2019

Misperception of Exponential Growth: Are People Aware of their Errors?

Henning Cordes

Bryan Foltice

Thomas Langer

Misperception of Exponential Growth: Are People Aware of their Errors?

Henning Cordes

Finance Center Münster, University of Münster, 48149 Münster, Germany, henning.cordes@wiwi.uni-muentster.de

Bryan Foltice

Lacy School of Business, Butler University, Indianapolis, Illinois 46208, bfoltice@butler.edu

Thomas Langer

Finance Center Münster, University of Münster, 48149 Münster, Germany, thomas.langer@wiwi.uni-muentster.de

Abstract: Previous research shows that individuals make systematic errors when judging exponential growth, which has harmful effects for their financial well-being. This study analyzes in how far individuals are aware of their errors and how these errors are shaped by arithmetic and conceptual problems. While arithmetic problems could be overcome by employing computational assistance like a pocket calculator, this is not the case for conceptual problems, a term we use to subsume other error drivers like a general misunderstanding of exponential growth or overwhelming task complexity. In an incentivized experiment, we find that participants strongly overestimate the accuracy of their intuitive judgment. At the same time, their willingness to pay for arithmetic assistance is too high on average, often much above the actual benefits a calculator provides. Using a multi-tier system of task complexity we can show that the willingness to pay for arithmetic assistance is hardly related to its benefits, indicating that participants do not really understand how the interplay of arithmetic and conceptual problems shape their errors in exponential growth tasks. Our findings are relevant for policy making and financial advisory practice and can help to design effective approaches to mitigate the detrimental effects of misperceived exponential growth.

Keywords: Exponential growth bias; Overconfidence; Bias awareness

1 Introduction

Exponential growth bias (EGB) refers to the systematic tendency to linearize exponential growth when assessing it intuitively (e.g., Wagenaar and Sagaria 1975, Jones 1979, Wagenaar and Timmers 1979). This has detrimental effects for financial decision making and well-being: More-biased individuals borrow more, save less, and favor shorter maturities (Stango and Zinman 2009). Various types of remedies have been suggested to mitigate the effects of EGB, such as teaching fundamental concepts and formulas, providing task-specific assistance by showing sample outcomes, and promoting the use of calculating devices.[1]

An important question for the design of just-in-time support (offering task-specific assistance right before individuals make a financial decision; Fernandes et al. 2014) is whether individuals believe that they can benefit from the offered assistance, i.e., whether they are aware of their errors when judging exponential growth. Research on overconfidence suggests that individuals overestimate the quality of their intuitive judgment (Glaser and Weber 2010, Levy and Tasoff 2017). As a consequence, optional support might not be utilized and more paternalistic interventions like a salient display of the long-term decision consequences could be advisable. On the other hand, if individuals understand how error-prone their judgment is and how much they can benefit from the offered assistance, less obtrusive interventions might be sufficient.

The study at hand has two related goals: Firstly, we explore whether individuals can properly assess the accuracy of their intuitive judgment in exponential growth-related decision problems. Secondly, we investigate whether they are well calibrated in their judgment of the benefits that a support tool will provide. We focus on a very simple type of support, a pocket calculator.

Following previous research on the value of a pocket calculator in exponential estimation tasks (e.g., Foltice and Langer 2017) we confront our participants with the same set of questions twice, once without any means of aid (Stage 1) and once equipped with a pocket calculator (Stage 2). The first set of answers allows to assess the overall error size, while the second set of answers helps to disentangle the extent to which the overall errors are driven by arithmetic and conceptual problems. The arithmetic component of the problems could be overcome by employing a pocket calculator, it thus merely reflects the inability to conduct the proper calculations in one's head. The remaining error is due to conceptual problems, a term we use to subsume all other error drivers, like a general misunderstanding of exponential growth or overwhelming task complexity, which increases the susceptibility for errors.

We conduct an incentivized experiment to answer our research questions. In Stage 1 of the experiment, participants provide their best estimate to 14 exponential growth-based household finance questions. Here, participants have to rely on their intuitive judgment, and both arithmetic and conceptual problems may induce errors in their answers. To assess the awareness of possible errors, we elicit the confidence that the given answer falls within a ±10% interval around the correct answer after each question. To assess the understanding of how the errors are shaped by arithmetic and conceptual problems, we also elicit the willingness to pay (WTP) for a pocket calculator after each question. Questions differ by their level of complexity. Our financially literate student participants should hardly have any conceptual problems when facing the simplest questions (with the calculator being of great value here), but can be expected to struggle when facing the most complicated questions (with the calculator being of little value here). The incentive structure is set up in a way that the WTP should reflect the anticipated benefit of calculator use. By relating the WTP to the actual benefits of the calculator (derived from a comparison of answers in the two stages), we can disentangle whether participants attribute their anticipated errors mainly to arithmetic problems or to conceptual problems.

We find strong overconfidence in the ability to judge exponential growth accurately, indicating that participants do not fully recognize the magnitude of their errors. At the same time, their WTP for the calculator is too high on average, in some cases much above the benefits the calculator provides. Overall, our participants seem to have just a vague understanding of how their errors are shaped by arithmetic and conceptual problems. The relation between the WTP for and the benefits of calculator use is positive but very weak. The strongest (negative) predictor of WTP is the confidence of our participants in the quality of their judgment in Stage 1. That is, they are particularly interested in employing calculator support when they perceive the task to be difficult, but fail to factor in that the calculator might be often of little help due to conceptual problems.

The remainder of this study is organized as follows: Section 2 derives our research hypotheses. Section 3 describes the experimental design. Section 4 presents the results. Section 5 concludes by summarizing the findings and discussing policy implications.

2 Framework and Expectations

Given the high monetary stakes involved in many exponential growth-related decisions and the general availability of support tools, it is puzzling that EGB has been found to be related to various types of subpar financial behavior (Stango and Zinman 2009). A plausible explanation is that individuals are unaware of the magnitude of their errors in the intuitive judgment of exponential growth and thus do not utilize the support.

To provide some formal framework for the following analysis, we define:

$$EP = \frac{x - x^*}{x^*} \tag{1}$$

to be the error percentage of an estimate, where x^* is the correct answer of a given exponential estimation task and x is the individual's estimate provided without calculator support. More important for our research questions is the absolute error percentage:

$$AEP = \frac{|x - x^*|}{x^*} \tag{2}$$

and our subjects' estimate $\widehat{AEP}$ of the magnitude of this quantity.[2] Building on overconfidence research showing that individuals overestimate the quality of their intuitive judgment (Glaser and Weber 2010, Levy and Tasoff 2017) we expect that participants are overconfident in the accuracy of their intuitive judgment of exponential growth, i.e. $\widehat{AEP} < AEP$.

For a profound discussion of proper decision support it is important to understand the individuals' introspection into the main drivers of their errors, arithmetic and conceptual problems. To this end, we define:

$$AEP_c = \frac{|x_c - x^*|}{x^*} \tag{3}$$

to be the absolute error percentage in estimations with a calculator available, i.e., x_c denotes our participants' answers in Stage 2 of the experiment. With the purely arithmetic problems resolved by the calculator, AEP_c can be interpreted as the component of the absolute error percentage that is due to conceptual problems. Finally, we define:

$$AEP_a = AEP - AEP_c \tag{4}$$

to be the component of the absolute error percentage that is due to arithmetic problems. While it seems natural to assume that AEP_a has to be non-negative, this is not necessarily the case. Foltice and Langer (2017) show that the availability of a calculator can also increase errors in specific exponential tasks when individuals overly rely on (flawed) formulas and neglect common sense. We are interested in $\widehat{AEP_a}$ for our research questions however, our participants' beliefs about the magnitude of their AEP_a. $\widehat{AEP_a}$ should not become negative as this would imply that participants know that the calculator deteriorates their judgment quality, and we therefore do not believe that negative values of $\widehat{AEP_a}$ affect our analyses to a relevant extent.

To measure $\widehat{AEP_a}$, we elicit the WTP for a pocket calculator, more precisely the WTP for receiving a remuneration for the Stage 2 answer (with calculator) instead of the respective answer in Stage 1 (without calculator). As the remuneration is a linear function of AEP (Stage 1) and AEP_c (Stage 2) in the relevant range of answers (for details see Subsection 3.2), the remuneration improvement is linear in AEP_a. The WTP for calculator use should correspond to the anticipated remuneration improvement and is thus linearly related to the anticipated error improvement $\widehat{AEP_a}$.

As we are not aware of any research that has previously explored individuals' awareness of arithmetic and conceptual problems (and their interaction) in EGB-related tasks, our analyses and expectations related to $\widehat{AEP_a}$ are rather explorative and based on intuition than derived from existing research or sound theory. Following our earlier prediction that our participants will underestimate the magnitude of AEP (i.e., $\widehat{AEP} < AEP$), it seems most natural to predict that AEP_a is also underestimated ($\widehat{AEP_a} < AEP_a$). This is because we are not aware of research showing a systematic bias in judging the relative contribution of AEP_c and AEP_a to AEP. Alternatively, one could have argued that participants are also overconfident in judging their ability to utilize the calculator, possibly even offsetting the underestimation of the Stage 1 error – but such a prediction would be at least as ad hoc as the one we chose.

We are further interested in the question of whether individuals have some understanding of how the arithmetic component of the error AEP_a and thus the value of the calculator varies with task complexity. To induce within-subject variation in the extent to which the calculator provides assistance, we ask questions in four levels of difficulty. We chose a sample of financially literate business students who should all have the conceptual knowledge to answer our simplest questions correctly when having access to a calculator, but who can be expected to struggle with the most complicated questions. Thus, the calculator should provide great benefits in the easy and familiar questions, but almost no benefits in the difficult and unfamiliar questions. Over all tasks and tiers we expect to observe a positive relation between $\widehat{AEP_a}$ and AEP_a. Translating the predictions about $\widehat{AEP_a}$ and AEP_a to testable hypotheses about the WTP for calculator use and its benefits, we predict that the WTP for the calculator is lower than the benefits it provides, yet that the WTP is at least positively related to the benefits it provides.

