Matthew Levy and Joshua Tasoff

Expotential-growth bias and overconfidence

Article (Accepted version) (Refereed)

Original citation: Levy, Matthew R. and Tasoff, Joshua (2017) *Exponential-growth bias and overconfidence.* Journal of Economic Psychology, 58 . pp. 1-14. ISSN 0167-4870
DOI: 10.1016/j.joep.2016.11.001

Exponential-Growth Bias and Overconfidence[a]

Matthew R. Levy[b]
London School of Economics

Joshua Tasoff[c]
Claremont Graduate University

November 3, 2016

Abstract

There is increasing evidence that people underestimate the magnitude of compounding interest. However, if people were aware of their inability to make such calculations they should demand services to ameliorate the consequences of such deficiencies. In a laboratory experiment, we find that people exhibit substantial exponential-growth bias but, more importantly, that they are overconfident in their ability to answer questions that involve exponential growth. They also exhibit overconfidence in their ability to use a spreadsheet to answer these questions. This evidence explains why a market solution to exponential-growth bias has not been forthcoming. Biased individuals have suboptimally low demand for tools and services that could improve their financial decisions.

Keywords: exponential-growth bias, overconfidence, financial literacy, overestimation, overprecision
JEL: D03, D14, D18

---

[a] We would like to gratefully acknowledge the financial support of a Fletcher Jones Foundation Faculty Research Grant from Claremont Graduate University. We are grateful to Masyita Crystallin, Peiran Jiao, Andrew Royal, Quinn Keefer, and Oliver Curtiss for research assistance. We thank Ananda Ganguly, Matthew Rabin, Paige Skiba, Justin Sydnor, Charles Thomas, Jonathan Zinman, and various seminar participants for helpful comments. Some results in this paper were previously circulated in a working-paper version of "Exponential-Growth Bias and Lifecycle Consumption". Institutional Review Board approval was obtained from OHRPP at UCLA [IRB #12-001092] and CGU [IRB #1591].

[b] London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Phone: +44 (0)20 7955 6596. Email: m.r.levy@lse.ac.uk

[c] Claremont Graduate University, Department of Economics, 160 E. Tenth St, Claremont, CA 91101, United States. Phone: (909) 621-8782. Fax: (909) 621-8460. joshua.tasoff@cgu.edu

1 Introduction

> *Chess ... was invented for the entertainment of a king who regarded it as a training in the art of war. The king was so delighted with the game that he offered the inventor any reward he chose to name. The latter said he only wished to have the amount of corn resulting from placing one grain on the first square, two on the second, and so on, doubling the number for each successive square of the sixty-four. This sum, when calculated, showed a total number of grains expressed by no less than twenty figures, and it became apparent that all the corn in the world would not equal the amount desired. The king thereupon told the inventor that his acuteness in devising such a wish was even more admirable than his talent in inventing [chess].* — A.A. Macdonell, "The Origin and Early History of Chess", *Journal of the Royal Asiatic Society*, January 1898, 30(1): pp. 117–141.

Exponential-growth bias (EGB) refers to the systematic tendency to underestimate compound growth processes (Stango and Zinman, 2009). The passage above highlights the unintuitive difficulty in perceiving exponential growth. But this misperception also implicates a second and equally important error; *the king is surprised by the magnitude of his misperception.* We refer to this mistake as overconfidence in exponential estimation, and it is the focus of this paper.

Misperceptions of fundamental financial processes may at first seem to lead to important inefficiencies. However, conventional economic thinking suggests that a well-functioning market should solve this problem. In the same way that people are not fully self-sufficient in modern society but acquire most needs through markets, a person could in principle do the same for their financial decisions. A financially unskilled agent could simply outsource financial decisions to an expert, and a competitive market for advice would eliminate the effect of EGB on financial decisions. But if people are not self-aware of their misperceptions they will exhibit insufficient demand for corrective tools and advice. Thus the presence of overconfidence in exponential estimation is fundamental to the welfare relevance of EGB.

In our experiment, subjects answer questions that involve exponential growth and are paid based on their accuracy. They may obtain either a spreadsheet or the true answer to improve their payment. We obtain incentive-compatible measures of subjects' willingness to pay (WTP) for the spreadsheet and for the correct answer. A risk-averse subject who expects to lose x on a question relative to the maximal earnings should be willing to pay at least x for the correct answer, and strictly more than x if she is strictly risk-averse over the experimental stakes. Any disutility from answering the questions without aid would further increase her WTP.

We find that subjects exhibit a high degree of EGB. We use the model of Levy and Tasoff (2016) to parameterize the accuracy of subjects' perceptions as α, where $\alpha = 1$ implies the person correctly perceives exponential growth and $\alpha = 0$ implies that a person perceives exponential growth as linear. The average α is 0.65 which is slightly higher than the distribution in a representative sample of the population US (Levy and Tasoff, 2016) and represents unawareness of 2/3 of the impact of compounding. Average performance earnings across the control group were \$11.33. Given that maximum earnings were \$25, and the correct answer, barring trembles, guarantees maximum earnings, the optimal average WTP should thus be at least \$13.67. Instead we find that the average WTP for the correct answer is only \$5.70. We construct a normalized measure of overconfidence defined as (Optimal WTP − Actual WTP)/\$25, where the Optimal WTP is defined conservatively as the earnings-maximizing WTP (\$25 − actual earnings without help). This gives us a measure on $[-1, 1]$ where 1 signifies that the person earned \$0 without help but believed he would earn the maximum, and -1 signifies that the person earned \$25 without help but believed he would earn \$0. The mean overconfidence is 0.31, and 86% of subjects exhibit overconfidence. Thus overconfidence may act as a significant damper on the market's ability to correct EGB through professional advice. Moreover we find evidence of "illusory superiority" (Kruger and Dunning, 1999), as less skilled individuals — i.e. those with lower α — tend to be the most overconfident. This suggests a pathological selection in the marketplace, whereby the people who need help the most have the lowest demand for it.

Given these findings, one may expect sub-optimally low demand for the spreadsheet as well. Surprisingly, we find the reverse. The spreadsheet had no significant positive impact on subject performance. Consequently, any positive WTP for the spreadsheet indicates a different type of overconfidence: overconfidence in one's ability to use the spreadsheet. The average WTP for the spreadsheet is \$4.59, indicating an average overconfidence in ability to use the spreadsheet of 0.165. We find that 75% of subjects in the spreadsheet group have significant overconfidence in their ability to use the spreadsheet. The experiment shows that people are willing to pay for tools that do not in practice improve their performance. That is, not only does this tool not improve performance, it lowers overall earnings because subjects are willing to pay for it.

Recent work has shown that EGB is widespread and correlated with important financial outcomes. Many psychology and economics papers, beginning with Wagenaar and Sagaria (1975), have shown robust evidence for EGB in the lab (Wagenaar and Timmers, 1979; Keren, 1983; Benzion, Granot and Yagil, 1992; MacKinnon and Wearing, 1991; Eisenstein and Hoch, 2007; McKenzie and Liersch, 2011). Stango and Zinman (2009) use the 1977 and 1983 Survey of Consumer Finances and find that those with a larger error on a question about interest rates have higher short-term debt to income ratios, lower stock ownership as a percentage of portfolios, lower savings rates, lower net worth, and no difference in long-term debt to income ratios. Goda et al. (2015) measure EGB in a representative sample of the US population and find that about one third are fully biased meaning that they perceive compound interest as simple interest. They find that all else equal, the un-biased type is associated with 40% more retirement savings at retirement age than a fully biased type than a fully biased type, or \$50,000 in absolute terms. It is therefore important to understand the mechanism underpinning the correlation.

While the evidence is mostly correlational, there are reasons to believe that some of these associations are causal in nature. First, theory predicts that EGB can lead to sub-optimally low savings. The model of Stango and Zinman (2009) predicts the correlations found in their empirical analysis, and the lifecycle-consumption model in Levy and Tasoff (2016) predicts that biased people will likely undersave. Second, there is observational evidence that a law designed to curb the negative effects of EGB heterogeneously affected consumers as a function of their bias. Stango and Zinman (2011) find that people who make larger errors on an interest rate question in the Survey of Consumer Finances have higher APR's on their personal loans prior to mandated APR disclosure. Mandated disclosure then compressed the interest rates on the loans of the relatively biased and unbiased people. Regulation seems to have prevented firms from price-discriminating on borrowers' cognitive biases.

These associations require some explanation for why market solutions have not eliminated the problem. This is the first paper to present evidence that people are unaware of exponential-growth bias, and therefore exhibit sub-optimally low demand for financial advice and tools that could improve their financial decisions. Overconfidence is a widespread cognitive bias but not universal. A common finding in the overconfidence literature is that people tend to be overconfident on hard tasks but under-confident on easy tasks (see Moore and Healy, 2008, for a discussion of the literature). Given that both over- and underconfidence are observed depending on the task, it is not at all obvious that people are overconfident about their exponential estimation. Lusardi and Mitchell (2014) find that confidence in financial literacy tends to be high even in samples that have low actual literacy rates. Souleles (2004), Puri and Robinson (2007), Hyytinen and Putkuri (2012) find that people's overoptimism predicts consumption, non-prudent financial behavior, and debt. Although these results are suggestive, the subjective Likert measures used to measure overconfidence and overoptimism in these studies are un-incentivized and cannot quantify magnitude due to the arbitrary scale of the Likert measures.[1] Moreover, these measures are silent on whether people are overconfident about their exponential estimation ability per se, or about other features of financial decision-making.

The extent of demand for advice appears insufficient to eliminate a cor-

[1]The exception is Puri and Robinson (2007), who measure overoptimism by comparing subjective life expectancy to actuarial projections.

relation between EGB and financial outcomes in cross-sectional analysis. Given the pervasiveness of EGB, if consumers were aware of their inability to evaluate financial products, there should be large demand for services and tools that could help biased consumers to make sound financial decisions. Although such services do exist it seems that consumer biases persist in the unfettered marketplace (Stango and Zinman, 2011). Agency problems may cause partial unraveling of advice markets. Mullainathan, Noeth and Schoar (2012) find that investment advisors encourage clients to incur unnecessary fees. But there is a supply of financial advisers that seem to offer sound honest advice to help clients reduce their debt, obtain low-interest loans, and save for retirement that appears to be free of significant agency problems. For example, some investment management companies have incentives well-aligned with their clients, and provide vast amounts of free guidance. The low take-up of these services is a puzzle that this paper addresses.

