

NBER WORKING PAPER SERIES

FINANCIAL EDUCATION, FINANCIAL COMPETENCE, AND CONSUMER WELFARE

Sandro Ambuehl  
B. Douglas Bernheim  
Annamaria Lusardi

Working Paper 20618  
<http://www.nber.org/papers/w20618>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2014

We thank Charles Sprenger, as well as participants at the Research Forum on the Effectiveness of Financial Education University of Arizona, the Stanford Institute for Theoretical Economics, and the Journées Louis-Andre Gerard-Varet in Aix-en-Provence for helpful comments and suggestions. Fulya Yuksel Ersoy provided excellent research assistance. This work was funded by the Department of Economics at Stanford University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Sandro Ambuehl, B. Douglas Bernheim, and Annamaria Lusardi. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Financial Education, Financial Competence, and Consumer Welfare  
Sandro Ambuehl, B. Douglas Bernheim, and Annamaria Lusardi  
NBER Working Paper No. 20618  
October 2014  
JEL No. D14,D91,I21

### **ABSTRACT**

We introduce the concept of financial competence, a measure of the extent to which individuals' financial choices align with those they would make if they properly understood their opportunity sets. Unlike existing measures of the quality of financial decision making, the concept is firmly rooted in the principles of choice-based behavioral welfare analysis; it also avoids the types of paternalistic judgments that are common in policy discussions. We document the importance of assessing financial competence by demonstrating, through an example, that an educational intervention can appear highly successful according to conventional outcome measures while failing to improve the quality of financial decision making. Specifically, we study a simple intervention concerning compound interest that significantly improves performance on a test of conceptual knowledge (which subjects report operationalizing in their decisions), and appears to counteract exponential growth bias. However, financial competence (welfare) does not improve. We trace the mechanisms that account for these seemingly divergent findings.

Sandro Ambuehl  
Department of Economics  
Stanford University  
Stanford, CA 94305  
<http://web.stanford.edu/~sambuehl/>  
[sambuehl@stanford.edu](mailto:sambuehl@stanford.edu)

B. Douglas Bernheim  
Department of Economics  
Stanford University  
Stanford, CA 94305-6072  
and NBER  
[bernheim@stanford.edu](mailto:bernheim@stanford.edu)

Annamaria Lusardi  
The George Washington University  
School of Business  
2201 G Street, NW  
Duques Hall, Suite 450E  
Washington, DC 20052  
and NBER  
[alusardi@gwu.edu](mailto:alusardi@gwu.edu)

An Online Appendix is available at:  
<http://www.nber.org/data-appendix/w20618>

# 1 Introduction

Low levels of financial literacy in the United States and the rest of the world raise serious questions about the general quality of financial decision making. Financial education aims to improve decision making by helping consumers acquire the basic knowledge and skills they need to understand the choices they face. A large and growing literature examines the effects of financial education on both financial literacy (as measured by test scores) and financial choices (such as saving).<sup>1</sup> Unfortunately, that literature sheds little objective light on the critical question of whether the behavioral effects of financial education are helpful or harmful. Discussions of these issues are typically colored by paternalistic judgments (for example, that people are better off with high saving and balanced portfolios) and/or strong preconceptions (for example, that a better understanding of choice options necessarily promotes better decisions). Yet it is also possible that financial education alters behavior through mechanisms that involve indoctrination, exhortation, deference to authority, social pressure, or psychological anchors, in which case it may induce people to act contrary to the preferences they themselves would reveal through choices if they properly understood the consequences of their actions. Because the existing literature does not distinguish between these mechanisms, it ultimately has little to say about the welfare consequences of financial education. (See Section 9 for a more detailed discussion of existing research that bears on this issue.)

This paper makes three main contributions. First, it introduces a new method for evaluating the quality of financial decision making and gauging a decision maker's degree of *financial competence*.<sup>2</sup> Our approach involves comparisons between the choices an individual makes in settings with objectively identical options but different presentations. Depending on the presentation, an understanding of specific conceptual principles either is or is not required to evaluate the options. Unlike existing measures of the quality of financial decision making, our measure of financial competence is firmly rooted in the principles of choice-based behavioral welfare analysis (Bernheim and Rangel, 2004, 2009). Moreover, because it relies on welfare criteria that are derived from an individual's own choices, it avoids the types of paternalistic judgments that are common in policy discussions.

The paper's second main contribution is to document the importance of assessing and analyzing financial competence, rather than relying exclusively on conventional outcome measures. It accomplishes that objective by demonstrating, through an example, that an educational intervention can appear highly successful according to those conventional measures while failing to improve the quality of financial decision making.

---

<sup>1</sup>See Lusardi and Mitchell (2014) and Hastings et al. (2013) for recent comprehensive reviews of this literature.

<sup>2</sup>The literature uses the term *financial capability* to signify the quality of financial decision making. Because that existing term lacks a precise definition, we avoid it. Our notion of *financial competence* is, of course, closely related, but it has a specific meaning.

The paper's third contribution is to explore the mechanisms that produce apparently divergent effects on conventional outcome metrics and financial competence. Because our analysis is confined to a single limited educational intervention, we caution against generalizing from our conclusions concerning these mechanisms. Still, we regard this portion of our investigation as an important step toward understanding the relationships between knowledge, motivation, and choice, and hence designing more effective and beneficial educational interventions.

Our approach is predicated on a distinction between decision problems with *complex framing* and *simple framing*. In a complexly framed problem, the consumer contemplates a menu of alternatives with components that she values not for themselves, but rather because they provide the means to secure the goods she actually desires. We call these components *consumption instruments*. A complexly framed version of the standard intertemporal choice problem might require the consumer to specify the levels of current consumption and current saving. Saving is then an instrument through which future consumption is achieved; the consumer cares about it for that purpose, and not intrinsically. As a general matter, other financial decisions are also complexly framed. For example, portfolios are the instruments through which consumers achieve state-contingent consumption bundles.

Typically, complexly framed choices lead to consumption through intermediate outcomes. In the case of a standard intertemporal decision problem, those outcomes involve the receipt of returns from saving and investments. In a simply framed version of the same decision problem, each complexly framed alternative is replaced with the intermediate outcome that it produces.

For every complexly framed decision problem, there is a substantively equivalent simply framed decision problem. However, there is no guarantee that the consumer will end up with the same consumption bundle in both of these problems. In each case, choices depend on the consumer's understanding of the relationship between intermediate outcomes and intrinsically valued consumption bundles, and on her preferences over those bundles. However, in complexly framed problems, choices also depend on her understanding of the relationship between consumption instruments and intermediate outcomes. If consumers misunderstand that relationship or fail to use what they understand when decisions have real consequences, they may well make different choices. The resulting divergence reflects the consumer's limited financial competence. If financial education concerning that relationship improves competence, it should bring the choices made in complexly framed problems into closer alignment with the ones they make in equivalent simply framed problems.

We operationalize this idea in the context of decision problems that require an understanding of compound interest, one of the fundamental concepts in financial decision making. Briefly, we have people perform paired valuation tasks involving real monetary consequences. Each pair consists of a simply framed task and a complexly framed task that are objectively equivalent. For example, if the

complexly framed task elicits the value of  $X$  for which a subject is indifferent between  $\$X$  immediately and  $\$10$  invested for 36 days at a return of 2% per day, the associated simply framed task would elicit the value of  $Y$  for which she is indifferent between  $\$Y$  immediately and  $\$20$  in 36 days. We evaluate the potential welfare loss resulting from limited financial competence by comparing  $X$ , the valuation of the future payoff with complex framing, to  $Y$ , the valuation of the same future payoff with simple framing.<sup>3</sup>

For the educational intervention, we use a section on compound interest from a popular and well-exposed personal finance book, presented as a narrated slide show. While the presentation begins with a basic explanation of compound interest, most of the substantive material concerns the “rule of 72,” which states that, to a close approximation, the percentage rate of interest times the amount of time required for the value of an investment to double equals 72. The text also contains rhetorical statements about the power of compounding, as well as exhortations to save.

As mentioned above, one of our objectives is to explore the mechanisms through which financial education influences behavior. Consequently, we examine three versions of this intervention: a “full intervention” consisting of the entire video, a “substance only intervention” which omits the rhetorical material, and a “rhetoric only intervention” which omits the explanation and applications of the rule of 72 (but includes the basic explanation of compound interest). We also created a “control intervention” based on an unrelated section from the same book.

Each subject in our experiment experienced one of the four interventions.<sup>4</sup> Then they performed ten simply framed valuation tasks, along with ten equivalent complexly framed valuation tasks. Once those tasks were complete, they took an incentivized test on compound interest, and answered questions concerning their decision-making strategies.

As mentioned above, our use of paired valuation tasks permits us to measure financial competence and make rigorous statements concerning welfare. Other aspects of our experimental design enables us to examine more conventional outcome measures, and hence to make comparisons between the conclusions that follow from these approaches. In particular, it is common in the literature to examine the effects of education on financial literacy, as measured by test scores. But improved knowledge may not imply better decision making, particularly if education also affects behavior through other channels. It is also common to examine *directional* effects on behavior, and to evaluate them in light of known or presumed biases. In the current context, it is well-established that people tend to suffer from *exponential growth bias*, the tendency to underestimate the growth of an investment when interest is compounded (see Wagenaar and Sagaria (1975), Eisenstein and Hoch (2007), Stango and

---

<sup>3</sup>A precise welfare measure derived from this comparison is presented and rigorously justified in Section V.

<sup>4</sup>As detailed below, subjects were recruited from the crowdsourcing platform Amazon Mechanical Turk, and were generally of working-age (20s and 30s), with lower-than-average incomes but higher-than-average education. Their *ex ante* understanding of compound interest was generally poor.

Zinman (2009), Almenberg and Gerdes (2012), Levy and Tasoff (2014)). Consequently, an educational intervention that increases valuations in our complexly framed tasks would be deemed successful according to this approach. But such an intervention could also be harmful if it causes people to overshoot and/or has the same effects on those who do not suffer from the aforementioned bias.

Our main findings are as follows. The full treatment substantially improves subjects' knowledge and conceptual understanding of compound interest (financial literacy), as measured by their incentivized test scores. Moreover, subjects report that they operationalize the newly gained knowledge in their consequential decisions. The fraction of subjects who say they make decisions based on numerical calculations rises sharply, and the fraction who say they seek external help does not decline, which suggests that the rule of 72 does not simply crowd out other potentially reliable approaches to the valuation tasks. Moreover, all of the aforementioned effects are primarily attributable to the substantive elements of the intervention, rather than to the rhetorical elements. The full treatment also significantly increases valuations for complexly framed choices, which is precisely what one would hope to find if it effectively counteracts exponential growth bias. In these respects, the intervention plainly has what appear to be the right effects for the right reasons. An analysis based on conventional outcome metrics such as financial literacy and/or directional effects on choice would therefore likely conclude that the intervention was highly successful and presumably welfare-enhancing. Yet our analysis of financial competence paints a much different picture. Indeed, the full intervention has *no effect whatsoever* on the average quality of financial decision making as measured by financial competence.

What accounts for this apparent divergence between the conclusions that follow from analyses of financial competence on the one hand, and conventional outcome measures on the other? Our analysis highlights two main issues. First, despite our findings concerning test scores and self-reported decision strategies, the effects of the full treatment on consequential choices (valuations) are primarily attributable to rhetoric, not to substance. The substance-only intervention does not have a significant impact on average valuations for complexly framed choices, while the effect of the rhetoric-only intervention is statistically indistinguishable from that of the full intervention.

Second, the effect of the full treatment on choices in complexly framed valuation tasks bears little or no relation to the subject's degree of exponential growth bias (as one might expect if the effect results from rhetoric rather than substance). Ideally, the intervention would increase valuations among those undervaluing complexly framed options, and decrease valuations by those overvaluing those options. Yet on average, it increases valuations in the complexly framed problems across the board (while leaving valuations in the simply framed problems unchanged).

These findings call into question the validity of efforts to evaluate the benefits of educational interventions through analyses that are confined to effects on financial literacy, directional changes in behavior, and/or changes in self-reported decision strategies. Our results also highlight the pitfalls of policy agendas that specifically target only the aforementioned objectives. At the same time, we offer a conceptually rigorous and practical alternative.

The remainder of the paper is organized as follows. Section 2 introduces and precisely defines the concept of financial competence. Section 3 describes our experiment, and section 4 presents summary statistics. Sections 5, 6, and 7 analyze the effects of the treatments on test scores, average choices, and consumer welfare, respectively. Section 8 analyzes the channels through which the interventions affect behavior. Section 9 relates our research to the existing literature. Section 10 discusses the policy implications of our research and concludes.

## 2 The Definition and Measurement of Financial Competence

In this section we formally define the concept of financial competence. In standard consumer theory, we think of a consumer as attaching intrinsic value to elements of a consumption set  $\mathbb{C}$ , and as making choices from an opportunity set,  $C \subseteq \mathbb{C}$ .<sup>5</sup> Yet in many settings, consumers must instead choose from opportunity sets containing bundles that include *consumption instruments* – derivative goods that are valued only because they provide the means to secure bundles of intrinsically valued goods. For example, they obtain future (as well as state-contingent) consumption by making decisions about saving and investments. We will use  $\mathbb{I}$  to denote the set of all possible instrumental alternatives potentially available to a decision maker at a particular point in time, and  $I \subseteq \mathbb{I}$  to denote an opportunity set. The consumer obtains an element of  $\mathbb{C}$  through a process that begins with the selection of some  $i \in I$ .<sup>6</sup>

In many settings, instrumental choices lead to consumption bundles through intermediate outcomes. For example, investments produce future monetary payoffs, which in turn govern consumption opportunities. We will use  $\mathbb{M}$  to denote the set of possible intermediate outcomes. Each option  $i \in \mathbb{I}$  leads to some outcome,  $f(i) \in \mathbb{M}$ . Elements of  $\mathbb{M}$  in turn map to elements of  $\mathbb{C}$  according to some

---

<sup>5</sup>For instance, elements of  $C \subseteq \mathbb{C}$  may involve time-dated and/or state-contingent consumption goods.