3 Experimental Design

3.1 General Setup

The experiment is fully computer-based, except for a short introduction that is handed out and read aloud by the experimenter prior to the start of the experiment (see Appendix A). Pen and paper are provided for notes and calculations; cell phones, calculators, etc. are prohibited.

The main experiment consists of four parts (Figure 1). First, comprehensive instructions explain the experimental task, the incentive structure, and the software interface (Appendix B contains the full instructions). We carefully check for understanding: Participants can only continue if they answer a set of check-up questions correctly.

Instructions	Explanation of experimental tasks, incentive structure, and the software interface.
Stage 1 (no calculator)	Estimation task: 14 exponential growth questions. Confidence task: Probability that AEP < 10%. WTP task: Willingness to pay for a calculator.
Stage 2 (with calculator)	Estimation task: Same 14 exponential growth questions. Confidence task: Probability that AEP_c < 10%.
Questionnaire	Demographics, personal financial situation, and standardized individual differences scales.

Figure 1: Course of the experiment.

In Stage 1, participants estimate the answers to 14 exponential growth-related household finance questions ("estimation task"; Appendix C contains the question texts). The sequence is randomized and different for each participants to control for ordering effects. To assess the perceived accuracy of answers, participants have to estimate the probability that their answer is within a ±10% interval around the correct answer directly after each question, i.e., AEP < 10% ("confidence task"). Further, participants have to state their WTP for a pocket calculator directly after each question ("WTP task"). The WTP is used to determine whether participants are paid according to their answer without calculator (Stage 1) or with calculator (Stage 2).

Participants are instructed to raise their hand after having completed Stage 1. Then, they receive a standard pocket calculator. To avoid operational issues, we explain the handling of the calculator prior to the start of the experiment and assist upon request during the experiment.

In Stage 2, participants answer the same 14 exponential growth questions as in Stage 1, but now with the help of the calculator. The ordering is again randomized, but different than in Stage 1. To assess the awareness of conceptual problems, participants have to estimate the probability that their answer falls within a ±10% interval around the correct answer directly after each question (i.e., AEP_c < 10%).

The exponential growth questions differ by their level of complexity (“Tiers”). Tier 1 questions cover fundamental issues involving exponential growth (e.g., estimating the future value of an investment). Tier 2 questions ask for additive combinations of two exponential processes (e.g., estimating the outperformance of stocks over bonds). Tier 3 questions ask for multiplicative combinations of two exponential processes (e.g., estimating the future value of an investment and adjusting it for fees). Tier 4 questions cover the additive combination of several exponential processes with different times to maturity (e.g., estimating the future value of annual savings plan contributions). Due to the high financial literacy of our student participants, the correct formula should be obvious and known to most participants in Tier 1 (conceptual knowledge high, calculator of great value). In contrast, the correct formula should be unknown to most participants in Tier 4 (conceptual knowledge low, calculator of little value).

MacKinnon and Wearing (1991) find systematic differences in the individuals’ ability to accurately assess exponential growth and asymptotic decline. To account for these differences, half of the questions is forward-looking (e.g., estimating the future value of an investment) and the other half is backward-looking (e.g., estimating the present value of an investment).

The experiment concludes with a questionnaire on demographics, the financial situation, and standardized scales on financial literacy (Lusardi and Mitchell 2017), numeracy (Weller et al. 2013), and cognitive reflection (CRT from Frederick 2005; REI-10 from Norris et al. 1998).

3.2 Incentive Structure

Participants can earn up to 29 Euro, consisting of three parts: A fix show-up fee (10 Euro), a variable payment based on the accuracy of answers in the estimation task (0-15 Euro), and a variable payment based on the confidence statements in the confidence task (0-4 Euro). The entire incentive structure was explained extensively in the instructions prior to the main experimental task. To keep the instructions short and concise, we did not prove incentive compatibility formally but assured to participants that they cannot benefit from consciously misspecifying their answers and that it is therefore in their best interest to state their beliefs and preferences truthfully.

The accuracy of answers in the estimation task is measured in absolute error percentages (i.e., AEP in Stage 1 and AEP_c in Stage 2, which we use synonymously for the remainder of Subsection 3.2). AEP reduces the payment from the estimation task linearly with the restriction that the payment cannot become negative:

$$\pi_{\text{estimation}} = 15 \text{ Euro} \cdot (1 - \min\{AEP; 1\}) \ . \tag{5}$$

A table illustrating the effects of AEP on the payment is handed out to participants prior to the start of the experiment (Appendix A). The implemented incentive structure in the estimation task implies that estimation errors are more costly when the correct answer (x^*) is lower (see Equations 1-3). This may impose an incentive to state answers in the estimation task which are below the participant's own expectation. We will address this concern in Section 4.4 and invalidate it based on the experimental data.

In the confidence task, participants state a probability p that their answer is within a ±10% interval around the correct answer. The payment is based on a quadratic scoring rule:

$$\pi_{\text{confidence}} = 4\text{ Euro} \cdot \begin{cases} (2 \cdot p - p^2) & \text{if } AEP < 10\% \\ (1 - p^2) & \text{otherwise.} \end{cases} \tag{6}$$

Strictly speaking, the quadratic scoring rule is incentive compatible only under risk neutral preferences (see e.g., Murphy and Winkler 1970 or Selten 1998 for detailed discussions). Hollard et al. (2010) however do not find meaningful differences in calibration between a quadratic scoring rule and two rules that do not depend on risk preferences in an experimental study. Therefore, we are confident that the quadratic scoring rule employed in our setup induces truthful response.

Participants are paid according to their answers in one of the 14 questions. To determine this question, each participant draws a ball out of a concealed urn with 14 numbered ping-pong balls after the experiment. This procedure aims at strengthening trust in how the payment is determined. We employ a standard Becker-DeGroot-Marschak (BDM) mechanism to decide whether estimate and confidence statement from Stage 1 (no calculator) or Stage 2 (with calculator) are used (Becker et al. 1964). The computer randomly selects a price for the calculator (0-15 Euro). If the WTP (also 0-15 Euro) is lower than the price, answer and confidence statement from Stage 1 (no calculator) determine the payment. If the WTP is higher than the price, answer and confidence statement from Stage 2 (with calculator) determine the payment, but the price for the calculator is subtracted (payments could not become negative). Thus, the WTP should reflect the anticipated benefits of the calculator, i.e., the difference in anticipated payments between Stage 1 (no calculator) and Stage 2 (with calculator).

A key feature of our experimental approach is the within-subject design, which allows exploring how the provision of arithmetic assistance affects the accuracy of answers and the confidence in providing a precise answer. This is crucial for the discussion of whether participants are aware of their errors and their specific drivers. Yet, the within-subject design requires special care when evaluating the behavioral impact of the incentives on our participants' judgment and decision making. To give an example: If a participant states a high WTP for the calculator, it is unlikely that the answer from Stage 1 (no calculator) is relevant for payment. This reduces the expected return to effort in Stage 1, which may reduce the quality of answers in Stage 1. Such a mechanism would compromise the interpretation of WTP and its correlation to the accuracy of answers. We will discuss such concerns in more detail in Subsection 4.4, and invalidate them with our experimental data. A second caveat related to our incentive structure is that it is not fully immune to strategic behavior. We designed incentives to be simple, intuitive and easy to grasp, such that participants do not worry about having misunderstood the experimental task and the incentive structure. This comes at the cost that participants who feel very unsure about their estimation abilities but are very insightful with respect to the incentive structure can strategically exploit the system. They could consciously provide very bad estimates to score at least for their perfectly calibrated prediction of having provided a very bad estimate. From previous experiments, we did not expect to see such strategic behavior of participants, and we will show in Subsection 4.4 that there is indeed no sign of it in the data.

3.3 Procedure

The experiment was conducted in the computer labs of a German university. Participants were second year undergraduate business students. As compound interest is taught extensively in the first year of the program, most participants should understand the general concept of exponential growth and know the basic formulas. Participation in the experiment was voluntary and N = 79 students took part. Given the complicated incentive structure, special attention was paid to the comprehension of the instructions. We exclude 3 participants from the main analysis because they revealed severe problems in the check-up questions after the instructions. It took participants about one hour to complete the entire experiment and the average payment was 17.26 Euro. Table 1 shows sample characteristics. As highlighted earlier, we deliberately recruited participants with high financial literacy and numerical abilities. This is reflected in the respective test scores. Albeit evidence amounts that financial literacy and (awareness of) misperceptions of exponential growth have distinct effects on financial decision making. Stango and Zinman (2009, p. 2845) conclude that: “[...] financial sophistication and bias have distinct effects on financial decisions. It may be useful to conceive of bias and awareness of bias as dimensions of sophistication”. Using a more sophisticated proxy for financial literacy, Almenberg and Gerdes (2012) also conclude that the two are distinct constructs measuring knowledge and cognitive skills respectively. This should be kept in mind when discussing the scope of our findings.

Table 1: Sample characteristics.

	Mean	Median	Min.	Max.
Gender (1=female)	35.53%	0	0	1
Age (in years)	22.59	22	19	30
Financial literacy score (4=best, Lusardi and Mitchell 2017)	3.93	4	3	4
Numeracy score (10=best, Weller et al. 2013)	9.30	10	7	10
Cognitive reflection score (3=highest, Frederick 2005)	2.12	2	0	3

Descriptive statistics on sample characteristics (N = 76 participants).

