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

Earn More Tomorrow: Overconfidence, Income Expectations, and Consumer Indebtedness

Journal of Money, Credit and Banking

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Grohmann, Antonia; Menkhoff, Lukas; Merkle, Christoph; Schmacker, Renke (2025) : Earn More Tomorrow: Overconfidence, Income Expectations, and Consumer Indebtedness, Journal of Money, Credit and Banking, ISSN 1538-4616, Wiley, Hoboken, NJ, Vol. 57, Iss. 5, pp. 1071-1102,
https://doi.org/10.1111/jmcb.13157

DOI: 10.1111/jmcb.13157

ANTONIA GROHMANN
LUKAS MENKHOFF
CHRISTOPH MERKLE
RENKE SCHMACKER

Earn More Tomorrow: Overconfidence, Income Expectations, and Consumer Indebtedness

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 (German Socio-Economic Panel–Innovation Sample [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

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.

Antonia Grohmann *is Associate Professor of Finance, Department of Economics and Business Economics, Aarhus University, and Research Fellow, Danish Finance Institute* (*E-mail*: agrohmann@econ.au.dk). Lukas Menkhoff *is Senior Research Associate, DIW Berlin, and Professor of Economics, School of Business and Economics, Humboldt-Universität zu Berlin, and IfW Kiel* (*E-mail*: lmenkhoff@diw.de). Christoph Merkle *is Associate Professor of Finance, Department of Economics and Business Economics, Aarhus University, and Research Fellow, Danish Finance Institute* (*E-mail*: cmerkle@econ.au.dk). Renke Schmacker *is Postdoctoral Researcher in Economics, WZB Berlin, and Research Associate, Public Economics Department, DIW Berlin* (*E-mail*: renke.schmacker@wzb.eu).

Received January 30, 2020; and accepted in revised form December 14, 2023.

Journal of Money, Credit and Banking, Vol. 57, No. 5 (August 2025)

consequences. In a robustness experiment, we rule out that overborrowing 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 codes: D14, D84, G40, G51
Keywords: debt, consumption, borrowing, overconfidence, income expectations

Consumer indebtedness is a core economic issue in modern societies. There are millions of individuals in every major OECD country who are considered overindebted. Their share is estimated to be between 5% and 10% in most developed countries; while in Germany, 10% of the adult population is considered over-indebted (Creditreform 2018), the United States 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 overborrowing 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. We study overconfidence in the form of overplacement, which refers to people's perception that they are better than others (following the terminology of Moore and Healy (2008), who identify three types of overconfidence). If people are overconfident about their relative ability, they may form overly high expectations about future income (Smith and Powell 1990, Reuben, Wiswall, and Zafar 2017, Buser et al. 2020). 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, overindebtedness 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).

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

We first analyze evidence for the relationship between general overconfidence, debt taking, and overindebtedness in a representative sample of German households (German Socio-Economic Panel [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 overplace 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 provide evidence of correlations.

In order to examine the *causality* of the relationship between overconfidence, income expectations, and debt taking, we conduct a laboratory experiment, in which we exogenously vary the expectations 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 overplace their relative performance in easy tasks and to underplace 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 overindebted, as most participants would prefer to avoid this cash penalty. This means that overconfident income expectations predict overindebtedness in the experiment. In the second and third experimental markets, 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 inflated income expectation 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 these low prices with higher prices that reflect real-world 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 overindebtedness. 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 suboptimal borrowing behavior and overindebtedness (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. Hyytinen 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 1 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 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 overtrading 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 overborrowing, they find an association between indebtedness in the experiment and "real life" debt indicators.

1. 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 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.

1.1 Overconfidence, Borrowing Behavior, and Repayment

We use two measures of overconfidence (overplacement) 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 overplacement, 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. 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 ($N = 87$).

