, they are asked how many of 100 randomly selected individuals from Germany would perform better than themselves in the tasks. We sort participants in deciles according to their expected performance and their actual performance. We then subtract their actual performance from their expected performance. We interpret the resulting difference as a measure for overconfidence and normalize it to the interval $[-1;1]$. We hereby obtain two separate variables for overconfidence related to the numbers task and the animal task, respectively. The overconfidence variables are available for 950 and 972 participants, respectively.

Average overconfidence (numbers) is 0.17, and average overconfidence (animals) is 0.16. The Pearson correlation between the two overconfidence measures is 0.36. The distributions of the two measures of overconfidence are shown in Figures A.1 and A.2. They show that a majority of participants are overconfident.

Similar to the experiment, the overconfidence measures are based on overplacement in relative comparisons (Moore and Healy, 2008; Merkle and Weber, 2011), and they have

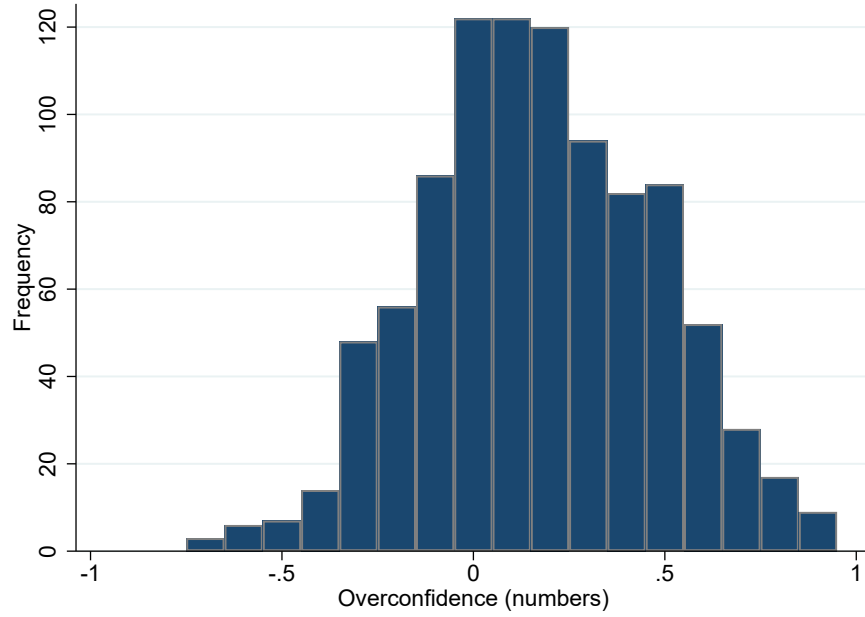


Figure A.1. Overconfidence in the numbers task
 Belief about relative performance minus actual performance compared to others.

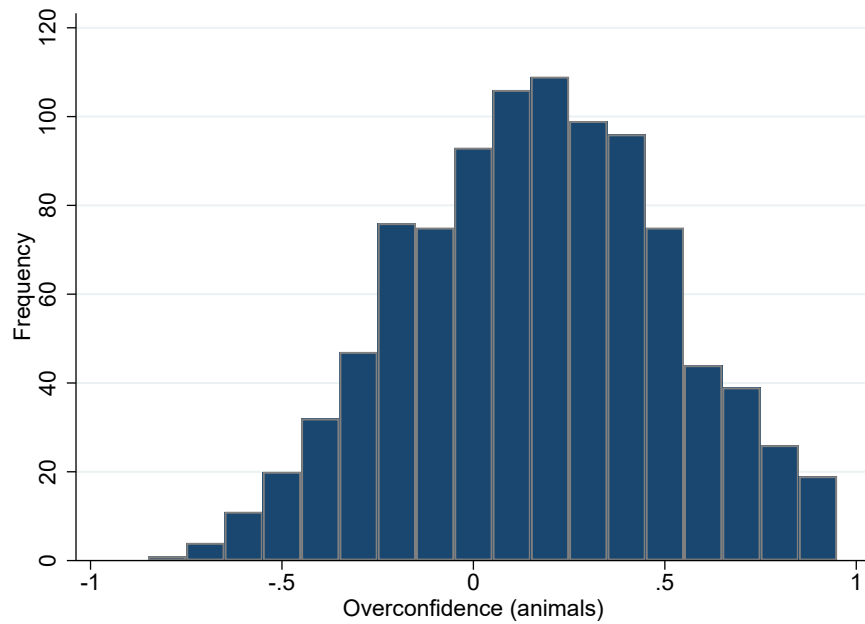


Figure A.2. Overconfidence in the animals task
 Belief about relative performance minus actual performance compared to others.

a hard or easy version. For the purpose of this paper, we pool both versions within each domain. There is no reason to expect an effect of task difficulty on actual borrowing behavior. Different from the experiment, overconfidence is elicited after the completion of the task and is unrelated to participants' income. Moreover, the performance in the task does not affect consumption opportunities.

Table A.2 shows regressions of both measures of overconfidence on socio-demographic variables. Few of these variables show a significant correlation with overconfidence and the

Table A.2. Overconfidence GSOEP

	Overconfidence (numbers) (1)	Overconfidence (animals) (2)
Female	−0.068*** (0.024)	−0.143*** (0.028)
Age	−0.001 (0.001)	0.000 (0.001)
Education	0.015* (0.008)	0.014 (0.009)
Works	0.009 (0.030)	−0.016 (0.033)
Log of income	−0.052** (0.026)	−0.032 (0.024)
Financial literacy	0.000 (0.008)	0.005 (0.009)
Risk tolerance (self-assessed)	−0.005 (0.006)	−0.000 (0.006)
Risk tolerance (choices)	−0.001 (0.002)	−0.001 (0.002)
Self control	−0.011 (0.015)	−0.017 (0.016)
Constant	0.670*** (0.209)	0.446** (0.201)
R ²	0.028	0.057
Observations	625	638

Notes: The table shows results of OLS regressions of overconfidence in the numbers and animal domain on control variables. All control variables are as defined in Table A.1. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

regressions have a relatively low R^2 . A robust finding is that women are less overconfident than men, both in the numbers and the animals domain. This is in line with the overconfidence literature. No other socio-demographic variable shows a robust statistically significant effect across both regressions.

A.3. Debt taking behavior

We focus on two questions associated with borrowing and repayment behavior. The first of these questions asks whether respondents currently use the overdraft facility of their checking account. In the German context, this is an ubiquitous form of debt as most households have access to overdrafts (in the sample 82% respond to have access to overdraft). It can be considered a problematic form of debt, as overdrafts are a very expensive form of debt (interest rates often exceed 10%). Overdrafts might occur unplanned and to bridge short-term liquidity needs. However, they might also lead to continuous indebtedness. Using overdraft can be considered quite similar to debt taking in the experiment, as both are typically used to finance consumption. After a question whether they have access to overdraft, GSOEP participants are asked “Are you currently using such an overdraft facility?” to which 11% respond yes.

To understand the link between overconfidence and over-indebtedness we further examine whether respondents expect to be able to repay their current loans. The survey includes the question “What do you think is the probability that you will be able to repay all your current loans until their maturity according to the original agreements?” We derive two variables from this question: one binary variable taking a value of one if a participant is certain to repay all debt on time and zero otherwise, and one continuous variable using the submitted

probability. In the main analysis, we focus on the binary variable as most participants (92%) express a probability of one to be able to repay. Our results are robust to using the continuous version.

Finally, we consider a subjective measure of debt problems, which is whether the respondents feel their debt payments to be a burden. “Can you pay the redemption respectively interest for your credits without problems? Please consider credits for purchases and other expenses.” This binary variable is reverse to the repayment question and focuses on current problems rather than expectations. The description of this item also narrows down relevant debt to mostly consumer credit. 11% of participants report debt payment problems conditional on indicating they have such debt. Debt problems are positively correlated with overdraft usage (0.15, $p < .05$) and negatively correlated with expected repayment ability (-0.14 , $p < .05$).

Online Appendix B. Experimental Materials

B.1. Paper instructions

The following instructions were distributed to participants on paper prior to the start of the experiment. Original instructions were in German. Written instructions remained available for participants throughout the experiment.

Instructions

This experiment is designed to study financial decision making.

You will receive 5 Euro show-up fee. You will receive this irrespective of your decisions during the experiment. In addition to this you will receive further payouts that depend on your performance and your decisions during the experiment. It is therefore very important that you pay attention and follow these instructions carefully.

During the experiment you are not allowed to use electronic devices or to talk to other participants. Please use only the programs and functions that are part of the experiment. If you have a question, please raise your hand. We will then come to you and answer your question quietly. Please do not ask your questions out loud. If the question is relevant for all participants, we will repeat it and answer it out loud. If you violate these rules, we must exclude you from the experiment and the payout.

At the beginning of the experiment you will see short comprehension questions on screen, which we ask you to answer. If you answer one or more of these questions incorrectly, one of the experimenters will come to you to clarify any open questions.

Experimental structure:

1. Measuring income expectations
2. Shopping round 1
3. Opportunity to earn income
4. Shopping round 2
5. Opportunity to earn income
6. Shopping round 3
7. Questionnaire

How is income earned?

You will receive a participation fee of 8 Euro for taking part in this experiment. In the course of the experiment you will have two more opportunities to earn additional income. During these two rounds 25% of the participants will receive 5 Euro, 4 Euro, 2 Euro and 1 Euro respectively. You can hence earn a maximum of 10 € or will get at least 2 €. Details of how you can earn income will be displayed on the screen later on.

You can use this income to buy products. You can take all purchased products with you at the end of the experiment. Unspent income will **not be paid out** (with the exception of the participation fee).

Shopping rounds

In the course of the experiment you have the possibility to choose from a range of ten goods three times. The products remain the same in every round. The prices of the goods rise in each

round. In the first round the goods are therefore cheaper than in the second round and in the second round they are cheaper than in the third round. From the very beginning, you will be shown all the prices for the other rounds, so you can think about whether to buy a product immediately or later.

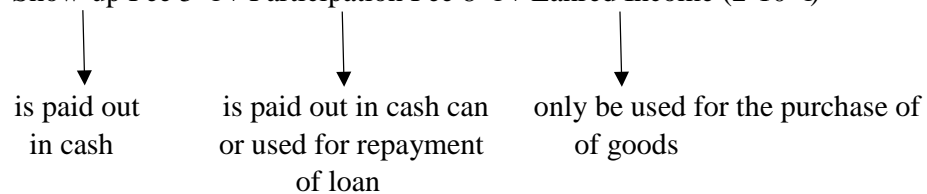
In the early rounds you have the opportunity to take out a loan, which you repay with the income you earn later on in the experiment. The maximum loan amount is 10€ in round 1 and 5€ in round 2. If you cannot repay the loan with income earned in the experiment, this amount will be deducted from your participation fee.