The next section presents the experimental design. Section 3 contains the results and Section 4 concludes.

2 Design

This study was conducted through the Center for Neuroeconomic Studies (CNS) at Claremont Graduate University on a sample of Claremont Colleges students in November and December 2011. There were 7 sessions and we recruited a total of 96 subjects. Like all laboratory-based experiments on university students, our results pertain to a potentially unrepresentative sample. Previous research has demonstrated that EGB is widely prevalent in the general population (Stango and Zinman, 2009; Levy and Tasoff, 2016; Goda et al., 2015), but it is possible that students are more or less overconfident. It was necessary to conduct the study under laboratory conditions rather than using a representative online panel, however, in order to exercise control over subjects' problem-solving resources. Subjects were not provided with any tools, and calculators/cell phones were expressly forbidden. The experimenter routinely checked upon the subjects to ensure that no one violated the rules. Time limits were not imposed for any part of the experiment. Most completed within 30-60 minutes. The design is a controlled abstraction that allows for a clean measure of people's expected performance.

To measure EGB, subjects answered questions that involved exponential growth, and to measure overconfidence subjects stated their WTP for a spreadsheet and for the answer, which could be purchased to improve responses. Overconfidence is measured by taking the difference between the WTP and the true value of the resource. The experiment began with comprehensive instructions and examples as well as a comprehension check that was necessary to correctly complete in order to proceed. The appendix contains the instrument. Figures A1–A4 display the instructions.

Subjects faced a series of 32 questions relating the growth of two hypothetical assets. The first asset's value was occluded by computations that included exponential growth, and the second asset always had a free variable X. The subjects' task was to estimate or compute the X that would make both assets equal after the specified number of periods. Subjects were paid based on the accuracy of a randomly-selected question.

Subjects were informed that one question would be chosen randomly by computer, and that they would receive an incentive payment based on the accuracy of their response to that question (in addition, subjects received a show-up fee of \$5.00). Subjects were informed of their payment on an exit screen at the end of the experiment; other than this, subjects received no feedback. The incentive payment used a quadratic scoring rule in accuracy, bounded below by zero. That is, if a subject i responded r_{ij} to a question on which the correct answer was c_j, then their payment would be:

$$\pi_{ij} = \max\left\{25 - 100 \cdot \left(1 - \frac{r_{ij}}{c_j}\right)^2, 0\right\} \tag{1}$$

Subjects were given examples of this payment rule in the instructions, and were provided with a table of payments corresponding to different percentage errors alongside every question (see Figure A2 for instance).

The 32 questions were divided across four domains, randomized first at the domain level and then within-domain. A list of all 32 questions is given in Table A1. The domains comprised the exponential, periodic savings domains, and fluctuating interest from Levy and Tasoff (2016), and an additional front-end load domain. Questions took the form:

- Exponential: "Asset A has an initial value of $\$P$, and grows at an interest rate of $i\%$ each period. Asset B has an initial value equal to $\$X$, and does not grow. What $\$X$ will make the value of asset A and B equal at the end of T periods?"

- Periodic savings: "At the beginning of each period, Asset A receives a $\$c$ contribution. These contributions earn $i\%$ interest every period, and Asset A includes both the contributions and the interest earned at the end. Asset B returns a fixed amount of $\$X$ at the end. What value of X will cause the two assets to be of equal value after T periods?

- Front-end load: "Asset A has an initial value of $\$P_A$, and grows at an interest rate of $i_A\%$ each period. Asset B has an initial value equal to $\$X$, and grows at an interest rate of $i_B\%$ each period. What $\$X$ will make the value of asset A and B equal at the end of T periods?"

- Fluctuating interest: "Asset A has an initial value of $\$P_A$, and grows at an interest rate of $i_{A1}\%$ in odd periods (starting with the first), and at $i_{A2}\%$ in even periods. Asset B has an initial value of $\$P_B$, and grows at $X\%$ per period. What value of X will cause the two assets to be of equal value after 20 periods?" In questions 29–32 (see Table A1) Asset B also experienced a fluctuating interest rate.

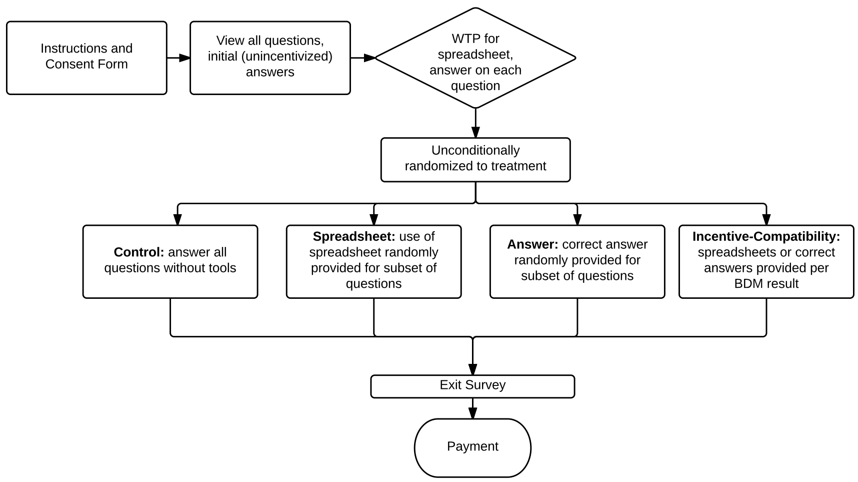
The variety of domains serves two purposes. First, it better captures the breadth of decisions that involve exponential growth in naturalistic environments. Second, the variety of domains also represents a range of more and less challenging questions. Whereas a spreadsheet may directly give the solution to the exponential domain, a bit more thinking is necessary to use it to solve the periodic-savings domain. This may make the value of the spreadsheet more noisy, just as it would be in naturalistic environments.

The experiment can be broken down into two phases. The design is presented in Figure 1. The purpose of the first phase was primarily to elicit subjects' WTP for the spreadsheet and the answer, but also recorded subjects' initial estimates of the value of an asset. Subjects indicated their WTP to receive the use of a spreadsheet and their WTP for the correct answer, on a question-by-question basis. The elicitation procedure was based on the Becker-DeGroot-Marshack mechanism to maintain incentive compatibility. Subjects were informed that with some exogenous probability, this mechanism would determine their receipt of help. Otherwise, they would either receive the correct answer, a spreadsheet, or no assistance on all questions independently of their stated WTP.

A variety of features were used to ensure subject comprehension. The instructions explicitly stated that the earnings-maximizing strategy was to enter the amount by which they expected having the correct answer would increase their earnings if the question were chosen, i.e. $\$25 - \mathrm{E}\left(100 \cdot \left(1 - \frac{r_{ij}}{c_j}\right)^2\right)$, and were given examples of how under-bidding and over-bidding were dominated strategies. Moreover, subjects could not proceed until they passed a comprehension test. Subjects were asked what bid would maximize their expected earnings if they thought their answer with no help would earn $9.50. Only once they correctly answered that a WTP of $15.50 would maximize expected earnings were they allowed to exit the instructions and proceed to the experiment. We would expect that, if anything, this would anchor their stated willingness-to-pay at $15.50 if subjects interpreted this example as containing information about their expected performance (this would have biased behavior away from overconfidence).

In the first phase subjects also gave an initial un-incentivized estimate of the value of the asset (see Figure A6). This was done to ensure they actively considered the question when deciding on their WTP for help. We will focus on subjects' incentivized responses from the second phase in our empirical analysis, but note that the initial responses may also be used to help measure treatment effects (i.e. the impact of the spreadsheet on performance), with the caveat that subjects understood these initial estimates would not count

Figure 1: Experiment Design



for payment.[2]

Prior to the second phase subjects were randomized into one of four treatments. Subjects in the control group were given neither the spreadsheet nor the correct answer, regardless of their WTP. Subjects in the spreadsheet treatment were given the spreadsheet, regardless of their WTP. Two auxiliary treatments were also used. To verify that subjects would in fact earn full payments when given the answers, a small number of subjects were allocated to the answer treatment and given the correct answer on some — though not all — questions, regardless of their WTP. Finally, to ensure strict incentive compatibility for the WTP-elicitation task, some subjects were randomized into the incentive-compatibility group and could purchase the spreadsheet or the correct answer at a randomly-drawn price X if it was below their indicated WTP, and receive no help and would not pay anything if X was

[2]An alternative design would have included incentivized, unaided questions for all subjects as a measure of baseline ability, and then elicited WTP for help on a different set of questions. In pilots, it was found that earnings varied considerably from question to question — in the final sample, average performance earnings by question varied from \$3.96 to \$19.98 — and so a baseline assessment on separate questions was deemed uninformative.

above their WTP.[3] Which resource could be purchased was randomly assigned. Subjects then proceeded to give final responses to the 32 questions, one of which would be selected for payment. Subjects who received help on some questions first gave final answers to questions for which they did *not* receive aid and then the questions for which they *did*. Subjects in the IC group faced identical conditions to those in the control until the first question for which they received help, when they could infer that they were indeed in the IC group.

Observing both the un-aided responses and initial willingness-to-pay enables us to calculate an index of overconfidence. What we would like to calculate is the difference between subjects' own expectations of their unaided earnings, and the rational expectation. The first ingredient, subjects' own expectations of their unaided earnings, is embedded in their willingness-to-pay: $E(p_{ij}) = \$25 - WTP_{ij}$, where p_{ij} denotes the earnings of subject i on question j.[4] To estimate the second ingredient, the "rational" expectation, our preferred method uses the earnings that would have resulted from that subject's actual response to a given question. While performance may be unexpectedly high or low on any particular question, averaging across all 32 questions will yield an unbiased estimate (if errors are uncorrelated)– since actual performance is what rational expectations must predict. Our overconfidence index is thus $\left[[\$25 - WTP_{ij}] - p_{ij}\right]/25$.[5]

The difficulty of a question will affect performance and could thereby affect overconficence. Moore and Healy (2008) show that people tend to

[3] To avoid anchoring subjects' responses, they were not informed of the distribution from which X was drawn. In practice, a uniform distribution over the [0,25] interval was used.