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 ($N = 807$). Participants also report whether they subjectively experience debt payments (installments and interest) as a problem. This feeling might be associated with overindebtedness, 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 ($N = 87$).[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. 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 sizable 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

2. The 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, overindebtedness, or repayment ability.

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

4. Sample 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.

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**		0.089**		−0.002	
	(0.045)		(0.039)		(0.056)	
Overconfidence (animals)		0.058		−0.021		−0.058
		(0.037)		(0.032)		(0.046)
Gender	0.022	0.024	0.004	−0.002	−0.027	−0.033
	(0.026)	(0.025)	(0.021)	(0.021)	(0.036)	(0.035)
Age	−0.001	−0.001*	0.001**	0.001**	−0.003**	−0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Education	0.001	0.001	−0.013	−0.010	0.008	0.010
	(0.008)	(0.008)	(0.008)	(0.007)	(0.012)	(0.012)
Works	0.020	0.025	−0.045**	−0.045**	−0.110**	−0.118**
	(0.031)	(0.031)	(0.023)	(0.022)	(0.048)	(0.048)
Log of income	0.006	0.002	0.042*	0.036	−0.016	−0.014
	(0.025)	(0.025)	(0.024)	(0.024)	(0.046)	(0.046)
Financial literacy	−0.012	−0.013	0.000	0.000	−0.051***	−0.051***
	(0.010)	(0.010)	(0.008)	(0.008)	(0.017)	(0.017)
Risk tolerance (self-assessed)	0.010	0.010	−0.000	−0.001	0.001	0.002
	(0.006)	(0.006)	(0.005)	(0.005)	(0.007)	(0.008)
Risk tolerance (choices)	0.000	−0.000	−0.000	−0.000	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Self-control	−0.012	−0.012	0.014	0.014	0.030	0.030
	(0.017)	(0.017)	(0.014)	(0.014)	(0.024)	(0.024)
Constant	0.131	0.166	0.544***	0.613***	0.668*	0.668*
	(0.218)	(0.221)	(0.193)	(0.189)	(0.398)	(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

NOTE: 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$; robust standard errors are shown in parentheses.

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 perceive their debt payments as problematic.

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.

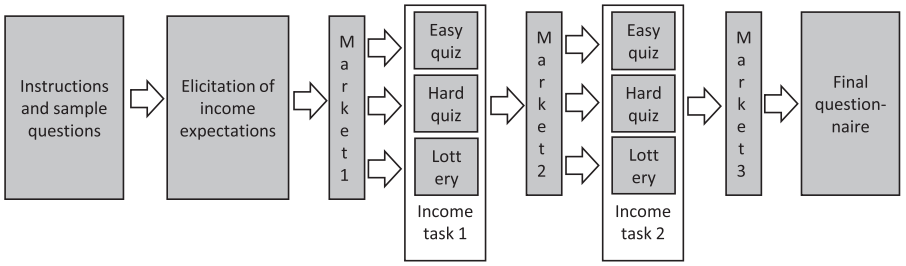

Fig 1. Sequence of the Experiment.

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.

2. 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.

2.1 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 10 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 5 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 overplace their relative position in easy tasks and underplace it in hard tasks. In the psychological literature, differences in beliefs between groups solving easy and hard tasks 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. While we operationalize overconfidence in the form of overplacement, Lawson, Larrick, and Soll (2023) find that all three forms of overconfidence share a common dispositional component.

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

5. In 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.

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.

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. As income in the experiment deterministically depends on the rank in the quiz task, income expectations directly reflect beliefs about placement in the task. A question in the comprehension check addresses whether people understand this relationship. 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]

2.2 Market for Consumption Goods

All three markets have the same structure. A total of 10 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 50% of the retail price in the first market, at about retail price in the second market, and 30% above the retail price 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.

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.

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

It follows from the design that participants can overborrow. 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 overborrowing 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 3.4 and 4).

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 4.

2.3 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).

2.4 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

7. In 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.

8. Laboratory 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 =

experiment without 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 Randomized Control Trial (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 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.

3. RESULTS

3.1 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

€8). To make this clear to participants, the two parts of the show-up fee are designated as "show-up fee" and "participation fee."

9. The 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, it did not reach the targeted number of participants. In this case, we used a smaller random group.

10. As 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.

TABLE 2

DEMOGRAPHIC AND CONTROL VARIABLES

			Mean by treatment			*F*-test
	n	Mean	Hard	Easy	Control	*p*-value
Gender (female = 1)	252	0.49	0.49	0.45	0.53	0.66
Age	252	22.65	21.70	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)	252	10.1	9.63	11.09	9.35	0.04
Risk tolerance (self-assessed)	252	4.75	4.55	5.22	4.45	0.03
Self-control	252	−0.03	−0.18	0.04	0.04	0.30

NOTE: 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 nonresponses. The *p*-values of one-way ANOVA *F*-tests for between group differences are reported.