Experimental procedure

After you have read on screen how you can earn income and have answered two short questions about your income expectations, the experiment begins with a shopping round. At this point, you have not yet earned any income. However, you can already buy products using the loan. You then repay the loan using income that you earn later on. If you do not earn enough income to repay this loan, the rest of the loan will be repaid using the participation fee.

Diagram

Total fee = Show-up Fee 5 € + Participation Fee 8 € + Earned Income (2-10 €)



B.2. Experimental screenshots

In this section we provide screenshots from the experiment. The order of the screenshots is as it has been in the experiment. If there are several similar screens such as for markets 1-3, one screenshot of the market stage is shown as an example. If screens are identical or very similar across treatments, only one example is shown.

Verständnisfragen

Bitte bearbeiten Sie den Verständnistest.

Wenn Sie Fragen haben, geben Sie bitte Handzeichen. Ein Experimentator wird dann zu Ihnen kommen.

Wenn alle Teilnehmer die Fragen richtig beantwortet haben, kann das Experiment fortfahren.

Wie viel Einkommen können Sie zusätzlich zu Erscheinungsgeld und Teilnehmergehalt mindestens erzielen?

- ☐ 1 €
- ☐ 2 €
- ☐ 6 €

Wie viel Einkommen können Sie in einer einzelnen Runde maximal erzielen?

- ☐ 2 €
- ☐ 5 €
- ☐ 10 €

Wie entwickelt sich der Preis der Produkte über die Runden des Experiments?

- ☐ Produkte werden teurer
- ☐ Produkte werden günstiger
- ☐ Preise bleiben konstant

Was geschieht mit nicht ausgegebenem Einkommen aus dem Experiment?

- ☐ wird zusammen mit Teilnehmergehalt bar ausgezahlt
- ☐ verfällt
- ☐ kann bei der Auszahlung noch gegen Produkte eingetauscht werden

Was geschieht, wenn ein aufgenommenen Kredit nicht aus dem Verdienst im Experiment zurückgezahlt werden kann?

- ☐ der offene Betrag muss vom Teilnehmer an den Experimentator bezahlt werden
- ☐ ein Teil der Produkte kann zurückgegeben werden
- ☐ der offene Betrag wird aus der Teilnehmergehalt beglichen

OK

Figure B.1. Screenshot of comprehension questions

The first screen after the welcome screen includes 5 comprehension questions: What income can you minimally earn in addition to the show-up and participation fee? What income can you maximally earn in one round of the experiment? How do the prices of goods evolve in the experiment? What happens to income that is not spend in the experiment? What happens if debt taken cannot be repaid from the income earned in the experiment?

Bestimmung des Einkommens

In diesem Experiment werden Sie mit 7 anderen Teilnehmern eine Gruppe von 8 Personen bilden. Ihre Bezahlung hängt von Ihrer Leistung in einem Quiz mit 10 Fragen zur Allgemeinbildung ab. Dabei bestimmt sich Ihr Verdienst durch Ihren Rangplatz im Vergleich zu den anderen Teilnehmern Ihrer Gruppe. Die Rangliste wird nach der Anzahl korrekter Antworten erstellt.

Rangplatz	Auszahlung
1 und 2	5 Euro
3 und 4	4 Euro
5 und 6	2 Euro
7 und 8	1 Euro

Im Falle eines Gleichstands an richtigen Antworten, entscheidet die niedrigere Antwortzeit über die Platzierung.

Sie nehmen an zwei Quizrunden teil, für die jeweils eine eigene Rangfolge erstellt wird. Sie können also maximal 10€ verdienen und erzielen mindestens 2€.

Übungsfragen

Hier sind vier Übungsfragen, die zufällig aus dem Set an Quizfragen ausgewählt wurden. Die Antworten auf die Übungsfragen gehen nicht in die Bewertung Ihres Ranges ein.

In welchem Jahrhundert fand der zweite Weltkrieg statt?

☐ 17. Jhdt.
☐ 18. Jhdt.
☐ 19. Jhdt.
☐ 20. Jhdt.

Welches Organ ist im menschlichen Körper für die Zirkulation des Blutes verantwortlich?

☐ Lunge
☐ Nieren
☐ Herz
☐ Leber

Wie heißt der aktuelle Bundespräsident der Bundesrepublik Deutschland?

☐ Horst Köhler
☐ Christian Wulff
☐ Frank-Walter Steinmeier
☐ Joachim Gauck

Astrid Lindgren hat welchen der folgenden Charaktere erfunden?

☐ Alice im Wunderland
☐ Pippi Langstrumpf
☐ Peter Pan
☐ Harry Potter

Experiment beginnen

Figure B.2. Screenshot of income task instructions.

On this screen people are informed about the income task, that they are part of a group of eight participants and earn income depending on their rank in this group. This is followed by four sample questions for the respective treatment. This screenshot shows the screen in the easy treatment. For hard treatment only questions were different. Translations of the questions used to prime participants in the hard and the easy treatment can be found in Online Appendix C.

xiii

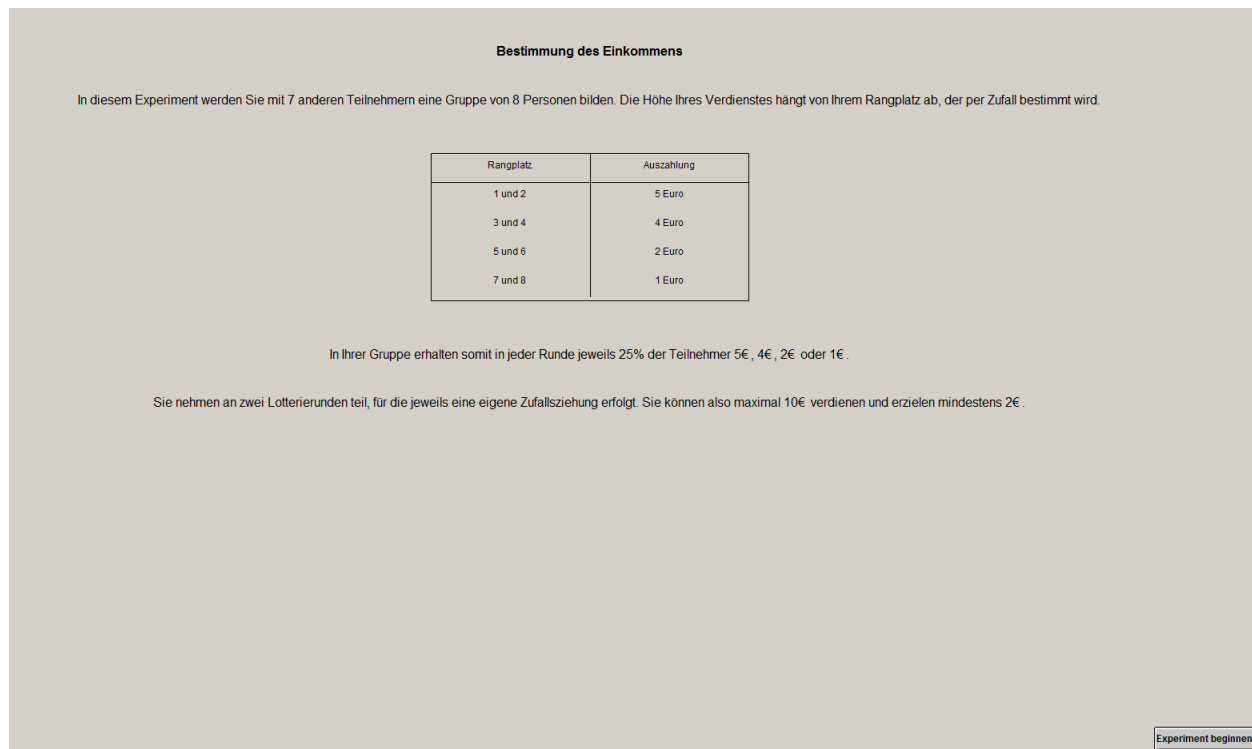


Figure B.3. Screenshot of income instructions in the random group. Participants in the random group are informed that they are paid according to a lottery in which they earn €5, €4, €2 or €1 with equal probability of 25%.

Sie haben zuvor erfahren welche Art von Fragen Ihnen in dem Quiz gestellt werden und wie sich daraus die Bezahlung ergibt.

Zur Erinnerung: Die besten 25% erhalten 5€ und die jeweils folgenden 25% erhalten 4€, 2€, und 1€. Es werden insgesamt 2 Runden gespielt.

Welchen Verdienst erwarten Sie insgesamt aus diesen zwei Runden?

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

Denken Sie an die erste der beiden Runden. Mit welcher Wahrscheinlichkeit erwarten Sie einen bestimmten Rang in Ihrer Gruppe zu erhalten?

Bitte geben Sie die Wahrscheinlichkeit in Prozent an, z.B. 0.5 für 50%. Bedenken Sie dabei, dass sich die Wahrscheinlichkeiten zu 1 addieren müssen.

Rang 1 oder 2	<input type="text"/>
Rang 3 oder 4	<input type="text"/>
Rang 5 oder 6	<input type="text"/>
Rang 7 oder 8	<input type="text"/>

OK

Figure B.4. Screenshot of questions to elicit income expectations.

After the instructions on the income task, income expectations are elicited. The first question asks participants to state how much they are expecting to earn on a scale from €2 to €10. The second question asks participants for the probability, which with they are expecting to earn each amount in one task selected by chance.



Figure B.5. Screenshot of the first market stage.

The screen displays pictures and names of the ten goods that are available for purchase. The prices of the goods in this round and in all following rounds are displayed below the pictures. With the “+” and “-” buttons, participants can select the number of items they want to purchase. At the top of the screen, the current balance and the maximal credit are displayed in red. The maximal amount still to spend and the current loan amount are displayed as well. The screens in the second and third market stages look very similar.