<sup>6</sup>As a concrete example, consider a working-age person engaged in life-cycle planning. Bundles specifying consumption throughout the life cycle, both before and after retirement, are elements of  $\mathbb{C}$ . Bundles specifying current consumption and current saving involve consumption instruments, and are therefore elements of  $\mathbb{I}$ .

function  $g$ , which may encompass the manner in which the individual makes additional choices.<sup>7</sup> Thus, a consumer who chooses  $i \in \mathbb{I}$  ends up with the consumption bundle  $g(f(i))$ .<sup>8</sup>

We will refer to a decision problem as *primitively framed* if it involves an opportunity set consisting of consumption bundles,  $C \subseteq \mathbb{C}$ , *simply framed* if it involves an opportunity set consisting of intermediate outcomes,  $M \subseteq \mathbb{M}$ , and *complexly framed* if it involves an opportunity set consisting of consumption instruments,  $I \subseteq \mathbb{I}$ . If  $M = f(I)$ , then the simply framed choice from  $M$  is substantively equivalent to the complexly framed choice from  $I$ .<sup>9</sup> Likewise, if  $C = g(M)$ , then the primitively framed choice from  $C$  is substantively equivalent to the simply framed choice from  $M$ .

To summarize the consumer's behavior, we define choice correspondences  $\phi_{\mathbb{C}}$ ,  $\phi_{\mathbb{M}}$ , and  $\phi_{\mathbb{I}}$  for primitively, simply, and complexly framed decision problems, respectively (where the functions are indexed by their domains). Each of these correspondences maps the pertinent type of opportunity sets into selections from those sets. Notice that  $g(f(\phi_{\mathbb{I}}(\cdot)))$  maps complexly framed decision problems (elements of  $\mathbb{I}$ ) into the intrinsically valued consumption bundles (elements of  $\mathbb{C}$ ) that the consumer selects through the choice of consumption instruments and subsequent decisions.<sup>10</sup>

If a consumer fully understands the relationships between consumption instruments and intermediate outcomes, then for any pair of equivalent opportunity sets  $M$  and  $I$  (for which  $f(I) = M$ ), we should observe  $\phi_{\mathbb{M}}(M) = f(\phi_{\mathbb{I}}(I))$ . However, a consumer who contemplates a menu of complexly framed options may misconstrue his opportunities in terms of intermediate outcomes (and hence also in terms of intrinsically valued goods). Indeed, the literature on financial literacy identifies a variety of settings in which common conceptual errors would likely cause  $\phi_{\mathbb{M}}(M)$  and  $f(\phi_{\mathbb{I}}(I))$  to diverge.<sup>11</sup> For example, a consumer may incorrectly evaluate the real returns to an investment by calculating simple interest rather than compound interest, or he may ignore inflation. Alternatively, he may avoid the most complex, difficult-to-evaluate alternatives, and focus instead on the simplest options.<sup>12</sup>

<sup>7</sup>To continue our example of life-cycle planning, the choice of current consumption and current retirement saving leads, in conjunction with subsequent decisions, to intermediate outcomes, such as the level of accumulated wealth at retirement. The function  $f$  captures this relationship. Retirement wealth then determines consumption during retirement according to the function  $g$ .

<sup>8</sup>We treat the function  $f$  as deterministic. This is without loss of generality. To incorporate uncertainty, one interprets elements of  $\mathbb{M}$  as state-contingent outcomes.

<sup>9</sup>We define  $f(I) = \{f(i) : i \in I\}$ .

<sup>10</sup>To illustrate, recall our life-cycle planning example. If the consumer's options are described as a set of intertemporal consumption bundles  $C$ , she chooses the bundle  $\phi_{\mathbb{C}}(C)$ . If her options are described as a set of alternatives  $M$  specifying pre-retirement consumption and wealth at retirement (an intermediate outcome), she chooses the alternative  $\phi_{\mathbb{M}}(M)$  and ends up with the intertemporal consumption bundle  $g(\phi_{\mathbb{M}}(M))$ . If her options are described as a set of alternatives  $I$  specifying current consumption and current saving (a consumption instrument), she chooses the alternative  $\phi_{\mathbb{I}}(I)$ , which leads to a particular realization of pre-retirement consumption and wealth at retirement,  $f(\phi_{\mathbb{I}}(I))$ , ultimately providing her with the intertemporal consumption bundle  $g(f(\phi_{\mathbb{I}}(I)))$ . If the consumer fully understands the relationships between current saving, future asset levels, and future consumption, then she should end up with the same consumption bundle regardless of whether she selects it directly from  $C$ , or indirectly from equivalent menus of alternatives involving either intermediate outcomes ( $M$ ) or consumption instruments ( $I$ ).

<sup>11</sup>See Lusardi and Mitchell (2014) for a review.

<sup>12</sup>Iyengar and Kamenica (2010) show both experimentally and in naturally occurring data that when subjects are confronted with large choice sets, they tend to focus on the simplest options.



Intuitively, we measure financial competence by assessing the magnitude of the difference between  $\phi_{\mathbb{M}}(M)$  and  $f(\phi_{\mathbb{I}}(I))$  for pairs of substantively equivalent simply and complexly framed choice problems. Formally, our measure is rooted in the approach to behavioral welfare economics developed by Bernheim and Rangel (2004, 2009). When a consumer’s choice is predicated on an incorrect understanding of the available opportunities, he is said to suffer from *characterization failure*.<sup>13</sup> Evidence of characterization failure provides justification for removing a particular choice from the set of choices one uses to evaluate welfare (the *welfare-relevant domain*).<sup>14</sup> Among other things, one can then evaluate the welfare loss (defined as equivalent or compensating variation) resulting from characterization failure.<sup>15</sup>

In our current context, we establish characterization failure by documenting the joint occurrence of (i) divergences between simply and complexly framed choices, and (ii) limited conceptual understanding of the relationships between intermediate outcomes and the consumption instruments used in the complexly framed problems. Accordingly, simply framed choices remain within the welfare-relevant domain, while complexly framed choices are excluded.

To simplify the measurement of welfare losses associated with complex framing, we use *paired valuation tasks*. A valuation task directly assesses the equivalent variation associated with some alternative. The difference between valuations assessed in simply and complexly framed decision problems indicates the potential welfare loss incurred as the result of complex framing. For example, if a consumer is willing to pay \$10 for an alternative when it is simply framed and \$15 when it is complexly framed, he may end up overpaying by as much as \$5 (e.g., if the complexly framed alternative is offered at a price of \$14.99). Conversely, if he is willing to pay \$15 for the alternative when it is simply framed but only \$10 when it is complexly framed, he may forgo opportunities to receive as much as \$5 of surplus (e.g., if the complexly framed alternative is offered at a price of \$10.01). Naturally, the expected welfare loss depends upon the process that generates the consumer’s opportunities; we assess it within the context of our experiment, where it is proportional to the square of the difference between the two valuations (see the discussion in Section 7).<sup>16</sup> We then ask how various educational interventions affect this quantitative measure of financial competence.

A possible objection to our approach is that the consumer may misunderstand not only the mapping  $f$  from consumption instruments to intermediate outcomes, but also the mapping  $g$  from intermediate outcomes to consumption bundles, believing instead that these relationships are governed by the

<sup>13</sup>This term first appears in Bernheim (2009), but the concept is present in Bernheim and Rangel (2004, 2009).

<sup>14</sup>Other recent applications of this framework include Chetty et al. (2009) and Bernheim et al. (2013).

<sup>15</sup>When choices within the welfare-relevant domain satisfy GARP, one uses the standard notions of equivalent or compensating variation. Otherwise, one uses the generalizations of these concepts introduced in Bernheim and Rangel (2009).

<sup>16</sup>Alternatively, one may want to consider the worst-case welfare loss. As explained above, this is given by the absolute value of the difference between the two valuations.

mappings  $f' \neq f$  and  $g' \neq g$ , respectively. If one is concerned with assessing actual welfare, the proper approach is then to delete both complexly and simply framed choices problems from the welfare-relevant domain, and to derive normative criteria based only on primitively framed choices alone (which usually are not observed). However, for our purposes, comparisons between simply and complexly framed choices remain highly instructive.

Suppose the objective is to evaluate educational interventions that are expected to help consumers distinguish usefully among the elements of  $\mathbb{I}$ , but not among the elements of  $\mathbb{M}$ . Then it may be reasonable to assume that the interventions leave  $g'$  unchanged.<sup>17</sup> In that case, restricting the welfare-relevant domain to simple choices allows us to determine what the effects of the interventions on welfare would be if the consumer's understanding of the relationship between intermediate outcomes and consumption bundles,  $g'$ , were actually correct. That standard is an appropriate one for evaluating the success of an intervention that aims to improve the subject's understanding of  $f$ . In effect, our evaluation strategy is to break the analysis of financial education into pieces; if consumers also misunderstand  $g$ , then plainly one would ideally want to consider interventions that correct those errors as well.

What if, contrary to the assumption made in the previous paragraph, an educational intervention targeted at consumers' understanding of  $f$  also affects their understanding of  $g$ ? If it were perfectly effective, it would still eliminate the differences between  $f$  and  $f'$ , and hence between  $f(\phi_{\mathbb{I}}(I))$  and  $\phi_{\mathbb{M}}(f(I))$ . It follows that the effect of an educational intervention on the gap between the valuations elicited from our complexly and simply framed tasks remains an intuitively useful measure of its success even if the conditions formally required to justify its interpretation as a welfare measure do not hold.

The preceding discussion implicitly assumes that subjects find it more natural to evaluate elements of  $\mathbb{M}$  than elements of  $\mathbb{I}$ , and consequently wish to convert the latter into the former before making choices. If they intrinsically value consumption instruments rather than consumption bundles or intermediate outcomes, the opposite would be true. In some settings, that alternative hypothesis may be implausible; it is always testable. If in fact consumers attempt to convert elements of  $\mathbb{I}$  into elements of  $\mathbb{M}$ , they will take longer to choose from opportunity sets involving instruments than from choice sets involving only intermediate outcomes.<sup>18</sup>

We note that our measure of financial competence is context-specific in two separate senses. First, it is specific to the nature of the consumption instruments used in the complexly framed decision problems. Two different complex framings (involving different instruments) of the same underlying

---

<sup>17</sup>Even without data on primitive choices, this assumption is (indirectly) testable: if it holds, then the interventions should have no effect on  $\phi_{\mathbb{M}}(\cdot)$ . In the current experiment, we verify that this condition is satisfied.

<sup>18</sup>In the current experiment, subjects take roughly two-and-a-half times as long on average to make complexly framed choices than to make simply framed ones.

opportunity set may lead to different choices and welfare losses, in which case they would be associated with different levels of financial competence. For example, consumers may make different investment decisions depending on whether they are provided with information about real or nominal interest rates; both types of choices involve complex framing, but the second raises the possibility that consumers may account for inflation incorrectly. Second, even fixing the nature of the consumption instruments, our measure of financial competence is specific to the opportunity sets under consideration. For example, consider two choice sets,  $I$  and  $I' \subseteq I$ , where  $I'$  excludes the (suboptimal) alternative chosen in  $I$ . Depending on the alternative to which the consumer switches in  $I'$ , the welfare loss could rise or fall.

It bears emphasis that our approach to measuring the quality of financial decision making offers several advantages over more conventional alternatives. First, it is *non-paternalistic*. The welfare criterion is derived from the choices the consumer would himself make if he fully understood his opportunity set. External judgments of consumers' choices, e.g., whether they are “sufficiently patient” or “saving enough,” are entirely avoided.

Second, our approach provides a *direct* and *quantitative* measure of the quality of financial decision making that is formally interpretable in terms of consumer welfare (within the Bernheim-Rangel framework). It is thus directly amenable to cost-benefit analysis. In contrast, existing measures of the quality of financial decisions do not generally provide quantitative measures of welfare losses (because the degree of sub-optimality is not apparent from measures of financial literacy or from complexly framed choices, either alone or in combination), and hence are not amenable to cost-benefit analysis.

Third, our approach imposes modest informational requirements. Because each simply framed valuation provides the normative benchmark for the decisions associated with the equivalent complexly framed valuation, we avoid the need for theoretical explanations and preference-based models of simply and complexly framed decisions. As an alternative, one could collect data on simple choices, estimate a preference-based model describing those choices, and then use the preferences recovered in this way to evaluate complexly framed choices.<sup>19</sup> A limitation of this alternative approach is that it requires one to identify the “right” model for simple choices. In many contexts, this may prove difficult. With the wrong model, good choices in complexly framed decision problems may be interpreted incorrectly as poor ones.

Fourth, because we make comparisons between simply and complexly framed valuations subject by subject, our approach explicitly accounts for population heterogeneity. Ignoring heterogeneity in the context of financial education is dangerous. Because different people start out with different levels of knowledge and even opposite biases, one would expect the effects of education to vary widely. Yet

---

<sup>19</sup>One can interpret Song (2012) as an example of this approach; see Section 9 for further discussion.

most previous studies ignore such heterogeneity, and those that allow for it do so in a comparatively crude manner.<sup>20</sup>

Fifth and finally, as long as the investigator can reduce complexly framed decision problems to simply framed ones, our approach is easily applied.

### 3 Experimental Design

We use our notion of financial competence to evaluate a narrow web-based financial education intervention focused on the concept of compound interest.<sup>21</sup> The structure of our experiment also allows us to evaluate the intervention according to conventional outcome measures.

The experiment consisted of three main stages. First, subjects watched one of four educational videos, selected at random. Second, they completed incentivized paired valuation tasks. Finally, they took a test on compound interest, and answered survey questions concerning the decision strategies they deployed in the second stage. Performance on the test was incentivized, and subjects knew this prior to watching the educational video. Additional explanation of each stage follows; for further details, see Appendix A.

**Education intervention** We used the section on compound interest from a popular investment guide, *The Elements of Investing: Easy Lessons for Every Investor*, by Malkiel and Ellis (2013). We selected this book because it is extremely well-exposed, widely read, and targets young adults who are beginning to think about long-term financial objectives, a group to which most of our subjects belong.

The text begins with a simple explanation of compound interest illustrated through an iterative calculation.<sup>22</sup> The remainder of the text consists of two components:

---

<sup>20</sup>There have been few previous attempts to measure heterogeneous treatment effects. One exception is Bernheim et al. (2001), who found that exposure to financial education in high school increased saving among people who described their parents as “not frugal,” but not among those with “frugal” parents. Cole and Shastry (2012), which replicated a portion that study using different data, failed to detect a significant overall effect on saving, but data limitations precluded them from examining treatment effects within the population sub-groups that were most likely to respond.

<sup>21</sup>We studied an intervention involving this topic for a number of reasons. First, as noted previously, it is associated with a well-documented bias, and hence is the natural focus of an educational intervention. Second, the design of suitable simply and complexly framed tasks is relatively straightforward. Third, this is a core topic in most financial education courses. Finally, the narrowness of the topic, and the corresponding brevity of the pertinent section of the investment guide discussed below, make it suitable for an intervention of limited duration.