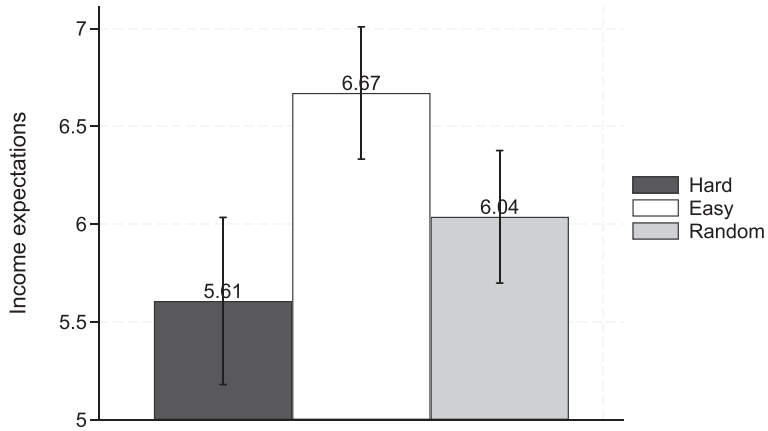

Fig 2. Income Expectations by Treatment before Market 1.

NOTES: Average income expectations in euro for both income tasks by treatment (possible range is between 2 and 10 euros). The 95% confidence interval is indicated.

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 ($p < 0.001$). Participants in the random group know that they are

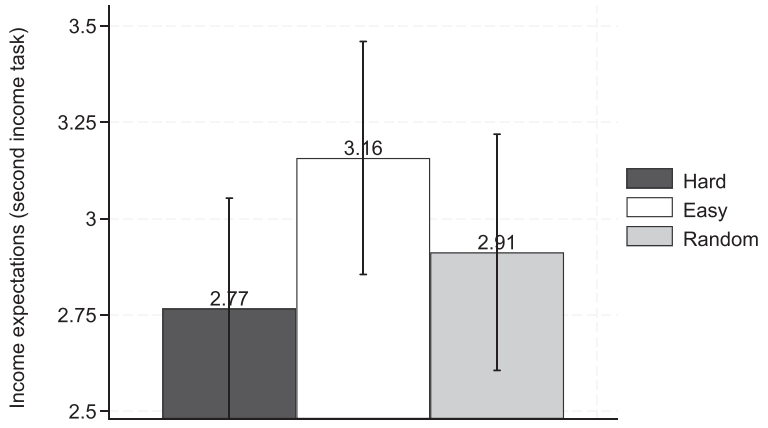

Fig 3. Income Expectations by Treatment before Market 2.

NOTES: Average income expectations in euro for the second income task by treatment group (possible range is between 1 and 5 euros). The 95% confidence interval is indicated.

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.

3.2 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

TABLE 3

OVERCONFIDENT INCOME EXPECTATIONS

Panel A: Overconfidence	*n*	Mean	Std. dev.	5p	95p	*p*-value
All participants	252	0.03	3.05	−5	5	0.89
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	0.79
Easy − hard		1.19				<0.01
Panel B: Overconfidence 2	***n***	**Mean**	**Std. dev.**	**5p**	**95p**	***p*-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	−0.19	2.24	−4	3	0.49
Easy − hard		0.31				0.29

NOTE: 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 nonresponses. 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.

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 expect too high 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 de-

sign, 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 Section 3.4, we will further investigate the consequences of overconfidence beyond the treatment effect.

3.3 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 2 and 3. There is some new borrowing in market 2 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 3 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 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 1. Coefficients have a natural interpretation in terms of euro. Participants in the easy treatment spend 73

11. The 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.

TABLE 4

BORROWING BEHAVIOR

Panel A: All participants	*n*	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	*n*	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	*n*	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	*n*	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

NOTE: 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**	0.600**	0.361*	0.111
	(0.366)	(0.297)	(0.192)	(0.073)
Random group	0.592*	0.313	0.169	0.028
	(0.345)	(0.305)	(0.194)	(0.074)
Constant	3.501***	2.128***	0.489***	0.310***
	(0.259)	(0.208)	(0.119)	(0.051)
R^2	0.019	0.016	0.013	0.010
Observations	252	252	252	252

NOTE: 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$; robust standard errors are shown in parentheses.

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 2: 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 (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 effect on the extensive margin of indebtedness, however, is not statistically significant.

To sum up, 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.

