

NBER WORKING PAPER SERIES

A METHOD FOR EVALUATING THE QUALITY OF FINANCIAL DECISION MAKING,
WITH AN APPLICATION TO FINANCIAL EDUCATION

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, Revised November 2017

Previously circulated as "The Effect of Financial Education on the Quality of Decision Making." We thank Charles Sprenger, Steven Sheffrin, GlenWeyl, as well as participants at the Research Forum on the Effectiveness of Financial Education at the University of Arizona, the Stanford Institute for Theoretical Economics, the Journées Louis- André Gérard-Varet in Aix-en-Provence, the New York University, the Murphy Institute's conference on Expanding the Frontiers in Behavioral Public Economics, the Cherry Blossom Financial Education Institute at the George Washington University, the Roybal Conference on Complexity in Decision Making at the University of Southern California, the Stanford Institute for Economic Policy Research, the TIAA Institute Fellows Symposium, the 2016 ASSA Meetings in San Francisco, the George Washington University Financial Literacy Seminar Series, and the Workshop on the Behavioral Economics of Financial Markets at the University of Zurich for helpful comments and suggestions. Fulya Yuksel Ersoy provided excellent research assistance. We appreciate funding from the TIAA-CREF and the Department of Economics at Stanford University. The experiment was approved in Stanford IRB protocol 29615. 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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.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.

A Method for Evaluating the Quality of Financial Decision Making, with an Application to Financial Education

Sandro Ambuehl, B. Douglas Bernheim, and Annamaria Lusardi

NBER Working Paper No. 20618

October 2014, Revised November 2017

JEL No. D14,D91,I21

ABSTRACT

We introduce a method for measuring the quality of financial decisions built around a notion of financial competence, which gauges the alignment between consumers' choices and those they would make if they properly understood their opportunities. We prove our measure admits a formal welfare interpretation even when consumers suffer from additional decision-making flaws, known and unknown, outside the scope of analysis. An application illuminates the pitfalls of the types of brief rhetoric-laden interventions commonly used for adult financial education: they affect behavior through unintended mechanisms, and hence may not improve decisions even when they perform well according to conventional metrics.

Sandro Ambuehl

Department of Management UTSC

Rotman School of Management

University of Toronto

105 St. George Street

Toronto, ON, M5S 3E6

CANADA

sandro.ambuehl@utoronto.ca

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

B. Douglas Bernheim

Department of Economics

Stanford University

Stanford, CA 94305-6072

and NBER

bernhelm@stanford.edu

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

*“A little learning is a dangerous thing; Drink deep, or taste not the Pierian spring:
There shallow draughts intoxicate the brain, And drinking largely sobers us again”*
– Alexander Pope, *An Essay on Criticism*, (1709)

1 Introduction

Low levels of financial literacy in the United States and the rest of the world raise doubts about the general quality of financial decision making. Financial education aims to improve decisions by helping consumers acquire the basic knowledge and skills they need to understand the choices they face. A large and growing literature finds mixed evidence that financial education interventions affect behavior (Hastings, Madrian and Skimmyhorn (2013), Lusardi and Mitchell (2014) provide reviews). Discussions of their welfare effects are typically informal and often colored by paternalistic judgments and preconceptions – for example, that people are better off with high saving and balanced portfolios, or that a better understanding of financial concepts necessarily promotes better decisions. Yet it is also possible that particular interventions alter behavior through mechanisms that involve indoctrination, exhortation, deference to authority, social pressure, or psychological anchors. If so, their benefits are unclear.

This paper makes two main contributions. First, we introduce a new method for measuring the quality of financial decision making. This contribution is important because rigorous analyses of decision-making quality are missing from most studies of financial education, likely due to the limitations of existing methods, as discussed in Section 2.6. The essence of our approach is to assess a consumer’s willingness to accept (WTA) for equivalent claims on future income. We design these claims so that a knowledge of targeted financial principles is required to understand that one is a simplified version of the other. Someone who both possesses and fully operationalizes that knowledge will consistently ascribe exactly the same value to the equivalent simply and complexly framed opportunities regardless of their preferences. When these WTAs differ systematically, the magnitude of the divergence provides a measure of *financial competence* with respect to the targeted principles: it indicates the extent to which a consumer’s incomplete operational command of those principles exposes her to decision error.