[4] This is a lower bound. Beliefs about earnings will be higher if subjects are risk-averse or have a disutility from effort from answering the questions.

[5] For example, consider someone who (correctly) expected to earn \$5 half of the time and \$25 half of the time, for \$15 on average, and stated a WTP on every question of \$15. We would characterize the person as "overconfident" by \$10 on the half of questions where she only earned \$5, as "underconfident" on the half of questions where they earned \$25, and overall as neither under- nor overconfident. As defined, overconfidence is bounded as a function of a person's performance. A person who answered perfectly on every question , and is thus revealed to be maximally competent could not exhibit overconfidence. Similarly, a person who earned \$0 on every question could not exhibit underconfidence.

overestimate their performance on difficult tasks and underestimate their performance on easy tasks. Therefore the degree of overconfidence measured in this paper might not reflect the degree of overconfidence found in the market. However, there is good reason to believe that the questions we use are a good deal easier than the typical and important financial decisions that consumers face thereby understating market overconfidence. Real-world decisions such as retirement savings, debt repayment, and portfolio choice all require multiple steps, multiple time horizons, and the incorporation of risk and uncertainty. Moreover, the extent to which the presentation of the problems in this experiment was unfamiliar to subjects should increase rather than decrease their WTP out of insurance motives.

The existence and randomization into the four treatment groups was clearly explained to subjects prior to the WTP elicitation task, although the probability of assignment into each treatment was not stated. The control group allowed for measuring performance, without help, on a random sample (i.e. without selecting those least willing to pay for help). The spreadsheet group served the same purpose, measuring performance with a spreadsheet on a random sample. These two groups are the focus of our analysis. The "answer" group was included to demonstrate that subjects could indeed use the correct answer if it were provided. Finally, the IC group was included so that subjects faced strict incentives to truthfully report their WTP in the first phase of the experiment. Because the IC were more likely to receive assistance on questions for which they had a greater demand, we exclude this self-selected sample from the main analysis.

The spreadsheet used was a custom application that was integrated into the experimental interface. It had cells and allowed for all arithmetic calculations needed to compute the correct answer and a user interface similar to Microsoft Excel, although it was based on an open-source alternative. Subjects were not given access to the spreadsheet before expressing their WTP. Practice computations could have primed subjects, especially if those practice computations involved expressions with exponentiation. The cost of avoiding such bias is that expectations about the utility of the spreadsheet are less well controlled. For this reason, overconfidence regarding the spreadsheet may be considered a secondary analysis, and the focus of this paper is squarely on overconfidence regarding the correct answer.

There were 52 subjects who were assigned to the control group. The spreadsheet treatment had 38 subjects, and the answer group had 2 subjects.[6] The remaining 4 subjects were assigned to the incentive-compatibility group. We drop 4 subjects in the control and 2 subjects in the spreadsheet group who had answers that did not allow for a convergent estimation of exponential-growth perception α leaving 48 in the control and 36 in the spreadsheet treatment; all results on overconfidence are robust to their inclusion.

At the end of the experiment there was a brief exit survey. Subjects were asked if they had, "taken at least 2 advanced math courses in college", "personally own stocks, bonds, or mutual fund shares", "use a credit card, but pay the balance in full each month", "use a credit card, but do not pay the balance in full each month", "am a current smoker", and "none of the above".

3 Results

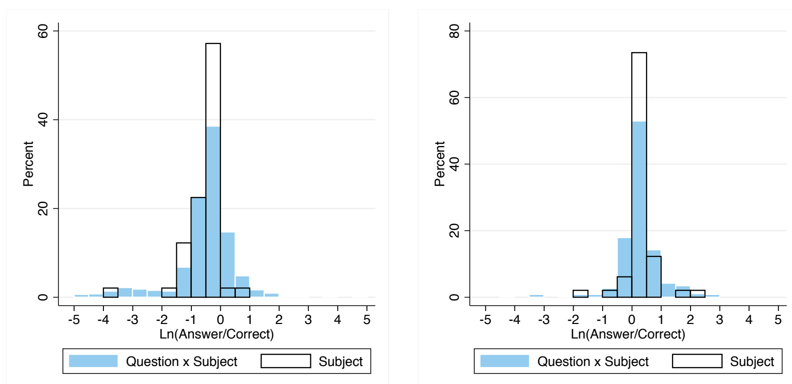
3.1 Exponential-Growth Bias

We first confirm that subjects are systematically biased in the direction predicted by exponential-growth bias. Figure 2 plots the distribution of log errors at the question$\times$subject and subject level for each of the 48 subjects' responses to each of the 32 questions. We are left with 1481 subject-question observations after dropping questions that subjects left blank. In panel (a), where under-estimation is predicted by EGB, the median at the question level (-0.34) and at the subject level (-0.42) are significantly negative ($p<0.01$), and in panel (b) where over-estimation is predicted the median

[6] The purpose of the answer group was to address the potential concern that subjects might not be competent to copy the answers into the response box. Ex ante we believed this highly unlikely. Hence we reserved as much of the subject pool as possible for the control and spreadsheet groups, treating the answer group as a nod toward common sense rather than formal statistical testing.

at the question level (0.19) and subject level (0.25) are significantly positive (p<0.01). The means are similarly significant (p<0.01), confirming that subjects' responses are systematically biased in the direction predicted by the theory.

Figure 2: Mistakes



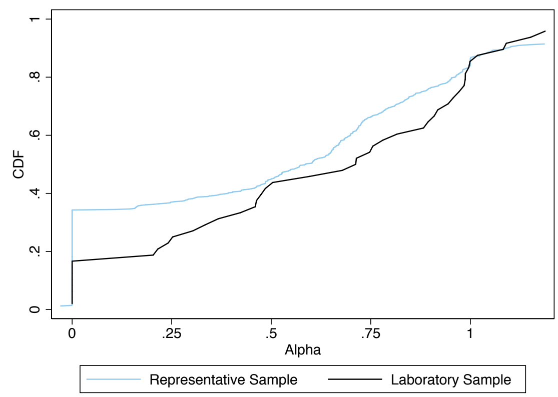
(a) Under-Estimation

(b) Over-Estimation

Notes: Underestimation on questions where EGB predicts a downward-biased answer (Panel a); overestimation on those where an upward bias is predicted (Panel b). The distribution of errors in predicted asset growth should be symmetric about zero if subjects' errors on a percentage basis are symmetric about zero. The means of all distributions are significantly different from zero at both the question and subject levels ($p < 0.01$)

We next use each subject's combined responses to estimate an individual-level degree of EGB corresponding to α in Levy and Tasoff (2016), where $\alpha = 0$ constitutes the fully-biased type and $\alpha = 1$ the fully rational type, and values of $\alpha \in (0, 1)$ constitute intermediate levels of bias. More specifically, an EGB agent of degree α mis-perceives the period-T value of \$1 invested at t and growing according to a vector of interest rates $\vec{i}$ as: $p(i, t; \alpha) = \prod_{s=t}^{T-1}(1 + \alpha i_s) + \sum_{s=t}^{T-1}(1 - \alpha) i_s$. We estimate for each subject the α which minimizes the sum of squared errors between their actual responses and the responses predicted for someone with degree α of EGB. We then compare the distribution of bias in our student sample to the general population in Figure 3. Subjects in the control group of our student sample had more accurate perceptions than representative samples of the US population (Levy and Tasoff, 2016; Goda et al., 2015). The mean α for the students is 0.65 with a median of 0.71 compared to 0.60 and 0.53 for the representative sample. Only 16% of our student sample is fully biased, in contrast to 33% in representative US samples. Overall, the students outperformed the representative sample despite the fact that the representative sample was allowed to use tools and get help and the student control-group sample was not.

Figure 3: CDF of Alpha



Notes: Cumulative distribution of EGB ($\alpha = 0$ is full-bias, $\alpha = 1$ no-bias). Laboratory sample estimates subject-level parameter for control subjects using all 32 questions. Representative sample is from Levy and Tasoff (2016)

3.2 Overconfidence in Exponential Estimation

Our main results establish that subjects overestimated both their accuracy and their precision. We begin by demonstrating that subjects systematically stated a willingness to pay for the correct answer that was below the ex-post optimal level. We then show that the elicited WTP measures are too low to be justified by the observed level of precision.

3.2.1 Main Results

Table 1 shows subjects' performance and demand for answers and the spreadsheet by treatment. The first column includes all observations in the control condition, while the second column includes all observations in the spreadsheet treatment.

Table 1: Behavior by Treatment

	Control (1)	Spreadsheet (2)	Difference (3)
First-pass Hypothetical Earnings	7.886	6.860	-1.026*
	(9.741)	(9.393)	[0.554]
Performance Earnings	11.33	11.47	0.140
	(10.24)	(10.68)	[1.101]
Performance − First-pass Earnings	3.439	4.606	1.166
	(12.93)	(12.56)	[0.939]
WTP Answer	5.699	5.451	-0.248
	(4.511)	(3.998)	[0.762]
WTP Spreadsheet	4.517	4.587	0.070
	(4.270)	(4.067)	[0.742]
α	0.652	0.750	0.098
	(0.444)	(0.518)	[0.108]
N	1471	1144	
Subjects	48	36	

Notes: "First-Pass Hypothetical Earnings" is earnings if the first-pass answer counted for pay. "Difference" columns show treatment group minus control group. "Spreadsheet" columns include all observations in the spreadsheet treatment in which the subject got the spreadsheet. Standard deviations in parentheses and robust standard errors clustered by subject in square brackets. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

The first row shows what the earnings would have been if the (unincentivized) first-pass responses were used for payment. For these questions subjects in all groups faced identical tasks. Thus one should expect that behavior here is the same across treatments. We use this as a proxy of ability and engagement with the task, though with the caveat that subjects knew their answers here would not count for payment. Control hypothetical earnings on average would have been \$7.89. The spreadsheet group performed slightly worse with hypothetical earnings at \$6.86. To test the difference between the groups we run an OLS regression of the outcome variable on a dummy for the two non-control groups clustering the standard errors by subject. The spreadsheet group does seem to do slightly worse by chance (p=0.07). The second row displays performance earnings that subjects would earn if a given question were selected for payment The control group earns \$11.33. The spreadsheet group only earns \$0.14 more and it is not significantly different from the control earnings. This indicates that the spreadsheets had no significant impact on performance. The third row shows the difference between subjects' earnings from their final answer (ignoring any price paid for the spreadsheet or answer) and the earnings that would have resulted from their first-pass answers. Subjects in both the control and spreadsheet groups significantly improved their second, incentivized, responses relative to their initial responses. The third column indicates that the extent of improvement was greater for subjects who received spreadsheets than those who did not but the difference is not statistically significant.