4 Findings

Before exploring our main hypotheses, we feel that it is important to validate some assumptions we have made with respect to our experimental setup (all experimental data can be obtained from the authors upon request). Given the rather complicated incentive structure, we first want to verify that our participants properly understood the experimental task – otherwise, many of our later arguments may lose their foundation. Participants seem to have understood the instructions well as can be learned from the number of errors in the check-up questions of the instructions: On median, participants answered 6 out of 7 questions correctly at first attempt (mean: 5.83). This is also reflected in the self-rated understanding: On a seven point Likert-type scale (where seven is best), the median participant assigned a rating of 7 for the understandability of the instructions in general (mean: 6.25) and a rating of 6 for the understandability of the incentive structure in particular (mean: 5.81). To address the concern that our results are driven by misunderstandings among participants, we conducted strict robustness checks and excluded participants with more than one error in the check-up questions from the analyses (N = 18 in total). All results reported in this study remain qualitatively unchanged.

As our participants are trained in assessing exponential growth, they may be sensitized towards its pitfalls. This could eliminate or even reverse the bias, which would compromise the external validity of our findings. For a quick analysis of the general pattern of EGB we look at the signed error percentage EP as defined in Section 2. EGB, operationalized as in Stango and Zinman (2009), predicts an underestimation of the correct answer for forward-looking questions ($EP < 0$) and an overestimation for backward-looking questions ($EP > 0$). Across all tiers, the median EP is -0.54 for forward-looking questions and 0.41 for backward-looking questions in Stage 1 (no calculator).[3] Thus, even our financially literate participants fall prey to EGB, suggesting that our experimental findings on absolute error percentages AEP are extendable beyond the specific subject pool.

The Tier-system is set up with the intention to produce a high level of conceptual knowledge in Tier 1 up to a low level in Tier 4. To evaluate whether the goal was achieved, we proxy conceptual knowledge by the fraction of “proper” answers in Stage 2, where judgment accuracy is only determined by conceptual knowledge. We define an answer to be “proper” if it is within a ±10%

interval around the correct answer, i.e. AEP < 10% (Stage 1) or AEP_c < 10% (Stage 2) respectively. We allow some deviation from the exact answer, because rounding errors can amplify in Tiers where multiple computational steps are necessary to derive the answer. The interval size is set to ±10% as it simplifies the calibration discussion later on (participants were confronted with ±10% intervals in their confidence statements). The fraction of proper answers in Stage 2 declines monotonically from 89.47% in Tier 1, 86.84% in Tier 2, 78.29% in Tier 3, down to only 9.87% in Tier 4 (Appendix E). Although we hoped for a more gradual decline, we conclude that the Tier-system produces variation in conceptual knowledge as expected.

4.1 Awareness of Errors when Judging Exponential Growth

We predicted that participants are overconfident in their ability to judge exponential growth intuitively, i.e. $\widehat{AEP} < AEP$. We explore this presumption by comparing the fraction of proper answers (AEP < 10%) to the average confidence participants state that their answer is proper ($\widehat{AEP} < 10\%$). If participants are overconfident, the average confidence in having provided a proper answer is higher than the actual fraction. Supporting our prediction, Table 2 shows that confidence is significantly higher than the fraction of proper answers in all Tiers of Stage 1. Thus, participants seem to be overconfident in the accuracy of their intuitive judgment.[4]

Table 2: Average confidence and fraction of proper answers in Stage 1 (no calculator).

Tier	*N*	Confidence in answer	Fraction of proper answers	Difference
All	1,064	37.26%	7.90%	29.36% ***
		(0.03)	(0.02)	(0.02)
Tier 1	456	41.77%	10.96%	30.81% ***
		(0.03)	(0.02)	(0.03)
Tier 2	152	32.09%	5.26%	26.83% ***
		(0.03)	(0.02)	(0.03)
Tier 3	304	33.58%	7.57%	26.01% ***
		(0.03)	(0.02)	(0.03)
Tier 4	152	36.24%	1.97%	34.26% ***
		(0.03)	(0.0)	(0.03)
Tier 4-1		-5.53% **	-8.99% ***	-3.46%
		(0.02)	(0.02)	(0.03)

Average confidence and fraction of proper answers in Stage 1 (no calculator), by Tiers. An answer is considered "proper" if AEP < 10% (Stage 1) or AEP_c < 10% (Stage 2) respectively. Cluster-robust standard errors are provided in parentheses. ***, **, * indicate statistical significance on the 1%, 5%, 10% level.

We next look at calibration curves for a more profound analysis of overconfidence. We assign confidence statements to 20% bins, and plot the average confidence against the fraction of proper answers within each bin. Panel A of Figure 2 depicts the calibration curve in Stage 1 (all Tiers combined). The curve is upward-sloping, but well below the identity line: Participants somewhat adapt their confidence to the accuracy of their answers, but they are generally overconfident.

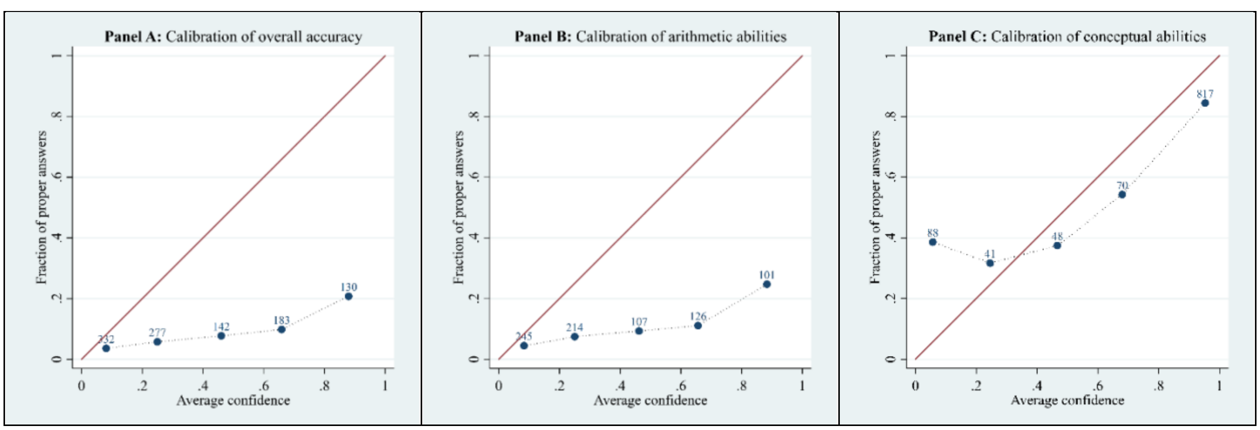

Figure 2: Panel A shows the calibration curve in Stage 1 (no calculator). Panel B shows the calibration curve where only Stage 1 answers are considered if the participant gave a proper answer to the same question in Stage 2 (with calculator). Panel C shows the calibration curve in Stage 2 (with calculator). An answer is considered "proper" if AEP < 10% (Stage 1) or AEP_c < 10% (Stage 2) respectively. Confidence statements are collapsed in the intervals [0;0.2], (0.2;0.4], (0.4;0.6], (0.6;0.8], (0.8;1.0]. Numbers above the scatters indicate the number of observations within the interval.

To assess overconfidence in arithmetic abilities, Panel B controls for conceptual problems. This is done by only considering Stage 1 answers that come from participants who have no conceptual problems (i.e., who gave a proper answer in Stage 2 with calculator). By visual inspection, the curve is marginally steeper compared to Panel A, but it remains well below the identity line. Thus, participants are overconfident in their arithmetic abilities in particular, indicating that they do not sufficiently grasp the computational component of their error.

To assess the confidence-accuracy calibration with respect to conceptual abilities, Panel C shows the calibration curve in Stage 2 (all Tiers combined). Albeit we did not formulate concrete expectations, analyzing this relationship promises to advance the understanding of the participants' introspection into the main drivers of their errors when judging exponential growth. Consistent to the hard-easy effect (e.g., Lichtenstein and Fischhoff 1977), participants are underconfident for low confidence levels (where the hard Tier 4 questions predominate) and somewhat overconfident for high confidence levels (where the easy Tier 1 questions predominate). Considering the fact that the vast majority of responses fall into the highest confidence bin (817 out of 1,064) where calibration is fairly good, comparing Panel C to Panel B suggests that calibration of conceptual abilities is better than calibration of arithmetic abilities. This implies that overconfidence in conceptual abilities is not the main driver of any misperceptions of the benefits of the calculator, which we will analyze in the following section.

Taken together, the data confirms our prediction: Participants are overconfident in their ability to judge exponential growth accurately, indicating that they are not fully aware of their errors. To reduce the detrimental effects of misperceptions of exponential growth on financial decision making, policy making could consider prompting individuals prior to making exponential growth-related decisions that their intuitive judgment is error-prone. We further find that participants are overconfident in their arithmetic abilities, whereas they are somewhat accurately calibrated regarding their conceptual abilities. The following subsection addresses this point in more detail by studying the WTP for a standard pocket calculator, which eliminates arithmetic problems but does not help at overcoming arithmetic abilities.

4.2 Relation between Willingness to Pay for and Benefits of Calculator Use

We now turn to the question whether participants sufficiently appreciate the opportunity to counteract the arithmetic component of their errors by employing computational assistance. Well-calibrated individuals should state a WTP that is equal to the benefits of the calculator, i.e., the improvement of payments between Stage 1 (no calculator) and Stage 2 (with calculator).

We predicted that the WTP is lower than the benefits of the calculator. The previous findings nurture this presumption and support our previous reasoning: Overconfidence in the quality of one's judgment should make assistance look less relevant.