Verbleibende Zeit [sec]: 10	
Wie viele Staaten sind Mitglied der EU?	<input type="radio"/> 8 <input type="radio"/> 18 <input type="radio"/> 28 <input type="radio"/> 38
Durch welches Land verlaufen die Alpen NICHT?	<input type="radio"/> Deutschland <input type="radio"/> Frankreich <input type="radio"/> Italien <input type="radio"/> Spanien
Wie viele Einwohner hat Deutschland?	<input type="radio"/> 53 Millionen <input type="radio"/> 63 Millionen <input type="radio"/> 73 Millionen <input type="radio"/> 83 Millionen
In welchem Jahr starb Adolf Hitler?	<input type="radio"/> 1944 <input type="radio"/> 1945 <input type="radio"/> 1946 <input type="radio"/> 1947
Wie heißt das nordkoreanische Staatsoberhaupt?	<input type="radio"/> Kim Jong-Il <input type="radio"/> Kim Jong-Un <input type="radio"/> Kim Il-Sung <input type="radio"/> Kim Sung-Un
Wer ist Regierungschef in Großbritannien?	<input type="radio"/> David Cameron <input type="radio"/> Boris Johnson <input type="radio"/> Theresa May <input type="radio"/> Gordon Brown
Wie heißt die Landeshauptstadt von Nordrhein-Westfalen?	<input type="radio"/> Köln <input type="radio"/> Stuttgart <input type="radio"/> Düsseldorf <input type="radio"/> Hannover
Wer wählt den Bundespräsidenten?	<input type="radio"/> Der Bundestag <input type="radio"/> Die Bundesversammlung <input type="radio"/> Der Bundesrat <input type="radio"/> Das Volk
Wofür ist Mao Zedong bekannt?	<input type="radio"/> Er war Staatspräsident der Volksrepublik China <input type="radio"/> Er war ein koreanischer Freiheitskämpfer <input type="radio"/> Er ist ein chinesischer Konzertpianist <input type="radio"/> Er war ein japanischer Feldherr
Der Name welches Helden der antiken Mythologie wird heute benutzt um eine Infahrt zu beschreiben?	<input type="radio"/> Odysseus <input type="radio"/> Paris <input type="radio"/> Ikarus <input type="radio"/> Achilles

Speichern

Figure B.6. Screenshot of the first income task.

The screenshot shows the first quiz task in the easy treatment and contains ten general knowledge questions. In the upper right corner, the remaining time is displayed. The screens for the hard treatment and for the second income task are the same. The translated questions for the income tasks in each treatment can be found in Online Appendix C.

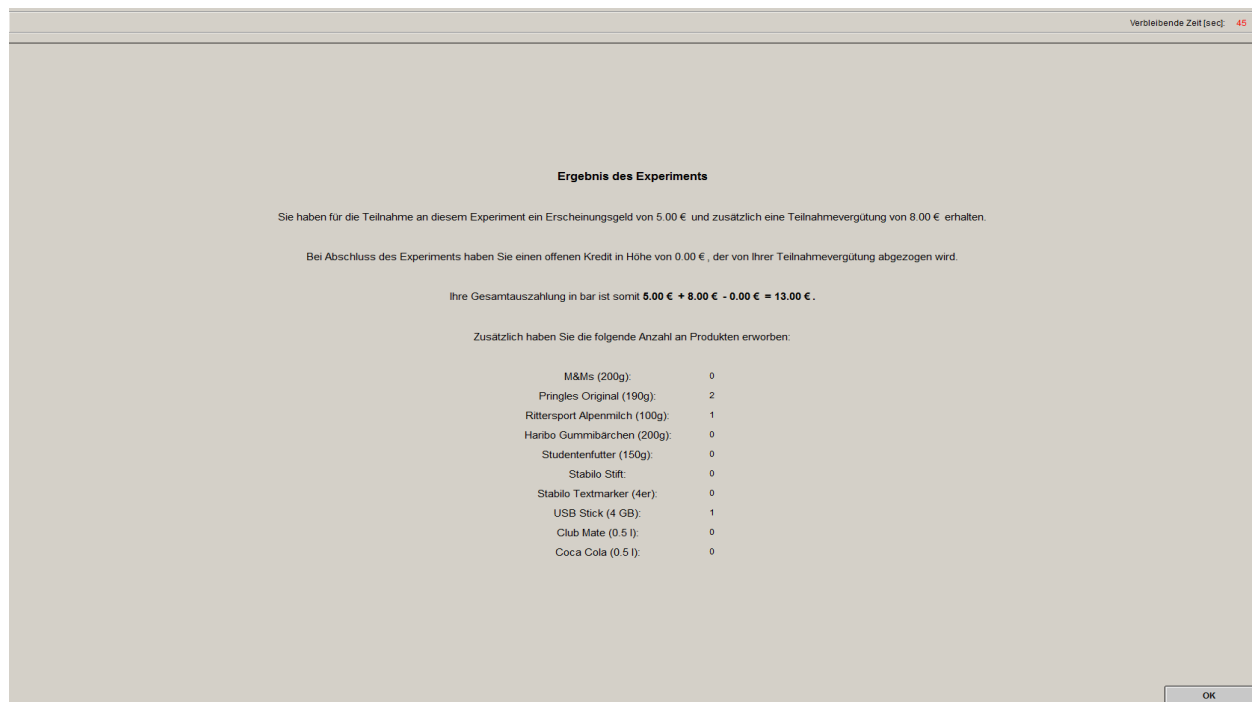


Figure B.7. Screenshot of the pay-off screen.

The screen displays the cash pay-off and the number of goods purchased in the experiment. The cash-payoff is calculated as the show-up fee plus participation fee minus open debt. The lab rules require that a minimum show-up fee of €5 is paid, which is why only the participation fee is subject to deductions. This is the last screen before the socio-demographic questionnaire.

Online Appendix C. Income Task

C.1. General knowledge questions in the easy treatment

C.1.1. Sample questions used to prime participants in the easy treatment

- | | |
|--|--|
| 1. During which century did the second world war take place? | 3. What is the name of the current President of the Federal Republic of Germany? |
| <input type="checkbox"/> 17th century | <input type="checkbox"/> Horst Köhler |
| <input type="checkbox"/> 18th century | <input type="checkbox"/> Christian Wulff |
| <input type="checkbox"/> 19th century | <input type="checkbox"/> Frank-Walter Steinmeier |
| <input type="checkbox"/> 20th century | <input type="checkbox"/> Joachim Gauck |
| 2. Which organ in the human body is responsible for blood circulation? | 4. Astrid Lindgren invented which of the following characters? |
| <input type="checkbox"/> Lung | <input type="checkbox"/> Alice in Wonderland |
| <input type="checkbox"/> Kidneys | <input type="checkbox"/> Pippi Longstocking |
| <input type="checkbox"/> Heart | <input type="checkbox"/> Peter Pan |
| <input type="checkbox"/> Liver | <input type="checkbox"/> Harry Potter |

C.1.2. Easy questions used in the two income tasks

- | | |
|--|---|
| 1. How many countries are members of the EU? | 3. Which country does NOT cover part of the Alps? |
| <input type="checkbox"/> 8 | <input type="checkbox"/> Germany |
| <input type="checkbox"/> 18 | <input type="checkbox"/> France |
| <input type="checkbox"/> 28 | <input type="checkbox"/> Italy |
| <input type="checkbox"/> 38 | <input type="checkbox"/> Spain |
| 2. Who is the head of government in Britain? | 4. What is the name of the state capital of North Rhine-Westphalia? |
| <input type="checkbox"/> David Cameron | <input type="checkbox"/> Cologne |
| <input type="checkbox"/> Boris Johnson | <input type="checkbox"/> Stuttgart |
| <input type="checkbox"/> Theresa May | <input type="checkbox"/> Düsseldorf |
| <input type="checkbox"/> Gordon Brown | <input type="checkbox"/> Hanover |

5. How many inhabitants does Germany have?
 - ☐ 53 million
 - ☐ 63 million
 - ☐ 73 million
 - ☐ 83 million
6. Who elects the Federal President in Germany?
 - ☐ The Bundestag
 - ☐ The Federal Assembly
 - ☐ The Federal Council
 - ☐ The people
7. In which year did Adolf Hitler die?
 - ☐ 1944
 - ☐ 1945
 - ☐ 1946
 - ☐ 1947
8. What is Mao Zedong known for?
 - ☐ He was President of the People's Republic of China
 - ☐ He was a Korean freedom fighter
 - ☐ He is a Chinese concert pianist
 - ☐ He was a Japanese general
9. What is the name of the North Korean head of state?
 - ☐ Kim Jong-Il
 - ☐ Kim Jong-Un
 - ☐ Kim Il-Sung
 - ☐ Kim Sung-Un
10. The name of which hero in ancient mythology is used today to describe an aimless trip?
 - ☐ Ulysses
 - ☐ Paris
 - ☐ Icarus
 - ☐ Achilles
11. In which city does an opera festival take place each summer, during which only operas by Richard Wagner are performed?
 - ☐ Montreux
 - ☐ Wacken
 - ☐ Bayreuth
 - ☐ Bad Segeberg
12. How high was the gross domestic product per capita in Germany in US-\$ in 2016?
 - ☐ 10,038
 - ☐ 17,901
 - ☐ 41,902
 - ☐ 103,199
13. Where did industrialization begin?
 - ☐ Flanders
 - ☐ France
 - ☐ Ruhr
 - ☐ Great Britain

14. In which galaxy is the Earth's solar system located?
 - ☐ Milky Way
 - ☐ Centaurus A
 - ☐ Cigar Galaxy
 - ☐ Andromeda Galaxy
15. What is the reason ice cubes do not sink in water?
 - ☐ The surface tension of the water
 - ☐ The air bubbles trapped in the ice
 - ☐ The lower density of the ice
 - ☐ The sublimation effect that creates an upward suction
16. What does the chemical compound H_2O stand for?
 - ☐ Water
 - ☐ Oxygen
 - ☐ Nitrogen
 - ☐ Iron
17. What is the title of a bestselling novel by Daniel Kehlmann?
 - ☐ "Das Gewicht der Welt"
 - ☐ "Volk und Welt"
 - ☐ "Die Vermessenheit der Forscher"
 - ☐ "Die Vermessung der Welt"
18. Which of the following characters does not appear in Goethe's Faust?
 - ☐ Heinrich Faust
 - ☐ Mephistopheles
 - ☐ Gretchen
 - ☐ Don Carlos
19. Which architectural style is NOT be found during the 20th century?
 - ☐ Bauhaus
 - ☐ Art Deco
 - ☐ Postmodernism
 - ☐ Gothic Art
20. Which country was a colonial power in India?
 - ☐ France
 - ☐ Pakistan
 - ☐ Germany
 - ☐ Great Britain

C.2. Hard treatment

C.2.1. Sample questions used to prime participants in the hard treatment

1. In which century did the Thirty Years War take place?
 - ☐ 16th century
 - ☐ 17th century
 - ☐ 18th century
 - ☐ 19th century
2. Which sensory cells in the human eye are responsible for colour vision?
 - ☐ Rods
 - ☐ Cones
 - ☐ Plugs
 - ☐ Buttons
3. Who was the second Federal President of the Federal Republic of Germany?
 - ☐ Walter Scheel
 - ☐ Theodor Heuss
 - ☐ Gustav Heinemann
 - ☐ Heinrich Lübke
4. James Joyce invented which of the following characters?
 - ☐ Leopold Bloom
 - ☐ Jarvis Lorry
 - ☐ Oliver Twist
 - ☐ Samuel Pickwick