The mean WTP for the correct answer was only \$5.70 in the control and \$5.45 in the spreadsheet group, indicating overconfidence. Mean WTP for the spreadsheet was \$4.52 in the control group, indicating overconfidence in ability to use the spreadsheet given that the actual value of the spreadsheet was not statistically different than zero. Mean WTP for the spreadsheet in the other groups is not statistically different than the control group.

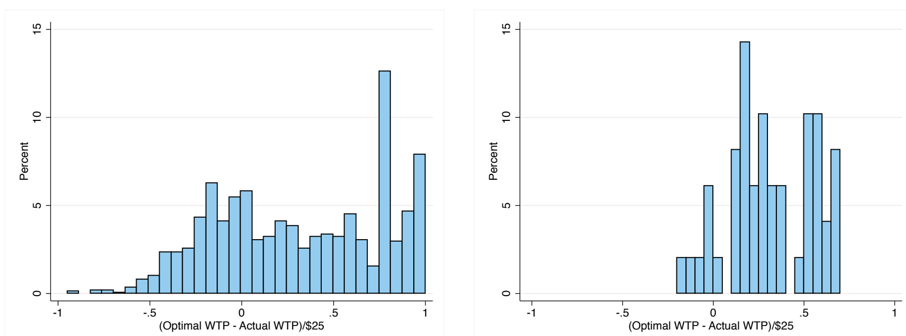
The last row in Table 1 shows the parametric estimate of subjects' degree of EGB, α. Subjects in the spreadsheet treatment displayed an average value of 0.75. This is higher than the 0.65 among controls, suggesting that the spreadsheet may have had some impact (particularly among those close to the correct answer, where earnings are relatively flat), although the difference is not significant.

The two subjects in the answer treatment are sufficient to quickly verify that subjects understood the task well enough to exploit the provided answers, although apparently not well enough to exclude some trembles (12 of the 14 observations achieved maximal earnings of \$25, and the average was \$24.57). Next we calculate overconfidence in exponential estimation, for every question in the control group, using the payment that a subject would have earned had that question been chosen for implementation according to the quadratic payment rule given by (1). Subjects answering exactly correctly would have an associated payment of \$25, while responses more than 50% from the correct answer would receive zero. The average associated payment across all 1481 subject-question pairs was \$11.28 (s.d. 10.25). If agents were risk-neutral over \$25 stakes, then the optimal strategy would be to state a willingness to pay for the correct answer of $WTP_{ij} = \$25 - E(p_{ij})$. Any concavity in utility would set this as a lower bound, as paying for the correct answer can be viewed as providing insurance for the earnings. Likewise, if subjects experience disutility from doing math then they should exhibit higher demand for the correct answer, and again this measure would be a lower bound for overconfidence. A simple test of whether subjects accurately predicted their performance is to compare the actual willingness to pay against this bound. Subjects may under- or over-pay on some questions, but by the law of large numbers the average willingness to pay across all questions should converge to $(\$25 - \bar{p}) = \$13.72$. Instead, the mean willingness to pay is significantly lower at \$5.76 ($p<0.01$). That is, subjects on average expect their answers to earn at least 40% more than they actually do.

Panel (a) of Figure 4 plots the distribution of overconfidence at the subject×question level. The depicted variable is the difference between the ex-post 'optimal' WTP (i.e. \$25 less the actual associated payment) and the stated willingness to pay for the answer, normalized by \$25. Thus a value of 1 indicates that a subject would pay \$0 for an answer to a question on which they would have earned no payment, and a value of -1 indicates that a subject would pay \$25 for an answer to a question on which they would have earned the full payment. This variable should be distributed about zero if subjects are risk-neutral, or some negative number if they are risk-averse. Instead, the distribution has a positive mean (0.318), and is skewed highly positive.

The second panel of Figure 4 helps establish that this result is driven by a large fraction of subjects being systematically overconfident across all questions. Panel (b) computes the mean of the under-payment variable from panel (a) at the subject level, and plots the distribution of this subject-level outcome. A subject who over-pays on some questions but under-pays on others would of course converge towards zero as we average over a large number of questions. Instead we find that both the mean (0.31) and median (0.28) are significantly overconfident ($p<0.01$), with 86% of subjects exhibiting positive overconfidence. This result thus also rules out an alternative hypothesis, that subjects correctly perceive the average ability but are unaware of their own. If that were the case, the subject-level overconfidence would again be symmetrically distributed around zero (those who are better than average being characterized as underconfident and vice-versa), which we reject.

Figure 4: Overconfidence — Underpayment for the Correct Answer



(a) Subject x Question Level

(b) Subject Level

Notes: "Optimal WTP" is defined as $25 less a subject's actual earnings on a question, and is therefore *ex post* optimal. Panel (a) shows the distribution of under-payment, and the mass weighted by the squared error should be equal on either side of 0 in the absence of systematic bias (or about some negative amount if subjects are risk-averse over $25 stakes). Panel (b) computes mean under-payment at the subject level, and should converge to a point mass at zero in the absence of systematic bias (or a mass at some negative amount if subjects are risk-averse).

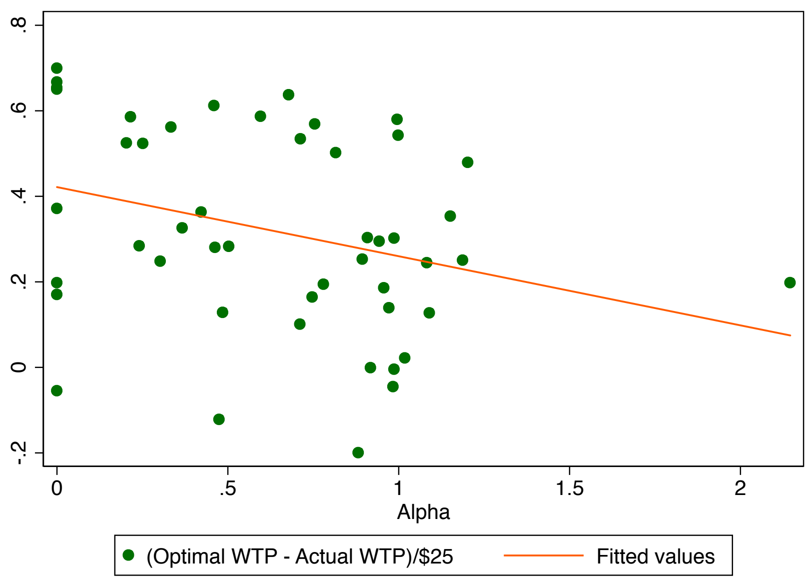
In Table 2, we explore the predictors of overconfidence. In column (2), we see that subjects who have taken advanced math courses and who personally owned stocks were significantly less overconfident (the effect decomposes

Table 2: Overconfidence — Underpayment for Correct Answer

	(1)	(2)	(3)	(6)
Constant	0.319***	0.442***	0.435***	0.559***
	(0.034)	(0.102)	(0.054)	(0.113)
Has Taken Adv. Math		-0.192**		-0.181**
		(0.092)		(0.084)
Owns Stocks		-0.138*		-0.117
		(0.078)		(0.080)
No Credit Card Balances		-0.083		-0.100
		(0.097)		(0.094)
Smoker		0.031		0.063
		(0.108)		(0.097)
Alpha			-0.177***	-0.174***
			(0.065)	(0.058)
N	1,471	1,471	1,471	1,471
Subjects	48	48	48	48

Notes: Dependent variable is question-level *ex post* overconfidence, defined as $(25 - p_{ij} - WTP_{ij})/25$. Alpha represents the degree of the subject's accuracy at exponential questions. Standard errors clustered by subject. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Figure 5: Subject-level Overconfidence as a Function of α



roughly half into higher earnings and half into higher WTP). A causal interpretation of this result is that experience with these calculations, be it through math courses or participation in the stock market, leads to better performance and greater self-awareness about one's errors. Alternatively, those who have an innate talent for math will be better at exponential questions as well as more aware of their own limitations and will be more likely to take advanced math classes and participate in the stock market. We focus on columns (3) and (4), which examine the relationship between the severity of a subject's EGB and his level of overconfidence. It is a common finding in the overconfidence literature that the most error-prone subjects are the most overconfident, and we find a similar pattern in our setting.[7] is not tautological that higher performers will have less overconfidence, as it is possible that those making the largest errors are self-awareor even pessimistic and choose a WTP weakly above the optimum, while those with the least errors believe they are perfect and choose a too-high WTP. We find that a fully-biased subject has a 13–14 percentage point greater level of overconfidence than an unbiased subject.

Figure 5 displays this visually, plotting subject-level average overconfidence against α. There is a clear pattern: more biased individuals are also more overconfident. This suggests a pathological selection in the market, whereby the least competent avoid advice the most. This reduces the incentives for market actors to correct the bias.

3.2.2 Unawareness vs. Overprecision

We next ask whether the overconfidence we observe comes only from the requirement that subjects cannot be aware of their systematic bias, or whether they are also overconfident about the precision of their errors even conditional on there being no systematic error. A subject who is not systematically biased in her responses, but nevertheless knows they sometimes overestimate and sometimes under-estimate, should have a positive willingness-to-pay for the correct answer. If subjects overestimate not only their accu-

[7] While the *bounds* of overconfidence are a function of performance, it

racy but also their precision (Moore and Healy, 2008), then this will further reduce their already too-low WTP for aid.