<sup>22</sup>The example is: “Stocks have rewarded investors with an average return close to 10 percent a year over the past 100 years. Of course, returns do vary from year to year, sometimes by a lot, but to illustrate the concept, suppose they return exactly 10 percent each year. If you started with a \$100 investment, your account would be worth \$110 at the end of the first year – the original \$100 plus the \$10 that you earned. By leaving the \$10 earned in the first year reinvested, you start year two with \$110 and earn \$11, leaving your stake at the end of the second year at \$121. In year three you earn \$12.10 and your account is now worth \$133.10. Carrying the example out, at the end of 10 years you would have almost \$260 – \$60 more than if you had earned only \$10 per year in simple interest.”

1. An explanation of the rule of 72, along with five illustrative applications. The rule of 72 is a method for approximating an investment’s doubling period; one can also use it to approximate the growth in an investment’s value over a fixed holding period. It states that the percentage interest rate on an investment multiplied by the number of periods required for its value to double equals 72 (approximately).
2. Rhetoric and exhortations. The section opens with the observation that “Albert Einstein is said to have described compound interest as the most powerful force in the universe.” It provides various anecdotes concerning small investments that grew to impressive sums (in some cases millions of dollars) over long time periods.<sup>23</sup> It also explicitly exhorts readers to behave frugally, asserting that “the power of compounding is why everyone agrees that saving early in life and investing is good for you,” and characterizing compounding as a “miracle.”

We employ a  $2 \times 2$  *between subjects* design to isolate the features of the educational intervention that drive changes in test-scores, self-reported decision strategies, choices, and welfare. In our *Full* treatment, subjects viewed a video covering all of the material, both substantive and rhetorical. In our *Substance-Only* treatment, they viewed a shorter video covering all of the substantive material, but omitting exhortations and atmospheric quotes. In cases where it was impossible to remove sentences containing rhetorical material, we substituted neutral language.<sup>24</sup> In contrast, for the *Rhetoric-Only* treatment, subjects viewed a video containing all of the rhetorical material and exhortations, as well as the introductory explanation of compound interest, but omitting all material on the rule of 72. Finally, subjects in the *Control* treatment viewed a stylistically similar video based on a section about index funds from the same investment guide. This section does not mention compound interest or the time value of money, and consequently we would not expect it to affect the types of choices that subjects were subsequently asked to make.

Subjects viewed videos of narrated slide presentations.<sup>25</sup> The narration was verbatim from the text (with a few minor adjustments), while the slides summarized key points. In format, the videos resemble those offered through the educational internet platform *Khan Academy*.<sup>26</sup> Since our study is internet-based, we took several precautionary measures to ensure that subjects were able to view the video and that they would pay attention to it. These are detailed in appendix A.

---

<sup>23</sup>These anecdotes do not include any computations, and hence are not helpful for understanding the mechanics of compound interest.

<sup>24</sup>For instance, the first example of compounding presented in the original text is preceded by the transitional question, “Why is compounding so powerful?” In the Substance-Only-treatment, we substituted the question, “How does compounding work?”

<sup>25</sup>We chose this approach because existing research indicates that financial education videos are generally more effective than written text (Lusardi et al. (2014)).

<sup>26</sup>[www.khanacademy.org](http://www.khanacademy.org).

**Paired valuation tasks** Subjects performed 10 paired valuation tasks. Each task elicited an equivalent current dollar value for a reward  $r$  to be received in either 36 or 72 days. With *simple framing*, the reward was described as follows: “We will pay you  $\$r$  in  $t$  days.” With *complex framing*, the same reward was described in terms of a return on an initial investment, as follows: “We will invest  $\$a$  at an interest rate of  $R\%$  per day. Interest is compounded daily. We will pay you the proceeds in  $t$  days.” Subjects made two sets of choices pertaining to each future reward, one with simple framing, the other with complex framing.<sup>27</sup> For each frame  $f$  (which includes the description of  $a$  and  $R$  for complex framing), we elicited a subject  $j$ ’s immediate dollar equivalent of a payment  $r$  received in  $t$  days,  $V_{j,r,t}^f$ , using the iterated multiple price list method with a resolution of  $\$0.20$  (Andersen et al. (2006)).<sup>28</sup>

Table I lists the parameters  $t$ ,  $r$ ,  $a$ , and  $R$  used for the paired valuation tasks. We chose these values to create variation in the number of times the initial investment doubles over the investment horizon. This allows us to investigate the cause of differences between valuations for complexly and simply framed rewards: subjects who erroneously compute simple rather than compound interest make larger mistakes when the investment horizon is a larger multiple of the doubling period. We chose time horizons of 36 and 72 days to simplify applications of the rule of 72.<sup>29</sup> Because our design is thereby skewed towards settings in which the substantive content of the intervention is potentially most useful, our study is biased in favor of finding *beneficial* behavioral effects. We randomized the order of the valuation tasks at the subject level. Subjects were not told that some of the tasks were substantively equivalent, and they typically did *not* perform equivalent simply and complexly framed tasks consecutively.

Subjects completed the paired valuation tasks at their own pace (subject to the restriction that they could not take more than 3 hours), and we recorded their response times. We intentionally placed no restriction on the use of other resources, such as calculators, the internet, and/or personal advice when making decisions.<sup>30</sup> We take the view that it is appropriate to allow subjects to decide for themselves whether to use such resources, in light of the fact that they always have that option when making real-world decisions. As detailed below, only a quarter of our subjects report exercising that option for answering the incentivized test questionnaire, a fraction that does not vary meaningfully across treatments. Significantly, that pattern mirrors findings concerning real financial decisions (Lusardi and Mitchell (2011)).

---

<sup>27</sup>We chose the parameters of the tasks so that the complexly framed task yielded the same future payment as the simply framed task according to the rule of 72. Since that rule is an approximation, future values actually differ by small amounts between the two frames.

<sup>28</sup>See appendix A for details.

<sup>29</sup>We used two different time frames so subjects would face a greater variety of decision problems, and hence would be less likely to consider successive problems highly similar.

<sup>30</sup>This feature differentiates our study from most of the literature on the effects of financial education (Hastings et al. (2013)). An exception is Levy and Tasoff (2014) who also conduct an internet-based study.

**Knowledge test and self-reports** Despite our focus on financial competence, we also gathered all data required to evaluate the intervention according to more conventional metrics.

Many studies have assessed the effectiveness of financial education interventions using tests of knowledge and understanding (e.g. Jump\$tart Coalition for Personal Financial Literacy (2006), Mandell (2009), Mandell and Klein (2009) Carpena et al. (2011), Heinberg et al. (2014), Lusardi et al. (2014), Walstad et al. (2010), Council for Economic Education (CCE) (2006), Collins (2010)). Accordingly, we administered a test consisting of the five questions about compound interest listed in table II, as well as five questions about the material covered in the video shown to the control group.<sup>31</sup> As mentioned above, performance on this test was incentivized.<sup>32</sup>

Previous studies have also examined self-reported decision strategies (for instance Heinberg et al. (2014), Lührmann et al. (2012), Carlin et al. (2014)). In the final stage of the experiment, we asked subjects whether they had used the rule of 72 in the complexly framed problems, and whether they had used it in the simply framed problems. We also elicited the number of complexly framed valuation tasks for which subjects explicitly calculated the future value of the investment, and asked whether they obtained help when taking the test on compound interest.<sup>33</sup>

**Discussion** It is worth emphasizing that we examine a narrow educational intervention that focuses on a particular skill, and evaluate it based on test questions and decision tasks that are directly connected to its substantive content. In contrast, much of the existing literature on financial education examines broad, highly composite, and often heterogeneous programs (such as high school classes or workplace seminars), as well as behaviors that the curricula may not explicitly address (such as rates of saving).<sup>34</sup> We suspect that these considerations may account, at least in part, for the literature’s mixed findings concerning the effects of financial education. We contend that a narrow focus makes it easier to determine which pedagogical approaches work and which do not, and to develop a nuanced understanding of the mechanisms through which such interventions influence behavior.

## 4 Implementation and Preliminary Analysis

Subjects participated in the experiment online rather than in person. An advantage of this feature is that it mirrors many real-world financial decisions, which have steadily migrated to online platforms. We recruited subjects through an online labor market, Amazon Mechanical Turk (AMT). For our

---

<sup>31</sup>We randomized the order of the questions at the subject level.

<sup>32</sup>Subjects were informed that their test results and choices in the paired valuation tasks would determine their rewards with 25% and 75% probabilities, respectively. For the test results, they received \$1 for each question they answered correctly.

<sup>33</sup>The questionnaire also addressed a small number of additional issues.

<sup>34</sup>See for instance Bernheim et al. (2001), Bernheim and Garrett (2003), Mandell (2009), and Cole and Shastry (2012).

purposes, an important feature of this population is that the typical member has a poor understanding of compound interest. Also, this group resembles the target populations for many financial education programs in terms of demographic characteristics such as age and income. Broadly, experience to date indicates that AMT provides a useful and reliable platform for many types of behavioral research in the social sciences.<sup>35</sup>

We ran eight sessions with a total of 504 subjects during April and May 2014.<sup>36</sup> We restricted participation to subjects who reside in the US and are at least 18 years of age. Subjects logged into our study from the AMT worker interface. They were welcomed by a two-and-a-half minute video recording of one of the authors (Bernheim), who vouched that we would pay subjects exactly the amount we promised them within at most two days of the promised date. The video invited subjects to click a link to the author’s homepage so they could verify the authenticity of the video.<sup>37</sup> Before participating in the main stages of the experiment, subjects completed an unincentivized questionnaire concerning demographics, as well as a standard battery of five questions designed to assess financial literacy.<sup>38</sup>

The average length of a session was 62 minutes (s.d. 22 minutes). Attrition was negligible and unrelated to the treatments.<sup>39</sup> On average, subjects earned \$22.86, including a fixed \$10 participation fee; earnings ranged from a low of \$10 to a high of \$30.47. In comparison, AMT participants typically earn about \$5 per hour.<sup>40</sup>

**Multiple switching** Any subject with coherent preferences will switch his choice from the immediate payment to the future reward at most once within a single price list. We did *not* impose this restriction on our subjects, but instead informed them that “most people begin a decision list by preferring the option on the left and then switch to the option on the right.” As a result, 7.7% of subjects (39 of 504) switched two or more times in at least one price list, and this number does not significantly differ across treatments ( $p = 0.85$ ). In other laboratory studies of risky choices by under-

---

<sup>35</sup>We turned to AMT after pilot experiments revealed that the concept of compound interest was already familiar to most Stanford undergraduates. For reviews on conducting behavioral research with AMT, see Horton et al. (2011), Mason and Suri (2012), and Peysakhovich et al. (2014).

<sup>36</sup>We ran all of the sessions on weekday mornings.

<sup>37</sup>We also invited subjects to click the link to the homepage of a graduate-student co-author (Ambuehl) in case they felt uncomfortable contacting and inconveniencing a professor.

<sup>38</sup>This test of financial literacy originated with van Rooij et al. (2011) and has been used in many other studies (Lusardi and Mitchell (2014)). We reproduce the five questions in appendix B.1. It is standard practice to administer this test without incentivization.

<sup>39</sup>Only four subjects who reached the stage at which they may have viewed a treatment video failed to complete the study. A larger number of subjects quit before reaching that stage, but that type of attrition is necessarily independent of the treatment, and hence largely innocuous; also, there is no reason to think that the pre-attrition sample is more representative of the general population than the post-attrition sample. Technical glitches may be responsible for both kinds of attrition. For example, a small number of subjects contacted us to report that the video failed to load on their computers.

<sup>40</sup>See Mason and Suri (2012).



graduate subjects, the comparable figure typically falls in the range of 10 to 15%.<sup>41</sup> We drop these subjects from the analysis; the results reported below are based on the 455 subjects who respected monotonicity.

**Demographics** We provide a detailed analysis of our subjects' demographic characteristics, broken down by treatment group, in Appendix B.1. While our subjects are not highly representative of the US population, neither are they highly unusual. On average, our sample is a somewhat poorer, better educated, and more likely to live in larger households than the average US citizen. While our sample mirrors the general population with respect to the prevalence of full-time employment, the fraction of respondents who describe themselves as working part-time is twice as high. Perhaps because we recruited our sample through the internet, our sample over-represents males, young adults, whites, urban residents, and people who have never been married.

We classify subjects as having *high financial literacy* if they correctly answered all three questions on (compound) interest in the initial questionnaire, and as having *low financial literacy* otherwise.<sup>42</sup> Roughly 65% of our subjects qualify as highly financially literate according to this criterion. Other studies tend to find somewhat lower levels for US subjects (Lusardi and Mitchell (2009)).<sup>43</sup>

Randomization into treatments was successful. Of the 34 *F*-tests we performed to test for differences in demographic characteristics across treatments (one for each characteristic), two are significant at the 5%-level, and two more are significant at the 10% level. These figures are well within the expected range.<sup>44</sup>

**Time preferences** For the sake of comparability across rewards of different sizes, we analyze subjects' choices in terms of *implied rates of time preference*. For individual  $j$ , reward  $r$ , time horizon  $t$ , and frame  $f \in \{simple, complex\}$ , this rate is given by

$$\delta_{j,r,t}^f = 100 \times V_{j,r,t}^f / r \tag{1}$$

where  $V_{j,r,t}^f$  is the subject's elicited valuation.<sup>45</sup>

---

<sup>41</sup>See, e.g., Holt and Laury (2002).

<sup>42</sup>These are questions FL1 - FL3 listed in appendix B.1. They are considerably easier than the questions on compound interest administered at the end of each session. Our subjects performed much less well on the latter questions.

<sup>43</sup>Lusardi and Mitchell (2009) report findings based on the American Life Panel, an online survey.

<sup>44</sup>See appendix B.1 for the results.

<sup>45</sup>As noted above, the price lists measure valuations in \$0.20 increments. Throughout, we set  $V_{j,r,t}^f$  equal to the midpoint of the pertinent interval.

Focusing on simply framed choices, subjects' mean rates of time preference are 76.7% and 70.6% for the 36 and 72 day horizons, respectively.<sup>46</sup> The distribution of subjects' mean rates of time preference is spread out between 0% and 100%, with only few subjects exhibiting rates exceeding 100%.

Notably, the treatments did not significantly affect the rates of time preference subjects expressed through their simply framed choices. Differences in average rates across the four treatments are not statistically significant at the 5% level.<sup>47</sup> (See Appendix C.1 for complete regression results.)