Panel A of Table 3 contrasts the WTP with the benefits of the calculator. Contradicting our prediction, the WTP is significantly higher than the benefits ($\Delta = 8.52 - 7.04 = 1.48$ Euro, $p < 0.01$). This is surprising, considering the pervasive overconfidence documented in Subsection 4.1. Looking at the relation between WTP and benefits across Tiers yields an intuition for this finding. While the benefits of the calculator decrease from 8.23 Euro in Tier 1 to 1.62 Euro in Tier 4 ($\Delta =$ -6.61 Euro, $p < 0.01$), the WTP remains constant across Tiers ($\Delta = 0.38$ Euro, $p = 0.20$). This results in an increasing gap between WTP and the benefits of the calculator, from 0.04 Euro in Tier 1 to 7.03 Euro in Tier 4 ($\Delta = 6.99$ Euro, $p < 0.01$).

Table 3: Willingness to pay (WTP) for the pocket calculator and benefits provided by the calculator.

		Panel A			Panel B		
	N	WTP	Benefits	Difference	Proper answers Stage 2	Confidence Stage 2	Difference
All	1,064	8.52	7.04	1.48 ***	74.53%	81.17%	-6.64% ***
		(0.46)	(0.33)	(0.48)	(0.02)	(0.02)	(0.02)
Tier 1	456	8.27	8.23	0.04	89.47%	88.63%	0.84%
		(0.45)	(0.39)	(0.48)	(0.02)	(0.03)	(0.03)
Tier 2	152	8.62	8.00	0.62	86.84%	83.39%	3.45%
		(0.51)	(0.50)	(0.68)	(0.03)	(0.03)	(0.03)
Tier 3	304	8.78	7.49	1.29 **	78.29%	79.86%	-1.57%
		(0.50)	(0.48)	(0.62)	(0.03)	(0.03)	(0.03)
Tier 4	152	8.65	1.62	7.03 ***	9.87%	59.19%	-49.32% ***
		(0.52)	(0.52)	(0.72)	(0.03)	(0.03)	(0.04)
Tier 4-1		0.38	-6.61 ***	6.99 ***	-79.60%	-29.44% ***	-50.16% ***
		(0.30)	(0.59)	(0.66)	(0.03)	(0.03)	(0.05)

Descriptive statistics on the WTP (in Euro), benefits of the calculator (in Euro), fraction of proper answers in Stage 2 (in %), and confidence in Stage 2 (in %). An answer is considered "proper" if $AEP < 10\%$ (Stage 1) or $AEP_c < 10\%$ (Stage 2) respectively. Cluster-robust standard errors are provided in parentheses. ***, **, * indicate statistical significance on the 1%, 5%, 10% level, based on two-sided t-tests.

Interestingly, the confidence in being able to give a proper answer in Stage 2 systematically decreases from Tier 1 to Tier 4 (Panel B of Table 3). The reduction in confidence does not match the actual reduction in the fraction of proper answers (from 89.47% in Tier 1 to 9.87% in Tier 4), but is significant (from 88.63% in Tier 1 to 59.19% in Tier 4, $\Delta = -29.44\%$, $p < 0.01$). The fact that the WTP does not adapt to the declining confidence in Stage 2 is puzzling and calls for a more general inspection of potential WTP determinants.

Before we do so in the next subsection, we shortly assess our prediction that the WTP for the calculator is positively related to the benefits it provides. Model 1 of Table 4 shows the results of an OLS regression with the benefits of calculator use (the difference between Stage 2- and Stage 1-payment for the same question) regressed on the WTP for the use of a calculator for this question. The model shows that the benefits of the calculator are positively correlated to WTP ($\beta = 0.08$, $p = 0.04$), supporting our prediction. It should be noted though that the effect is small: WTP increases by only 0.08 Euro for every 1 Euro the calculator improves payments, while it should increase by 1 Euro ($\beta = 1$). Model 2 of Table 4 verifies that a larger effect is not masked by individuals who falsely believe to know the correct formula and therefore misperceive the benefits of the calculator; even after controlling for conceptual knowledge, provided benefits affect WTP only weakly (β = -0.0002 + 0.12 = 0.12, p = 0.03).

Table 4: Drivers of the willingness to pay for the pocket calculator (WTP).

	(1)	(2)	(3)	(4)	(5)
Provided benefits	0.08 **	-0.0002			
	(0.04)	(0.07)			
Know formula (1=yes)		-0.63		1.06	
		(0.71)		(1.14)	
Know formula x provided benefits		0.12 *			
		(0.07)			
Shortfall payments Stage 1			0.11 **	0.14	
			(0.05)	(0.09)	
Know formula x shortfall payments Stage 1				-0.02	
				(0.09)	
Confidence Stage 1					-5.57 ***
					(1.42)
Confidence increase Stage 2					1.85 **
					(0.90)
Intercept	7.96 ***	8.14 ***	7.41 ***	6.43 ***	9.78 ***
	(0.55)	(0.58)	(0.71)	(1.22)	(0.94)
Adj. R^2	0.01	0.02	0.01	0.02	0.19
N	1,064	1,064	1,064	1,064	1,064

Effects on WTP from OLS regressions. “x” denotes an interaction. Cluster-robust standard errors are provided in parentheses. ***, **, * indicate statistical significance on the 1%, 5%, 10% level.

4.3 Other Determinants of WTP

Given that the actual benefits of calculator use are hardly related to the WTP for calculator access, we should explore some other potential determinants of the WTP. First we examine whether WTP is simply driven by the magnitude of the errors in Stage 1. Participants might well grasp how much potential there is for improvement but not factor in that conceptual problems will limit the benefits of calculator assistance. To proxy the magnitude of errors, we use the shortfall of payments in Stage 1 (no calculator) from the maximum payment in the estimation task (15 Euro), which participants receive if they know the correct formula and state the correct answer in Stage 2 (with calculator). We use this indirect and restricted measure of the error magnitude, because the relationship between WTP and the shortfall of payments is better interpretable than the relationship between WTP and more direct error measures. Because higher errors translate into higher shortfalls, we would thus predict a positive correlation between WTP and the shortfall of payments. Model 3 of Table 4 supports this prediction: For every 1 Euro that the error reduces payments in Stage 1, the WTP for the calculator increases by $\beta = 0.11$ Euro ($p = 0.03$). It is noteworthy though that the effect is again small and that the major part of the WTP cannot be explained by the shortfall of payments.

To obtain a clearer picture of how the WTP is shaped by the different components of the error, we again control for conceptual knowledge. Participants who give a proper answer in Stage 2 (with calculator) know the correct formula and have no conceptual problems. For those participants, the error in Stage 1 (no calculator) is entirely owed to their arithmetic problems (where the calculator helps) and they should state a WTP of 1 Euro for every 1 Euro the payment in Stage 1 falls short of the maximum (i.e., the effect should be $\beta = 1$). Participants who do not give a proper answer in Stage 2 (with calculator) most likely do not know the correct formula and cannot benefit from the calculator; they should state a WTP of 0 Euro and the effect should be $\beta = 0$. Model 4 of Table 4 checks this proposition by interacting the shortfall of payments from Stage 1 (no calculator) with a dummy variable indicating whether participants gave a proper answer in Stage 2 (i.e., whether they know the formula). Participants who know the formula increase their WTP by $\beta = 0.14 – 0.02 = 0.12$ Euro ($p = 0.03$) for every 1 Euro the payment in Stage 1 falls short of the maximum payment. The effect is well below the benchmark of well-calibrated individuals ($\beta = 1$), illustrating that participants do not sufficiently grasp the magnitude of their arithmetic problems. Interestingly, neither the effect of the payment shortfall nor the WTP base-level differ significantly between participants who know and do not know the formula (difference in effect-size: $\Delta_\beta = -0.02$, $p = 0.83$; difference in base-level: $\Delta_\alpha = 1.06$, $p = 0.36$). This suggests that participants do not understand how conceptual problems affect the accuracy of their judgment.

With the main part of the WTP still unexplained by the factors that should be expected to determine it (Models 1-4) we finally explore whether the WTP is simply determined by how little confidence participants have in their judgment. Model 5 of Table 4 relates the WTP for the calculator to the confidence in providing a proper answer in Stage 1 (no calculator) and the confidence-increase due to the availability of the calculator (difference in confidence between Stages 2 and 1). Interestingly, the confidence in providing a proper answer in Stage 1 (no calculator) has a strong negative impact on WTP: The difference in WTP between participants who are 0% and 100% confident is 5.57 Euro on average ($p < 0.01$). This speaks in favor of a notion that there is some general aversion towards decision making in a situation with little confidence involved and in particular less confidence than necessary. This is supported by the fact that the confidence increase due to the availability of the calculator also has a strong (positive) impact on WTP ($\beta = 1.85$, $p = 0.04$).

To sum up, the demand for arithmetic assistance is mostly unrelated to the benefits of the assistance (Model 1-2), the error size (Model 3) and arithmetic and conceptual problems (Model 4). Instead, the demand seems to be mainly determined by a lack of confidence in the own intuitive judgment, and by the anticipated increase in confidence due to the availability of assistance (Model 5). This suggests that individuals are keen to accept any form of assistance, independent of the actual benefit it provides – at least if they are put into a position where they recognize how little confidence they have in their intuitive judgment and how it compares to the (supposedly) higher confidence in a decision situation where assistance is available.

4.4 Alternative Explanations

Subsection 3.2 discussed how our experimental design may impose potential problems to the interpretation of our data. This subsection is dedicated to discussing these concerns in more detail, and invalidate them based on our experimental data.

The within-subject design may systematically bias the WTP and thereby its correlation to other variables. Stating a high WTP for the calculator reduces the probability that the answer from Stage 1 (no calculator) becomes relevant for payment. Thus, increasing the WTP reduces the expected return to effort in Stage 1, which may reduce the quality of estimates in Stage 1 (because participants rely on their "second chance" of providing a good estimate in Stage 2). This could bias the correlation between WTP and the benefits of the calculator (Table 3), as well as the correlation between the WTP and the shortfall of payments (Table 4). To address this concern, we regress the WTP on the time in minutes participants spend on answering a question (as a proxy for the effort put into the answer). Model 1 and 2 of Table 5 show that WTP does not affect the time spent on answering a question in Stage 1 and 2 respectively. Therefore, it seems unlikely that WTP systematically affects the effort and thereby the quality of answers.