3.4 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

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	0.240**		0.277***		0.113**		0.029	
	(0.093)		(0.082)		(0.053)		(0.019)	
Overconfidence		0.077		0.176***		0.184***		0.075***
		(0.047)		(0.040)		(0.029)		(0.009)
Female	−0.821***	−0.917***	−0.388	−0.506**	−0.294*	−0.351**	−0.096	−0.113*
	(0.314)	(0.314)	(0.246)	(0.242)	(0.175)	(0.156)	(0.068)	(0.058)
Age	−0.002	0.004	0.008	0.014	0.002	0.002	−0.006	−0.006
	(0.046)	(0.045)	(0.041)	(0.039)	(0.025)	(0.023)	(0.009)	(0.008)
Bachelor	0.027	0.076	−0.296	−0.230	−0.189	−0.150	−0.053	−0.040
	(0.385)	(0.386)	(0.310)	(0.303)	(0.202)	(0.183)	(0.087)	(0.077)
Masters	0.373	0.357	0.627	0.603	0.467	0.451	0.207	0.201
	(0.919)	(0.919)	(0.736)	(0.716)	(0.418)	(0.386)	(0.164)	(0.152)
Works	0.062	0.157	−0.191	−0.108	0.011	0.010	−0.051	−0.059
	(0.324)	(0.333)	(0.263)	(0.260)	(0.180)	(0.158)	(0.070)	(0.061)
Log of income	−0.102	−0.096	−0.102	−0.131	−0.051	−0.107	0.025	−0.000
	(0.141)	(0.145)	(0.102)	(0.103)	(0.076)	(0.074)	(0.024)	(0.022)
Financial literacy	0.193*	0.186*	0.217***	0.239***	0.039	0.086	−0.009	0.012
	(0.111)	(0.112)	(0.083)	(0.085)	(0.061)	(0.061)	(0.022)	(0.020)
Risk tolerance	0.015	0.027	0.013	0.022	0.033	0.030	0.001	−0.001
(choices)	(0.031)	(0.031)	(0.029)	(0.027)	(0.025)	(0.021)	(0.007)	(0.006)
Risk tolerance	0.085	0.114	0.105*	0.134**	−0.033	−0.027	−0.001	−0.001
(self−assessed)	(0.078)	(0.078)	(0.061)	(0.060)	(0.039)	(0.034)	(0.016)	(0.014)
Self control	0.075	0.063	−0.012	−0.016	−0.079	−0.067	−0.013	−0.006
	(0.142)	(0.143)	(0.113)	(0.112)	(0.067)	(0.061)	(0.029)	(0.027)
Constant	1.918	2.998**	−0.333	1.088	0.008	0.807	0.256	0.515**
	(1.426)	(1.390)	(1.109)	(1.086)	(0.647)	(0.659)	(0.273)	(0.244)
R^2	0.14	0.12	0.16	0.18	0.08	0.25	0.04	0.25
Observations	248	248	248	248	248	248	248	248

NOTE: 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$; robust standard errors are shown in parentheses.

analyses of their total effect using ordinary least squares (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.

3.4.1 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 24 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 0.22 (22% 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 overindebted. 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).

3.4.2 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 an excluded instrument. In the first stage, we estimate:

$$\text{overconfidence}_i = \alpha_0 + \alpha_1 T_i + \alpha_2 X_i + \varepsilon_i, \tag{1}$$

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:

$$\text{debt}_i = \beta_0 + \beta_1 \widehat{\text{overconfidence}}_i + \beta_2 X_i + \varepsilon_i, \tag{2}$$

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 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 too high expected income, participants borrow 52 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 sizable at 32 cents per euro of too high expected 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 participants are unable to repay their debt by 12 percentage points (per €) in the IV results. We thus observe a positive effect of overconfidence on the intensive and extensive margin of overindebtedness (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 AR 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 LATEs, 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] As the IV estimations

12. The 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.

13. Moreira (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 2024). 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.