C.2.2. Hard questions used in the two income tasks

1. How many states are members of the UNO?
 - ☐ 173
 - ☐ 183
 - ☐ 193
 - ☐ 203
2. Who is the head of government in Finland?
 - ☐ Mari Kiviniemi
 - ☐ Alexander Stubb
 - ☐ Juha Sipilä
 - ☐ Jyrki Katainen
3. Which country does NOT cover part of the Atlas Mountains?
 - ☐ Morocco
 - ☐ Algeria
 - ☐ Libya
 - ☐ Tunisia
4. What is the name of the capital of the Austrian province of Styria?
 - ☐ Graz
 - ☐ Innsbruck
 - ☐ Linz
 - ☐ Sankt Pölten

5. How many inhabitants does the EU have?
 - ☐ 499 million
 - ☐ 508 million
 - ☐ 516 million
 - ☐ 523 million
6. According to the German Basic Law, who determines the “politic guidelines” in Germany?
 - ☐ The Federal Chancellor
 - ☐ The Federal Constitutional Court
 - ☐ The Bundestag
 - ☐ The people
7. Which of these leading National Socialists was already dead at the beginning of the Second World War?
 - ☐ Ernst Röhm
 - ☐ Hjalmar Schacht
 - ☐ Julius Streicher
 - ☐ Wilhelm Keitel
8. How did Mao Zedong expand his power in China?
 - ☐ Great Jump
 - ☐ Wide Field
 - ☐ Heavenly Peace
 - ☐ Long March
9. Which of these Western orchestras gave a concert in North Korea in 2008?
 - ☐ Vienna Philharmonic Orchestra
 - ☐ Berlin Philharmonic Orchestra
 - ☐ New York Philharmonic Orchestra
 - ☐ Academy of St. Martins in the Field
10. In an ancient tragedy, a woman murders her children to take revenge on her unfaithful husband. What is the woman’s name?
 - ☐ Aida
 - ☐ Electra
 - ☐ Pamela
 - ☐ Medea
11. In which year did the Richard Wagner Festival first take place?
 - ☐ 1867
 - ☐ 1876
 - ☐ 1882
 - ☐ 1924
12. What was the average GDP per capita in Tanzania in US-\$ in 2016?
 - ☐ 480
 - ☐ 638
 - ☐ 970
 - ☐ 1,170

13. What historical significance do manufactories have?
 - ☐ In manufactories means of production belong to workers
 - ☐ Manufactories were originally prison workshops
 - ☐ Manufactories were forerunners of industrial mass production
 - ☐ Guilds exercised their monopoly in the manufactories
14. What does a so-called red shift suggest?
 - ☐ Vision loss
 - ☐ Expansion of the universe
 - ☐ High blood alcohol level
 - ☐ Beginning of autumn
15. A rocket trying to escape the Earth's gravitational pull must reach a certain escape velocity. How fast does it have to be?
 - ☐ 15,000 km/h
 - ☐ 26,000 km/h
 - ☐ 36,000 km/h
 - ☐ 40,000 km/h
16. What does the chemical compound NH_3 stand for?
 - ☐ Ammonia
 - ☐ Ethanol
 - ☐ Nicotine
 - ☐ Hydrochloric acid
17. Which of the following works is NOT by Herta Müller?
 - ☐ "Niederungen"
 - ☐ "Atemschaukel"
 - ☐ "Heute wäre ich mir lieber nicht begegnet"
 - ☐ "Wunschloses Unglück"
18. From which work is the quote: "Ich wei, sie tranken heimlich Wein und predigten öffentlich Wasser"?
 - ☐ Heine, "Germany. A Winter's Tale"
 - ☐ Goethe, "Faust"
 - ☐ Brecht, "Mutter Courage"
 - ☐ Lessing, "Minna von Barnhelm"
19. Who was one of the architects who built the Hamburg Elbphilharmonie?
 - ☐ Norman Forster
 - ☐ Zaha Hadid
 - ☐ Jacques Duke
 - ☐ Peter Zumthor
20. Which country was a colonial power in Iraq during the last century?
 - ☐ France
 - ☐ Iran
 - ☐ Germany
 - ☐ Great Britain

Online Appendix D. Evidence on the Hard-easy Effect

In this section, we collect evidence on the hard-easy effect to justify its usage as an experimental manipulation of beliefs. First, we show cumulative distribution functions (CDFs) for overconfidence in both treatments in the main experiment and robustness experiment.

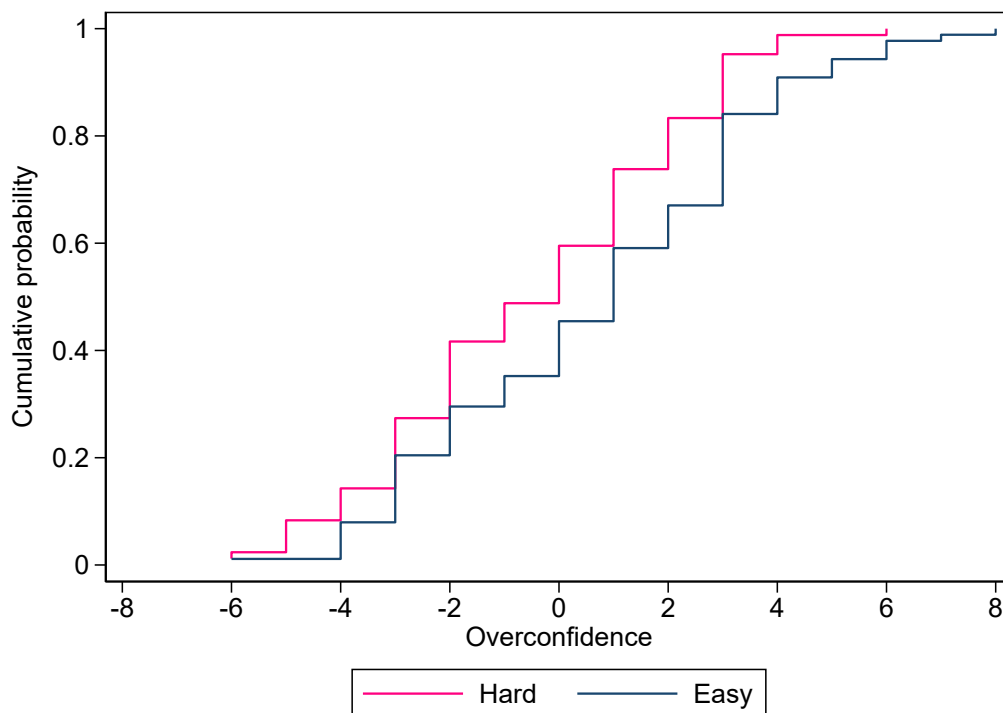


Figure D.1. CDFs of overconfidence by treatment (main experiment)

Notes: Cumulative distribution functions (CDFs) are displayed for the hard and easy treatment in the main experiment. Overconfidence is defined as expected income minus actual income (in €).

Figure D.1 shows the strong treatment effect in the main experiment, the entire CDF for the easy treatment is to the right of the CDF for the hard treatment, which means that higher overconfidence is observed in the easy treatment. The hard-easy effect is less strong in the robustness experiment with partly overlapping CDFs (see Figure D.2).

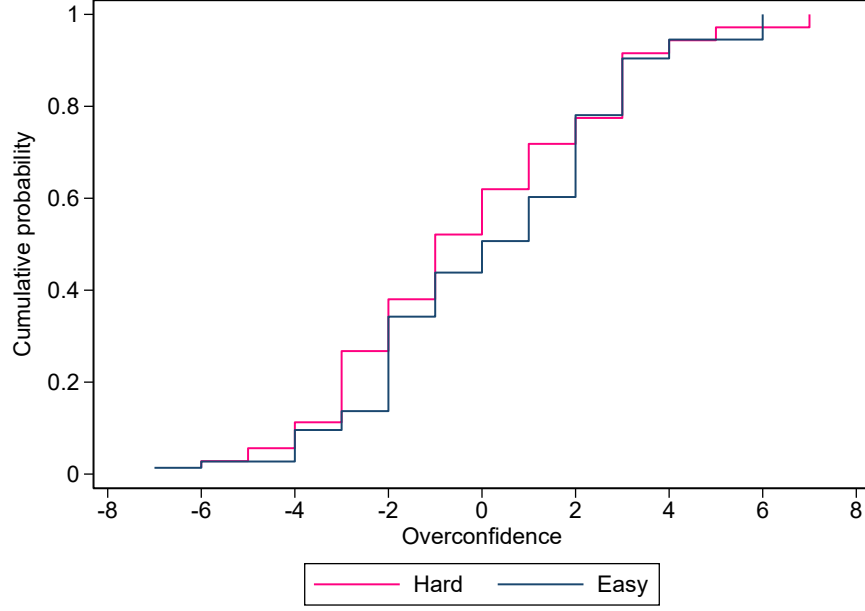


Figure D.2. CDFs of overconfidence by treatment (robustness experiment)

Notes: Cumulative distribution functions (CDFs) are displayed for the hard and easy treatment in the robustness experiment. Overconfidence is expected income minus actual income (in €).

Our experiments rely on the hard-easy effect to produce sufficient treatment differences. Table D.1 summarizes results of five of the most cited psychology papers on the effect. We collect original data for the studies by Moore and Small (2007) and Moore and Healy (2008). For the remaining papers, we use results as reported in the paper.

All studies find a strongly significant hard-easy effect ($p < 0.001$), and the average Cohen's d is 0.88. An effect size of $d = 0.8$ is commonly considered a strong effect (Lakens, 2013). Typically, a majority of participants views themselves as better than average in the easy task, while only a minority does so in the hard task. We take the strengths and reliability of the effect as justification to use it as manipulation in the experiments.

In more recent research, a number of economics papers has used the hard-easy effect in a similar way as our study to manipulate beliefs. We obtain replication files from Dargnies,

Table D.1. Psychology articles on the (reversed) hard-easy effect

	Google cites	N	t-value hard-easy	Cohen's d	BTA easy	BTA hard
Kruger (1999), Exp 2	1257	94	5.63	1.11	63%ile	42%ile
Moore and Kim (2003), Exp 1	325	78	3.53	0.80	n/a	n/a
Moore and Kim (2003), Exp 2	325	88	3.61	0.77	73%ile	57%ile
Burson et al. (2006), Exp 1	444	90	3.64	0.77	62%ile	48%ile
Moore and Small (2007), Exp 1	385	255	8.48	1.06	62% (62%ile)	20% (37%ile)
Moore and Healy (2008)	2698	82	7.20	0.74	66%	27%

Notes: The table summarizes results by Kruger (1999), Moore and Kim (2003), Burson, Larrick and Klayman (2006), Moore and Small (2007), and Moore and Healy (2008). The table focuses on experiments within each paper that involve a trivia quiz or similar task. Preference is given to the baseline version of the experiment without further manipulations. For Moore and Small (2007) and Moore and Healy (2008) original data were available on Don Moore's webpage. For the other articles, reported statistics were used for calculations where possible. Google cites is the number of cites as of 5/10/2023. N is the number of participants in the easy and hard treatments (excluding participants in control treatments or medium treatments, if any). t-value hard-easy is the t-value of a standard t-test for the equality of group means (for Moore and Healy (2008) the t-value is adjusted for repeated measurements). Cohen's d is the standardized effect size for the treatment effect. BTA easy and BTA hard report the percentage of participants that view themselves as above average. If this percentage is not available, the average percentile participants believe they are in is reported.