Table 5: Invalidation of alternative explanations.

	(1) Time spent on answering a question in Stage 1	(2) Time spent on answering a question in Stage 2	(3) Confidence in providing a proper answer in Stage 1	(4) Confidence in providing a proper answer in Stage 2
Willingness to pay	-0.03	0.02	-0.03 ***	0.002
	(0.02)	(0.01)	(0.01)	(0.004)
Constant	2.02 ***	1.21 ***	0.59 ***	0.80
	(0.22)	(0.08)	(0.05)	(0.04)
Adj. R^2	0.01	0.003	0.18	0.001
N	1,064	1,064	1,064	1,064

Effect of WTP on the total time in minutes spent on answering a question in Stage 1 (Model 1) and in Stage 2 (Model 2), and the confidence of providing a proper answer in Stage 1 (Model 3) and in Stage 2 (Model 4) from OLS regressions. Cluster-robust standard errors are provided in parentheses. ***, **, * indicate statistical significance on the 1%, 5%, 10% level.

WTP may also bias the confidence in providing a proper answer in Stage 1: Stating a high WTP reduces the probability that the stated confidence in the Stage 1-answer is relevant for payment. Thus, participants who state a high WTP know that the confidence in the Stage 1-answer is probably irrelevant for payment, and may consciously overstate their confidence (e.g., due to social desirability). In Stage 2, knowing that their answer now matters, they state the correct confidence. Those participants would be approximately maximizing their payments, and they may even be correctly calibrated. Yet, the measures reported in Table 2 and Figure 2 would classify them as being overconfident in Stage 1. Model 3 of Table 5 shows that WTP negatively predicts the confidence in providing a proper answer in Stage 1. This speaks directly against the mechanism outlined above, and supports our original notion that WTP expresses a lack of confidence in providing a proper answer in Stage 1 (see also Model 5 of Table 4). Participants who state a low WTP may follow the opposite pattern (i.e., correct confidence in Stage 1 and overly high confidence in Stage 2 due to social desirability). Our measures would classify those participants as being overconfident in Stage 2. Yet, WTP does not affect stated confidence in Stage 2 (Model 4 of Table 5). All this suggests that participants do not systematically state very high confidence levels (e.g., due to social desirability) if their answer is likely not relevant for payment. This lends support to the validity of our findings on overconfidence.

Our experimental incentive structure is a compromise between strict incentive compatibility and understandability. Existing research shows that performance-based incentives can improve diligence and judgement quality (Camerer and Hogarth 1999). Given the repetitive nature of our experimental task (participants answered 2x14 questions), we believe that linking participant remuneration to accuracy generally improves diligence. Yet, enforcing strict incentive compatibility comes at the cost of a much more complicated incentive structure. As we were concerned that participants worry about their misconception of the experimental task, we refrained from enforcing strict incentive compatibility. The implemented incentive structure in the estimation task implies that estimation errors are more costly when the correct answer (x^*) is lower (see Equations 2, 3, and 5). This imposes an incentive to state answers in the estimation task which are below the own expectation. Yet, we believe that the vast majority of participants does not exploit this subtle loophole by systematically stating answers below their expectation. Existing literature on EGB provides indirect evidence for this presumption, as similar levels of inaccuracy have been observed in experiments with very different incentive structures, from offering no compensation at all to enforcing strict incentive compatibility.[5] To address this presumption directly with our experimental data, we regressed the correct answer (x^*) on a dummy variable indicating whether a participant underestimated the correct answer in an (unreported) probit model; we find that the magnitude of the correct answer positively affects the probability of *underestimating* the correct answer ($p < 0.01$). In other words: The lower the correct answer, the lower the probability of underestimating the correct answer. This speaks directly against the concern that the linear relation between error size and payments from the estimation task induces a downward bias in given answers when the correct answer is particularly low.

Also for reasons of simplicity, the quadratic scoring rule in the confidence task only takes the probability for providing an answer within the ±10% interval around the correct answer into account (see Equation 6), rather than the whole distribution of errors. This may impose an incentive to state more predictable answers. For example, a participant who thinks there is a 50% chance her estimate is within the ±10% interval expects a payment of 3 Euro in the confidence task from reporting her confidence truthfully. However, she could revise her estimate to one that is definitely wrong, report a confidence of 0%, and earn 4 Euro in the confidence task for sure (note that participants who truly behave strategically would not state a nonsensical answer in the estimation task and then a confidence of e.g. 1% or 2%). Albeit this is not a dominant strategy (stating a nonsensical answer reduces the expected payoff in the estimation task), some participants may engage in such strategic decision making. The experimental data shows that this is not a widespread behavior among participants. Only 4.23% of answers in Stage 1 and 2.44% of answers in Stage 2 report a confidence of 0%, and almost none of them appeared in combination with completely nonsensical estimates.

In sum, we find no concrete evidence that our within subject design or our incentive structure induces participants to systematically and strategically misspecify their answers. This lends support to the validity of our findings regarding the awareness of errors in judging exponential growth and their specific drivers.

5 Conclusion

This study explores whether individuals are aware of their errors when answering exponential growth-related household finance questions, and whether they understand how these errors are shaped by arithmetic and conceptual problems.

The arithmetic problems stem from the mere inability to conduct the proper calculations in one's head and could be overcome by employing a pocket calculator. The errors remaining even with a pocket calculator stem from conceptual problems, a term we use to subsume all other error drivers (like a general misunderstanding of exponential growth or overwhelming task complexity), which increases the susceptibility for errors.

Our participants are overconfident in their ability to judge exponential growth intuitively, indicating that they are not (fully) aware of the magnitude of their errors. Interestingly, they reveal at the same time a strong preference for obtaining arithmetic assistance (a simple calculator) to improve their judgment and are willing to pay a price for such assistance that outweighs its benefits. Our experimental design allows to conclude that the WTP is hardly related to the benefits the assistance provides or the magnitude of the errors. The strongest (negative) predictor of WTP is the confidence of our participants in the quality of their judgment in Stage 1. That is, they are particularly interested in employing calculator support when they perceive the task to be difficult and mostly neglect the fact that conceptual problems limit the benefits of the assistance.

Our findings advance the understanding of the motives and drivers of individual behavior in exponential growth-related decision situations and will stimulate further research on effective decision support. Albeit individuals generally underestimate the magnitude of their errors, they are keen on using assistance – even if it is offered at a significant price or involves significant effort. The interest in such support is not driven by a profound introspection into its benefits, not to speak of an understanding of the interplay of arithmetic and conceptual problems but rather by a general aversion towards decision making in a situation where they have less confidence in their judgment than necessary.

Our findings are also important to guide policy making in its attempts to improve financial decision making of private households in exponential growth-related decision problems. They suggest that assistance does not need to be mandatory when individuals make (exponential growth-based) financial decisions. They seem to have a strong interest in using it when they realize how little confidence they have into their intuitive judgment. The latter can be achieved by making the shortcomings of the intuitive judgment salient as we do in our experiment. The question which of our experimental design features (the explicit offering of support, the confidence elicitation for intuitive judgment, or something else) is the strongest driver for the phenomenon is not explicitly addressed in this paper but left for future research.[6] The fact that we explored the behavior of financially literate business students and find hardly any relation between their interest in decision support and its actual benefits further supports the relevance of our research. Less literate individuals can be expected to rely even more on general heuristics instead of sensible cost-benefit considerations.

Endnotes

1 For instance, teaching the "rule of 72" (Eisenstein and Hoch 2007) or basic compound interest formulas (Foltice and Langer 2017) improves the judgment of exponential growth. Sampling experiences with exponential growth, either in real life (Keren 1983) or in laboratory settings (MacKinnon and Wearing 1991), seems to train the intuitive understanding of exponential growth and mitigate misperceptions. Foltice (2015) shows that teaching formal or informal rules as well as experience sampling improves the judgment of exponential growth immediately after the intervention, but is not very persistent.

2 In a more elaborate error analysis, $\widehat{AEP}$ should be modeled as a distribution of beliefs about the size of AEP. But as we are not interested in the specifics of this distribution, we collapse it into a single best estimate. Note that $\widehat{AEP}$ measures error awareness, but not EGB awareness in its typical directional sense. Our participants' best guess with respect to their signed error, i.e. $\widehat{EP}$, is bound to be zero if participants properly react to the given incentives.

3 Appendix D contains cumulative distribution functions of EP. In Stage 1 (no calculator), median answers across all Tiers are biased in the direction as predicted by EGB; negative sign for forward-looking questions and positive sign for backward-looking questions. An exception are backward-looking Tier 2 questions, where EGB predicts a positive sign for very high levels of biasedness and a negative sign for moderate levels of biasedness (which we find).

4 Strictly speaking, this pattern indicates overprecision, which is considered a specific dimension of overconfidence (e.g., Griffin and Brenner 2007). For sakes of simplicity, we do not make this subtle distinction.

[5] Most EGB studies do not link remuneration to accuracy. Stango and Zinman (2009) use unincentivized survey data. McKenzie and Liersch (2011) offer their student participants partial course credit (no grades) for participation (experiments 1-3). Eisenstein and Hoch (2007) offer their participants a fix monetary payment for participation (experiments 2-3). Only few EGB studies link remuneration to accuracy. Similar to our approach, Foltice and Langer (2017) link remuneration linearly to the estimation error. Levy and Tasoff (2015) link remuneration piecewise linearly to the estimation error. Levy and Tasoff (2017) link remuneration to a quadratic scoring rule in accuracy.