We make the following parametric assumptions for this exercise. Suppose subject i believes that his response to question j, r_{ij}, is noisy. That is $r_{ij} = (b_i + \eta_{ij}) \cdot c_j$, where c_j is the correct response and η_{ij} is drawn i.i.d. from an exponential distribution: $F_\eta(y) = 1 - e^{-\lambda y}$. Here b_i represents their perceived degree of bias. An agent believing himself to be unbiased must have expectations $E[r_{ij}] = c_j$, and therefore $E[b+\eta] = 1$. Given that $E[\eta] = 1/\lambda$ for an exponential distribution, this is equivalent to the restriction $b = 1 - 1/\lambda$. This rational expectations condition on b does not hold for a biased agent. Given the multiplicative structure, however, we can estimate the λ parameter from the variance of r_{ij}/c_j without having to estimate b, since $Var[r/c] = Var[\eta] = 1/\lambda^2$. The mean value across all subjects for $1/\lambda$ is 1.07. We can then simulate the subject's earnings under the counter-factual restriction that η_{ij} is still exponentially distributed according to λ_i, but imposing the restriction that $b = 1 - 1/\lambda_i$ to simulate an unbiased agent with the same noise as the biased agent.

We perform this simulation exercise separately for questions on which exponential-growth bias predicts a positive and a negative bias. In both cases, the simulated responses are associated with higher earnings than subjects' actual answers: \$13.87 (s.d. 0.23) and \$14.23 (s.d. 0.38), respectively, as compared to actual means of \$10.93 and \$12.28. Subjects who were aware of the noise in their answers, but not the systematic bias, therefore ought to have a willingness to pay for the correct answer of between \$10.77 and \$11.13 (or more if risk-averse). This is still significantly above the observed willingness to pay of our subjects, which indicates that they must be overoptimistic about the precision of their answers in addition to being unaware of their bias. Indeed, the low willingness to pay is rationalized only if the variance of η_i is one-quarter of its true value, suggesting that subjects greatly overestimated their precision in addition to their accuracy.

It is possible that subjects' poor calibration may be due to a lack of experience with these tasks. However, this cannot explain the systematic downward bias in people's WTP for the answer. In subjective expected utility theory, a person must have a prior distribution for her earnings. A lack of experience should lead to a more diffuse prior (i.e. mean-preserving spread). This implies that not getting the answer is more risky. An agent who is risk-averse should therefore increase her valuation of the answer as her prior becomes more diffuse. Thus lack of experience should result in the exact opposite result — a WTP that is *above* the optimal WTP. If one instead modeled the setting as ambiguous, so that subjects did not have a single well-defined prior, the prediction remains the same: less experience implies more ambiguity, which should also increase the WTP for the correct answer.

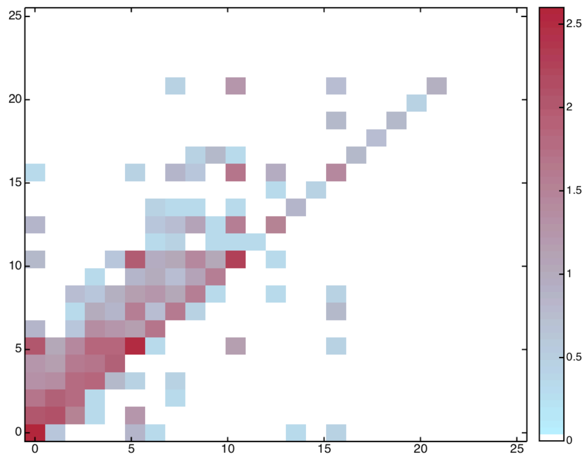
3.3 Overconfidence in Being Able to Use a Spreadsheet

Subjects were overconfident with their answers and they were also overconfident in their ability to use a spreadsheet. Receiving the spreadsheet had no effect on their errors. If the spreadsheet group had indeed paid their WTP for the spreadsheet, their earnings averaged over all questions would have been \$5.40 (p=0.003) less than the control group's earnings.

Figure 6 displays the joint distribution of the WTP for the answer and the WTP for the spreadsheet. As one can see, the WTP for the answer almost always weakly exceeds the WTP for the spreadsheet. There is considerable mass where WTP for the two are equal, 46.7% in the control group and 44.4% in the full sample. This implies that on close to half of questions, subjects believe that getting a spreadsheet is just as good as getting the correct answer. This indicates confidence in one's ability to use the spreadsheet effectively and no perceived cost or inconvenience.

We next measure overconfidence in being able to use a spreadsheet. Unlike the general overconfidence measure we found in the previous section, we cannot measure overconfidence in spreadsheet ability on a within-subject basis. The performance under the counterfactual that a subject receives the correct answer is approximately the maximal possible earnings \$25. Here, though, we have two possible approaches given the between-subjects design. We can predict counterfactual earnings of subjects in the spreadsheet group

Figure 6: WTP for Spreadsheet vs WTP for Answer



had they not had access to the spreadsheet, and counterfactual earnings for the control group had they had access.[8] In both cases, we predict performance on the question level using the regression

$$PE_{ij} = \beta_0 + \beta_1 FPHE_{ij} + \beta_2 FPHE_{ij}^2 + \delta \cdot q_j + e_{ij}. \qquad (2)$$

The outcome variable is the performance earnings (what the subject would have earned if the question were selected to count) for individual i on question j. $FPHE_{ij}$ is the first-pass hypothetical earnings discussed above and $FPHE_{ij}^2$ is the square of this term. Question dummies are in the vector q_j, and e_{ij} is the error term. We estimate the coefficients using one group, and use the coefficients to predict counterfactual earnings for the other group. The predictive models perform well, F(33, 48)=24.81, $r^2 = 0.240$ among controls and F(33,36) = 90.52, $r^2 = 0.178$ among the spreadsheet group.

First, we predict performance earnings for the spreadsheet treatment, under the counterfactual that they did not get the spreadsheet. The differ-

[8]A third approach would be to use treated subjects' unincentivized estimates to estimate within-subject treatment effects of the spreadsheet. It is undesirable to use these responses directly, however, as they are unincentivized and, in general, are associated with substantially lower earnings than final answers.

ence between earnings with the spreadsheet and predicted earnings without the spreadsheet is our estimate for the value of the spreadsheet. We define overconfidence in being able to use the spreadsheet as as the WTP minus the estimated value of getting the spreadsheet, divided by \$25: (WTP- Predicted Gain)/\$25. Thus if the person values the spreadsheet at \$25 but the benefit was 0 then overconfidence is 1, and if the person values the spreadsheet at \$0 but its benefit is \$25 then overconfidence is -1. If the spreadsheet hurts performance then values of overconfidence greater than 1 are possible.

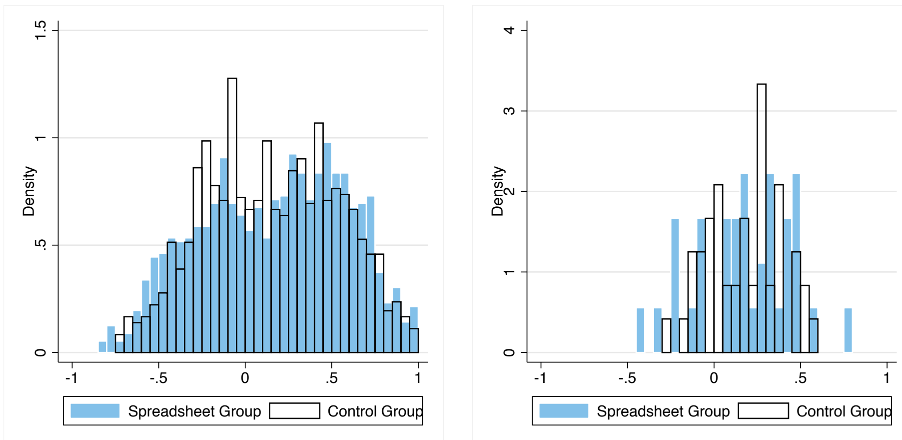
Second, we predict performance earnings for control subjects under the counterfactual that they did have access to the spreadsheet. Again we estimate equation (2), this time using the spreadsheet group, and use the coefficients to predict earnings for control subjects. The estimated value of the spreadsheet for controls is their counterfactual performance earnings with the spreadsheet, minus their actual earnings. Overconfidence is again defined as subjects' WTP minus the estimated value of the spreadsheet, divided by \$25.

Figure 7 panel (a) shows the distribution of overconfidence in being able to use a spreadsheet using both approaches (i.e. counterfactuals for the spreadsheet sample and the control sample). We restrict attention to the interval [-1,1].[9] The mean estimated spreadsheet overconfidence in the spreadsheet sample is 0.198 with 62.2% of responses exhibiting overconfidence. This is significantly different from zero (clustered standard errors, p=0.001). The mean estimated spreadsheet overconfidence in the control sample is 0.178 with 57.63% of responses exhibiting overconfidence. A Mann-Whitney test fails to reject equality of the two distributions (p=0.35), providing reassurance that the two methods are indeed measuring the same thing. Panel (b) shows subject level overconfidence. The population is significantly overconfident (clustered standard errors, p<0.001 for both samples) with 70.3% and 74.5% exhibiting overconfidence on average. A Mann-Whitney test again does not reject equality of the distributions (p=0.75).

Overconfidence in exponential estimation conceptually differs from over-

[9]This includes 97.9% of the sample. The remaining 2.1% have overconfidence greater than 1 implying a prediction the spreadsheet may have hurt their performance.

Figure 7: Overconfidence In Using the Spreadsheet



(a) Subject x Question Level

(b) Subject Level

confidence in being able to use the spreadsheet. The former measures the gap between the expected performance and actual performance, whereas the latter measures the gap between the expected value added of the spreadsheet and the actual value added. A person who has $\alpha = 1$ could not be overconfident in exponential estimation, but would necessarily have overconfidence in being able to use the spreadsheet if she had a strictly positive valuation for it.

An alternative explanation is that, if people believe that the spreadsheet reduces the amount of effort involved in making calculations, then they may pay for the spreadsheet to reduce effort even if it does not improve earnings. If this motivation is true, then it would apply even more when paying for the answer, yet there we find that people significantly underpay — that is, our estimates of overconfidence there would be understated. Thus if people are not truly overconfident in their spreadsheet ability because they use it to avoid effort, then they must be even more overconfident about their raw performance.