Hakimov and Kübler (2019) and Barron and Gravert (2022), as well as summary statistics via personal communication from Klühs, Koch and Stein (2019) and Bruhin, Petros and Santos-Pinto (2022). Statistics for Colzani and Santos-Pinto (2021) are calculated using reported data. Table D.2 shows summary statistics for these papers.

Results reveals that effect sizes are smaller in these studies (average $d=0.47$, commonly regarded as medium effect size). Treatment effects remain mostly significant, which is partly due to a higher number of observations. We find that our experiment 2 is no outlier among these studies. Overall the hard-easy effect appears more fragile in these more recent papers.

Table D.2. Economic studies using the hard-easy effect as treatment manipulation

	N	t-value hard-easy	Cohen's d	BTA easy	BTA hard
This paper, Exp 1	172	3.86	0.59	53%	29%
This paper, Exp 2	144	1.87	0.31	56%	30%
Dargnies, Hakimov and Kübler (2019)	160	2.12	0.33	56%	40%
Barron and Gravert (2022)	100	3.71	0.74	86%	66%
Klühs, Koch, and Stein (2019)	471	1.93	0.18	79%	74%
Colzani and Santos-Pinto (2021)	190	4.40	0.64	70%	45%
Bruhin et al. (2023)	240	3.97	0.51	55%	23%

Notes: The table summarizes results by us, Dargnies, Hakimov and Kübler (2019), Barron and Gravert (2022), Klühs, Koch and Stein (2019), Colzani and Santos-Pinto (2021), and Bruhin, Petros and Santos-Pinto (2022). The table focuses on the first stage of the experiments (treatment manipulation) that involves a trivia quiz or similar task. For Dargnies, Hakimov and Kübler (2019) and Barron and Gravert (2022) original data were obtained from replication files. For Klühs, Koch and Stein (2019) and Bruhin, Petros and Santos-Pinto (2022) relevant statistics were obtained via personal communication. For Colzani and Santos-Pinto (2021) reported statistics in the paper were used. For this paper standard errors were approximated based on visual inspection of confidence intervals. N is the number of participants in the easy and hard treatments (excluding participants in control treatments, if any). t-value hard-easy is the t-value of a standard t-test for the equality of group means. Cohen's d is the standardized effect size for the treatment effect. BTA easy and BTA hard report the percentage of participants that view themselves as above average.

Online Appendix E. Robustness Tests for the Main Experiment

In this section we report additional tables and robustness tests performed on the data of the main experiment. In order to increase transparency, we include Table E.1 that shows the reduced form relationship between being in the easy treatment and the outcome variables.

As robustness checks, we first discuss the exclusion of participants based on lack of comprehension or non-standard preferences. We then consider alternative measures of overconfidence. Finally, we provide evidence for the intensive margin of debt taking.

E.1. Sample exclusions

We include comprehension questions at the beginning of the experiment to make sure that participants understand the experimental design. We exclude participants with two or more errors, which is a compromise between preserving a substantial sample and reducing noise. As this exclusion was not pre-registered, we provide robustness on its effects (we did not anticipate the substantial language problems by some participants).

In Table E.2, we compare socio-demographics of excluded participants with participants that are part of the main analysis. The two groups are mostly similar, except that people in the excluded sample are slightly older and are more likely to have a bachelor degree. As the intake of foreign students in Berlin is higher at the Masters level, this is in line with our observation that some participants were not in command of sufficient German to participate in the experiment. We further observe lower levels of financial literacy among excluded participants, which speaks in favor of poor comprehension (linguistically or otherwise).

Table E.1. Indebtedness over the course of the experiment (OLS)

	Debt after market 1		Debt after market 2		Final debt		Has debt	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Easy	0.728** (0.366)	0.587 (0.370)	0.600** (0.297)	0.465 (0.302)	0.361* (0.192)	0.370* (0.192)	0.111 (0.073)	0.126 (0.080)
Bachelor		0.324 (0.518)		0.042 (0.360)		0.028 (0.223)		0.016 (0.105)
Masters		0.657 (1.427)		0.882 (0.558)		0.187 (0.435)		0.185 (0.247)
Works		0.710 (0.482)		0.538 (0.370)		0.303 (0.240)		0.042 (0.090)
Female		-1.112** (0.430)		-0.528 (0.332)		-0.271 (0.229)		-0.097 (0.087)
Age		-0.053 (0.065)		-0.059 (0.046)		-0.041 (0.030)		-0.023* (0.013)
Log of income		-0.139 (0.304)		-0.262 (0.254)		-0.135 (0.166)		-0.027 (0.055)
Financial literacy		0.075 (0.155)		0.160 (0.112)		-0.025 (0.082)		-0.014 (0.030)
Risk tolerance (choices)		0.014 (0.037)		0.012 (0.033)		0.044 (0.029)		-0.002 (0.008)
Risk tolerance (self-assessed)		0.070 (0.102)		0.126* (0.076)		-0.003 (0.042)		0.012 (0.019)
Self control		0.132 (0.180)		0.081 (0.139)		0.042 (0.076)		-0.001 (0.037)
Constant	3.501*** (0.259)	5.011** (2.265)	2.128*** (0.208)	3.619** (1.669)	0.489*** (0.119)	1.937* (1.041)	0.310*** (0.051)	1.042** (0.437)
R ²	0.02	0.12	0.02	0.13	0.02	0.11	0.01	0.06
Observations	172	165	172	165	172	165	172	165

Notes: The table shows results of OLS regressions of the debt level after market 1, 2, and at the end of the experiment on income expectations and overconfidence. Debt is the amount in Euro that participants spend of their maximum initial credit of €10. Income expectations is the expected income in Euro from both income tasks. Overconfidence is the difference between the expected income and the actual income. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

Table E.2. T-test comparing excluded participants with full sample

	Mean excluded sample	Mean used sample	t-statistic	p-values
Female	0.58	0.49	0.95	0.35
Age	24.79	22.65	2.66	0.01
Bachelor	0.33	0.18	2.04	0.04
Masters	0.12	0.06	1.34	0.18
Works	0.27	0.31	-0.48	0.64
Log of income	6.39	6.33	0.36	0.72
Financial literacy	4.03	5.27	-4.37	0.00
Risk tolerance (choices)	11.41	10.07	1.42	0.16
Risk tolerance (self-assessed)	5.09	4.75	0.84	0.40
Self control	-0.02	-0.03	0.05	0.96
<i>N</i>	33	252		

Notes: The table shows descriptive statistics for participants included into the final sample and participants excluded due to limited comprehension. T-statistics and p-values are reported for the differences between groups.

Table E.3 shows the income expectations for different sub-samples. The treatment effect on income expectations is about the same for the used sample and the full sample. The excluded sample shows a larger effect, which attains marginal significance for this very small sample ($n=20$). The excluded participants seem to fall for the reversed hard-easy effect more strongly. In this group, participants even believe to outperform in the random lottery condition (another sign for insufficient understanding). We confirm this general tendency by contrasting participants who answer all comprehension questions correctly with those who make at least one error.¹ Again, the treatment effect on income expectations is stronger for those with less comprehension.

We next test for the presence of the treatment effect in the sub-samples. Table E.4 shows the treatment effect in the used sample, the full sample, and for participants who answer all comprehension questions correctly. The treatment effect considerably weakens for the full

¹One error is relatively common ($n=115$), which is why we refrain from excluding this group but formulate the less strict exclusion restriction of two or more errors.

Table E.3. Participant comprehension and income expectations

Income expectations	n	Mean	Mean by treatment				p -value
			Hard	Easy	Control	Easy–Hard	
Used sample	252	6.12	5.61	6.67	6.04	1.06	<0.01
Full sample	285	6.17	5.61	6.72	6.19	1.10	<0.01
Excluded sample	33	6.64	5.67	7.25	7.15	1.58	0.096
All correct	137	5.88	5.39	6.39	5.76	1.00	<0.01
One or more errors	148	6.45	5.78	7.09	6.60	1.31	<0.01

Notes: The table shows income expectations for different sub-samples in the main experiment. The used sample is the sample for which results are reported in the main text, the full sample are all participants, the excluded sample contains participants with two or more errors in the comprehension test, “all correct” are participants who answer all comprehension questions correctly. The table reports the number of observations, the mean, the mean by treatment, the difference between the easy and the hard treatment, and the p -value of a two-sided t -test for a difference between the the easy and the hard treatment.

sample. We also find that the correlation between income expectations and consumption is negative for the excluded participants. This raises the concern that they misunderstood how the earned income can be used in the experiment. The results for the participants with the highest comprehension corroborates this assumption, as the treatment effect is stronger in this sub-sample.

A second impediment to the experiment might not be limited understanding, but preferences that contradict basic assumptions of the experimental design, i.e., a preference for goods over the cash reward. In this case, a positive debt level at the end of the experiment would not be unintended. We discuss this issue in subsection 4.4 when justifying the interpretation of final debt as over-indebtedness. A question in the experiment asks whether participants prefer the goods over the cash reward (the exact wording is: “Imagine you received an additional Euro, would you prefer to have it paid out in cash or to purchase additional goods at the prices of the first market stage?”). Panel A of Table E.5 shows the treatment effect for the sample excluding these participants. Results confirm that the treat-

Table E.4. Participant comprehension and the treatment effect

Panel A: Used sample	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	0.728** (0.366)	0.600** (0.297)	0.361* (0.192)	0.110 (0.072)
Random group	0.592* (0.345)	0.313 (0.305)	0.169 (0.194)	0.029 (0.075)
Observations	252	252	252	252
Panel B: Full sample	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	0.483 (0.357)	0.356 (0.296)	0.230 (0.194)	0.073 (0.071)
Random group	0.445 (0.339)	0.289 (0.314)	0.202 (0.210)	0.011 (0.070)
Observations	285	285	285	285
Panel C: All questions correct	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	1.245** (0.487)	0.918** (0.377)	0.563** (0.234)	0.095 (0.101)
Random group	0.898* (0.493)	0.538 (0.399)	0.361 (0.251)	0.061 (0.104)
Observations	137	137	137	137

Notes: The table shows results of regressions of debt levels during the experiment on treatment dummies for different sub-samples. Panel A shows results for the sample used in the experiment, Panel B shows results for the full sample, and Panel C shows results for participants who answered all comprehension questions correct. The regression specifications are as in Table 5. Coefficients are significant at $*p < .10$, $**p < .05$, $***p < .01$; robust standard errors are shown in parentheses.

ment effect is stronger among participants preferring cash. They are the relevant sample, as the experimental structure of borrowing and repaying debt relies on an incentive to avoid over-indebtedness (which is absent if one prefers the goods anyway).