14. An 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.

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**	0.515*	0.502**	0.426*	0.303*	0.318**	0.093	0.118*
	(0.366)	(0.346)	(0.282)	(0.258)	(0.157)	(0.149)	(0.057)	(0.060)
Female		−1.111**		−0.522		−0.252		−0.091
		(0.460)		(0.334)		(0.210)		(0.075)
Age		−0.003		−0.019		−0.012		−0.013
		(0.080)		(0.054)		(0.032)		(0.014)
Bachelor		0.182		−0.072		−0.037		−0.013
		(0.649)		(0.446)		(0.251)		(0.107)
Masters		−0.090		0.332		0.054		0.102
		(1.760)		(1.007)		(0.510)		(0.260)
Works		0.569		0.415		0.249		0.010
		(0.518)		(0.364)		(0.206)		(0.081)
Log of income		−0.281		−0.387*		−0.273**		−0.066
		(0.281)		(0.212)		(0.136)		(0.053)
Financial literacy		0.331		0.355**		0.116		0.034
		(0.226)		(0.152)		(0.103)		(0.037)
Risk tolerance		0.014		0.012		0.043**		−0.002
(choices)		(0.040)		(0.028)		(0.020)		(0.007)
Risk tolerance		0.018		0.078		−0.036		−0.002
(self-assessed)		(0.122)		(0.088)		(0.044)		(0.018)
Self control		0.111		0.064		0.036		−0.005
		(0.200)		(0.141)		(0.072)		(0.035)
Constant	3.821***	4.003*	2.391***	2.984**	0.647***	1.784**	0.358***	0.952***
	(0.218)	(2.076)	(0.168)	(1.411)	(0.093)	(0.817)	(0.035)	(0.337)
AR p-value	0.048	0.083	0.045	0.079	0.060	0.040	0.131	0.069
Observations	172	168	172	168	172	168	172	168
Panel B: First stage								
Easy	1.194***	1.190**	1.194***	1.190**	1.194***	1.190**	1.194***	1.190**
	(0.440)	(0.467)	(0.440)	(0.467)	(0.440)	(0.467)	(0.440)	(0.467)
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
1st stage F	7.37	6.50	7.37	6.50	7.37	6.50	7.37	6.50
Observations	172	168	172	168	172	168	172	168

NOTE: 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Robust standard errors are shown in parentheses. AR p-values indicate Anderson–Rubin p-values shown at the bottom of Panel A. First-stage F is the F-statistic of the first stage.

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

Several observations support this assumption: first, we ask people in the postexperimental 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] Second, 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.

4. 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 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 low price levels. 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 10 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

15. In 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 4, we discuss a robustness test with prices at or above retail prices.

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*	0.167	0.515	0.097
	(0.305)	(0.219)	(0.485)	(0.312)
Random group	0.064	−0.222	0.016	−0.094
	(0.309)	(0.220)	(0.506)	(0.386)
Constant	5.732***	2.915***	−0.282	−0.028
	(0.237)	(0.149)	(0.352)	(0.258)
R^2	0.02	0.01	0.01	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	0.427	0.046	0.019
	(0.406)	(0.346)	(0.289)	(0.079)
Random group	−0.482	−0.522	−0.464**	−0.079
	(0.400)	(0.319)	(0.235)	(0.083)
Constant	3.842***	2.243***	0.818***	0.324***
	(0.283)	(0.254)	(0.201)	(0.056)
R^2	0.02	0.04	0.02	0.01
Observations	193	193	193	193

NOTE: 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$; robust standard errors are shown in parentheses.

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

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**		0.252***		0.125		0.009	
	(0.112)		(0.094)		(0.081)		(0.019)	
Overconfidence		0.088		0.237***		0.272***		0.077***
		(0.071)		(0.058)		(0.051)		(0.010)
Constant	2.322***	3.866***	0.771	2.291***	−0.027	0.740***	0.258**	0.317***
	(0.660)	(0.169)	(0.555)	(0.133)	(0.450)	(0.101)	(0.118)	(0.030)
R^2	0.04	0.01	0.05	0.12	0.02	0.24	0.00	0.22
Observations	193	193	193	193	193	193	193	193

NOTE: 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 $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$; robust standard errors are shown in parentheses.

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 (columns (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 (Fiedler and Prager 2018, Camerer, Dreber, and Johannesson 2019), 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 and the correlation between elicited rank beliefs and income expectations is 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 $p = 0.06$). 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 (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 due to chance. 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 techni-

16. The 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.

cally 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).

5. CONCLUSION

Household indebtedness is a problem in many countries and the fraction of households that have to be considered overindebted rises. Economic shocks and structural changes in lending technology (Livshits, Mac Gee, and Tertilt 2016) as classic explanations for overindebtedness 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.