[6] Levy and Tasoff (2017) do not elicit the (confidence in the) own judgment prior to eliciting the interest in decision support tools in a similar experimental task. They do not find such a strong interest in decision support tools, indicating that the confidence elicitation in our experimental setup raises the participants' awareness of the vagueness of their intuitive judgment and in turn the WTP for the calculator.

References

Almenberg J, Gerdes C (2012) Exponential growth bias and financial literacy. *Appl. Econom. Lett.* 19(17):1693-1696.

Becker GM, Degroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. *Behavioral Sci.* 9(3):226-232.

Camerer CF, Hogarth RM (1999) The effects of financial incentives in experiments: A review and capital-labour-production framework. *J. Risk Uncertainty* 19(1-3):7-42.

Eisenstein EM, Hoch SJ (2007) Intuitive compounding: Framing, temporal perspective, and expertise. Working paper, Fox School of Business, Temple University, Philadelphia.

Fernandes D, Lynch JG, Netemeyer RG (2014) Financial literacy, financial education, and downstream financial behaviors. *Management Sci.* 60(8):1861-1883.

Foltice B (2015) How to decrease the amortization bias: Experience vs. rules. *J. Financial Education* (forthc.).

Foltice B, Langer T (2017) In equations we trust? Formula knowledge effects on the exponential growth bias in household finance decisions. *Decision Anal.* 14(3):170-186.

Frederick S (2005) Cognitive reflection and decision making. *J. Econom. Perspectives* 19(4):25-42.

Glaser M, Weber M (2010) Overconfidence. Baker HK, Nofsinger JR, eds. *Behavioral finance: Investors, corporations, and markets*, 1st ed. (John Wiley & Sons, Hoboken, NJ), 241-258.

Griffin D, Brenner L (2004) Perspectives on probability judgment calibration. Koehler DJ, Harvey N, eds. *Blackwell Handbook of Judgment and Decision Making*, 1st ed. (Blackwell Publishing, Hoboken, NJ), 157-176.

Hollard G, Massoni S, Vergnaud JC (2010). Comparing three elicitation rules: The case confidence in own performance. Working paper, CES, Université Paris 1, Paris.

Jones GV (1979) A generalized polynomial model for perception of exponential series. *Attention Perception Psychophysics* 25(3):232-234.

Keren, G (1983) Cultural differences in the misperception of exponential growth. *Perception Psychophysics* 34(3):289-293.

Levy MR, Tasoff J (2015) Exponential-growth bias and lifecycle consumption. *J. Eur. Econom. Assoc.* 14(1):545-583.

Levy MR, Tasoff J (2017) Exponential-growth bias and overconfidence. *J. Econom. Psychol.* 58(1):1-14.

Lichtenstein S, Fischhoff B (1977) Do those who know more also know more about how much they know? *Organ. Behav. Human Performance* 20(1):159-183.

Lusardi A, Mitchell OS (2017) How ordinary consumers make complex economic decisions: Financial literacy and retirement readiness. *Quart. J. Finance* 7(3):1-32.

Mackinnon AJ, Wearing AJ (1991) Feedback and forecasting of exponential change. *Acta Psych.* 76(2):177-191.

McKenzie CRM, Liersch MJ (2011) Misunderstanding savings growth: Implications for retirement savings behavior. *J. Marketing Res.* 48(SPL):1-13.

Murphy AH, Winkler RL (1970) Scoring rules in probability assessment and evaluation. *Acta Psych.* 34(1):273-286.

Norris P, Pacini R, Epstein S (1998) The rational-experiential inventory, short form. Working paper, University of Massachusetts at Amherst, Amherst.

Selten R (1998) Axiomatic characterization of the quadratic scoring rule. *Experim. Econom.* 1(1):43-61.

Stango V, Zinman J (2009) Exponential Growth Bias and household finance. *J. Finance* 64(6):2807-2849

Wagenaar WA, Sagaria SD (1975) Misperception of exponential growth. *Attention Perception Psychophysics* 18(6):416-422.

Wagenaar WA, Timmers H (1979) The pond-and-duckweed problem: Three experiments on the misperception of exponential growth. *Acta Psych.* 43(3):239-251.

Weller JA, Dieckmann NF, Tusler M, Mertz CK, Burns WJ, Peters E (2013) Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, 26 (2), 198-212.

Appendix A: Written introduction.

Page 1:

Dear experiment participant,

Welcome to our experiment. We would like to thank you in advance for your participation. Participants of previous experiments have let us know that they perceived economic experiments as very exciting. We hope that you will share this feeling.

Our experiment is divided into three sections. In section 1, you will have to estimate the correct answer to 14 mathematical questions only using a pen and a paper. After having completed section 1, we will provide you with a standard pocket calculator which you are allowed to use in sections 2 and 3. In section 2, you will have to estimate the correct answer to the same 14 mathematical questions as in section 1. Section 3 is a concluding questionnaire.

You will receive a fixed payment of 10 Euro for participating in our experiment. In addition to that, you will receive a variable payment of up to 19 Euro. The variable payment is based, among other things, on the precision of your answers in sections 1 and 2.

Specifically, the variable payment will be based on the precision of your answer to one single answer and not the average precision of all your given answers. Therefore, it is in your best interest to think carefully about all answers that you give in this experiment – even if you had troubles answering some of the previous questions. For determining which question will be used for determining the variable payment, you will draw a numbered ball out of an urn with 14 balls after having completed the experiment. A more detailed explanation of the exact payment rules will follow at the relevant stages during the experiment.

Please note that is strictly forbidden to use any devices that you brought along (e.g. pocket calculators, cell phones) and the internet during this experiment. You are also not allowed to talk to fellow students during the experiment or to look at other screens. A violation of these rules will cause an immediate exclusion without pay from the experiment.

In a few moments, you will receive the login details that are needed for starting the experiment together with a pen and a piece of paper from us.

We will explain the directly following subtasks before each section starts. It is strictly required that you carefully read and comprehend the instructions and tutorials for every section.

Please raise your hand if you have any questions during the experiment. The experimenter will immediately come and assist you.

Page 2:

The variable payment

Additionally to the 10 Euro show-up fee, you will receive a variable payment upon completing this experiment. The variable payment is based on the precision of your answer to one randomly chosen question.

Precision is measured in absolute error percentages (AEP):

$$AEP = \frac{|\text{your answer} - \text{correct answer}|}{\text{correct answer}}$$

Your variable payoff is determined by the formula:

$$\text{your payoff} = 15\text{ Euro} - 15\text{ Euro} \cdot AEP$$

The following table gives an overview of what your variable payoff looks like for various answers (assuming that the correct answer is 100):

Correct answer	**Your answer**	$\boldsymbol{AEP}$	**Your payoff**
100	< 0	100%	0.00 Euro
100	0	100%	0.00 Euro
100	10	90%	1.50 Euro
100	20	80%	3.00 Euro
100	30	70%	4.50 Euro
100	40	60%	6.00 Euro
100	50	50%	7.50 Euro
100	60	40%	9.00 Euro
100	70	30%	10.50 Euro
100	80	20%	12.00 Euro
100	90	10%	13.50 Euro
100	**100**	**0%**	**15.00 Euro**
100	110	10%	13.50 Euro
100	120	20%	12.00 Euro
100	130	30%	10.50 Euro
100	140	40%	9.00 Euro
100	150	50%	7.50 Euro
100	160	60%	6.00 Euro
100	170	70%	4.50 Euro
100	180	80%	3.00 Euro
100	190	90%	1.50 Euro
100	200	100%	0.00 Euro
100	> 200	100%	0.00 Euro

Please note that the AEP measure is restricted in the experiment to not exceed 100%.

Appendix B: Experimental instructions.

Screen 1: Introduction

Welcome and thank you for participating in this experiment. Your assistance is highly appreciated. We are interested in your uninfluenced judgment and decision making. Therefore, talking to other participants, looking at other people's screens, using the internet or personal devices that you brought along (e.g., cellphones) is strictly prohibited and will disqualify you from the experiment without pay. Please click on "Continue".

Screen 2: Introduction

The experiment consists of three sections:

- Section 1: 14 mathematical questions without calculator
- Section 2: 14 mathematical questions with calculator
- Section 3: Concluding questionnaire

In Section 1, you are supposed to estimate the answer to the 14 mathematical questions. You may use the pen and paper provided by the experimenter. No calculators are allowed for Section 1. In Section 2, we will ask you the same 14 mathematical questions, only this time you are allowed to use a calculator. Please notify the experimenter by raising your hand once you have completed Section 1. The experimenter will come to your place and give you a standard pocket calculator. Please click on "Continue".

Screen 3: How you will be paid

Upon the completion of the experiment, you will be paid a 10 Euro show-up fee. Additionally, you will receive a second payment according to the precision of your answers to the mathematical questions. The following formula is used to determine your payoff:

$$\text{your payoff} = 15\text{ Euro} - 15\text{ Euro} \cdot AEP$$

AEP stands for Absolute Error Percentage and is a measure for the accuracy of your answer. It is determined by the formula:

$$AEP = \frac{|\text{your answer} - \text{correct answer}|}{\text{correct answer}}$$

Please click on "Continue".

Screen 4: How you will be paid

Let's assume that the correct answer is 100.00 for the following examples:

- If your answer is 50.00, your absolute deviation from the correct answer is 50% (i.e., AEP = 50%) and your payment will therefore be 15 Euro – 15 Euro · 50% = 7.50 Euro.
- If your answer is 100.00, your absolute deviation from the correct answer is 0% (i.e., AEP = 0%) and your payment will therefore be 15 Euro – 15 Euro · 0% = 15.00 Euro.
- If your answer is 150.00, your absolute deviation from the correct answer is 50% (i.e., AEP = 50%) and your payment will therefore be 15 Euro – 15 Euro · 50% = 7.50 Euro.