Another way to look at the same issue is to exclude participants who already in market stage one spend more than the total income they expect to earn in the experiment. These

Table E.5. Other exclusion criteria and the treatment effect

Panel A: Preferring cash	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	0.942** (0.397)	0.717** (0.318)	0.409** (0.203)	0.132* (0.079)
Random group	0.633* (0.379)	0.335 (0.337)	0.223 (0.215)	0.021 (0.080)
Observations	215	215	215	215
Panel B: Consumption \leq income exp.	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	0.668** (0.325)	0.558** (0.271)	0.257 (0.173)	0.082 (0.077)
Random group	0.576* (0.305)	0.245 (0.268)	0.067 (0.152)	0.041 (0.079)
Observations	205	205	205	205
Panel C: Positive consumption	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy treatment	0.719** (0.352)	0.634** (0.303)	0.379* (0.205)	0.114 (0.077)
Random group	0.433 (0.333)	0.242 (0.312)	0.150 (0.204)	0.013 (0.077)
Observations	236	236	236	236

Notes: The table shows results of regressions of debt levels during the experiment on treatment dummies for different sub-samples. Panel A shows results excluding participants who prefer the goods over the cash reward, Panel B shows results excluding participants who consume more than their income expectations in the first market stage, and Panel C shows results excluding participants who do not consume in the first market stage. The regression specifications are as in Table 5. Coefficients are significant at $*p < .10$, $**p < .05$, $***p < .01$; robust standard errors are shown in parentheses.

participants seem not to worry about ending up in debt as they literally set themselves up for it (and 60% indeed end in debt). Interestingly, this group has little overlap with the one preferring goods over cash, even though this appears to be the only reasonable explanation for this behavior. As Panel B shows, the results are not stronger using this exclusion. Although

the excluded participants behave in a way that is at odds with consumption depending on income expectations, their overspending seems not to interfere with the treatment effect. Additional evidence for the effect of spending motives comes from the robustness experiment reported in section 5 and Online Appendix F, which removes overly cheap products as a borrowing motive.

The other extreme is participants who do not like the offered products at all. We aim at a menu of products that appeals to everyone, but some participants might still not find something they regard worthwhile to purchase.² As a proxy for such preferences, we exclude participants who do not spend anything in the first market. Participants know they will earn at least €2, which allows for some risk-less consumption if products have positive utility. As shown in Panel C of Table E.5, the treatment effect is robust against this exclusion.³

The results for income expectations and overconfidence are robust to the above reported exclusion criteria (comprehension and preferences). In particular, the effect of income expectations on borrowing behavior becomes stronger when removing participants with non-standard preferences.

E.2. Alternative overconfidence measures

Results reported in the paper mostly use a parsimonious measure of overconfidence based on participants' income expectations at the beginning of the experiment. We also consider alternative measures of overconfidence, such as overconfidence for income in the second round

²Some participants, e.g., complained that they do not eat sweets. However, they could still choose from the non-food items. Two participants did not buy any product (one each in the main experiment and the robustness experiment).

³Exclusions based on spending behavior are evenly distributed across treatments. Participants who state they prefer goods are more frequent in the hard treatment. Due to their lower consumption they might assign higher marginal utility to additional consumption.

(*overconfidence 2*). Participants are again asked for income expectations after the realization of income in the first round.

These two overconfidence measures are based on point estimates of expected performance, which have been criticized as not necessarily reflecting true overconfidence (Benoît and Dubra, 2011). Although Merkle and Weber (2011) show that this is rather a theoretical than empirical concern, we nevertheless use their method to elicit a probability distribution of quiz task performance. Participants are asked with what probability they believe to obtain each pair of ranks in their randomly assigned group of eight participants (we ask for pairs of ranks as the payoff is the same for ranks, see screenshot B.4). This allows to calculate an expected income based on probabilities. Unsurprisingly, it is strongly correlated with the income expectation elicited as a point estimate (Pearson correlation 0.61, $p < 0.001$).

Table E.6 shows correlations between socio-demographic variables and all measures of overconfidence. There are few significant correlations between the different measures of overconfidence and socio-demographic indicators. People with higher financial literacy seem to be less overconfident. Other than in the GSEOP we find no consistent relationship between gender and overconfidence. Importantly, experimental overconfidence is manipulated by the treatments which may weaken relationships between socio-demographics and overconfidence measures.

We further investigate whether our results are influenced by the measure of overconfidence used. Panel A of Table E.7 shows results for overconfidence for the income in the second income task (*overconfidence 2*). This measure is relevant only for the later stage of the experiment as it is elicited after the first income task and first feedback. The findings confirm

Table E.6. Correlates of Overconfidence

	(1) Overconfidence Market 1	(2) Overconfidence Market 2	(3) Overconfidence Market 1	(4) Overconfidence Prob. Market 2
Gender	0.050 (0.073)	-0.130* (0.071)	0.140* (0.073)	-0.093 (0.078)
Age	0.010 (0.011)	-0.001 (0.011)	0.011 (0.010)	-0.013 (0.011)
Bachelor	0.005 (0.097)	0.026 (0.096)	-0.158* (0.092)	0.096 (0.103)
Masters	-0.125 (0.165)	0.120 (0.176)	-0.084 (0.171)	0.205 (0.178)
Works	-0.032 (0.076)	0.027 (0.073)	-0.048 (0.075)	0.104 (0.082)
Log of income	0.034 (0.047)	-0.028 (0.045)	0.011 (0.050)	0.015 (0.050)
Financial literacy	-0.037 (0.024)	-0.058** (0.023)	-0.007 (0.024)	-0.028 (0.024)
Risk tolerance (choices)	0.006 (0.007)	0.001 (0.007)	-0.004 (0.007)	0.000 (0.007)
Risk tolerance (self-assessed)	-0.005 (0.016)	-0.016 (0.016)	0.004 (0.016)	-0.004 (0.018)
Self control	-0.002 (0.032)	-0.046 (0.030)	-0.021 (0.031)	-0.056 (0.036)
Constant	0.160 (0.346)	1.018*** (0.342)	0.226 (0.362)	0.829** (0.374)
R ²	0.03	0.05	0.04	0.04
Observations	241	241	241	207

Notes: The table shows results of OLS regressions of socio-demographics on four measures of overconfidence. Overconfidence 1 and 2 are expected income minus realized income in market stage 1 and 2, respectively. Overconfidence probability is based on the alternative overconfidence measure (see Section E.2) in market 1 and 2, respectively. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

the impact of overconfident income expectations on final debt levels and the likelihood to remain in debt.

Economically, the coefficients are even somewhat larger than for initial overconfidence (as displayed in Table 6). This finding could be interpreted in the sense that those who do not

Table E.7. Results for alternative overconfidence measures

Panel A: Overconfidence 2		Final debt		Has debt		
	(1)	(2)	(3)	(4)	(5)	(6)
Overconfidence 2	0.230*** (0.044)	0.225*** (0.044)	0.230*** (0.044)	0.104*** (0.014)	0.103*** (0.014)	0.110*** (0.015)
Easy treatment		0.300 (0.208)	0.292 (0.202)		0.074 (0.073)	0.073 (0.077)
Random group		0.134 (0.206)	0.162 (0.194)		0.025 (0.071)	0.025 (0.071)
Constant	0.739*** (0.090)	0.592*** (0.129)	0.637 (0.814)	0.376*** (0.030)	0.342*** (0.050)	0.431 (0.295)
R ²	0.107	0.115	0.185	0.179	0.183	0.216
Observations	218	218	214	218	218	214
Control variables	no	no	yes	no	no	yes
Panel B: Income (probabilites)		Debt market 1		Debt market 2		Final debt
		(1)	(2)	(3)	(4)	(5) (6)
Income expectations (probabilities)	0.862*** (0.265)	0.664** (0.285)	0.867*** (0.229)	0.832*** (0.241)	0.179 (0.163)	0.161 (0.148)
Easy treatment		0.281 (0.387)		0.059 (0.316)		0.287 (0.207)
Random group		0.477 (0.352)		0.107 (0.301)		0.172 (0.197)
Constant	1.306 (0.796)	1.059 (1.586)	-0.213 (0.691)	-1.412 (1.273)	0.121 (0.497)	0.100 (0.693)
R ²	0.048	0.143	0.067	0.164	0.007	0.076
Observations	252	248	252	248	252	248
Control variables	no	yes	no	yes	no	yes
Panel C: Overconfidence (probabilites)		Debt market 1		Debt market 2		Final debt
		(1)	(2)	(3)	(4)	(5) (6)
Overconfidence (probabilities)	0.171* (0.093)	0.195** (0.087)	0.428*** (0.073)	0.445*** (0.069)	0.237*** (0.051)	0.242*** (0.052)
Easy treatment		0.452 (0.360)		0.161 (0.281)		0.220 (0.179)
Random group		0.584* (0.334)		0.186 (0.268)		0.145 (0.177)
Constant	3.932*** (0.144)	3.137** (1.410)	2.409*** (0.114)	1.507 (1.052)	0.653*** (0.077)	0.910 (0.677)
R ²	0.016	0.140	0.132	0.256	0.093	0.169
Observations	252	248	252	248	252	248
Control variables	no	yes	no	yes	no	yes

Notes: The table shows in Panel A results of OLS regressions of the debt level after the final market (columns (1) to (3)) and propensity to remain in debt (columns (4) to (6)) on income expectations, overconfidence for the second income task, and control variables. Panels B and C show results of OLS regressions of the debt level after market stage 1, 2 and at the end of the experiment as the dependent variable. Panel B includes income expectations and Panel C includes overconfidence which are based on elicited income probabilities. Treatment indicators and control variables are as defined in Tables 2 and 6. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

learn from feedback are in particular trouble. However, also the levels of income expectations (in €) are smaller, as they refer only to the remaining income task. As a consequence, the level of *overconfidence 2* is as well smaller than the level of initial overconfidence, which (more than) offsets the effect of the larger coefficients.

Results displayed in Panel B of Table E.7 are based on the alternative measure of income expectations. We confirm the importance of income expectations for early borrowing, which is reduced when approaching the end of the experiment. By subtracting the actual income of participants from their expectation one obtains an alternative measure of overconfidence based on the probability information. The correlation of this variable with initial overconfidence based on point estimates is 0.69. Unsurprisingly, results for borrowing behavior using this alternative overconfidence variable reflect the importance of overconfidence in the later stage of the experiment (see Panel C). Different than in the results presented in Table 6, overconfidence based on probabilities is already significant for initial debt levels.