**Attention** A concern with studies conducted on internet platforms is that some subjects may pay insufficient attention to the experimental tasks. We motivated subjects to attend by providing monetary incentives that were large relative to the wages for which they ordinarily work, and by emphasizing the broader value of understanding the material covered in the videos. Several findings suggest that we were successful. First, choice patterns are coherent, both with respect to time preferences, and with respect to our educational interventions. Second, the extremely low rate of attrition (mentioned above) indicates that subjects were highly engaged. Indeed, many subjects provided us with unsolicited positive feedback concerning the educational interventions, in some cases asking for copies of the videos.<sup>48</sup> Third, in response to the exit survey, the overwhelming majority of subjects reported paying the highest level of attention to the video and to their choices.<sup>49</sup> Fourth, in Appendix D.1, we consider the possibilities that inattention may have caused subjects to make noisy choices, or to be unresponsive to varied stimuli. We show that our conclusions are robust with respect to both considerations.

## 5 Financial literacy and Self-Reported Decision Strategies

Studies that evaluate financial education interventions frequently focus on financial literacy and/or self-reported behavioral outcomes. (See section 3 for references.) One can draw conclusions about the welfare consequences of financial education from such studies if one is willing to assume that financial education affects behavior only through its effects on the understanding of financial concepts, and that such understanding necessarily promotes better decision making. We begin our analysis by examining effects on a collection of such variables.

---

<sup>46</sup>Note that our typical subject discounts the future rather heavily, that the mean rate of time preference for the longer horizon is lower, and that the relative magnitudes of these rates across horizons are inconsistent with exponential discounting. These three findings are common in studies that elicit time preferences over short horizons (Frederick et al. (2002)).

<sup>47</sup>The Rhetoric-Only treatment, however, raised rates of time preference by 5.78 percentage points relative to Control (significant at the 10% level). Splitting our sample according to subjects' financial literacy scores, we find that this effect is driven by the subjects with high financial literacy, for whom the difference is significant at the 5% level.

<sup>48</sup>In order to control dissemination and exposure of potential future subjects to the videos, we did not provide them in response to these requests.

<sup>49</sup>However, in light of other results reported below, our confidence in the accuracy of self-reports is not high.

We measure treatment effects by estimating the following regression:

$$y_{j,k} = \beta_{Control} + \beta_{Full}Full_j + \beta_{Substance}Substance_j + \beta_{Rhetoric}Rhetoric_j + \epsilon_{j,k} \quad (2)$$

Here,  $j$  indexes individuals and  $k$  indexes decisions;  $y_{j,k}$  is an outcome variable, and  $Full_j$ ,  $Substance_j$ , and  $Rhetoric_j$  are treatment dummies. Hence the intercept  $\beta_{Control}$  measures the average level of  $y_{j,k}$  in the Control condition, and the parameters  $\beta_{Full}$ ,  $\beta_{Substance}$ , and  $\beta_{Rhetoric}$  measure the effect of the corresponding treatment on that average. We assume  $\epsilon_{j,k}$  is orthogonal to the treatment dummies. Whenever a regression includes multiple observations per subject, we cluster standard errors at the subject level.

Column 1 of table III shows the effects of the various treatments on subjects' test scores for the five questions pertaining to compound interest. In the Control condition, the average subject answers just under two of five questions correctly. The Full intervention dramatically increases the average score, by about 1.4 additional correct answers. To put this effect in perspective, the average percentage test score rises from 39% to 68%. When the rhetoric is removed from the intervention (the Substance-Only treatment), the effect is only slightly smaller, and the difference is not statistically significant. In contrast, when material on the rule of 72 is removed (the Rhetoric-Only treatment), the effect declines dramatically, to roughly 0.5 (or 10 percentage points).<sup>50</sup> Thus, according to standard measures, the substantive interventions are highly effective at promoting financial literacy.<sup>51</sup>

Column 2 of table III shows the effects of the various treatments on subjects' test scores for the five test questions pertaining to topics covered in the Control video. Notice that the Control subjects gave slightly more than one additional correct answer on average than the other groups, an improvement of more than 20 percentage points. This finding is notable because it rules out the possibility that differences in test performance between the Control group and the treatment groups are due to effects on general motivation.

Column 3 of table III examines the effects of the various educational interventions on the (self-reported) extent to which subjects employ external help. A natural concern is that education may simply displace the use of reference materials or reliance on knowledgeable friends. Such displacement could in principle dampen the effects of the interventions on test scores and choices. In fact, the differences across the treatments are small and statistically insignificant. If anything, reliance on external help appears to be slightly higher for the Substance-Only and Rhetoric-Only treatments than for the Control (though the differences are not statistically significant at the 10% level).

---

<sup>50</sup>Recall that the Rhetoric-Only treatment includes the simple explanation of compound interest, illustrated through an iterative calculation. Accordingly, it is not surprising to find some effect on measured financial literacy.

<sup>51</sup>Appendix C.2 details the effects of the treatments on individual test questions.

The results in Column 4 of table III show that subjects reported operationalizing the knowledge they acquired from the substantive interventions. Only 13% of subjects in the Control reported using the rule of 72 when making complexly framed choices. In sharp contrast, the corresponding figure exceeded 70% for the Full and Substance-Only treatments. Somewhat surprisingly, we also find an increase – albeit a much smaller one – for the Rhetoric-Only treatment.<sup>52</sup>

As shown in Column 5 of table III, we find a qualitatively similar pattern for self-reported operationalization of the rule of 72 in simply framed choices; however, the frequencies and treatment effects are all considerably smaller than for complexly framed choices. Because subjects may report using the rule of 72 in simply framed problems for a variety of reasons,<sup>53</sup> this finding is not entirely unexpected.

In principle, the increased use of the rule of 72 could crowd out other types of calculations, such as iterative computations, applications of the compound interest formula, or (inappropriate) evaluations of simple interest. Depending on the nature of the displaced approach, such crowding out could dampen the effect of education on test scores and behavior. In fact, the results in Column 6 of table III show that the Full and Substance-Only interventions significantly increased the average number of complexly framed decision tasks for which subjects reported making explicit calculations, from roughly 6.4 to 8.1 out of 10 (i.e., by approximately 27%). For the Rhetoric-Only treatment, the effect was much smaller and statistically insignificant. Thus, the educational interventions did not simply increase (self-reported) reliance on the rule of 72 by migrating subjects from other methods of explicit calculation.

We are of course mindful that changes in self-reported behavior could involve experimenter-demand effects. Indeed, that limitation is an important reason for developing and implementing an objective measure of financial competence. We are much less concerned that similar considerations could account for the measured effects on incentivized test scores, which likely reflect actual knowledge.<sup>54</sup>

Several lessons emerge from this analysis. First, the Full intervention successfully and significantly increased financial literacy. Second, it was successful for the right reason: removing rhetorical material and retaining substance leaves the effect on financial literacy almost unchanged, whereas removing substantive material and retaining rhetoric reduces it dramatically. Third, according to self-reports, the Full intervention successfully motivates subjects to operationalize their newly obtained knowledge

---

<sup>52</sup>There are two possible explanations for this finding. One is that some subjects already know the rule of 72 but apply it only when they are sufficiently motivated. The other is that rhetorical exhortation motivates subjects to misrepresent their knowledge and use of the rule.

<sup>53</sup>Subjects may apply the rule inappropriately, they may discount future rewards to the present at a market interest rate, or they may misrepresent their actual decision processes.

<sup>54</sup>Subjects were plainly motivated to perform well on the incentivized test for their own benefit. Indeed, we received a large number of comments from subjects who complained that they had been tested on material not covered in the intervention video. (Recall that the test covers the substantive material in both the Full video and the Control video, and that each subject views only one video.)

in their decisions. Fourth, removing rhetorical material and retaining substance leaves the effect on self-reported operationalization of knowledge almost unchanged, whereas removing substantive material and retaining rhetoric reduces it dramatically. Fifth, according to self-reports, the use of new quantitative tools does not simply crowd out reliance on other resources or other computational methods.

Based on these results, one would expect to find that the Full intervention unambiguously improves the quality of financial decision making, and that this effect is driven by substantive material rather than rhetoric. However, the results presented in the following sections paint a much different picture.

## 6 Financial Choices

Much of the literature on financial education is concerned with measuring effects on behavior. (This literature originated with Bernheim and Garrett (2003) and Bernheim et al. (2001); other examples include Duflo and Saez (2003), Bayer et al. (2009), Goda et al. (2012), Cole et al. (2011), Skimmyhorn (2012), Gartner and Todd (2005), Servon and Kaestner (2008), Collins (2010), Lührmann et al. (2014)). Some studies also make casual inferences concerning welfare by asking whether these effects directionally counteract presumed biases and thereby redress deficiencies in decision making. For instance, financial education interventions are often deemed successful if they increase contributions to retirement savings accounts.

For the types of decisions we examine in this study, there is a strong presumption that people typically underestimate the power of compound interest (Wagenaar and Sagaria (1975), Eisenstein and Hoch (2007), Stango and Zinman (2009), Almenberg and Gerdes (2012), Levy and Tasoff (2014)), a phenomenon known as *exponential growth bias*. Consequently, following the approach adopted in the literature, one would deem an intervention welfare-improving if it increased the average valuations of the complexly framed rewards.

One natural approach would be to study the effects of the various interventions on rates of time preference,  $\delta_{j,r,t}^c$ , implied by complexly framed valuations,  $V_{j,r,t}^c$ . We adopt a slight variant of that approach, normalizing the rate of time preference with complex framing using the corresponding rate of time preference with simple framing. Formally, we define the *framing distortion* for reward  $r$ , time horizon  $t$ , and individual  $j$ , as:<sup>55</sup>

$$d_{j,r,t} = \delta_{j,r,t}^c - \delta_{j,r,t}^s$$

---

<sup>55</sup> $\delta_{j,r,t}^f$  is subject  $j$ 's implied rate of time preference for reward  $r$  to be received at time  $t$  in framing  $f$ . See equation 1 in section 4.

In previous studies, the data required to perform this normalization have not been available. Because simply framed valuations do not vary significantly across treatments, examining the framing distortion rather than the implied rate of time preference has very little effect on the measured treatment effects.<sup>56</sup> The advantage of analyzing the framing distortion is that it renders magnitudes more easily interpretable. For instance, if an educational intervention eliminates a tendency for subjects to underestimate compound interest, we would find that the average framing distortion is significantly negative for the Control group and zero for the pertinent treatment group.

Once again we assess treatment effects by estimating the regression model specified in equation (2), in this case using  $d_{j,r,t}$  as the dependent variable. To explore the robustness of our findings, we present multiple versions of this model in Table IV. The versions shown in the first two columns use all of our data. The first column exhibits results for the basic specification; for the second column, we add demographic controls.<sup>57</sup> Both regressions pool observations across subjects with high and low levels of financial literacy, and across investments with different horizons. Because the effects of financial education may differ according to the subjects' initial knowledge, we present versions of the basic model estimated separately for subjects with high and low financial literacy (columns 3 and 4). Likewise, we present versions estimated separately for observations involving decisions with time horizons of 36 and 72 days (columns 5 and 6).

Consistent with other evidence on exponential growth bias, results for the Control group show that subjects on average substantially underestimate the benefits of compound interest: in the basic specification, complex framing reduces the apparent rate of time preference by more than 13 percentage points (column 1). With demographic controls (column 2), the corresponding figure represents the average framing distortion with all the demographic variables zeroed out.<sup>58</sup> We review detailed results concerning demographic correlates of framing distortions in Appendix B.2. Exponential growth bias is more severe for subjects with low financial literacy than for those with high financial literacy, but it is substantial even for the latter group (columns 3 and 4). Finally, differences in the severity of the bias across decisions with different time horizons are minor (columns 5 and 6).

Focusing again on the basic specification, we see that the estimated effect for the Full treatment (13.91 percentage points) is extremely close to the absolute value of the average framing distortion for the Control group (-13.31 percentage points).<sup>59</sup> Thus, the Full treatment not only drives valuations in the right direction, but also essentially *eliminates* the average framing distortion. The same conclusion

---

<sup>56</sup>Appendix D.3 presents results for the unnormalized rates of time preference, and shows that our substantive conclusions are generally unaffected. One exception is that, without the normalization, we find a significant positive effect for the Substance-Only intervention on subjects with low financial literacy.

<sup>57</sup>Specifically, we add the term  $\gamma'X$  to equation (2).  $X$  consists of income, as well dummies for gender, age, rural/urban residence, race, education, employment, marital status, household size, stock ownership, and financial literacy.

<sup>58</sup>Since some of the demographic variables, such as income, are generally positive, this parameter does not have a straightforward interpretation.

<sup>59</sup>Appendix B.2 analyzes how demographic characteristics modulate treatment effects.

holds for each of the other specifications: the absolute difference between  $\beta_{Control}$  and  $\beta_{Full}$  ranges from a low of 0.09 percentage points (for decisions with 36 day time horizons) to 1.11 percentage points (for decisions with 72 day time horizons).<sup>60</sup>

Before proceeding, it is worth reviewing some of the key results presented so far. We have found that the Full treatment substantially increases measured understanding of compound interest, the frequency with which subjects report operationalizing this understanding when making decisions, and the extent to which they report making any explicit calculations. Moreover, we find no evidence that the treatment displaces the use of other reliable resources. These findings lead us to conjecture that the Full treatment will not only influence behavior, but will do so for the right reasons. Consistent with that expectation, we find that the Full treatment increases complexly framed valuations, thereby presumably offsetting exponential growth bias. Indeed, the treatment essentially eliminates the average framing distortion.

On the basis of these findings, one would be strongly inclined to conclude that the Full intervention was highly effective and beneficial. Moreover, because the Substance-Only intervention has essentially the same effects on measured financial literacy and self-reported decision strategies, one might well conjecture that it must be equally successful and beneficial, and that rhetoric is an unnecessary distraction. Surprisingly, all of these conclusions are incorrect.

Focusing again on the basic specification, we see that the estimated effect for the Substance-Only treatment (4.00 percentage points) is statistically indistinguishable from zero, and significantly smaller than that of the Full treatment (13.91 percentage points,  $p = 0.00184$ ). In contrast, the estimated effect for the Rhetoric-Only treatment (13.22 percentage points) is almost identical to that of the Full treatment, and we do not reject equality ( $p = 0.827$ ). Accordingly, despite demonstrable effects on comprehension and subjects' statements concerning their proclivities to operationalize substantive knowledge in their decisions, *the behavioral effects of the Full treatment are traceable almost entirely to its rhetorical components*. These findings are robust across all of the specifications reported in Table IV.