We demonstrate that our measure of financial competence—the divergence between WTAs for the types of equivalent claims we consider—has a precise welfare interpretation: it indicates the extent to which the consumer’s incomplete operational command of the principles that govern the equivalence exposes her to losses. Reliable welfare analysis is potentially challenging because consumers may suffer

from additional decision-making flaws falling outside the scope of analysis, such as incorrect expectations about the mapping from current and future income flows to consumption, and/or evaluative “biases.” In principle, such considerations could render all observed choices unreliable as guides to welfare. Yet we prove that our measure of financial competence admits a formal welfare interpretation even when the consumer suffers from other decision biases, known and/or unknown.

Our measure of financial competence has several additional virtues. First, as a welfare measure, it is non-paternalistic. The types of external judgments of consumers’ choices that are common in policy discussions, such as whether they are “sufficiently patient” or “save enough,” are entirely avoided. Second, it imposes modest information requirements. By comparing a consumer’s choices for equivalent tasks, we avoid the need for parametric models of decision making. Moreover, as mentioned above, the welfare interpretation of our financial competence measure is robust with respect to decision-making flaws outside the scope of analysis, which do not require modeling. Third, it is simple, intuitive, and easily implemented. As we explain in the next section, our method also offers important advantages over existing approaches to measuring the quality of financial decision making, including the examination of dominated choices (Ernst et al., 2004; Calvet et al., 2007, 2009; Agarwal et al., 2009; Baltussen and Post, 2011; Choi et al., 2011), the evaluation of WARP-consistency (Choi et al., 2014), and structural modeling (Song, 2015).¹

Our second main contribution is to document, through an experiment, the potential pitfalls of the types of brief rhetoric-laden interventions that are commonly used for workplace financial education, and to demonstrate that conventional methods of evaluation may fail to detect their deficiencies. Workplace interventions provide the lion’s share of adult financial education in the U.S.² Employers effectively treat *brevity* as a design constraint: thorough educational programs are not only costly but also time-consuming, which makes them unappealing to workers.³ To compensate for brevity, these programs generally focus on simple heuristics accompanied by highly motivating messages. The intent is to make the substantive material engaging, memorable, and actionable. Yet compelling rhetoric may also distract from substance and promote a one-size-fits-all response, which may be excessive for some and even directionally inappropriate for others.

¹A handful of other studies undertake comparisons between simply and complexly framed choices (Hastings and Tejeda-Ashton, 2008; Bertrand and Morse, 2011; Abeler and Jäger, 2015; Kalayc and Serra-Garcia, 2016), but none uses equivalent valuation tasks to infer the welfare losses resulting from complex framing, or the effect of educational intervention on those losses.

²In a 2013 survey of 407 retirement plan sponsors covering more than 10 million workers by Aon Hewitt, 77% of providers offered on-site financial education seminars or meetings (Austin and Evens, 2013). In the 2015 FINRA National Financial Capability Study, 40.24% of respondents aged 20 - 65 who have received financial education did so through an employer.

³A meta-analysis by Fernandes, Lynch Jr. and Netemeyer (2014) finds that the average financial education program involves only 9.7 hours of instruction. That time is usually divided among a long list of complex topics. For example, Skimmyhorn (2016) reports that a financial education program used by the U.S. military covers compound interest, the focus of our current study, along with a collection of several more complex topics – retirement concepts, the Thrift Savings Plan, military retirement programs, and investments – all within a single two-hour session.

Our experimental intervention focuses on compound interest, one of the fundamental concepts in personal finance. It resembles typical employer-sponsored interventions with respect to its brevity, as well as its emphasis on heuristics and motivational messages. It also appears to be highly effective according to conventional outcome measures: treated subjects perform substantially better on an incentivized financial literacy test, they report applying their newly gained knowledge when performing the decision tasks we assign them, and their average WTAs for interest-bearing assets change in a direction that counteracts the previously documented tendency to underestimate compounding, a phenomenon known as *exponential growth bias* (Wagenaar and Sagaria, 1975; Eisenstein and Hoch, 2007; Stango and Zinman, 2009; Almenberg and Gerdes, 2012; Levy and Tasoff, 2016). Nevertheless, using our approach, we find that the intervention does not, on average, improve the quality of decision making.