E.3. The intensive margin of debt levels

When we study debt levels in the main experiment, we include participants with no debt. To zoom in on the intensive margin of debt taking, we provide robustness results excluding participants who do not take debt at all (the extensive margin was already analyzed using the indicator “has debt”). Table E.8 shows the treatment effect under this restriction. Results hold for the level of debt and coefficients are economically comparable to the regressions including participants not taking debt. As many participants repay their debt before the end of the experiment the number of observations is low for final debt (column (3)). Therefore, the coefficient does not attain significance in this specification.

We next examine the influence of income expectations and overconfidence on debt levels (see Table E.9). Income expectations have a strong impact on initial debt levels. This relationship is not 1:1, though, as an additional Euro of income expectations results in an

Table E.8. Intensive margin of debt and the treatment effect

	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)
Easy treatment	0.719** (0.352)	0.623** (0.300)	0.443 (0.383)
Random group	0.433 (0.333)	0.259 (0.312)	0.369 (0.443)
Constant	3.820*** (0.253)	2.416*** (0.215)	1.578*** (0.286)
R ²	0.019	0.019	0.014
Observations	236	226	90

Notes: The table shows results of regressions of debt levels during the experiment on treatment indicators. Participants without debt are excluded. The regression specifications are as in Table 5. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

additional 30 cents of debt, when looking at the OLS regression. The influence of income expectations levels off in later stages of the experiment as already observed in the main regressions in Table 6. The coefficients for the intensive margin are slightly higher than in these regressions. An interpretation would be that income expectations are less relevant for the decision to borrow or not (as also the relatively weak results for the debt indicator “has debt” show), but more relevant for the amount of debt taken. Income expectations determine how much debt a participant can subjectively afford.⁴

⁴In real financial markets the supply side contributes to this effect by limiting debt levels to individual debt capacity. In the experiment, we abstract from this as all participants receive the same credit line.

Table E.9. Intensive margin of debt and mechanisms

Panel A: Income expectations	Debt market 1		Debt market 2		Final debt	
	(1)	(2)	(3)	(4)	(5)	(6)
Expectations	0.327*** (0.094)	0.729** (0.365)	0.261*** (0.089)	0.680* (0.348)	0.203 (0.135)	0.348 (0.286)
Constant	2.181*** (0.567)	-0.292 (2.244)	1.117** (0.526)	-1.483 (2.176)	0.561 (0.828)	-0.351 (1.769)
R2	0.07	.	0.07	.	0.06	0.03
Observations	159	159	153	153	63	63
1st stage F-stat.		11.71		9.48		7.53
Panel B: Overconfidence	Debt market 1		Debt market 2		Final debt	
	(1)	(2)	(3)	(4)	(5)	(6)
Overconfidence	0.085 (0.059)	0.024 (0.066)	0.115** (0.053)	0.047 (0.062)	0.189*** (0.042)	0.209*** (0.048)
Constant	4.184*** (0.176)	4.177*** (0.177)	2.714*** (0.146)	2.715*** (0.149)	0.718*** (0.093)	0.695*** (0.092)
R2	0.01	0.00	0.03	0.00	0.17	0.18
Observations	159	154	153	147	153	147

Notes: The table shows results of OLS and IV regressions of the debt level after market stage 1, 2 and at the end of the experiment as the dependent variable (excluding participants with no debt). Panel A includes income expectations (instrumented using treatment assignment as an instrument) and Panel B includes overconfidence as independent variable (using the method described in the main part of this paper for the IV estimation). Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

The results for overconfidence provided in Panel B are strong and significant for the later debt markets, even though these regressions have few observations. The coefficients are higher than in the main regressions. In contrast to income expectations, overconfidence was also strongly significant for the extensive margin (see Table 6). This implies that overconfidence not only increases the risk to become over-indebted, but also has a positive effect on the level of indebtedness.

Online Appendix F. Robustness Experiment

Table F.1 contains descriptive statistics on participants in the robustness experiment. As participants are recruited from the same subject pool, we do not expect systematic differences. The demographic information displayed in the table confirms this expectations.⁵ The differences between the participants in the robustness experiment and the main experiment are insignificant except for marginally higher income. Between treatment differences are also insignificant except for risk tolerance based on the lottery choices.

⁵Due to technical difficulties, in one session not all demographic data was saved (gender, age, income, and self-assessed risk tolerance). This explains the lower number of observations for these variables.

Table F.1. Demographic and control variables in robustness sample

			t-test	Mean by treatment			F-test
	<i>n</i>	Mean	<i>p</i> -value	Hard	Easy	Control	<i>p</i> -value
Gender (female=1)	197	0.47	0.50	0.51	0.46	0.42	0.54
Age	197	24.7	0.96	24.9	24.6	24.7	0.96
Bachelor degree	219	0.23	0.44	0.24	0.20	0.25	0.73
Masters degree	219	0.10	0.12	0.13	0.08	0.12	0.55
Works	219	0.33	0.63	0.31	0.38	0.29	0.52
Income	194	776	0.07	817	766	734	0.50
Financial literacy	219	5.30	0.23	5.33	5.30	5.24	0.94
Risk tolerance (self-assessed)	197	4.86	0.72	5.04	4.82	4.68	0.59
Risk tolerance (choices)	213	10.39	0.72	11.95	9.35	9.76	<0.01
Self control	219	0.04	0.43	0.11	0.08	-0.10	0.40

Notes: The table shows means of demographic variables and controls for the participants in the robustness sample. All variables are also shown by treatment group. The variables are as defined in Table 2. Different numbers of observation reflect non-responses and the fact that due to technical difficulties, in one session not all demographic data was saved. The *p*-values of t-tests comparing the robustness sample to the main experiment, and the *p*-values of one-way ANOVA F-tests for between group differences are reported.

We are unable to perform IV regressions using the treatment manipulation as an instrument for overconfidence because the treatment manipulation has not significantly shifted income expectations in the robustness experiment. Therefore, we use a different approach and perform OLS regressions with varying control variables (also including treatment assignment) to examine the potential mechanism behind the results shown in the main part of the paper.

Table F.2 shows results for the debt levels after market one in the robustness test for different sets of control variables. For initial debt there only is an effect of income expectations, as already reported in Table 9. In the specifications with full controls, part of this effect is picked up by the treatment effect. The design is not as clean as an IV regression, as exogenous and endogenous effects of income expectations are not clearly separated. In columns (1) to (4), the total effect of income expectations and overconfidence is examined, while columns (5) and (6) control for exogenous treatment effects. Table F.3 shows results for the debt levels after market two in the robustness test for different sets of control variables. A strong positive effect of overconfidence on borrowing is visible in these regressions (in line with Table 9). We also observe a marginally significant treatment effect. The effect of income expectations is smaller compared to initial debt taking. Tables F.4 and F.5 show results for the final debt level and the propensity to remain in debt at the end of the experiment. The results confirm the important role of overconfidence for over-indebtedness. In the robustness experiment, there is no strong role for the treatment effect.

We finally show that the results of Tables 5 and 6 generally hold, if we pool the main experiment and the robustness experiment. The corresponding tables are Tables F.6 and

Table F.2. Robustness experiment: Indebtedness after market stage 1

	Debt level after market stage 1					
	(1)	(2)	(3)	(4)	(5)	(6)
Income expectations	0.258** (0.112)		0.226** (0.113)		0.186 (0.124)	
Overconfidence		0.088 (0.071)		0.119 (0.073)		0.102 (0.076)
Easy treatment					0.668 (0.447)	0.747* (0.426)
Random group					0.061 (0.424)	0.068 (0.425)
Gender			-0.256 (0.351)	-0.442 (0.346)	-0.248 (0.344)	-0.401 (0.338)
Age			0.082 (0.056)	0.079 (0.059)	0.081 (0.054)	0.078 (0.056)
Bachelor			0.355 (0.435)	0.437 (0.438)	0.424 (0.429)	0.496 (0.428)
Masters			-0.383 (1.059)	-0.310 (1.099)	-0.252 (1.031)	-0.184 (1.064)
Works			-0.240 (0.386)	-0.333 (0.380)	-0.334 (0.391)	-0.418 (0.378)
Log of income			-0.218 (0.194)	-0.193 (0.192)	-0.186 (0.190)	-0.162 (0.187)
Financial literacy			0.131 (0.119)	0.187 (0.120)	0.137 (0.119)	0.186 (0.118)
Risk tolerance (self-assessed)			-0.026 (0.121)	-0.011 (0.116)	-0.020 (0.120)	-0.009 (0.115)
Risk tolerance (choices)			0.067 (0.045)	0.073 (0.044)	0.078* (0.045)	0.084* (0.044)
Self-control			0.009 (0.213)	0.026 (0.211)	-0.008 (0.216)	0.006 (0.213)
Constant	2.322*** (0.660)	3.866*** (0.169)	0.768 (1.660)	1.692 (1.640)	0.382 (1.632)	1.074 (1.663)
R ²	0.037	0.011	0.106	0.095	0.122	0.117
Observations	193	193	171	171	171	171

Notes: The table shows results of OLS regressions of the debt level after market one on income expectations, overconfidence, and control variables. Income expectations is the expected income in Euro for the two income tasks. Overconfidence is the difference between the expected income and the actual income. Easy treatment and control group are indicator variables for the respective treatment (with the hard treatment as the omitted category). All control variables are as defined in Table F.1. Ln income is the natural logarithm of participants external income. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

Table F.3. Robustness experiment: Indebtedness after market stage 2

	Debt level after market stage 2					
	(1)	(2)	(3)	(4)	(5)	(6)
Income expectations	0.252*** (0.094)		0.205** (0.089)		0.161 (0.098)	
Overconfidence		0.237*** (0.058)		0.226*** (0.060)		0.209*** (0.061)
Easy treatment					0.613* (0.368)	0.602* (0.329)
Random group					−0.264 (0.338)	−0.261 (0.325)
Gender			−0.098 (0.290)	−0.333 (0.268)	−0.107 (0.283)	−0.311 (0.261)
Age			0.057 (0.042)	0.052 (0.046)	0.056 (0.039)	0.051 (0.042)
Bachelor			0.213 (0.374)	0.264 (0.368)	0.293 (0.357)	0.326 (0.351)
Masters			−0.176 (0.874)	−0.228 (0.857)	−0.052 (0.880)	−0.126 (0.851)
Works			−0.019 (0.298)	−0.073 (0.292)	−0.145 (0.293)	−0.178 (0.286)
Log of income			−0.274* (0.154)	−0.243* (0.146)	−0.239 (0.155)	−0.212 (0.147)
Financial literacy			−0.083 (0.095)	0.032 (0.094)	−0.080 (0.094)	0.028 (0.094)
Risk tolerance (self-assessed)			0.080 (0.098)	0.048 (0.093)	0.090 (0.095)	0.053 (0.091)
Risk tolerance (choices)			0.033 (0.036)	0.038 (0.033)	0.041 (0.036)	0.044 (0.033)
Self-control			−0.058 (0.149)	0.006 (0.145)	−0.077 (0.150)	−0.015 (0.147)
Constant	0.771 (0.555)	2.291*** (0.133)	1.099 (1.167)	1.849 (1.152)	0.877 (1.152)	1.451 (1.194)
R ²	0.051	0.119	0.099	0.159	0.136	0.196
Observations	193	193	171	171	171	171