You immediately see that you get the highest payment for giving the correct answer. Over- and underestimations lead to a lower payment. Please note that the AEP measure is restricted by the program to not exceed 100%. Therefore, over- or underestimating the correct answer by more or less than 100% will not result in a negative payment. Please refer to the printed table on your desk to see how other AEPs will affect your payment. Please click on "Continue".

Screen 5: How you will be paid

As noted earlier, we will ask you the same 14 questions in Sections 1 and 2. In Section 1, you are not allowed to use a calculator. In Section 2, you are allowed to use a calculator. We will not pay you according to the precision of your answer to all 14 questions, but only according to the precision of your answer to one question. The relevant question will be determined randomly after the experiment, so think carefully about all your answers. To determine the question that will be paid out to you, you will draw a ball out of an urn with 14 balls (numbered from 1 to 14). Please click on "Continue".

Screen 6: How you will be paid

Once you have drawn the ball out of the urn, we still have to determine whether you will be paid according to your Section 1 answer (without calculator) or your Section 2 answer to the same question (with calculator). To do so, we will ask you after each question in Section 1 for a maximum fee (in Euro) that you are willing to pay in order to be paid according to your answer to the same question in Section 2 (when the calculator is available) rather than to be paid according to the answer that you just gave in Section 1 (when no calculator is available). The fee that you state may lie between 0 and 15 Euro. After the experiment, the computer will randomly select an effective fee that also lies between 0 and 15 Euro (all numbers are equally likely to occur).

- If the effective fee that was selected by the computer is lower than the fee that you have stated, you will be paid according to your answer with the calculator. In this case, the effective fee will be subtracted from your payment.
- If the effective fee that was selected by the computer is higher than the fee that you have stated, you will be paid according to your answer without the calculator. In this case, the effective fee will not be subtracted from your payment.

If the effective fee that was selected is higher than the variable payment that you would receive for your answer to the mathematical question, you will receive a variable payment of 0 Euro (i.e., the payment cannot become negative). Please click on "Continue".

Screen 7: How you will be paid

We ask for your maximum acceptable fee to learn about the strength of your preference:

- If you state a low fee, this indicates that you are quite certain that your answer in Section 1 (without calculator) is almost as accurate as the answer that you will give in Section 2 (with calculator).
- If you state a high fee, this indicates that you are quite certain that your answer in Section 1 (without calculator) is much less accurate than the answer that you will give in Section 2 (with calculator).

The payout mechanism used in this experiment is called the Becker-DeGroot-Marschak mechanism. It is known to be incentive compatible and it is therefore in your best interest to truthfully state your maximum acceptable fee. We will not go into much detail here, but you can trust us that you cannot gain from consciously misspecifying your preference. It is always in your best interest to give truthful answers to all questions. Please click on "Continue".

Screen 8: Confidence in your answers

After each question, we will additionally ask you about your confidence in the answer that you just gave. Specifically, we will ask you: *What do you think is the probability that your answer falls within a ±10% interval around the correct answer?* You may answer this question by stating a probability between 0 and 100%. You will also get additional money for answering this question. The amount of money that you will receive is determined by a so-called Quadratic-Scoring-Rule. Please click on "Continue".

Screen 9: Confidence in your answers

If your answer falls within a ±10% interval around the correct answer, you will additionally receive:

$$4 \text{ Euro} \cdot (2 \cdot p - p^2)\ ,$$

where p is the probability that you have stated. If your answer does not fall within a ±10% interval around the correct answer, you will additionally receive:

$$4 \text{ Euro} \cdot (1 - p^2)\ .$$

The overall payout for your stated probability ranges from 0 to 4 Euro and is paid on top of all other payments. We will again not go into much detail here, but you can trust us that you cannot gain from consciously mis-specifying your beliefs. You maximize your expected payment if you truthfully state the probability that you believe. Please click on "Continue".

Screen 10: Summary

To sum up, you will be paid in the following ways:

- You will receive a 10 Euro show-up fee for completing the experiment.
- After the experiment, you will draw a ball out of an urn with 14 balls to determine the question that will be paid out to you. This variable payment may be up to 15 Euro and will be paid out in addition to the 10 Euro show-up fee.
- After each question in Section 1 (without calculator), we will ask for your maximum acceptable fee in order to be paid according to your answer in Stage 2 (with calculator) rather than your answer to the same question in Section 1 (without calculator). Then, the computer randomly selects an effective fee: If your stated fee is higher than the effective fee, you will be paid based on your Section 2 answer (minus the effective fee). If your stated fee is lower than the effective fee, you will be paid based on your Section 1 answer (the fee is irrelevant in this case).
- Finally, we will ask you after each question in Section 1 and 2 about the confidence in your answer. You will get up to 4 Euro based on the probability that you have stated and whether your answer falls within the ±10% interval around the correct answer

Please click on "Continue".

Screen 11: Section 1

You have now completed the tutorial. Please click on "Continue" to start Section 1.

Screens 12-25: Questions Section 1 (without calculator)

Question texts are specified in Appendix C.

Screen 26: Section 2

You have now completed Section 1. Please notify the experimenter by raising your hand to receive a calculator. After having received a calculator, click on "Continue" to begin with Section 2.

Screens 27-40: Questions Section 2 (with calculator)

Question texts are specified in Appendix C.

Screen 41: Questionnaire

You have now completed Section 2. Please click on “Continue” to begin the concluding questionnaire.

Screens 42: Questionnaire

Now we want to learn more about you. Your answers will help us to draw more meaningful conclusions from the answers that you gave in Sections 1 and 2. It is very important for us that you take your time to truthfully answer all questions. Your answers will be treated strictly confidential and will be used for research purposes only. Please click on “Continue”.

Screen 43-54: Questionnaire

Questions on...

- Age, gender, nationality, major, personal financial situation
- Financial literacy core items (Lusardi and Mitchell 2017)
- Numeracy test (Weller et al. 2013)
- Cognitive reflection test (Frederick 2005)
- Rational-experiential inventory 10 item test (Norris et al. 1998)
- Understandability of experimental instructions

Possibility to provide comments and feedback (free text).

Appendix C: Question texts.

Block 1: Tier 1 and Tier 2, forward-looking:

Tier 1: Today, you invest 10,000 Euro in a stock mutual fund that is expected to return a constant annual return of 10% for the next 32 years. Assume no additional deposits of withdrawals. Growth is compounded annually and reinvested into the account. Based on the above information, estimate the value of this stock mutual fund after 32 years.

Tier 2: Alternatively, a bond mutual fund grows at a constant annual return of 5% over the same 32-year period of time. Growth is compounded annually and reinvested into the account. Assume that you invest in only the stock fund over this 32-year period, estimate how much more money (in absolute terms) you would have in your account compared to investing exclusively in the bond fund.

Block 2: Tier 1 and Tier 2, backward-looking:

Tier 1: Your goal is to have 100,000 Euro saved 30 years from today. Today, you invest in a stock mutual fund that is expected to return a constant annual return of 10% for the next 30 years, before expenses. Assume no additional deposits or withdrawals. Interest is compounded annually and reinvested into the account. How much do you need to invest today in order to reach your savings goal in 30 years?

Tier 2: Alternatively, a bond mutual fund grows at a constant annual return of 5% over this same 30-year period of time. Growth is compounded annually and reinvested into the account. Assume that you invest in only the stock fund over this 30-year period, estimate how much less money you would need to invest today in order to reach your savings goal in 30 years compared to investing exclusively in the bond fund.

Block 3: Tier 1 and Tier 3, forward-looking:

Tier 1: Today, you invest 10,000 Euro in a mutual fund that is expected to return a constant annual interest rate of 10% for the next 36 years. Assume no additional deposits of withdrawals. Interest is compounded annually and reinvested into the account. Based on the above information, estimate the final portfolio value after 36 years.

Tier 3: Now, assume that this mutual fund charges a 2% management fee per year (this fee is applied to the total value of the portfolio at the end of each year). Based on all of the above information, estimate the final portfolio value after 36 years.

Block 4: Tier 1 and Tier 3, backward-looking:

Tier 1: Your goal is to have 100,000 Euro saved 35 years from today. Today, you invest in a mutual fund that is expected to return a constant annual interest rate of 10% for the next 35 years, before expenses. Assume no additional deposits or withdrawals. Interest is compounded annually and reinvested into the account. How much do you need to invest today in order to reach your savings goal in 35 years?

Tier 3: Now, assume that this mutual fund charges a 2% management fee per year (this fee is applied to the total value of the portfolio at the end of each year). Based on all of the above information, how much do you need to invest today in order to reach your savings goal in 35 years?

Block 5: Tier 1 and Tier 3, forward-looking:

Tier 1: Assume that 1 apple cost 1 Euro today. Due to inflation, prices of apples will rise by 4% per year over the next 36 years. What will be the price of 1 apple in 36 years?

Tier 3: Assume that you can invest the 1,000 Euro in a savings account that will yield a constant return of 8% over the next 36 years. How many apples will you be able to buy in 36 years with the money that you invested into the account?

Block 6: Tier 1 and Tier 3, backward-looking:

Tier 1: Assume that 1 apple cost 1 Euro today. Due to inflation, prices of apples rose by 3% per year over the next 36 years. What was the price of 1 apple 36 years ago?

Tier 3: Assume that you have 1,000 Euro in your savings account today. Money in the savings account grew by 6% per year over the past 36 years. You did not make any deposits in or withdrawals out of that savings account during the past 36 years. How many apples were you able to buy with the money that was in that savings account 36 years ago?

Block 7: Tier 4, forward-looking:

Tier 4: Today, you decide to contribute 1,000 Euro per year in an account that earns a constant annual rate of 9%. Assume no additional deposits or withdrawals. Interest is compounded annually and reinvested into the account. Based on the above information, estimate your total account balance after 36 years.