Since the spreadsheet was a customized feature of the interface, it may be that subjects expected the common software application Microsoft Excel and were upset when they found an open-source application. This may explain why subjects had a high WTP. This seems unlikely given that the application could be used straightforwardly to answer all the exponential questions in the same method used for Microsoft Excel. Nonetheless, we cannot fully eliminate the possibility that subjects expected a different application and were discouraged when they saw the open-source application.

4 Discussion and Conclusion

A growing literature has shown that EGB is a prevalent phenomenon. This does not imply that there is a market failure. Just as the inability to make finely crafted widgets does not preclude one from obtaining them in a competitive marketplace, EGB should not preclude individuals from obtaining the advice or tools to make good financial decisions. A second error is necessary for this to happen. Agents who are overconfident in their exponential estimation will exhibit sub-optimally low demand for such advice and tools. Our laboratory sample exhibited high degrees of EGB and overconfidence. While subjects believed that they earned at least \$19.24 on average, they actually only earned \$11.28. Individuals with greater EGB exhibited greater overconfidence. This suggests pathological selection in the marketplace. Those who would benefit the most from advice or tools have lower demand for these things. Ironically, subjects exhibit too much demand for a spreadsheet that does not help them. The average WTP for the spreadsheet in the full sample is \$5.78 yet it had no statistically significant impact on performance. The provision of costly tools that require significant skill to use effectively could have easily made the vast majority of our subjects worse off.

EGB is conceptually related to broader concepts of financial literacy. Hung et al. (2009), reviewing the literature, define financial literacy as the "knowledge of basic economic and financial concepts, as well as the ability to use that knowledge and other financial skills to manage financial resources effectively for a lifetime of financial well-being." According to the definition, comprehension and ability to compute compounding is certainly a component of financial literacy. Perhaps the most common measurement is the 3-question inventory developed by Lusardi and Mitchell (2009). It includes simple questions on interest rates (not about compounding), on inflation, and on diversification. Almenberg and Gerdes (2012) find that EGB is associated with broader measurements of financial literacy. Goda et al. (2015) show that EGB predicts retirement savings while controlling for 3-question inventory of financial literacy, IQ, education, time preferences, risk preferences, and many standard controls. Investment firms often remind consumers about the power of compound growth to increase their savings. Overconfidence about EGB in particular is thus likely to be important and require specifically designed policy interventions.

A possible countervailing force against consumer overconfidence is firms' profit motive. A common argument against the persistence of biases in the marketplace is that firms will inform consumers about their biases in order to sell them advice or financial tools. To the contrary, firms may have incentives to keep consumers ignorant. For example, a lender may wish to keep her clients biased so that they underestimate the costs of a loan. But a classic pro-market argument states that an honest firm could undercut the deceptive firms by informing consumers. Gabaix and Laibson (2006) show that this need not be the case. When there are "shrouded" add-on goods, firms will not de-bias consumers in equilibrium because it would cause them to earn strictly lower profits. Similarly, Heidhues, Kőszegi and Murooka (2016) show that firms may deceive naive consumers about socially wasteful products in equilibrium. Thus if people are overconfident in their exponential estimation, firms may have little incentive to de-bias them and much incentive to sustain the bias. The exploitative lender loses its ability to exploit by informing clients, and providing information will not attract more clients if exploitative competitors subsidize informed clients using the extra revenue generated from exploiting the biased clients.

Our results and those in the literature lead us to question whether ad hoc interventions can ameliorate confusing choice architectures. We show that providing costly but perfect solutions will not work because people have insufficient demand, and that providing costly tools that require skill may make people worse off. Ambuehl, Bernheim and Lusardi (2014) further show that educational interventions can change behavior but still fail to improve welfare. Perhaps a more promising approach to address EGB is to design choice architectures that make it irrelevant. Explicit statements about the time it takes to pay off a loan, or the amount of retirement income generated from savings as in Goda, Manchester and Sojourner (2014) are steps in this direction. Similarly, Royal and Tasoff (2014) show that if tools are a complement with ability (tools are more valuable to high ability people), not only will overconfident people exhibit overly high demand for the tools, but the mere opportunity to obtain tools can induce more over-entry into skilled tasks thus making them strictly worse off. In the context of EGB, spreadsheets and other financial software may induce over-entry into tasks that require computation of exponential growth such as active trading and speculative real-estate investment. However, this remains an open question.

References

Almenberg, Johan and Christer Gerdes, "Exponential Growth Bias and Financial Literacy," *Applied Economics Letters*, 2012, *19*, 1693–1696.

Ambuehl, Sandro, B. Douglas Bernheim, and Annamaria Lusardi, "Financial Education, Financial Competence, and Consumer Welfare," October 2014. NBER Working Paper No. 20618.

Benzion, Uri, Alon Granot, and Joseph Yagil, "The Valuation of the Exponential Function and Implications for Derived Interest Rates," *Economic Letters*, 1992, *38*, 299–303.

Eisenstein, Eric M. and Stephen J. Hoch, "Intuitive Compounding: Framing, Temporal Perspective, and Expertise," December 2007. WorkingPaper.

Gabaix, Xavier and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Goda, Gopi Shah, Colleen Flaherty Manchester, and Aaron Sojourner, "What Will My Account Be Worth? Experimental Evidence on How Retirement Income Projections Affect Saving," *Journal of Public Economics*, 2014, *119*, 80–92.

__ , **Matthew Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff**, "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings," August 2015. NBER Working Paper 21482.

Heidhues, Paul, Botond Kőszegi, and Takeshi Murooka, "Inferior Products and Profitable Deception," *Review of Economic Studies*, Forthcoming 2016.

Hung, Angela A., Andrew M. Parker, and Joanne K. Yoong, "Defining and Measuring Financial Literacy," September 2009. RAND Labor and Population Working Paper Series.

Hyytinen, Ari and Hanna Putkuri, "Household Optimism and Borrowing," May 2012. Bank of Finland – Working Paper.

Keren, Gideon, "Cultural Differences in the Misperception of Exponential Growth," *Perception and Psychophysics*, 1983, *34* (3), 289–293.

Kruger, Justin and David Dunning, "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *Journal of Personality and Social Psychology*, 1999, *77* (6), 1121–1134.

Levy, Matthew and Joshua Tasoff, "Exponential-Growth Bias and Lifecycle Consumption," *Journal of the European Economic Association*, June 2016, *14* (3), 545–583.

Lusardi, Annamaria and Olivia S. Mitchell, "How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement Readiness," September 2009. Working Paper.

__ **and** __ , "The Economic Importance of Financial Literacy: Theory and Evidence," *Journal of Economic Literature*, March 2014, *52* (1), 5–44.

MacKinnon, Andrew J. and Alexander J. Wearing, "Feedback and the Forecasting of Exponential Change," *Acta Psychologica*, 1991, *76*, 177–191.

McKenzie, Craig R.M. and Michael J. Liersch, "Misunderstanding Savings Growth: Implications for Retirement Savings Behavior," *Journal of Marketing Research*, 2011, *48*, S1–S13.

Moore, Don A. and Paul J. Healy, "The Trouble with Overconfidence," *Psychological Review*, 2008, *115* (2), 502–517.

Mullainathan, Sendhil, Markus Noeth, and Antoinette Schoar, "The market for financial advice: An audit study," March 2012. NBER Working Paper 17929.

Puri, Manju and David T. Robinson, "Optimism and Economic Choice," *Journal of Financial Economics*, 2007, *86*, 71–99.

Royal, Andrew and Joshua Tasoff, "Overconfidence and Capital," November 2014. Working Paper.

Souleles, Nicholas S, "Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys," *Journal of Money, Credit and Banking*, Feb 2004, *36* (1), 39–72.

Stango, Victor and Jonathan Zinman, "Exponential Growth Bias and Household Finance," *Journal of Finance*, December 2009, *64* (6), 2807–2849.

_ and _, "Fuzzy Math, Disclosure Regulation, and Market Outcomes: Evidence from Truth-in-Lending Reform," *Review of Financial Studies*, 2011, *24* (2), 506–534.

Wagenaar, Willem A. and Han Timmers, "The Pond-and-Duckweed Problem: Three Experiments on the Misperception of Exponential Growth," *Acta Psychologica*, 1979, *43*, 239–251.

_ and Sabato D. Sagaria, "Misperception of Exponential Growth," *Perception and Psychophysics*, 1975, *18* (6), 416–422.

5 Appendix

Table A1: Experiment Questions

Domain	Question	Asset A: T	Asset B: T	Asset A: P_0	Asset B: P_0	Asset A: i	Asset B: i
Exponential	1	10	10	100	X	10%	0%
	2	20	20	200	X	5%	0%
	3	10	10	90	X	25%	0%
	4	35	35	30	X	7%	0%
	5	30	30	120	X	4%	0%
	6	10	10	500	X	-8%	0%
	7	20	20	1000	X	-4%	0%
	8	8	8	1000	X	-15%	0%
Periodic Savings	9	30	30	+10/period	X	5%	0%
	10	30	30	+15/period	X	2.5%	0%
	11	20	20	+12/period	X	10%	0%
	12	20	20	+20/period	X	5%	0%
	13	15	15	+20/period	X	6%	0%
	14	20	20	+20/period	X	6%	0%
	15	25	25	+4/period	X	5%	0%
	16	50	50	+6/period	X	5%	0%
Front-end Load	17	25	25	100	X	5%	8%
	18	25	25	100	X	5%	10%
	19	20	20	100	X	10%	13%
	20	10	10	200	X	10%	20%
	21	25	25	40	X	8%	5%
	22	25	25	100	X	8%	5%
	23	20	20	50	X	13%	10%
	24	30	30	100	X	5%	4%
Fluctuating i	25	10	10	100	100	40% in odd; 0% in even	X%
	26	14	14	50	50	30% in odd; 0% in even	X%
	27	10	10	100	100	50% in odd; 0% in even	X%
	28	6	6	50	50	100% in odd; 0% in even	X%
	29	20	20	100	100	10%	-20% in odd; X% in even
	30	6	6	300	100	0%	-40% in odd; X% in even
	31	30	30	30	30	10%	-10% in odd; X% in even
	32	20	20	20	20	20%	-20% in odd; X% in even

Figure A1: Payment Instructions

Logged in as PGw689z.

Logout

This portion of the experiment is divided into 32 independent questions. Each question will describe one or more financial asset, and ask you about their value over time.