Notes: The table shows results of OLS regressions of the debt level after market two on income expectations, overconfidence, and control variables. All variables are as defined in Tables F.1 and F.2. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

Table F.4. Robustness experiment: Indebtedness at the end of the experiment

	Debt level at the end of the experiment					
	(1)	(2)	(3)	(4)	(5)	(6)
Income expectations	0.125 (0.081)		0.096 (0.066)		0.083 (0.073)	
Overconfidence		0.272*** (0.051)		0.253*** (0.052)		0.250*** (0.052)
Easy treatment					0.100 (0.299)	0.002 (0.231)
Random group					−0.302 (0.254)	−0.305 (0.229)
Gender			0.204 (0.212)	0.011 (0.193)	0.189 (0.214)	0.000 (0.196)
Age			0.035 (0.030)	0.030 (0.033)	0.035 (0.029)	0.030 (0.032)
Bachelor			0.168 (0.301)	0.162 (0.278)	0.195 (0.302)	0.176 (0.278)
Masters			0.251 (0.661)	0.079 (0.624)	0.275 (0.644)	0.080 (0.603)
Works			−0.177 (0.225)	−0.165 (0.202)	−0.230 (0.227)	−0.200 (0.201)
Log of income			−0.191 (0.124)	−0.164 (0.108)	−0.180 (0.131)	−0.159 (0.113)
Financial literacy			−0.096 (0.074)	0.038 (0.062)	−0.097 (0.074)	0.035 (0.064)
Risk tolerance (self-assessed)			0.098 (0.083)	0.026 (0.071)	0.103 (0.084)	0.029 (0.071)
Risk tolerance (choices)			−0.012 (0.023)	−0.010 (0.021)	−0.012 (0.024)	−0.012 (0.022)
Self-control			−0.086 (0.108)	0.003 (0.102)	−0.093 (0.111)	−0.001 (0.104)
Constant	−0.027 (0.450)	0.740*** (0.101)	0.533 (0.754)	0.774 (0.803)	0.604 (0.816)	0.870 (0.892)
R ²	0.019	0.240	0.083	0.272	0.096	0.280
Observations	193	193	171	171	171	171

Notes: The table shows results of OLS regressions of the debt level after the final market on income expectations, overconfidence, and control variables. All variables are as defined in Tables F.1 and F.2. Coefficients are significant at $*p < .10$, $**p < .05$, $***p < .01$; robust standard errors are shown in parentheses.

Table F.5. Robustness test: Propensity to remain in debt at the end of the experiment

	Indicator whether in debt at the end of the experiment					
	(1)	(2)	(3)	(4)	(5)	(6)
Income expectations	0.009 (0.019)		0.007 (0.019)		0.002 (0.021)	
Overconfidence		0.077*** (0.010)		0.075*** (0.012)		0.074*** (0.011)
Easy treatment					0.070 (0.088)	0.023 (0.071)
Random group					−0.035 (0.089)	−0.038 (0.079)
Gender			0.130* (0.074)	0.084 (0.064)	0.128* (0.074)	0.084 (0.065)
Age			0.020*** (0.007)	0.018*** (0.006)	0.020*** (0.007)	0.018*** (0.006)
Bachelor			0.053 (0.085)	0.041 (0.076)	0.062 (0.085)	0.045 (0.076)
Masters			−0.006 (0.170)	−0.076 (0.136)	0.008 (0.175)	−0.072 (0.138)
Works			−0.069 (0.074)	−0.054 (0.064)	−0.084 (0.075)	−0.061 (0.065)
Log of income			−0.082** (0.039)	−0.076** (0.036)	−0.078** (0.039)	−0.074** (0.037)
Financial literacy			−0.022 (0.024)	0.019 (0.022)	−0.021 (0.024)	0.019 (0.023)
Risk tolerance (self-assessed)			−0.011 (0.020)	−0.038** (0.019)	−0.009 (0.021)	−0.037* (0.019)
Risk tolerance (choices)			0.013* (0.008)	0.013* (0.008)	0.014* (0.008)	0.013* (0.007)
Self-control			−0.048 (0.037)	−0.019 (0.036)	−0.050 (0.037)	−0.020 (0.037)
Constant	0.258** (0.118)	0.317*** (0.030)	0.281 (0.266)	0.256 (0.243)	0.258 (0.265)	0.249 (0.247)
R ²	0.001	0.217	0.110	0.276	0.118	0.279
Observations	193	193	171	171	171	171

Notes: The table shows results of a linear probability model (OLS) of an indicator variable whether a participants remains in debt at the end of the experiment. Explanatory variables are income expectations, overconfidence, and control variables. All variables are as defined in Tables F.1 and F.2. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

F.7. Treatment effects are somewhat weaker in the pooled results. However, OLS results are strong for all debt measures.

Table F.6. Debt taking and treatment effect (pooled results)

	Debt after market 1 (1)	Debt after market 2 (2)	Final debt (3)	Has debt (4)
Easy	0.563** (0.271)	0.521** (0.225)	0.217 (0.168)	0.069 (0.054)
Control	0.157 (0.262)	-0.013 (0.226)	-0.097 (0.155)	-0.014 (0.055)
Constant	3.657*** (0.191)	2.181*** (0.161)	0.639*** (0.113)	0.316*** (0.037)
R ²	0.01	0.02	0.01	0.01
Observations	445	445	445	445

Notes: The table shows results of regressions of debt taking variables on treatment indicators as described in Table 5. Observations are pooled from the main experiment and the robustness experiment. Coefficients are significant at $*p < .10$, $**p < .05$, $***p < .01$; robust standard errors are shown in parentheses.

Table F.7. Indebtedness over the course of the experiment (OLS, pooled results)

	Debt after market 1		Debt after market 2		Final debt		Has debt	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Expectations	0.241*** (0.070)		0.243*** (0.060)		0.111*** (0.042)		0.018 (0.014)	
Overconfidence		0.098** (0.040)		0.198*** (0.034)		0.216*** (0.028)		0.076*** (0.007)
Gender	-0.574** (0.230)	-0.696*** (0.231)	-0.291 (0.186)	-0.438** (0.180)	-0.098 (0.134)	-0.197 (0.122)	0.016 (0.050)	-0.011 (0.043)
Age	0.035 (0.037)	0.034 (0.038)	0.029 (0.030)	0.027 (0.031)	0.022 (0.020)	0.019 (0.020)	0.007 (0.006)	0.005 (0.006)
Bachelor	0.068 (0.285)	0.129 (0.288)	-0.031 (0.240)	0.023 (0.238)	-0.024 (0.184)	-0.007 (0.170)	-0.017 (0.062)	-0.017 (0.055)
Masters	0.120 (0.734)	0.154 (0.750)	0.324 (0.588)	0.323 (0.577)	0.381 (0.430)	0.337 (0.392)	0.097 (0.125)	0.076 (0.108)
Works	-0.055 (0.250)	-0.050 (0.255)	-0.043 (0.205)	-0.045 (0.200)	-0.027 (0.150)	-0.038 (0.132)	-0.053 (0.053)	-0.058 (0.046)
Log of income	-0.096 (0.187)	-0.081 (0.193)	-0.251 (0.156)	-0.235 (0.156)	-0.198* (0.113)	-0.190* (0.098)	-0.025 (0.037)	-0.023 (0.033)
Financial literacy	0.144* (0.080)	0.160* (0.082)	0.095 (0.064)	0.154** (0.065)	-0.015 (0.050)	0.068 (0.047)	-0.007 (0.017)	0.024 (0.016)
Risk tolerance (choices)	0.042* (0.025)	0.051** (0.025)	0.022 (0.022)	0.028 (0.021)	0.018 (0.018)	0.017 (0.015)	0.007 (0.005)	0.005 (0.005)
Risk tolerance (self-assesed)	0.029 (0.067)	0.056 (0.067)	0.094* (0.053)	0.104** (0.052)	0.013 (0.039)	-0.005 (0.035)	-0.003 (0.013)	-0.014 (0.011)
Self control	0.096 (0.120)	0.087 (0.118)	-0.021 (0.092)	-0.006 (0.089)	-0.085 (0.059)	-0.046 (0.054)	-0.030 (0.023)	-0.013 (0.021)
Constant	1.214 (1.268)	2.321* (1.319)	0.764 (0.989)	1.824* (1.026)	0.597 (0.635)	1.010 (0.637)	0.217 (0.255)	0.263 (0.223)
R ²	0.11	0.09	0.11	0.15	0.05	0.24	0.02	0.23
Observations	406	406	406	406	406	406	406	406

Notes: The table shows results of OLS regressions of the debt level after market 1, 2, and at the end of the experiment on income expectations and overconfidence. Debt is the amount in Euro that participants spend of their maximum initial credit of €10. Income expectations is the expected income in Euro from both income tasks. Overconfidence is the difference between the expected income and the actual income. Coefficients are significant at * $p < .10$, ** $p < .05$, *** $p < .01$; robust standard errors are shown in parentheses.

Appendix References

- Benoît, Jean-Pierre, and Juan Dubra. 2011. “Apparent Overconfidence.” *Econometrica*, 79(5): 1591–1625.
- Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2016. “The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences.” IZA Discussion Paper Series No. 9674.
- Horneber, Sophie. 2017. “CAPI Questionnaire: Inno Gesamtfragebogen 2017.” GSOEP-IS Codeplan 2017.
- Merkle, Christoph, and Martin Weber. 2011. “True Overconfidence: The Inability of Rational Information Processing to Account for Apparent Overconfidence.” *Organizational Behavior and Human Decision Processes*, 116(2): 262–271.
- Moore, Don A., and Paul J. Healy. 2008. “The Trouble with Overconfidence.” *Psychological Review*, 115(2): 502–517.
- Richter, David, and Jürgen Schupp. 2012. “SOEP Innovation Sample (SOEP-IS)-Description, structure and documentation.” SOEP Papers on Multidisciplinary Panel Data Research No. 463.
- Tangney, June P., Roy F. Baumeister, and Angie L. Boone. 2004. “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success.” *Journal of Personality*, 72(2): 271–324.