So far, we have focused exclusively on the average framing distortion. Figure 1 shows the cumulative distribution of  $d_{j,r,t}$  for each treatment. While framing distortions are clustered near 0, there is substantial variation in  $d_{j,r,t}$  within each treatment. Notice that the Substance-Only treatment shifts the CDF for the Control group slightly to the right. Both the Full and Rhetoric-Only interventions yield much larger rightward shifts. This pattern is worrisome, inasmuch as an effective educational intervention would shift valuations upward for those with negative values of  $d_{j,r,t}$ , and *downward* for those with positive values of  $d_{j,r,t}$ . Thus, one would hope to see the treatment CDFs *crossing* the

---

<sup>60</sup>The difference appears to be larger for the specification with demographic controls, but recall that the "Level in Control" coefficient zeroes out all demographic variables, including income.

Control CDF at  $d_{j,r,t} = 0$ . Instead, the increase in valuation appears to be independent of the initial bias.

In light of these additional findings, there is reason to question whether the Full treatment actually improves the quality of subjects' decisions. By analyzing financial competence, we can address that issue formally and with quantitative precision.

## 7 Financial Competence

We now turn to the welfare effects of the educational interventions. As far as we know, Song (2012) is the only existing study to attempt a formal, quantitative welfare analysis of a financial education program.<sup>61</sup>

As discussed in section 2, we measure welfare (financial competence) by calculating the expected monetary loss from poor decision making in complex frames, using simply framed choices as the normative benchmarks. To illustrate, if the willingness to pay (WTP) for a future reward is \$10 with simple framing and \$15 with complex framing, and if the reward is available for \$12, the loss is \$2 (because the subject pays \$2 more than the valuation he attaches to the reward when he understands his opportunity). If the WTPs are reversed (\$10 with complex framing and \$15 with simple framing), the loss is \$3 (because the subject fails to take advantage of an opportunity that he would recognize as providing \$3 in surplus if he understood it). In either case, if the reward is instead available for \$8 or \$17, there is no loss (because the subject makes the same decision with both simple and complex framing).<sup>62</sup> When the price of the future reward is uniformly distributed (as it is in our experiment), the probability of a loss is proportional to the difference between the simply and complexly framed valuations, while the expected value of a loss conditional on its occurrence is exactly half of that difference. Thus, the expected loss is proportional to the square of the difference in valuations. Normalizing by the size of the reward, we use  $W = -(\delta_{j,r,t}^s - \delta_{j,r,t}^c)^2$  as our welfare measure.

As explained in section 2, our approach assumes that simply framed choices provide appropriate normative benchmarks for evaluating equivalent complexly framed choices, which are subject to characterization failure. Two separate findings provide support for this assumption. First, subjects

<sup>61</sup>Both Hastings et al. (2013) and Lusardi and Mitchell (2014) note that there have been no carefully crafted cost-benefit analyses of financial education interventions.

<sup>62</sup>Formally, the "losses" described in this paragraph are interpretable as equivalent and/or compensating variations. Suppose the subject starts out with the outcome of the simply framed choice. For the first example given in the text (for which the WTPs are \$10 with simple framing and \$15 with complex framing, and the reward is available for \$12), \$2 is the compensating variation associated with switching to the outcome of the complexly framed choice. For the second example given in the text (for which the WTPs are reversed), \$3 is the equivalent variation associated with the same switch. In using our welfare measure, we implicitly assume that people exhibit negligible income effects in their simply framed choices over the relevant range; e.g., an agent who considers \$12 today and \$0 in the future just as valuable as \$0 today and \$R in the future is also indifferent between \$15 today and \$0 in the future on the one hand, and \$3 today and \$R in the future on the other hand. With that assumption, equivalent and compensating variations coincide.



take 59 seconds on average to complete one complex valuation task, compared with only 22 seconds for an equivalent simple task ( $p < 0.001$ ). This difference suggests that the complexly framed tasks require additional cognitive effort, presumably because subjects attempt to reduce them to simply framed tasks. Second, the treatment interventions have significant effects on the complexly framed valuations, but leave the simply framed valuations unchanged on average.<sup>63</sup>

Table V presents our estimate of welfare effects for the same collection of specifications as in Table IV. The coefficients in the row labeled “Level in Control” measure the expected welfare loss absent any intervention.<sup>64</sup> Comparing columns (3) and (4), we see that the ambient level of welfare is lower for those with less financial literacy ( $p = 0.028$ ). Consequently, our findings are consistent with the hypothesis that greater financial literacy is at least *correlated* with better decision making. Comparing columns (5) and (6), we see that the ambient quality of decision making is unrelated to the time horizon.

Turning to the effects of the interventions, we see that the Full treatment has essentially no effect on welfare. In five of the six specifications, average welfare is actually lower in the Full treatment than in the control, but the differences are both small in magnitude and statistically insignificant. For the Substance-Only and Rhetoric-Only treatments, the estimated effects on welfare are uniformly positive, but relatively small and, in most cases, statistically insignificant. Indeed, for the Substance-Only treatment, we cannot reject the hypothesis that the effect on welfare is zero in any specification. For the Rhetoric-Only treatment, the effects are somewhat larger and estimated with comparable precision. Indeed, the increase in welfare is statistically significant for subjects with high financial literacy and (marginally) for decisions with 36-day horizons. Thus, ironically, we find that the least substantive intervention is most effective at raising welfare.

We reach these stunning conclusions despite the fact that the Full treatment significantly enhances financial literacy, induces people to operationalize their knowledge in their decisions without reducing reliance on other resources (according to self-reports), increases the frequency with which people report using decision strategies that involve explicit calculations, and brings average complexly framed valuations into almost perfect alignment with average simply framed valuations. The failure of the Full treatment to increase welfare despite these effects is apparently attributable to the two surprising findings noted in the preceding section – that its behavioral effects are in fact driven by rhetoric, and are indiscriminate (in the sense that they are unrelated to the initial framing bias).

To provide further insight concerning the welfare effects of the Full treatment, we separate our aggregate welfare measure into two components, based on the sign of the framing distortion  $d_{j,r,t}$ .

---

<sup>63</sup>Note, however, that by the symmetry of the welfare function, our analysis would proceed unchanged if instead we used the complexly framed choices as normative benchmarks for evaluating the equivalent simply framed choices.

<sup>64</sup>As before, the “Level in Control” coefficient for the specification with demographic controls is not directly comparable to the corresponding coefficients for the other specifications.

Specifically, we use  $W^- = -(\min\{d_{j,r,t}, 0\})^2$  to capture the part of the welfare loss that is due to *underestimation* of the power of compound interest, and  $W^+ = -(\min\{0, -d_{j,r,t}\})^2$  to capture the part that is due to *overestimation*. Note that  $W = W^- + W^+$ . Column (1) of Table VI (which uses our basic specification) reports the effects of the various treatments on  $W^-$ . As one would expect based on all of the preceding, the Full and Rhetoric-Only treatments significantly reduce welfare losses from underestimating compound interest. (The Substance-Only treatment leaves  $W^-$  unaffected.) The explanation for the overall null effect of the Full treatment on  $W$  is readily apparent from column (2): because its behavioral effects are unrelated to the initial framing bias, it *increases* the welfare loss associated with overestimation of compound interest. These two opposing effects roughly offset. For the Rhetoric-Only treatment, the welfare loss associated with overestimation of compound interest also grows, but by a smaller amount. (The Substance-Only treatment also leaves  $W^+$  unaffected).

These findings are precisely what one would expect in light of Figure 1. An ideal intervention would help each subject recognize the direction and magnitude of his error and adjust accordingly. Yet as noted in the previous section, the Full and Rhetoric-Only treatments appear to raise  $\delta_{j,r,t}^c$  indiscriminately, irrespective of whether an individual initially underestimates or overestimates the benefits of compound interest.<sup>65</sup> Visually, both treatments shift the entire distribution of framing distortions to the right. Neither truly resolves exponential growth bias; instead, they appear to introduce countervailing biases. The Rhetoric-Only treatment is more beneficial than the Full treatment because its influence is more highly correlated with the severity of the initial exponential growth bias: it induces a larger rightward shift for small framing distortions, and a smaller rightward shift for large framing distortions (so that the CDFs for the Full and Rhetoric-Only treatments cross). Unfortunately, because the reasons for this fortuitous correlation are unknown, we are unable to draw general conclusions from the apparent superiority of the Rhetoric-Only treatment.

As explained above, we use a quadratic welfare function because it yields a measure of the expected loss resulting from complex framing, given the stochastic choice environment that subjects actually face in this experiment. Naturally, this measure is context specific; for example, a change in the relative probabilities of implementing different choices would change the subjects' expected losses. It is therefore important to assess whether our conclusions are robust with respect to the use of alternative welfare functions that aggregate across choice tasks using different weights. We consider two alternatives: the (negative of the) absolute value, and the cubed absolute value, of the framing distortion. Compared with the quadratic welfare function used above, the absolute value places more

---

<sup>65</sup>Levy and Tasoff (2014) also find that a substantial fraction of subjects overestimate compound interest. Song (2012) concludes that the effect of an educational intervention concerning compound interest conducted in rural China on his subjects' retirement savings contributions is largely independent of the extent to which the their saving differed from an optimal benchmark derived from a life-cycle consumption model. In particular, he finds that the education intervention induces some subjects to oversave.

weight on small decision errors, and the cubed absolute value places more weight on large ones. Notice that the absolute value of the difference in valuations is proportional to the largest possible loss. Results appear in columns (3) and (4) of Table VI. The effect remains statistically insignificant for the Full and Substance-Only treatments in all specifications, and becomes statistically significant for the Rhetoric-Only treatment only with the absolute value welfare function. Thus our central conclusions are qualitatively unaffected.

So far, we have proceeded as if each subject’s choice mapping is deterministic.<sup>66</sup> Stochastic choice patterns would affect our results through two separate channels. First, were we to substitute the substantively equivalent simply framed valuation task for each complexly framed valuation task in our experiment, so that subjects performed each simply framed task twice, we would likely find that these paired choices would differ in some cases. Applying the method implemented above, we would then measure a positive welfare loss in simply framed choices, even though we would be using (other) simply framed choices as normative benchmarks. As we show in Appendix D.2, explicit recognition of this consideration does not alter our main conclusions concerning welfare. Intuitively, noisy choice inflates the overall *level* of the welfare losses measured with our method, but does not materially affect the *relative* magnitudes of the measured welfare losses for the various interventions.

Second, even though stochasticity in choice would not obscure effects on average framing distortions, it could hamper efforts to detect improvements in welfare.<sup>67</sup> We discount this concern because the choices of our typical subject display a high degree of internal consistency.<sup>68</sup> In an abundance of caution, we also address it by performing additional subject-level analyses using a slightly modified welfare measure: for each subject, we compute the average value of  $d_{j,r,t}$  over the ten pairs of valuation tasks, square it, and change signs.<sup>69</sup> Averaging before squaring substantially attenuates the effects of choice stochasticity on measured welfare. Appendix D.2 shows that our results remain largely unchanged.<sup>70</sup>

---

<sup>66</sup>All analyses in this section are based on the assumption that subjects’ choices derive from well-defined preferences that satisfy GARP (and are possibly implemented with noise). Appendix C.3 presents an analysis of our data based on the Bernheim-Rangel framework which does not require that assumption.

<sup>67</sup>More precisely, if we actually observe  $D_{j,r,t} = d_{j,r,t} + \eta_{j,r,t}$ , where  $\eta_{j,r,t}$  is random noise, we will encounter no bias in measuring the mean of  $D$ , or how it changes across treatments. Suppose, however, that the distribution of  $d_{j,r,t}$  for the control group is highly concentrated with a mean of  $-m < 0$ , and that a treatment shifts that distribution upward by a constant amount  $c < m$ . Under these assumptions,  $-(d_{j,r,t})^2$  rises for the vast majority of subjects. However, if the variance of  $\eta_{j,r,t}$  is sufficiently large,  $D_{j,r,t}$  may be positive for many subjects in the control group, and  $-(D_{j,r,t})^2$  will fall for those subjects. With sufficient noise, the latter effect may obscure beneficial effects on welfare.

<sup>68</sup>Appendix C.4 shows that subjects who underestimate (overestimate) compound interest in some decisions tend to do so in all decisions, and by comparable amounts. Accordingly, the degree of idiosyncratic randomness in individual choice tasks must be limited.

<sup>69</sup>In other words, for each subject  $j$ , we calculate  $-\left(\frac{1}{10} \sum_{r,t} d_{j,r,t}\right)^2$ .

<sup>70</sup>One change is that the effect of the Rhetoric-Only treatment on welfare is now statistically significant, presumably because subjects who systematically overestimate compound interest are less common than individual choices reflecting overestimation of compound interest.

## 8 Mechanisms

There is clear evidence that at least two of our three treatments significantly alter behavior. In this section, we attempt to shed some additional light on the associated mechanisms. Our findings so far exclude the possibility that the interventions led subjects to operationalize substantive knowledge properly. Here we address several additional questions. First, did they at least suppress reliance on simple interest calculations? Second, did they induce subjects to think more deliberately about their choices (which presumably would take more time)? Third, were the substantive interventions more effective at improving decision-making quality for problems to which the rule of 72 is more easily applied? An affirmative answer to all three questions would provide hope that more extensive training might yield unambiguous benefits. Unfortunately, that is not what we find. Reliance on simple interest calculations is dramatically reduced, and the substance-only treatment appears to foster more careful deliberation. However, with the addition of rhetoric, the latter effect disappears. Discouragingly, there is no indication that the benefits of the substantive interventions differ according to the difficulty of applying the rule of 72.

We begin by investigating reliance on simple interest calculations. Previous research indicates that many people believe investment values grow linearly rather than exponentially (Eisenstein and Hoch (2007), McKenzie and Liersch (2011)).<sup>71</sup> Here we ask whether our interventions induced those who otherwise would have employed the logic of simple interest to abandon it in favor of some other alternative (one that is apparently inappropriate in most cases). There are at least two other possibilities: those who employed simple interest calculations may have continued to do so with an *ad hoc* adjustment to account for their downward bias, or the behavioral effects of the interventions may be attributable to those subjects who did not rely on the logic of simple interest to begin with.

We differentiate between these hypotheses as follows. Let  $FV_{r,t}^{SI}$  and  $FV_{r,t}^{CI}$  denote the future value of an investment calculated according to simple and compound interest, respectively. Then  $\frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}}$  represents the degree to which simple interest understates the investment's true value. If subject  $j$ 's choices are guided by the simple interest formula, then this ratio should correlate with his valuation ratio,  $\frac{V_{j,r,t}^c}{V_{j,r,t}^s}$ . In contrast, if  $j$ 's choices are consistent with correct compounding, then his valuation ratio should equal one.