A possible explanation for this finding is that subjects may interpret motivational rhetoric as substantive advice and, even when their tested knowledge improves, emerge with an insufficient *operational* understanding of financial concepts to make appropriate adjustments. To explore this hypothesis, we implement two additional variants of the intervention, one that retains its substantive elements but omits the motivational rhetoric, and another that retains the motivational rhetoric but omits almost all of the substance. We show that the effects on financial literacy and self-reported decision strategies are primarily attributable to the substantive elements of instruction, as one would hope. However, in sharp contrast, the effects on financial choices are primarily attributable to the non-substantive elements. In particular, the intervention's motivational rhetoric increases subjects' WTA for interest-bearing assets regardless of the extent to which any particular individual initially understates or overstates the effects of compounding.⁴ This indiscriminate response is beneficial in some cases and harmful in others; on average, there is no benefit.⁵ When stripped of motivational rhetoric, exclusively substantive instruction has some effect on behavior, and it does reduce reliance on simple interest calculations (the most common type of mistake), but it fails to promote reliance on correct compound interest calculations, instead increasing the prevalence of other mistakes. As a result, its impact on WTAs for interest-bearing assets is directionally haphazard and, on average, welfare-neutral.

Thus, while financial literacy undoubtedly plays an important role in decision making (as shown by Lusardi and Mitchell, 2011), the associated mechanisms are complex and mediated by a variety of other factors. Educational interventions that achieve similar improvements in tested comprehension

⁴As in Goda et al. (2015) and Levy and Tasoff (2017), we document considerable heterogeneity with respect to the perceived benefits of compounding.

⁵Song (2015) also offers evidence that an educational intervention of involving compound interest has an indiscriminate impact: the effect on measured saving is not closely related to the gap between actual and optimal rates implied by a parameterized life-cycle consumption model.

may have dissimilar effects on behavior, depending on the particular manner in which each intervention motivates participants, and whether it helps them learn to internalize and operationalize conceptual knowledge rather than directional imperatives. Accordingly, one would expect to find sharp differences between the effects of adult financial education programs and high school courses: as we have noted, the former typically compensate for brevity with simple heuristics and motivational rhetoric; in contrast, the latter often span a full semester, permitting a more expansive and in-depth treatment of subject matter, as well as more effective pedagogy, including practice and discussion. While the literature studies these two settings separately (beginning with Bernheim, Garrett and Maki, 2001, and Bernheim and Garrett, 2003), it has only recently begun to explore the heterogeneity of approaches within each category, and to examine how the effects of an intervention depend on its design and constituent components. Consistent with our findings, recent work by Brown et al. (2014) shows that the effects of high school financial education on behavior are most pronounced when schools offer full courses taught by trained teachers. More generally, the considerations highlighted in the current study may help to explain why different authors reach different conclusions about the effects of financial education when studying different programs; see in particular Duflo and Saez (2003), Bayer, Bernheim and Scholz (2009), Goda, Manchester and Sojourner (2014), Cole and Shastry (2010), Cole, Sampson and Zia (2011), Skimmyhorn (2012), Servon and Kaestner (2008), Collins (2013), Lührmann, Serra-Garcia and Winter (2015a), Mandell (2009), Drexler, Fischer and Schoar (2014), Carlin, Jiang and Spiller (2014), Heinberg et al. (2014), Lusardi et al. (2015), and Bertrand and Morse (2011).

The remainder of the paper is organized as follows. Section 2 precisely defines the concept of financial competence, discusses its measurement, and explains its formal connection to consumer welfare. It also compares our approach to other methods for assessing the quality of financial decision making. Section 3 describes the design of our experiment and section 4 discusses its implementation. Section 5 analyzes the effects of the treatments on standard outcomes measures, including test scores, self-reported decision strategies, and average choices. Section 6 examines effects on the quality of financial decision making. We address the important issue of generalizability at some length in Section 7. Section 8 discusses the implications of our research and concludes.