The values of the assets will be calculated according to the description in the questions. **Your payment** will be dependent on the accuracy of your response according to the quadratic equation, payment= $25 - 100(1 - \frac{r}{v})^2$, where r is your response and v is the correct value. All negative payments will be treated as zero.

Your payment is maximized when your response equals the correct value, $r = v$. Below is a graph that explains your payments as a function of accuracy.

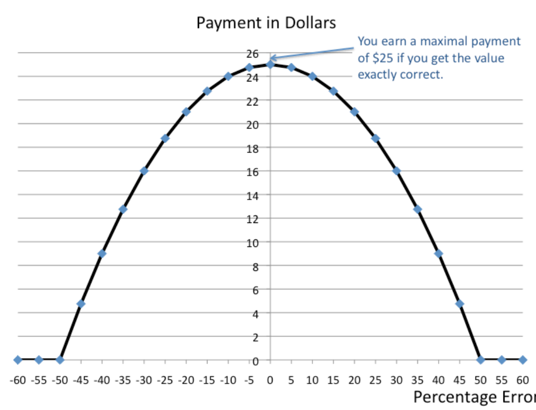


Figure A2: Payment Instructions, continued

This formula gives the exact way we will calculate your earnings, but it may be useful to think about a few examples. For your reference:

If your answer is this far from the correct value,	*then you will earn:*
0%	\$25.00
5%	\$24.75
10%	\$24.00
15%	\$22.75
20%	\$21.00
25%	\$18.75
30%	\$16.00
35%	\$12.75
40%	\$9.00
45%	\$4.75
50% or more	\$0.00

Example 1:

Asset A is worth \$100 + 330 − 20. Asset B is worth \$X. What value of X makes Asset A and Asset B worth the same amount?

The true value of this Asset A is \$410.

- Suppose Leia states that X is \$410. Then Leia will earn \$25 if this question is selected.
- However, Chewy states that X is \$390. Then Chewy will earn $25 - 100(1 - \frac{390}{410})^2 = \$24.76$ if this question is selected.
- If Chewy had stated it is worth \$300 he would have only earned $25 - 100(1 - \frac{300}{410})^2 = \$17.80$.
- If Chewy had stated that it is \$700 then he would have earned \$0 because $25 - 100(1 - \frac{700}{410})^2 < 0$.

In this example, how much would an answer of \$800 have earned?

Continue

Figure A3: Treatment Instructions

Logged in as PGw689z.

Logout

First Phase

There are two phases in the experiment. In the first phase you will face a series of questions about the values or interest rates of assets. For each question you will be asked for three responses: your initial estimate, and the maximum you are willing to pay for each of two services that may be available in the second phase. In the second phase you will answer all the questions a second time but possibly with the assistance of one of the aforementioned services.

The first service is a spreadsheet capable of basic arithmetic, exponentiation, and summation. The second service will provide you with the correct answer directly. Both of these services may help you in calculating the value of the asset for questions in the second phase, thereby increasing your expected earnings.

You will be asked to state the maximum you are willing to pay (cutoff) for each of these two services on every question. On any given question in the second phase, you may ultimately receive either the spreadsheet or the correct answer, or neither, but never both.

After you provide an answer, and state your cutoffs for each service on each question, you will begin the second phase.

Second Phase

In the second phase you will answer the exact same questions a second time possibly with the help of a service. You will be randomly assigned to one of three possible tracks which determine the service you receive and how you receive it.

- **Track A:** Whether you receive one of these services and how much you pay for it is determined in the following manner. For a given question, we will first randomly determine which service is available. We will then randomly determine a price p for that service that is independent of your stated cutoff. In other words, your stated cutoff will have no effect on the random price. However, it will determine whether you purchase the service or not. If p is equal to or above your stated cutoff (p ≥ cutoff) then you will not get the service. If p is less than your stated cutoff (p < cutoff) then you will get the service for that question. At the end of the experiment, you will pay $p (subtracted from your final earnings although you will always earn at least $5 for participation) if and only if that particular question is selected for payment. This procedure is then repeated to determine which service you get for every question.

  To maximize your expected earnings you should estimate **how much each service would improve your earnings** and then state that as your cutoff. If the price p is then less than your estimate, you will get the service and will expect to earn more money in total by buying the service. If the price p is greater than your estimate, you will not buy the service which is a good thing because you believe the service would cost more than it earns you (meaning your total earnings would be lower if you had bought it).

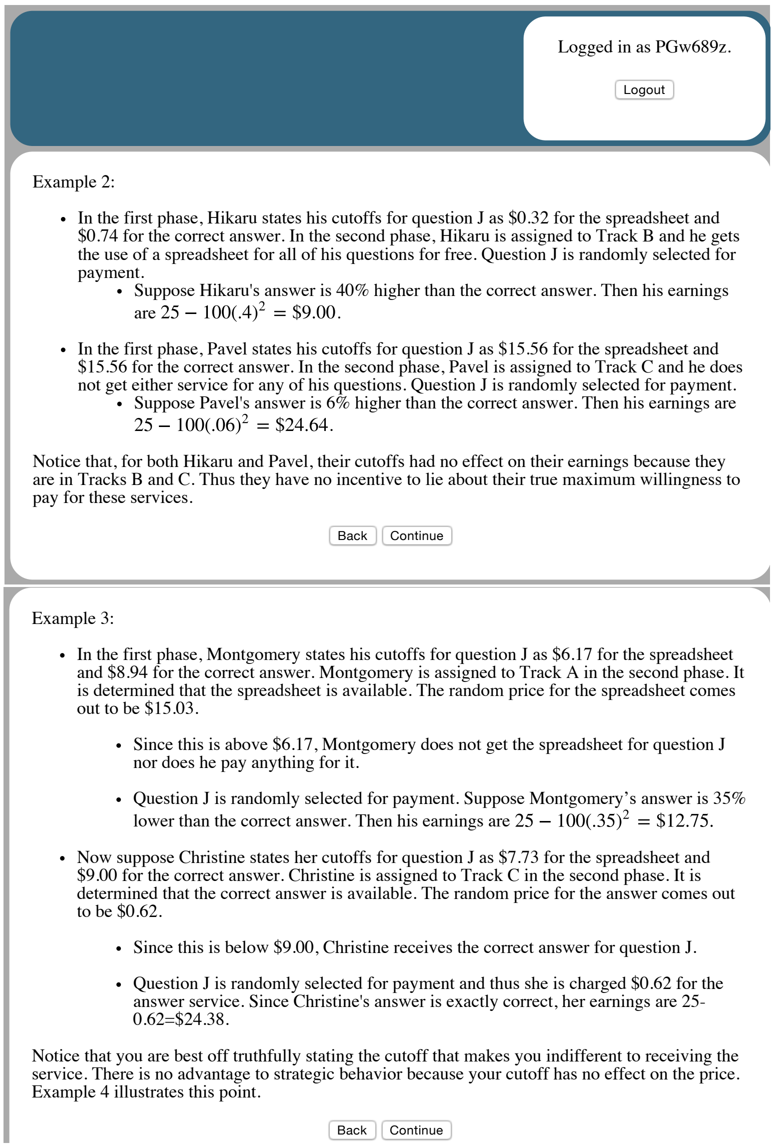
- **Track B:** you will receive one of the two services for **free** on all questions.

- **Track C:** you will **not** receive any of the services nor will you pay for them.

You will then proceed to answer the questions a second time with the services that have been granted to you.

Back Continue

Figure A4: WTP Examples


Logged in as PGw689z.

Logout

Example 2:

- In the first phase, Hikaru states his cutoffs for question J as \$0.32 for the spreadsheet and \$0.74 for the correct answer. In the second phase, Hikaru is assigned to Track B and he gets the use of a spreadsheet for all of his questions for free. Question J is randomly selected for payment.
  - Suppose Hikaru's answer is 40% higher than the correct answer. Then his earnings are $25 - 100(.4)^2 = \$9.00$.
- In the first phase, Pavel states his cutoffs for question J as \$15.56 for the spreadsheet and \$15.56 for the correct answer. In the second phase, Pavel is assigned to Track C and he does not get either service for any of his questions. Question J is randomly selected for payment.
  - Suppose Pavel's answer is 6% higher than the correct answer. Then his earnings are $25 - 100(.06)^2 = \$24.64$.

Notice that, for both Hikaru and Pavel, their cutoffs had no effect on their earnings because they are in Tracks B and C. Thus they have no incentive to lie about their true maximum willingness to pay for these services.

Back Continue

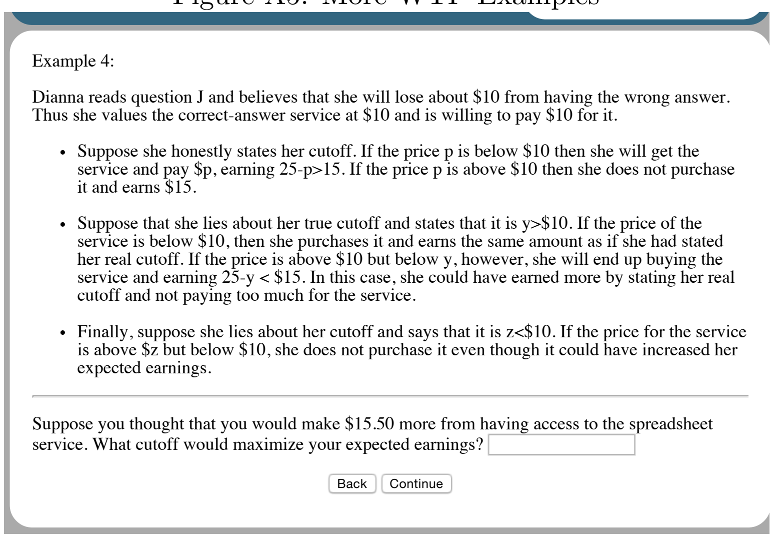
Example 3:

- In the first phase, Montgomery states his cutoffs for question J as \$6.17 for the spreadsheet and \$8.94 for the correct answer. Montgomery is assigned to Track A in the second phase. It is determined that the spreadsheet is available. The random price for the spreadsheet comes out to be \$15.03.
  - Since this is above \$6.17, Montgomery does not get the spreadsheet for question J nor does he pay anything for it.
  - Question J is randomly selected for payment. Suppose Montgomery’s answer is 35% lower than the correct answer. Then his earnings are $25 - 100(.35)^2 = \$12.75$.
- Now suppose Christine states her cutoffs for question J as \$7.73 for the spreadsheet and \$9.00 for the correct answer. Christine is assigned to Track C in the second phase. It is determined that the correct answer is available. The random price for the answer comes out to be \$0.62.
  - Since this is below \$9.00, Christine receives the correct answer for question J.
  - Question J is randomly selected for payment and thus she is charged \$0.62 for the answer service. Since Christine's answer is exactly correct, her earnings are 25-0.62=\$24.38.