Formally, we estimate the following regression model:

$$\frac{V_{j,r,t}^c}{V_{j,r,t}^s} = \sum_{\tau \in T} \left[ \beta_0^\tau + \beta_1^\tau \frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}} \right] \mathbb{I}_j(\tau) + \epsilon_{j,r,t} \quad (3)$$

---

<sup>71</sup>Likewise, we conducted a pilot study on Amazon Mechanical Turk in which we asked subjects to calculate the future value of four different investments, with given interest rates and time frames. The pilot revealed that the most common modes of calculation are evaluation of simple interest and (correct) evaluation of compound interest. Of course, many subjects fell into neither of these categories.

where  $T = \{Control, Full, Substance, Rhetoric\}$  is the set of all treatments, and  $\mathbb{I}_j(\tau)$  is an indicator function that equals 1 if subject  $i$  is in treatment  $\tau$ .<sup>72</sup> In this specification,  $\beta_1^\tau$  gauges the prevalence of simple interest calculations. Suppose for example that all subjects compute future value according to either the simple or compound interest formula. Then  $\beta_0^\tau + \beta_1^\tau = 1$ , and we can interpret  $\beta_1^\tau$  as the fraction of decisions that are consistent with simple rather than compound interest calculations in treatment  $\tau$ . In the extreme, if all subjects correctly calculate future value, we should find  $\beta_0^\tau = 1$  and  $\beta_1^\tau = 0$ , and if all subjects use the simple interest formula, we should find  $\beta_0^\tau = 0$  and  $\beta_1^\tau = 1$ .

We estimate (3) pooling data for all of our subjects, as well as separately for subjects with high and low financial literacy. In each case, we pool data across all valuation tasks.<sup>73</sup> Having observed that the distribution of the dependent variable is highly skewed due to the presence of observations with values of  $V_{j,r,t}^s$  is close to 0, we chose to estimate the model using median regression.

Results appear in Table VII. We interpret our basic specification as indicating that roughly 30% of subjects use the simple interest formula for valuation tasks in the Control group. Not surprisingly, this method appears to be far more prevalent among those with low financial literacy (for which the comparable figure is roughly 49%) than among those with high financial literacy (for which the comparable figure is roughly 20%). The Substance-Only treatment reduces reliance on simple interest calculations to roughly 9% overall (roughly 29% for those with low financial literacy, and roughly 6% for those with high financial literacy). Notably, both the Full and Rhetoric-Only treatments essentially eliminate dependence on simple interest calculations for both groups (though the effect of the Rhetoric-Only treatment on subjects with low financial literacy is estimated imprecisely). Hence, *all* of our treatments successfully discourage reliance on the logic of simple interest.

For all three specifications and every treatment group,  $\beta_0^\tau + \beta_1^\tau$  is extremely close to unity, suggesting that our model is well-specified.<sup>74</sup> Absent other evidence, one might therefore be tempted to conclude that subjects make either simple interest or (correct) compound interest calculations, and that the interventions successfully push them toward the latter. However, in light of our findings concerning welfare, it is clear that, even though all of the interventions discourage the use of the simple interest formula, they do *not* succeed in fostering the correct calculation of compound interest.

---

<sup>72</sup>Note that  $\frac{V_{j,r,t}^c}{V_{j,r,t}^s}$  is likely independent of subject  $i$ 's time preferences: If subject  $i$  perceives future values  $FV_{j,r,t}^f$  in frame  $f$ , and  $V_{j,r,t}^f = \tilde{\delta} FV_{j,r,t}^f$ , then  $\frac{V_{j,r,t}^c}{V_{j,r,t}^s}$  is independent of  $\tilde{\delta}$ .

<sup>73</sup>In particular, our regressions employ data for valuation tasks with both 36 and 72 day horizons. As discussed later in this section, there is reason to think that subjects may be more likely to compute compound interest with 72 day horizons, at least in the treatments that teach the rule of 72. If the time horizon were systematically related to the values of  $\frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}}$ , our estimates of 3 could confound the effects of the future value ratio with the effects of the time horizon. This is not a problem, however, because we have chosen the parameters of the valuation tasks so that the values of  $\frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}}$  are the same for both time horizons. In any case, as shown below, the time horizon does not appear to have much of an effect on the valuation ratio in practice.

<sup>74</sup>We fail to reject the hypothesis that  $\beta_0^\tau + \beta_1^\tau = 1$  in all cases with  $p > 0.3$ .

We can gain further insight into the mechanisms through which the interventions affect behavior by studying subjects' response times,  $\tau_{j,r,t}^f$ . Column (1) of table VIII shows how the various treatments affect response times for complexly framed choice tasks. The Substance-only treatment significantly increases the time that subjects spend on those tasks, while the Full and Rhetoric-only treatments do so to a much lesser extent. Thus, the provision of substantive information appears to induce greater effort and deliberation, but simplistic rhetorical assertions concerning the power of compound interest seem to negate that effect, perhaps because they point to a less cognitively demanding heuristic. Column (2) shows that this effect is indeed limited to the complexly framed tasks; response times for the simply framed tasks do not differ significantly across treatments.

Conceivably, our failure to promote the correct use of compound interest calculations in decision making could reflect the complexity of the choice tasks, particularly given the limited nature of the interventions. To investigate this possibility, we examined the effects of the various interventions on welfare, differentiating between tasks according to the difficulty of applying the rule of 72. As a general matter, the rule is easiest to apply when the investment in question doubles only once over the time horizon, more difficult to apply when it doubles an integer number of times, and most difficult to apply when it doubles a non-integer number of times. Accordingly, we re-estimated the basic specification from Table V separately for valuation tasks with a single doubling, two to four doublings, and 2.5 doublings. Results appear in columns (3) - (5) of table (VIII).

Under the hypothesis that the ease of applying the rule of 72 should improve the success of interventions that teach the rule of 72, we should see systematic differences in the *relative* welfare effects of the substantive and Rhetoric-Only interventions across these three categories of valuation tasks.<sup>75</sup> Thus we would expect to find that the *difference* between the effect of the Full (or Substance-Only) treatment and the Rhetoric-Only treatment decreases as we move from column (3) (which pertains to valuation tasks with single doubling) to columns (4) and (5) (which pertain to valuation tasks with multiple and non-integer doublings, respectively). In fact, no such pattern is observed. Indeed, we cannot reject the hypothesis that the difference between the welfare effects of the Full and Rhetoric-Only treatments is the same for all three classes of valuation tasks ( $p > 0.10$  for all pairwise comparisons). The same is true of the difference between the welfare effects of the Substance-Only and Rhetoric-Only treatments ( $p > 0.10$  for all pairwise comparisons).<sup>76</sup> Thus, one cannot attribute

---

<sup>75</sup>One's first instinct is to evaluate whether our substantive treatments were more successful in the context of simpler valuation tasks by comparing their effects across these specifications. However, such comparisons would be inappropriate. For each of these classes of valuation tasks, the initial exponential growth bias is systematically different. Accordingly, a treatment that mechanically increases complex valuations by the same fixed amount in all valuation tasks will manifest different welfare effects across these classes of tasks. In particular, the treatment will appear less beneficial in the context of tasks for which the average exponential growth bias is initially smaller (e.g., because the gap between simple and compound interest is smaller), simply because its tendency to increase overestimation will receive more weight.

<sup>76</sup>We note that  $(\beta_{Substance} - \beta_{Rhetoric})$  is significantly different across non-integer and integer doublings ( $p < 0.05$ ). However, the actual sign of this difference is opposite the hypothesized sign.

the poor performance of our substantive interventions in terms of welfare to the difficulty of applying the rule of 72 in our valuation tasks.

## 9 Related Literature

Hastings et al. (2013) and Lusardi and Mitchell (2014) provide detailed comprehensive reviews of the literature on financial literacy and financial education. Here we focus on the portions of that literature that are most closely related to the more novel aspects of our study, as well as some other pertinent areas of inquiry.

**Financial education and well-being** To our knowledge, Song (2012) is the only existing study that conducts an explicit welfare analysis of a financial education intervention. Farmers in rural China received instruction concerning compound interest, and were then given opportunities to adjust their contributions to a state sponsored retirement savings plan. To evaluate welfare effects, Song employed a life-cycle consumption model parameterized to reflect risk and time preferences elicited from the subjects. He concluded that the intervention improved welfare on average even though its effect on behavior was indiscriminate.<sup>77</sup> While Song’s approach allows him to assess the welfare effects of changes in life-cycle consumption plans, it requires him to make a collection of strong assumptions – most notably, that he has the “right” model of life-cycle consumption, and that his preference-elicitation procedure parametrizes it appropriately. In contrast, our approach employs much weaker assumptions.<sup>78</sup>

While avoiding formal welfare analyses, other previous studies of financial education and the quality of financial decision making have examined various outcome measures that arguably serve as reasonable proxies for non-paternalistic notions of well-being. The general strategy is to examine the effect of financial education on the frequency with which people make choices that are objectively poor irrespective of their preferences (i.e., dominated choices). Studies that have taken this approach include Ernst et al. (2004), Calvet et al. (2007, 2009), Agarwal et al. (2009), Baltussen and Post (2011), and Choi et al. (2011). Choi et al. emphasize the (presumed) dominance relation by using the evocative phrase “dollar bills on the sidewalk.”

This approach to evaluating welfare effects has at least three important limitations. First, in naturally occurring settings, dominance is extremely hard to establish, and rationalizations for allegedly poor choices are often possible to imagine. For example, Ernst et al. (2004) point to the use of payday loans by people who have liquidity on credit cards with lower interest rates. But it may be rational for

---

<sup>77</sup>Specifically, actual changes in saving were not closely related to the optimal changes prescribed by the life cycle model, and the education intervention induced some subjects to oversave.

<sup>78</sup>Of course, unlike Song’s approach, it does not provide measures of life-cycle welfare.

those individuals to avoid depleting all forms of instant liquidity, e.g., to provide for various possible emergencies. Second, even a reduction in dominated choices does not necessarily imply an increase in welfare. For example, indoctrination may help people avoid dominated choices, but only by ignoring their own preferences among undominated alternatives. Indeed, we have seen in the current study that suppressing a particular type of dysfunctional behavior (here, the use of simple interest calculations) need not improve welfare, because new choice patterns may be equally problematic. Third and finally, one cannot translate effects on the frequency of dominated choices into standard welfare measures such as compensating or equivalent variations, which are needed for cost-benefit analyses of potential interventions.

**Experimental evaluation of narrow financial education interventions** Much of the literature on financial education studies interventions that are highly compound and heterogenous, often using naturally occurring data. We add to a burgeoning literature that demonstrates the importance of investigating the narrow constituent parts of such interventions using experimental methods. Goda et al. (2012) conduct a large field experiment in which employees of the University of Minnesota are provided with information about voluntary retirement savings plans. In one treatment, different savings levels are projected into assets at retirement. Another treatment adds projections of retirement income.<sup>79</sup> Goda et al. find significant increases in contributions when income projections are provided, but none when projections are limited to assets at retirement. Like the current paper, their research demonstrates that seemingly minor differences in the provision of information sometimes have large behavioral effects, and that identifying the drivers of behavioral change is critical for the design of effective interventions. In a similar vein, Drexler et al. (2014) provided a group of micro-entrepreneurs in the Dominican Republic with basic training in accounting, and compared the effectiveness of this intervention with training that emphasizes rules of thumb and heuristics. For less highly skilled subjects, the rules-of-thumb training led to significantly greater improvements in firms' financial practices, objective reporting of quality, and revenues. In an unincentivized experiment, Carlin et al. (2014) found that subjects' propensity to choose the best credit card from a list of options is significantly enhanced when explicit information about the location of the pertinent information is added to an educational intervention. Finally, Heinberg et al. (2014) administered an educational intervention concerning five basic financial concepts, in some cases through written narratives and in others through videos. They find that, of the two approaches, videos more effectively improve motivation and perceived self-efficacy regarding financial decision making. Likewise, Lusardi et al. (2014) find that videos are more effective at improving financial literacy.

---

<sup>79</sup>In both treatments, projections are customized to the recipient's financial situation.



**Other research on imperfect decision making** As mentioned in Section 2, the current study is an application of behavioral welfare economics, and therefore can be read as a contribution to that burgeoning literature. It is also related to other research that explores aspects of imperfect decision making.

Methodologically, the current paper develops and implements a new tool for measuring and quantifying the quality of decision making. Quantitative measures of decision-making quality have been developed and deployed in other contexts. The Afriat (1972) critical cost efficiency index is one well-known example; Choi et al. (2014) use it to show that conformance with the generalized axiom of revealed preference (GARP) is correlated with wealth, even controlling for measures of cognitive ability and conscientiousness. However, unlike our approach, such measures are not designed to assess the welfare costs of imperfect decision making.<sup>80</sup>

Our study is also related to a handful of papers that investigate the effects of complex framing on decision making. Hastings and Tejeda-Ashton (2008) investigate hypothetical choices among investment funds by financial illiterate Mexican workers, and show that subjects are more inclined to select funds with lower fees when those fees are presented as pesos rather than annual percentage rates. In a field experiment, Bertrand and Morse (2011) find that providing borrowers with information that reinforces the adding-up of dollar fees incurred when rolling over loans reduces the take-up of future payday loans by 11%. Kalaci and Serra-Garcia (2012) conduct an experiment in which subjects have to choose from a set of options that entail both costs and benefits. They find that complex presentation of the costs increases subjects' propensity to choose the highest gross-benefit option (which differs from the highest net-benefit option), whereas complex presentation of the benefits induces more random choice (rather than increasing subjects' propensity to choose the lowest-cost option). Abeler and Jaeger (2014) study subjects' effort choices in a piece rate task involving taxes and subsidies that are framed either simply or complexly. They find that complex framing reduces the magnitude of responses to changes in tax rates, compared with simple framing. In contrast to our study, they evaluate complexly framed choices in relation to a theoretical benchmark, rather than in relation to each subject's own simply framed choices.<sup>81</sup>

---

<sup>80</sup>Echenique et al. (2011) provide an alternative measure of divergences from GARP. One can interpret their measure as the *maximal* amount of money that one can extract from a decision maker with specific violations of GARP.

<sup>81</sup>Each of their subjects makes either complexly or simply framed choices, not both. Hence, their design precludes the use of a subject's simply framed choices as normative benchmarks for their complexly framed choices. They do, however, compare subjects' earnings in the complexly framed treatment not only to the theoretical benchmark, but also to other subjects' earnings in the simply framed treatment, and they argue that differences in effort costs across subjects are unlikely to affect behavior in their setting.