Block 8: Tier 4, backward-looking:

Tier 4: Today, you decide to contribute money each year in an account that earns a constant annual rate of 9%. Assume no additional deposits of withdrawals. Interest is compounded annually and reinvested into the account. If your goal is to have 1,000,000 Euro in your account in 32 years, how much do you need to contribute each year into this account, assuming that you contribute the same amount each year?

Appendix D: CDFs of EP in Stages 1 (no calculator) and 2 (with calculator).

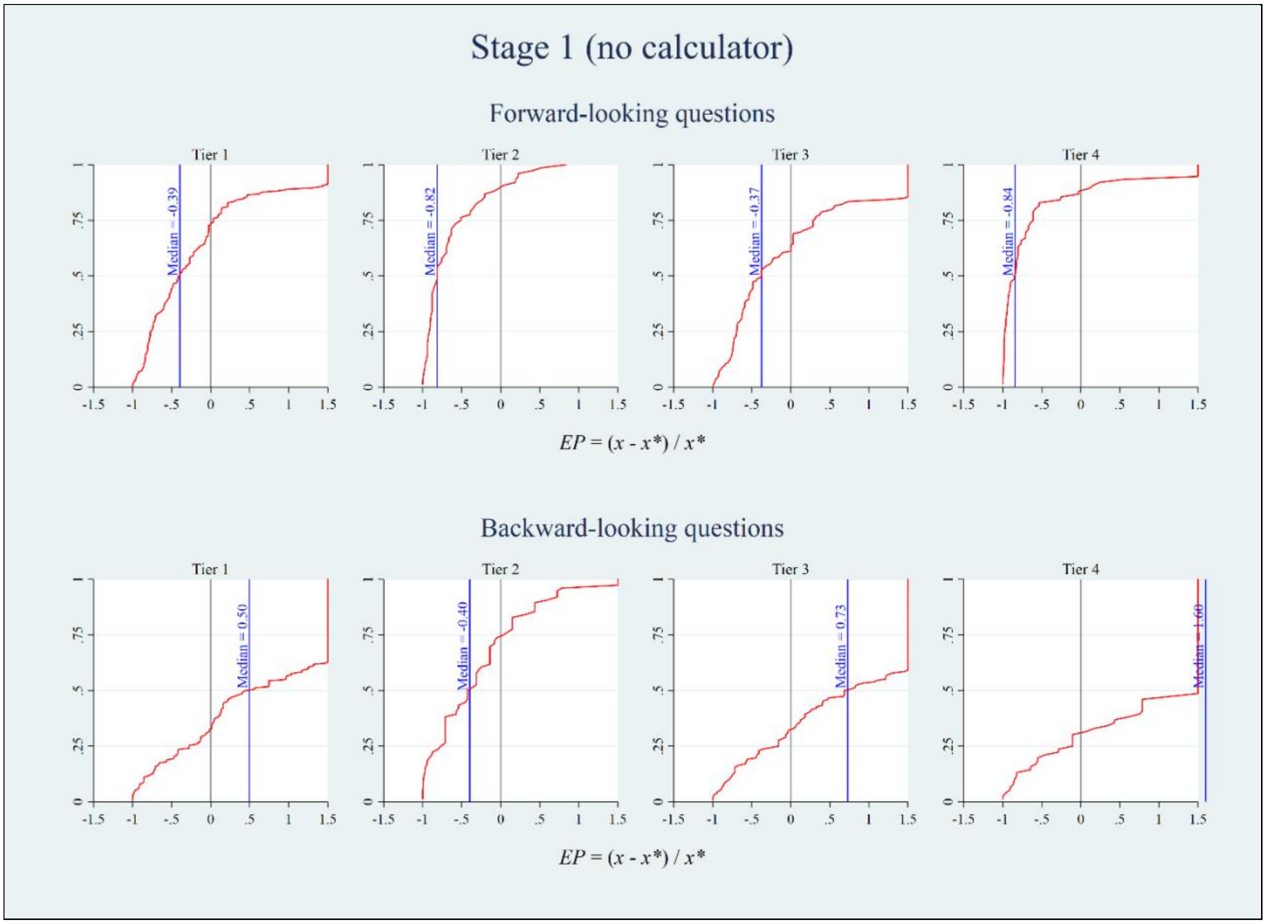

Figure D.1: The fugure shows cumulative distributions of EP in Stage 1 (without calculator). The distribution is winsorized between -1.5 and 1.5 in this figure.

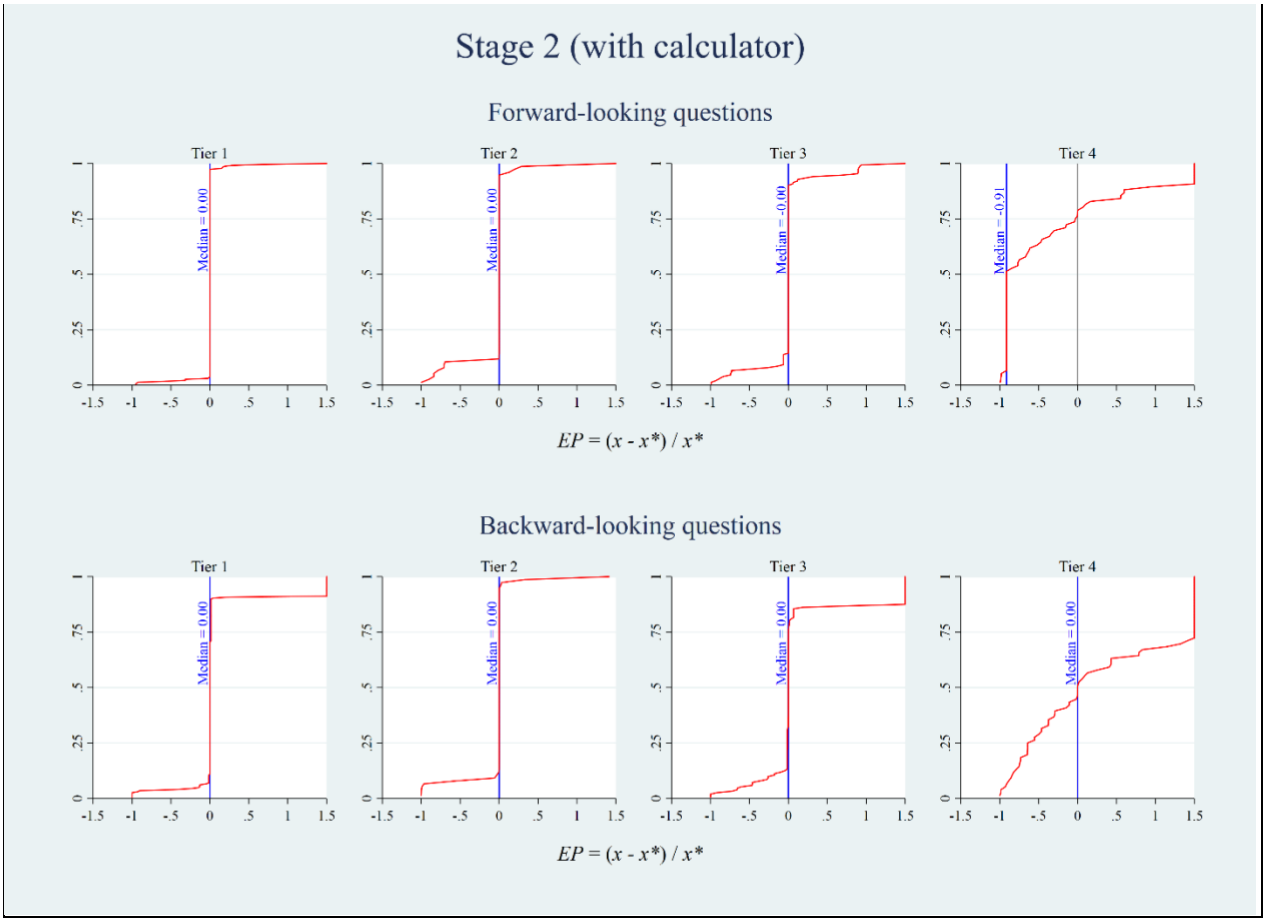

Figure D.2: The fugure shows cumulative distributions of EP in Stage 2 (with calculator). The distribution is winsorized between -1.5 and 1.5 in this figure.

Appendix E: Descriptive statistics of key variables.

Table E.1: AEP, proper answers, and confidence levels in Stages 1 (no calculator) and 2 (with calculator).

		Panel A: Median AEP			Panel B: Fraction proper answer			Panel C: Average confidence		
Tier	N	Stage 1	Stage 2	Diff.	Stage 1	Stage 2	Diff.	Stage 1	Stage 2	Diff.
All	1,064	76.15%	0.00%	-76.15%	7.90%	74.53%	66.63%	37.26%	81.17%	43.91%
				(0.00)			(0.00)			(0.00)
1	456	73.95%	0.00%	-73.95%	10.96%	89.47%	78.51%	41.77%	88.63%	46.86%
				(0.00)			(0.00)			(0.00)
2	152	71.28%	0.00%	-71.28%	5.26%	86.84%	81.58%	32.09%	83.39%	51.30%
				(0.00)			(0.00)			(0.00)
3	304	73.03%	0.10%	-72.93%	7.57%	78.29%	70.72%	33.58%	79.86%	46.28%
				(0.00)			(0.00)			(0.00)
4	152	94.12%	91.35%	-2.77%	1.97%	9.87%	7.90%	36.24%	59.19%	22.95%
				(0.00)			(0.00)			(0.00)

Median AEP, fraction of proper answers, and average confidence, by Tiers. An answer is considered “proper” if AEP < 10% (Stage 1) or AEP_c < 10% (Stage 2) respectively. Two-sided p-values are provided in parentheses and are based on Wilcoxon signed rank-tests, McNemar-tests, and paired Student t-tests respectively.