We contribute to a literature finding that optimistic forecast errors predict debt taking (Souleles 2004, Hyytinen and Putkuri 2018, Cocco, Gomes, and Lopes 2021). While these papers provide valuable insights, their results cannot be readily interpreted as causal. One reason is that there might be unobserved characteristics simultaneously affecting both the propensity to make a forecast error and borrowing behavior. Thus, we complement this literature by conducting a laboratory experiment, in which we employ the reversed hard–easy gap (Kruger 1999, Burson, Larrick, and Klayman 2006, Moore and Small 2007) to shift self-confidence and, thereby, beliefs about future income in our setting.

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 counseling, which may help consumers to obtain a more realistic picture of their debt and repayment capacity.

ACKNOWLEDGMENTS

Open access funding enabled and organized by Projekt DEAL.

LITERATURE CITED

Alan, Sule, Mehmet Cemalcilar, Dean Karlan, and Jonathan Zinman. (2018) "Unshrouding: Evidence from Bank Overdrafts in Turkey." *Journal of Finance*, 73, 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, 727–53.

Angrist, Joshua, and Michal Kolesár. (2024) "One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV." *Journal of Econometrics*, 240(2): 105398.

Angrist, Joshua D., and Jörn-Steffen Pischke. (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Barber, Brad M., and Terrance Odean. (2000) "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *Journal of Finance*, 55, 773–806.

Barron, Kai, and Christina Gravert. (2022) "Confidence and Career Choices: An Experiment." *Scandinavian Journal of Economics*, 124, 35–68.

Becker, David, Lena Jaroszek, and Martin Weber. (2017) "Borrower Heterogeneity and the (Ir)Rational Demand for Short-Term Credit." Working Paper, University of Mannheim.

Ben-David, Itzhak, John R. Graham, and Campbell R. Harvey. (2013) "Managerial Miscalibration." *Quarterly Journal of Economics*, 128, 1547–84.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian. (2018) "Behavioral Household Finance." In *Handbook of Behavioral Economics: Foundations and Applications 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, pp. 177–276. Amsterdam: Elsevier.

Bruhin, Adrian, Fidel Petros, and Luıs Santos-Pinto. (2022) "The Role of Self-Confidence in Teamwork: Experimental Evidence." Working Paper, WZB Berlin.

Brunnermeier, Markus K., and Jonathan A. Parker. (2005) "Optimal Expectations." *American Economic Review*, 95, 1092–118.

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, 60–77.

Buser, Thomas, Gianluca Grimalda, Louis Putterman, and Joël van der Weele. (2020) "Overconfidence and Gender Gaps in Redistributive Preferences: Cross-Country Experimental Evidence." *Journal of Economic Behavior & Organization*, 178, 267–86.

Camerer, Colin F., Anna Dreber, and Magnus Johannesson. (2019) "Replication and Other Practices for Improving Scientific Quality in Experimental Economics." In *Handbook of Research Methods and Applications in Experimental Economics*, edited by Arthur Schram and Aljaz Ule, Chapter 5, pp. 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, 1433–36.

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, 920–80.

Cocco, Jo ao F., Francisco Gomes, and Paula Lopes. (2021) "Evidence on Expectations of Household Finances." Working Paper, London School of Economics.

Colzani, Paola, and Luis Santos-Pinto. (2021) "Does Overconfidence Lead to Bargaining Failures?" Working Paper, University of Lausanne.

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." www.creditreform.de/nc/aktuelles/news-list/details/news-detail/schuldneratlas-deutschland-2018.html.

Dargnies, Marie-Pierre, Rustamdjan Hakimov, and Dorothea Kübler. (2019) "Self-Confidence and Unraveling in Matching Markets." *Management Science*, 65, 5603–18.

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*, 69, 4977–91.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. (2011) "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association*, 9, 522–50.

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, Jacqeline Kirshenbaum, Richard A. Klein, Richard E. Lucas, Christopher J.N. Lustgraaf, Daniel Martin, Madhavi Menon, Mitchell Metzger, Jaclyn M. Moloney,

Patrick J. Morse, Radmila Prislin, Timothy Razza, Daniel E. Re, Nicholas O. Rule, Donald F. Sacco, Kyle Sauerberger, Emily Shrider, Megan Shultz, Courtney Siemsen, 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, 115–24.

Fischbacher, Urs. (2007) "z-tree - Zurich Toolbox for Readymade Economic Experiments." *Experimental Economics*, 10, 171–78.