Notice that you are best off truthfully stating the cutoff that makes you indifferent to receiving the service. There is no advantage to strategic behavior because your cutoff has no effect on the price. Example 4 illustrates this point.

Back Continue

Figure A5: More WTP Examples



Example 4:

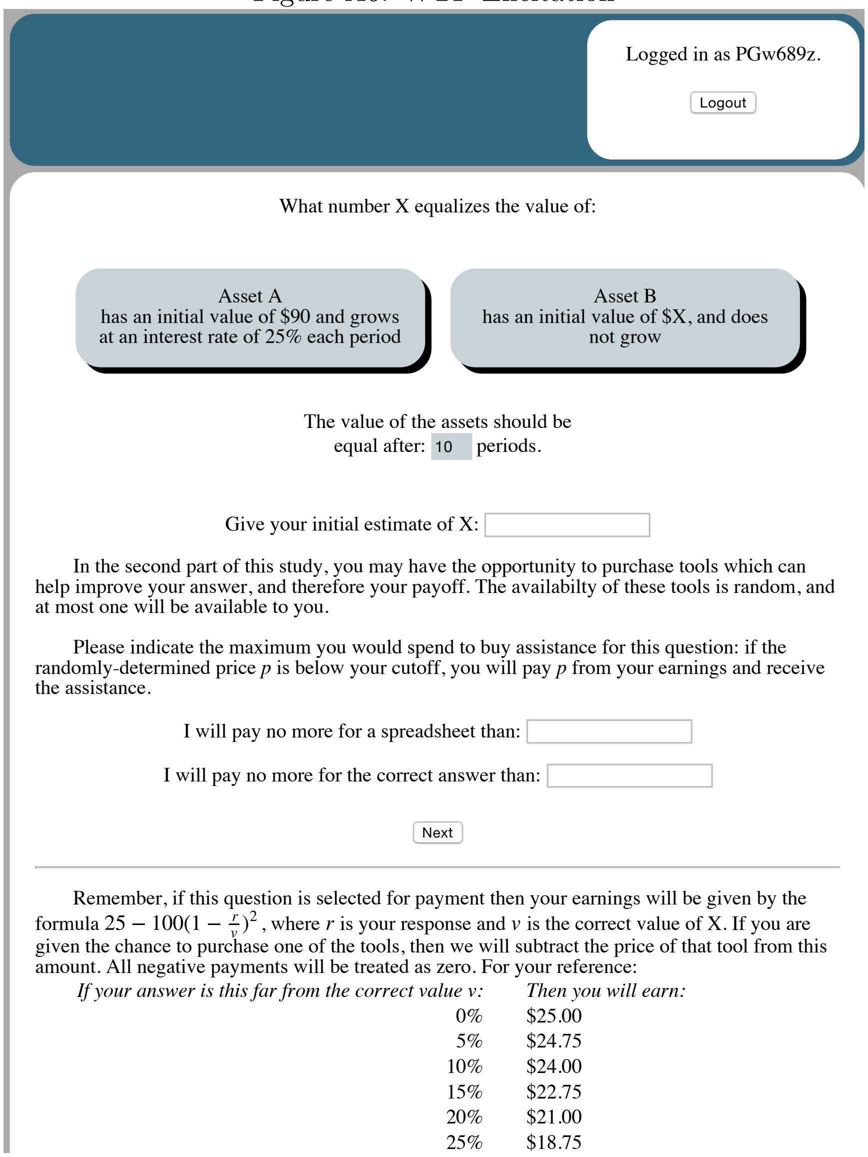
Dianna reads question J and believes that she will lose about $10 from having the wrong answer. Thus she values the correct-answer service at $10 and is willing to pay $10 for it.

- Suppose she honestly states her cutoff. If the price p is below $10 then she will get the service and pay $p, earning 25-p>15. If the price p is above $10 then she does not purchase it and earns $15.
- Suppose that she lies about her true cutoff and states that it is y>$10. If the price of the service is below $10, then she purchases it and earns the same amount as if she had stated her real cutoff. If the price is above $10 but below y, however, she will end up buying the service and earning 25-y < $15. In this case, she could have earned more by stating her real cutoff and not paying too much for the service.
- Finally, suppose she lies about her cutoff and says that it is z<$10. If the price for the service is above $z but below $10, she does not purchase it even though it could have increased her expected earnings.

Suppose you thought that you would make $15.50 more from having access to the spreadsheet service. What cutoff would maximize your expected earnings?

Back Continue

Figure A6: WTP Elicitation



Logged in as PGw689z.

Logout

What number X equalizes the value of:

Asset A
has an initial value of \$90 and grows at an interest rate of 25% each period

Asset B
has an initial value of \$X, and does not grow

The value of the assets should be equal after: 10 periods.

Give your initial estimate of X:

In the second part of this study, you may have the opportunity to purchase tools which can help improve your answer, and therefore your payoff. The availabilty of these tools is random, and at most one will be available to you.

Please indicate the maximum you would spend to buy assistance for this question: if the randomly-determined price p is below your cutoff, you will pay p from your earnings and receive the assistance.

I will pay no more for a spreadsheet than:

I will pay no more for the correct answer than:

Next

Remember, if this question is selected for payment then your earnings will be given by the formula $25 - 100(1 - \frac{r}{v})^2$, where r is your response and v is the correct value of X. If you are given the chance to purchase one of the tools, then we will subtract the price of that tool from this amount. All negative payments will be treated as zero. For your reference:

If your answer is this far from the correct value v:	*Then you will earn:*
0%	\$25.00
5%	\$24.75
10%	\$24.00
15%	\$22.75
20%	\$21.00
25%	\$18.75

Figure A7: Control Group

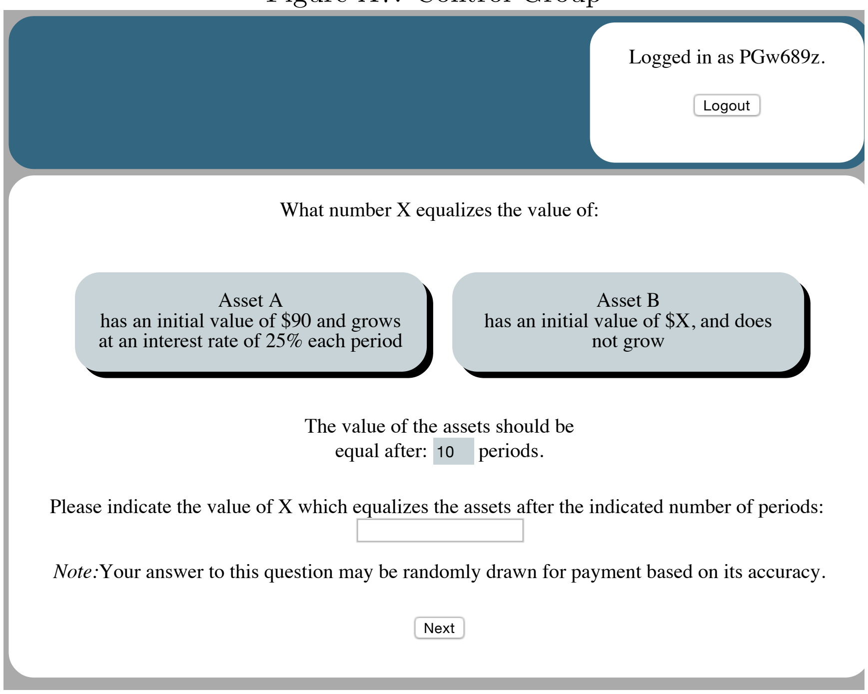


Figure A8: Spreadsheet Treatment

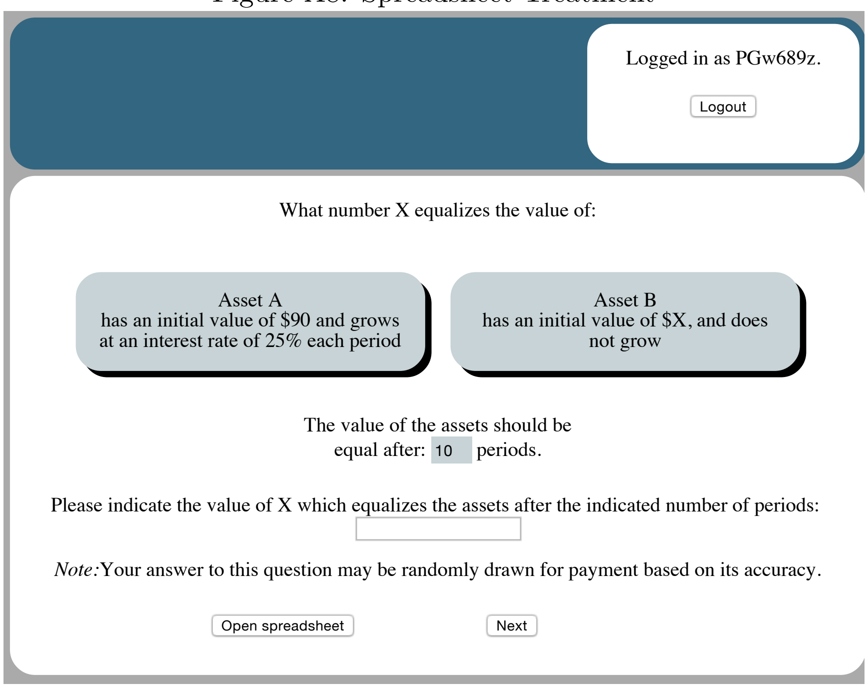


Figure A9: Answer Treatment