## 10 Conclusion

In this paper, we have introduced the notion of *financial competence*, and used it to analyze the effects of a narrow financial education intervention concerning compound interest. The intervention significantly improves measured financial literacy, and subjects report that they operationalize their improved knowledge when making choices. Indeed, on average, the intervention eliminates exponential growth bias. However, financial competence (which measures welfare) does *not* improve. An examination of two additional interventions (one without rhetoric, one with limited substance) reveal that while the effects on measured financial literacy and self-reported choice strategies are attributable to the substantive components of the intervention, changes in behavior are almost *entirely* attributable to the rhetorical components. As a result, despite the intervention’s success in discouraging subjects from employing simple interest calculations, it does not induce them to evaluate compound interest correctly when making decisions (even though tests reveal that their ability to compute compound interest improves dramatically). Thus, while the intervention appears highly successful according to conventional measures, it is not actually beneficial.

In generalizing from these results, one must of course exercise caution, especially since the interventions we study are so limited. The general lesson to be drawn from this analysis is not that a particular intervention had certain effects, but rather that it is possible for financial education to be highly successful according to conventional outcome measures while failing to improve the quality of financial decision making. Thus, we provide a decidedly negative answer to an important open question identified in the literature review by Hastings et al. (2013): “whether test-based measures provide an accurate measure of actual financial capability.” While we remain convinced that financial literacy is important, it does not by itself guarantee financial competence.

A curious feature of our findings is that financial education improves some consequential choices (*incentivized* test responses) but not others (complexly framed valuation tasks). One potentially important distinction is that the former involve questions with right and wrong answers, whereas the latter also involve expressions of preference over timing.<sup>82</sup> We conjecture that the presence of the preference element somehow derails an analytic problem-solving approach to decision making.

Our main findings pose serious challenges for public policy regarding financial education. At the strategic level, we can imagine three broad alternative approaches. The first is to devise educational methods that more effectively lead people to put pertinent knowledge into practice, and to do so correctly, when they make decisions. Implementing this strategy will require extensive research into

---

<sup>82</sup>This is not the only distinction. For example, the incentives for answering test questions correctly may be easier to understand, while the incentives for performing complexly framed valuation tasks are significantly stronger. In particular, compared with test questions, the complex valuation tasks involve stakes that are more than twice as high for a subject with the average rate of time preference, and up to 3.5 times as high for subjects who are significantly more or less patient. (To calculate these stakes, we compare optimal choices to random choices.)

the effects of alternative pedagogical approaches not only on financial literacy, but also on financial competence. A second approach is to deploy educational programs targeted at populations known to manifest particular biases in order to create countervailing biases (in effect accomplishing the right objective for the wrong reason). For example, in the current study, we have found that the most beneficial intervention is the one with the least substance and the most emphasis on rhetoric. Presumably, we could enhance its aggregate benefit by limiting its deployment to subjects whose demographic characteristics and initial test scores indicate a high degree of susceptibility to exponential growth bias. We are skeptical about the practical value of this “debiasing” approach, because it seems likely that any success in balancing countervailing biases will be context-specific, and consequently not necessarily indicative of how any particular individual would make a broad range of real-world decisions involving the pertinent financial concepts. A third approach is to develop better tools to assist with real-world decision making. Using our terminology, the object would be to turn naturally occurring complexly framed decision problems into simply framed problems. In principle this is a promising approach, but its effective deployment will require much additional research.

Having developed a framework for answering practical questions about financial competence, we envision many directions for subsequent research, some of which we are already pursuing. One important task is to extend our methods to other types of financial decisions such as insurance and portfolio allocation, involving concepts such as risk taking, inflation, and management fees. It is also important to study other populations, as well as other types of educational interventions, particularly ones that are used in practice. Accordingly, we anticipate using these methods to evaluate actual educational interventions in the workplace, as well as in high schools. Research on pedagogical design will, however, at least initially require extensive study of more narrowly focused interventions in the laboratory. Indeed, we have emphasized that a focus on narrow educational interventions makes it easier to determine which pedagogical approaches work and which do not, and to develop a nuanced understanding of the mechanisms through which such interventions influence behavior. For these reasons, we think the call by Hastings et al. (2013) for studies of “large scale interventions” may be premature. The effective design of large-scale interventions likely requires a much more comprehensive micro-level understanding of financial education than we currently possess. An initial focus on narrowly focused, small-scale interventions is, in our view, the best route to developing that understanding.

In principle, our methods could be used to evaluate other types of educational interventions that aim to provide people with a better understanding of their choice’s consequences. Applications to problems involving health and nutrition are worth exploring.

STANFORD UNIVERSITY

STANFORD UNIVERSITY

THE GEORGE WASHINGTON UNIVERSITY

## References

- Abeler, J. and Jaeger, S. (2014). Complex tax incentives. *American Economic Journal: Economic Policy*.
- Afriat, S. N. (1972). Efficiency estimation of production function. *International Economic Review*, 13(3):568–98.
- Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2009). The age of reason: Financial decisions over the life cycle and implications for regulation. *Brookings Papers on Economic Activity*, Fall:51–101.
- Almenberg, J. and Gerdes, C. (2012). Exponential growth bias and financial literacy. *Applied Economics Letters*, 19(17).
- Andersen, S., Harrison, G. W., Lau, M. I., and Rutstrom, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, 9:383–405.
- Baltussen, G. and Post, G. T. (2011). Irrational diversification: An examination of individual portfolio choice. *Journal of Financial and Quantitative Analysis*, 5:1463 – 1491.
- Bayer, P. J., Bernheim, B. D., and Scholz, J. K. (2009). The effects of financial education in the workplace: Evidence from a survey of employers. *Economic Inquiry*, 47(4):605–624.
- Bernheim, B. D. (2009). Behavioral welfare economics. *Journal of the European Economic Association*, 7(2-3):267–319.
- Bernheim, B. D. and Garrett, D. M. (2003). The effects of financial education in the workplace: evidence from a survey of households. *Journal of Public Economics*, 87.
- Bernheim, B. D., Garrett, D. M., and Maki, D. M. (2001). Education and saving: The long-term effects of high school financial curriculum mandates. *Journal of Public Economics*, 80:435–465.
- Bernheim, B. D., Popov, I., and Fradkin, A. (2013). The welfare economics of default options in 401(k) plans. *NBER Working Paper*, 17587.
- Bernheim, B. D. and Rangel, A. (2004). Addiction and cue-triggered decision processes. *American Economic Review*, 94(5):1558–90.
- Bernheim, B. D. and Rangel, A. (2009). Beyond revealed preference: Choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51–104.

- Bertrand, M. and Morse, A. (2011). Information disclosure, cognitive biases, and payday borrowing. *The Journal of Finance*, LXVI(6):1865–93.
- Bricker, J., Kennickell, A. B., Moore, K. B., and Sabelhaus, J. (2012). Changes in u.s. family finances from 2007 to 2010: Evidence from the survey of consumer finances. *Federal Reserve Bulletin*, 98(2).
- Calvet, L. E., Campbell, J. Y., and Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy*, 115(5):707–47.
- Calvet, L. E., Campbell, J. Y., and Sodini, P. (2009). Measuring the financial sophistication of households. *American Economic Review: Papers & Proceedings*, 99(2):393–398.
- Carlin, B. I., Jiang, L., and Spiller, S. A. (2014). Learning millennial-style. *Working Paper, Anderson School of Business, UCLA*.
- Carpena, F., Cole, S., Shapiro, J., and Zia, B. (2011). Unpacking the causal chain of financial literacy. *The World Bank Policy Research Working Paper*, (5798).
- Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–1177.
- Choi, J. J., Laibson, D., and Madrian, B. C. (2011). \$100 bills on the sidewalk: Suboptimal investment in 401(k) plans. *Review of Economics and Statistics*, 93(3).
- Choi, S., Kariv, S., Mueller, W., and Silverman, D. (2014). Who is (more) rational? *American Economic Review*, 104(6):1518–1550.
- Cole, S., Sampson, T., and Zia, B. (2011). Prices or knowledge? what drives demand for financial services in emerging markets? *The Journal of Finance*, 66(6):1933–1967.
- Cole, S. and Shastri, G. K. (2012). Is high school the right time to teach self-control? the effect of education on financial behavior. *Mimeo, Harvard University*.
- Collins, J. (2010). The impacts of mandatory financial education: evidence from a randomized field study. *Working Paper, Center for Financial Security, University of Wisconsin-Madison*.
- Council for Economic Education (CCE) (2006). Financing your future (dvd). <http://financingyourfuture.councilforeconed.org/>.
- Drexler, A., Fischer, G., and Schoar, A. (2014). Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics*, 6(2):1–31.

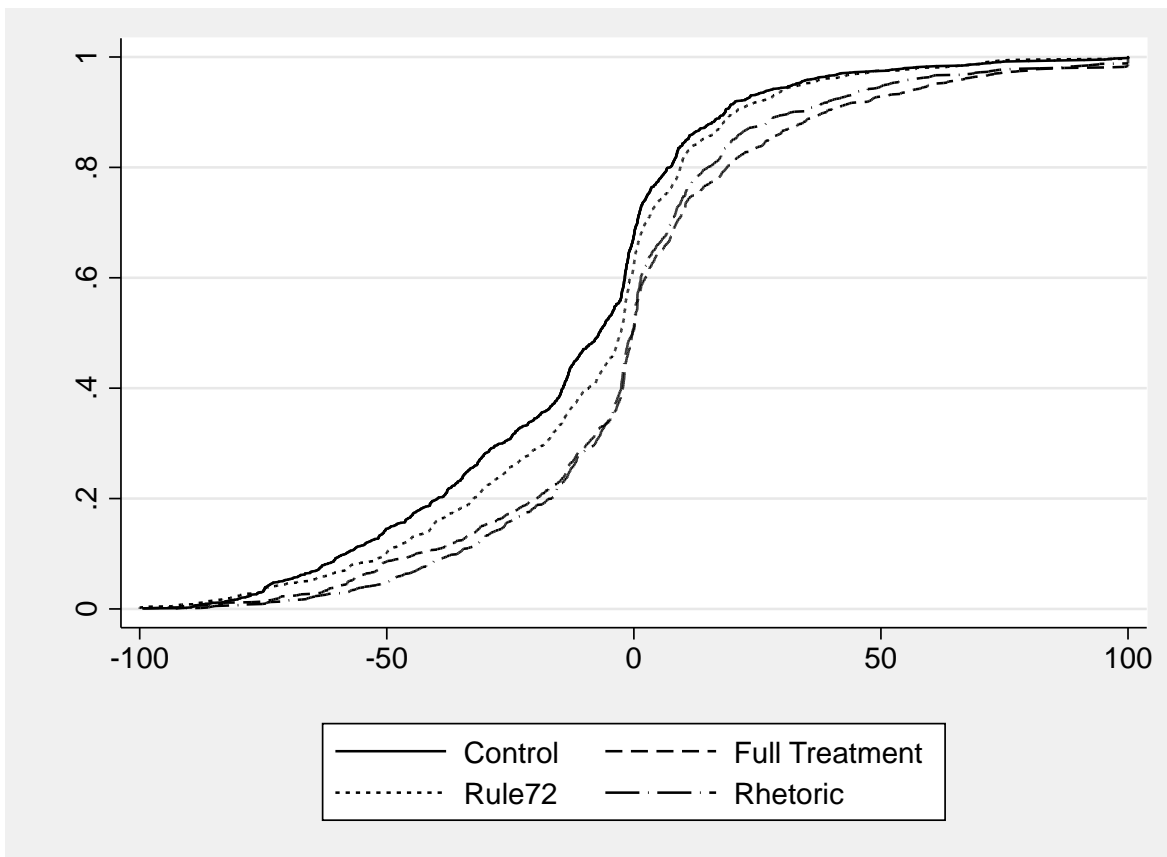
- Dufo, E. and Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *Quarterly Journal of Economics*, 118(3).
- Echenique, F., Lee, S., and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223.
- Eisenstein, E. M. and Hoch, S. J. (2007). Intuitive compounding: Framing, temporal perspective, and expertise. *Mimeo*.
- Ernst, K., Farris, J., and King, U. (2004). Quantifying the economic cost of predatory payday lending. Technical report, Center for Responsible Lending.
- Frederick, S., Loewenstein, G., and O’Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.
- Gartner, K. and Todd, R. M. (2005). Effectiveness of online early intervention financial education programs for credit-card holders. *Working Paper, Federal Reserve Bank Chicago*.
- Goda, G. S., Manchester, C. F., and Sojourner, A. (2012). What will my account really be worth? an experiment on exponential growth bias and retirement saving. *NBER working paper*, 17927.
- Hastings, J. S., Madrian, B. C., and Skimmyhorn, W. L. (2013). Financial literacy, financial education, and economic outcomes. *Annual Review of Economics*, 5:347–373.
- Hastings, J. S. and Tejada-Ashton, L. (2008). Financial literacy, information, and demand elasticity: Survey and experimental evidence from Mexico. *NBER Working Paper*, 14538.
- Heinberg, A., Hung, A. A., Kapteyn, A., Lusardi, A., Samek, A. S., and Yoong, J. K. (2014). Five steps to planning success. experimental evidence from U.S. households. *NBER Working Paper*, 20203.
- Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644 – 1655.
- Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics*, 14:399–425.
- Iyengar, S. S. and Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94:530–539.
- JumpStart Coalition for Personal Financial Literacy (2006). Financial literacy shows slight improvement among nation’s high school students. *Report*, Washington, D.C.

- Kalaci, K. and Serra-Garcia, M. (2012). Complexity and biases: An experimental study. *Mimeo*.
- Kline, P. (1999). *Handbook of Psychological Testing*. Routledge, London and New York, 2 edition.
- Levy, M. R. and Tasoff, J. (2014). Exponential growth bias. *mimeo*.
- Lührmann, M., Serra-Garcia, M., and Winter, J. (2012). Teaching teenagers in finance: does it work? *Munich Discussion Paper*, 24.
- Lührmann, M., Serra-Garcia, M., and Winter, J. (2014). The impact of financial education on adolescents' intertemporal choices. *mimeo, University of Munich*.
- Lusardi, A. (2011). Americans' financial capability. *NBER Working Paper*, 17103.
- Lusardi, A. and Mitchell, O. (2009). How ordinary consumers make complex economic decisions: Financial literacy and retirement readiness. *NBER Working Paper*, 15350.
- Lusardi, A. and Mitchell, O. (2011). Financial literacy and planning: Implications for retirement well-being. In Lusardi, A. and Mitchell, O., editors, *Financial Literacy. Implications for Retirement Security and the Financial Marketplace*, pages 17–39. Oxford University Press.
- Lusardi, A. and Mitchell, O. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature*, 52(1):5–44.
- Lusardi, A., Samek, A. S., Kapteyn, A., Glinert, L., Hung, A., and Heinberg, A. (2014). Visual tools and narratives: New ways to improve financial literacy. *NBER Working Paper*, 20229.
- Mandell, L. (2009). The financial literacy of young american adults: Results of the 2008 national jumpstart coalition survey of high school seniors and college students. Washington, D.C.(Jumpstart Coalition).
- Mandell, L. and Klein, L. S. (2009). The impact of financial literacy education on subsequent financial behavior. *Journal of Financial Counseling and Planning*, 20(1):15–24.
- Mason, W. and Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavioral Research*, 44:1–23.
- McKenzie, C. R. M. and Liersch, M. J. (2011). Misunderstanding savings growth: Implications for retirement savings behavior. *Journal of Marketing Research*.
- Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications*, 5.