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." *Quarterly Journal of Economics*, 121, 505–40.

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, 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, 844–75.

Glaser, Markus, and Martin Weber. (2007) "Overconfidence and Trading Volume." *GENEVA Risk and Insurance Review*, 32, 1–36.

Graham, John R., Campbell R. Harvey, and Hai Huang. (2009) "Investor Competence, Trading Frequency, and Home Bias." *Management Science*, 55, 1094–106.

Greiner, Ben. (2015) "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE." *Journal of the Economic Science Association*, 1, 114–25.

Heidhues, Paul, and Botond Köszegi. (2010) "Exploiting Naïviete about Self-Control in the Credit Market." *American Economic Review*, 100, 2279–303.

Hyytinen, Ari, and Hanna Putkuri. (2018) "Household Optimism and Overborrowing." *Journal of Money, Credit and Banking*, 50, 55–76.

Imbens, Guido, and Joshua Angrist. (1994) "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62, 467–76.

International Monetary Fund (IMF). (2012) "Dealing with Household Debt." In *World Economic Outlook: Growth Resuming, Dangers Remain*, Chapter 3, pp. 89–124. Washington, D.C.: IMF.

International Monetary Fund (IMF). (2017) "Household Debt and Financial Stability." In *Global Financial Stability Report: Is Growth at Risk?* Chapter 2, pp. 53–89. Washington, D.C.: IMF.

Keese, Matthias. (2012) "Who Feels Constrained by High Debt Burdens? Subjective vs. Objective Measures of Household Debt." *Journal of Economic Psychology*, 33, 125–41.

Klühs, Theres, Melanie Koch, and Wiebke Stein. (2019) "Don't Expect too Much: The Effect of Biased Expectations on (Over)-Indebtedness." CRC TRR 190 Working Paper No. 200.

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, 221–32.

Lawson, Asher, Richard P. Larrick, and Jack B. Soll. (2023) "Forms of Overconfidence: Reconciling Divergent Levels with Consistent Individual Differences." Working Paper, Duke University.

Livshits, Igor, James C. Mac Gee, and Michèle Tertilt. (2016) "The Democratization of Credit and the Rise in Consumer Bankruptcies." *Review of Economic Studies*, 83, 1673–710.

Lusardi, Annamaria, and Peter Tufano. (2015) "Debt Literacy, Financial Experiences, and Overindebtedness." *Journal of Pension Economics and Finance*, 14, 332–68.

Melzer, Brian T. (2011) "The Real Costs of Credit Access: Evidence from the Payday Lending Market." *Quarterly Journal of Economics*, 126, 517–55.

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, 262–71.

Money Advice Service. (2017) "Levels of Over-Indebtedness in the UK: Technical Report." www.moneyadviceservice.org.uk/en/corporate/a-picture-of-over-indebtedness-in-the-uk.

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, 972–89.

Moore, Don A., and Paul J. Healy. (2008) "The Trouble with Overconfidence." *Psychological Review*, 115, 502–17.

Moreira, Marcelo J. (2009) "Tests with Correct Size When Instruments Can Be Arbitrarily Weak." *Journal of Econometrics*, 152, 131–40.

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?" *Journal of Finance*, 53, 1775–89.

Odean, Terrance. (1999) "Do Investors Trade Too Much?" *American Economic Review*, 89, 1279–98.

OECD. (2018) "Wealth: OECD.Stat." stats.oecd.org/index.aspx?datasetcode=Wealth.

Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. (2017) "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender." *Economic Journal*, 127, 2153–86.

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, 457–90.

Smith, Herbert L., and Brian Powell. (1990) "Great Expectations: Variations in Income Expectations among College Seniors." *Sociology of Education*, 63, 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, 39–72.

Stango, Victor, and Jonathan Zinman. (2009) "Exponential Growth Bias and Household Finance." *Journal of Finance*, 64, 2807–49.

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." *Quarterly Journal of Economics*, 132, 1755–817.

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, 271–324.

van Rooij, Marten, Annamaria Lusardi, and Rob Alessie. (2011) "Financial Literacy and Stock Market Participation." *Journal of Financial Economics*, 101, 449–72.

Zinman, Jonathan. (2015) "Household Debt: Facts, Puzzles, Theories, and Policies." *Annual Review of Economics*, 7, 251–76.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Data S1