- Servon, L. and Kaestner, R. (2008). Consumer financial literacy and the impact of online banking on the financial behavior of lower-income bank customers. *Journal of Consumer Affairs*, 42:271–305.
- Skimmyhorn, W. L. (2012). *Essays in behavioral household finance*. PhD thesis, Harvard Kennedy School, Cambridge, MA.
- Song, C. (2012). Financial illiteracy and pension contributions: A field experiment on compound interest in china. *Mimeo*.
- Stango, V. and Zinman, J. (2009). Exponential growth bias and household finance. *Journal of Finance*, 64(6).
- van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101:449–472.
- Wagenaar, W. M. and Sagaria, S. D. (1975). Misperception of exponential growth. *Perception and Psychology*, 18(6):416–422.
- Walstad, W. B., Rebeck, K., and Macdonald, R. A. (2010). The effects of financial education on the financial knowledge of high school students. *The Journal of Consumer Affairs*, 44(2).

Figure 1: C.D.F. OF FRAMING DISTORTION, BY TREATMENT.



For better visibility, data are truncated at -100 and at 100.

Table I: DECISION PROBLEMS

Future Reward $r$	Investment Amount $a$	Daily Interest Rate $R$	Number of Doublings
Duration: 72 days			
\$20	\$10	0.01	1
\$18	\$4.5	0.02	2
\$16	\$2	0.03	3
\$14	\$0.9	0.04	4
\$12	\$2	0.025	2.5
Duration: 36 days			
\$20	\$10	0.02	1
\$18	\$4.5	0.04	2
\$16	\$2	0.06	3
\$14	\$0.9	0.08	4
\$12	\$2	0.05	2.5

*Number of doublings* is the number of times the initial investment doubles over the investment horizon according to the rule of 72. Final amounts are calculated using the rule of 72. Exact final amounts differ by no more than \$0.80, except for the 4% interest rate over 72 days, where the rule understates the future value by \$1.16. Our analysis controls for these differences.

Table II: TEST QUESTIONS

Q1. If the interest rate is 10% per year (interest is compounded yearly), how many years does it take until an investment doubles?

*7 years, 7.2 years, 7.4 years, 7.8 years, 8 years*

Q2. If somebody tells you an investment should double in four years, what rate of return (per year) is he promising?

*15%, 16%, 17%, 18%, 19%, 20%*

Q3. If the interest rate is 7% per year (interest is compounded yearly), about how long does it take until an investment has grown by a factor of four (i.e. is four times as large as it was originally)?

*About 5 years to about 40 years, in steps of 5 years.*

Q4. Paul had invested his money into an account which paid 9% interest per year (interest is compounded yearly).

After 8 years, he had \$500. How big was the investment that Paul had made 8 years ago?

*\$200 to \$400 in steps of \$10*

Q5. If an investment grows at 8 percent per year (interest is compounded yearly), by how much has it grown after

4 years?

*By 30%, to by 40% in steps of one percentage point.*

Questions were presented in random order and intermingled with the questions concerning material covered in the Control video.

Table III: TEST SCORES AND SELF-REPORTED BEHAVIOR.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Test score	Test score	External	Uses rule	Uses rule	Explicit
	compounding	control	help	in complex	in simple	calculation
				framing	framing	
Level in Control	1.963*** (0.139)	3.284*** (0.103)	0.220*** (0.0416)	0.128*** (0.0396)	0.0917** (0.0386)	6.404*** (0.354)
<i>Treatment effects</i>						
Full	1.442*** (0.197)	-1.058*** (0.146)	-0.0126 (0.0592)	0.579*** (0.0564)	0.172*** (0.0550)	1.738*** (0.504)
Substance-Only	1.271*** (0.189)	-1.339*** (0.140)	0.0611 (0.0565)	0.637*** (0.0539)	0.260*** (0.0525)	1.737*** (0.482)
Rhetoric-Only	0.492** (0.195)	-1.079*** (0.144)	0.0655 (0.0584)	0.104* (0.0556)	0.0600 (0.0542)	0.418 (0.497)
$P(\beta_{Substance}=\beta_{Rhetoric})$	3.79e-05	0.0618	0.937	0.000	0.000146	0.00606
$P(\beta_{Full}=\beta_{Rhetoric})$	1.73e-06	0.885	0.184	0.000	0.0403	0.00871
$P(\beta_{Substance}=\beta_{Full})$	0.368	0.0467	0.196	0.285	0.0994	0.999
$P(\text{joint insignificance})$	0.000	0.000	0.400	0.000	3.17e-06	0.000201
Observations	455	455	455	455	455	455
$R^2$	0.139	0.187	0.007	0.320	0.061	0.043

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

The dependent variable in columns 1, 2, 3, 4, and 5 are the mean number of exam question answered correctly (1 to 5), the self-reported answer to whether the subjects used external help in the exam, the answer to the question whether the rule of 72 was used in the framed problems, the answer to the question whether the rule of 72 was used in the simply framed problems, and the self-reported number of complexly framed problems (out of 10) for which the subject explicitly calculated the future reward.

Table IV: FRAMING DISTORTION.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$100 \times d_{j,r,t}$					
Sample	all	all	high FL	low FL	36 days	72 days
Demographic controls	No	Yes	No	No	No	No
Level in Control	-13.31*** (2.221)	-10.38 (7.663)	-11.97*** (2.404)	-15.62*** (4.401)	-14.38*** (2.283)	-12.24*** (2.334)
<i>Treatment effects</i>						
Full	13.91*** (3.332)	13.59*** (3.202)	12.68*** (3.688)	15.95** (6.774)	14.47*** (3.489)	13.35*** (3.487)
Substance-Only	4.002 (2.961)	3.733 (2.908)	3.304 (3.208)	5.279 (5.819)	4.910 (3.185)	3.095 (2.984)
Rhetoric-Only	13.22*** (2.952)	13.29*** (2.903)	11.58*** (3.132)	16.02*** (5.892)	13.33*** (3.075)	13.11*** (3.063)
$P(\beta_{Substance}=\beta_{Rhetoric})$	0.000903	0.000837	0.00497	0.0510	0.00568	0.000258
$P(\beta_{Full}=\beta_{Rhetoric})$	0.827	0.924	0.748	0.991	0.734	0.942
$P(\beta_{Substance}=\beta_{Full})$	0.00184	0.00143	0.00802	0.0976	0.00581	0.00139
$P(\text{joint insignificance})$	0.000269	0.000	0.0195	8.75e-06	3.55e-06	0.000
$P(\beta_0 = 0)$	4.27e-09	0.176	1.10e-06	0.000506	7.09e-10	2.43e-07
$P(\beta_0 + \beta_{Full} = 0)$	0.808	0.662	0.798	0.949	0.972	0.667
$P(\beta_0 + \beta_{Substance} = 0)$	2.71e-06	0.383	5.88e-05	0.00732	2.46e-05	1.25e-06
$P(\beta_0 + \beta_{Rhetoric} = 0)$	0.965	0.692	0.847	0.918	0.612	0.659
Observations	4,550	4,460	2,920	1,630	2,275	2,275
Number of subjects	455	446	292	163	455	455
$R^2$	0.033	0.065	0.035	0.032	0.033	0.034

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

The last four lines display the  $p$ -values of tests that mean framing distortion in the corresponding treatment is zero. Demographic controls consist of the variables used in appendix B.2. Estimates with demographic controls exclude 9 subjects who preferred not stating their income. Standard errors clustered by subject.

Table V: FINANCIAL COMPETENCE

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	$-100 \times (d_{j,r,t})^2$					
Sample	all	all	high FL	low FL	36 days	72 days
Demographic controls	no	yes	no	no	no	no
Level in Control	-11.69*** (1.232)	-15.34*** (4.076)	-9.816*** (1.298)	-14.93*** (2.425)	-11.82*** (1.245)	-11.56*** (1.375)
<i>Treatment effects</i>						
Full	-0.155 (2.035)	-0.471 (1.917)	0.119 (2.359)	-2.120 (3.564)	-0.171 (2.325)	-0.139 (2.099)
Substance-Only	1.461 (1.669)	0.894 (1.632)	2.056 (1.868)	0.714 (3.054)	0.848 (1.800)	2.074 (1.766)
Rhetoric-Only	2.546 (1.700)	2.469 (1.615)	3.402** (1.680)	1.398 (3.400)	3.055* (1.745)	2.037 (1.860)
$P(\beta_{Substance}=\beta_{Rhetoric})$	0.505	0.313	0.433	0.821	0.217	0.982
$P(\beta_{Full}=\beta_{Rhetoric})$	0.177	0.117	0.144	0.321	0.164	0.282
$P(\beta_{Substance}=\beta_{Full})$	0.413	0.457	0.417	0.378	0.665	0.253
$P(\text{joint insignificance})$	0.390	0.332	0.179	0.771	0.287	0.469
Observations	4,550	4,460	2,920	1,630	2,275	2,275
Number of subjects	455	446	292	163	445	445
$R^2$	0.002	0.043	0.005	0.002	0.003	0.002

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Welfare loss is measured by compensating variation in present value, with welfare function  $W(d_{j,r,t}) = -(d_{j,r,t})^2$ . Standard errors clustered by subject. 46

Table VI: ALTERNATIVE WELFARE FUNCTIONS

	(1)	(2)	(3)	(4)
Welfare function	$100 \times W^+$	$100 \times W^-$	$-100 \times  d_{j,r,t} $	$-100 \times  d_{j,r,t} ^3$
Level in Control	-9.241*** (1.176)	-2.452*** (0.647)	-24.45*** (1.633)	-7.252*** (1.010)
<i>Treatment effects</i>				
Full	4.183*** (1.470)	-4.337*** (1.653)	1.584 (2.386)	-1.585 (2.119)
Substance-Only	1.572 (1.580)	-0.111 (0.891)	2.436 (2.172)	0.750 (1.405)
Rhetoric-Only	5.240*** (1.351)	-2.694** (1.236)	4.651** (2.155)	1.035 (1.570)
$P(\beta_{Substance}=\beta_{Rhetoric})$	0.00347	0.0344	0.270	0.854
$P(\beta_{Full}=\beta_{Rhetoric})$	0.339	0.375	0.171	0.238
$P(\beta_{Substance}=\beta_{Full})$	0.0585	0.0103	0.706	0.268
$P(\text{joint insignificance})$	0.000251	0.0103	0.178	0.639
Observations	4,550	4,550	4,550	4,550
Number of subjects	455	455	455	455
$R^2$	0.019	0.010	0.005	0.002

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Standard errors clustered by subject.



Table VII: USE OF SIMPLE INTEREST FORMULA

VARIABLE	(1)	(2)	(3)
	all	high FL	low FL
$\beta_1^{Control}$	0.304*** (0.100)	0.197** (0.0908)	0.489** (0.212)
$\beta_1^{Full}$	0.00928 (0.0309)	0.0133 (0.0269)	-0.00397 (0.134)
$\beta_1^{Substance}$	0.0883** (0.0380)	0.0600* (0.0348)	0.294* (0.178)
$\beta_1^{Rhetoric}$	0.0234 (0.0329)	0.000435 (0.0316)	0.0821 (0.108)
$\beta_0^{Control}$	0.721*** (0.0849)	0.814*** (0.0797)	0.527** (0.208)
$\beta_0^{Full}$	0.993*** 0.0227	0.994*** 0.0191	0.984*** 0.102
$\beta_0^{Substance}$	0.906*** 0.0298	0.930*** 0.0262	0.730*** 0.149
$\beta_0^{Rhetoric}$	0.983*** 0.0199	1.000*** 0.0169	0.926*** 0.0823
$P(\beta_1^{Control} = \beta_1^{Full})$	0.00489	0.0527	0.0496
$P(\beta_1^{Control} = \beta_1^{Substance})$	0.0436	0.160	0.481
$P(\beta_1^{Control} = \beta_1^{Rhetoric})$	0.00771	0.0412	0.0880
Observations	4,550	2,920	1,630
Subjects	455	292	163

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Estimated using median regression. Standard errors clustered by subject.

Table VIII: RESPONSE TIMES AND EFFECT OF DOUBLINGS

	(1)	(2)	(3)	(4)	(5)
VARIABLES	$\tau_{j,r,t}^c$	$\tau_{j,r,t}^s$	$-100 \times (d_{j,r,t})^2$	$-100 \times (d_{j,r,t})^2$	$-100 \times (d_{j,r,t})^2$
Doublings			1	[2 3 4]	2.5
Level in Control	50.81*** (2.675)	22.41*** (1.073)	-7.036*** (1.435)	-13.00*** (1.409)	-12.42*** (1.909)
<i>Treatment effects</i>					
Full	2.904 (4.471)	-0.768 (1.724)	0.302 (1.889)	1.578 (2.079)	-5.808 (3.983)
Substance-Only	19.51*** (7.069)	-0.428 (1.472)	1.663 (1.616)	1.914 (1.919)	-0.102 (2.688)
Rhetoric-Only	10.08** (4.599)	0.941 (2.197)	2.428 (1.589)	4.601** (1.829)	-3.502 (3.310)
$\beta_{Full} - \beta_{Rhetoric}$			-1.946	-3.023	-2.306
$\beta_{Substance} - \beta_{Rhetoric}$			-0.765	-2.687	3.400
Observations	4,550	4,550	910	2,730	910
Subjects	455	455	455	455	455
$R^2$	0.005	0.000	0.006	0.007	0.006

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Columns (1) and (2) show the effect of the treatments on mean response times for the complexly and simply framed problems, respectively. Columns (3) - (5) show the effect on welfare for problems that vary depending on how many times the investment amount doubles over the investment period in the complex framing. Standard errors clustered by subject.