2 The Definition and Measurement of Financial Competence

Our first objective is to devise a general framework for assessing the quality of financial decision making. We seek to formalize the intuitive notion that a good decision maker is one who avoids mistakes. This objective requires us to depart from classical consumer theory: if all choices reveal preferences, then none are mistaken, and any apparent inconsistencies must reflect our own misconceptions about the consumer's aims.

As a general matter, we say that a consumer displays financial competence with respect to targeted financial principles if they make equivalent choices from equivalent opportunity sets in contexts where the targeted principles govern the equivalence. The essence of our approach is to compare a consumer’s willingness to accept (WTA) for two equivalent claims on future income, where one is a simplified version of the other. The simple version states the future claim transparently. The complex version packages the claim as an income-generating asset. We design the asset so that a knowledge of targeted financial principles is required to infer the claim, and hence to understand the equivalence between the simple and complex versions. Someone who both possess and fully operationalizes that knowledge will consistently ascribe the same value to both claims regardless of their preferences and/or other decision biases. Thus, when a consumer’s WTAs for equivalent claims differ systematically, the magnitude of the divergence provides an intuitively appealing measure of her competence to make good decisions in contexts involving the pertinent principles.

To illustrate, say we are concerned that people poorly understand the concept of compound interest, and that this limitation causes them to make suboptimal investment decisions. To evaluate this possibility, we might assess the consumer’s WTA for pairs of equivalent claims such as the following: the complex claim represents a \$10 investment that promises a return of 6% per day compounded daily for 15 days while the simple claim simply promises \$24 in 15 days. Ordinarily, a consumer will be willing to choose each asset over a fixed sum of money if and only if the sum does not exceed some threshold value, call it p^* for the first claim and q^* for the second. A quick calculation reveals that the two claims are equivalent, subject to rounding. Thus, swapping out one for the other in a decision problem changes framing while leaving opportunities intact. As a general matter, any education intervention that successfully provides subjects with an operational understanding of compound interest should bring p^* and q^* into closer alignment.

As we explain below, the divergence between WTAs for the types of equivalent claims we consider has a precise welfare interpretation: it indicates the extent to which the consumer’s incomplete operational command of the principles that govern the equivalence exposes her to losses. Significantly, we prove that our measure of financial competence admits a useful welfare interpretation even when the consumer suffers from other decision biases, known and/or unknown.

2.1 Setting

We study choices for which each alternative involves either an immediate payoff, x , or a delayed payoff, y . The consumer evaluates these payments according to the indirect utility functions $V_0(x)$ and $V_1(y)$, respectively. For the moment, we will assume she maximizes the mathematical expectation of these functions when the choice involves risk (but we impose no restriction on how the functions V_1 and V_2

are derived). Appendix A explains how our analysis generalizes beyond this case.

We call the utility functions V_1 and V_2 “indirect” because the consumer separately chooses how to deploy their income – when to spend it, and on which goods. These functions capture the consumer’s expectation about this deployment, as well as the manner in which she evaluates the anticipated outcome at the time of her decision. We impose no assumptions on the shape of V_1 and V_2 other than that they are increasing (“more money is better”). For notational simplicity, we will normalize x and y so that the baseline for both (absent payments other than those discussed below) is 0.

The typical claim on future income is packaged as a financial instrument, z .⁶ Instruments generate future income according to a CDF, $F_z(y)$. The distinction between describing an alternative in terms of F_z or in terms of z is of no consequence for someone who fully appreciates the relationships between instruments and payoffs and consequently evaluates z according to the the value of $\int V_2(y)dF_z(y)$. However, in the context of financial decisions, those relationships are governed by principles that many people demonstrably do not understand (see, for example, Lusardi and Mitchell (2011)). We are concerned here with detecting and evaluating mistakes emanating from such misunderstandings. Accordingly, we assume the consumer acts as if she believes the returns to instrument z are governed by the CDF $G_z(y, \theta)$, where θ is a policy variable such as financial education. Thus, she evaluates z according to the value of $\int V_2(y)dG_z(y, \theta)$.

As an example, consider our application in the current paper, which focuses on financial competence with respect to the concept of compound interest and involves an appropriately selected class of instruments. Each instrument, z , promises to pay some fixed interest rate per day, $r(z) >> r^M$, where r^M is the market rate of interest, for a specified number of days, $t(z)$, on a fixed initial investment of $\$m(z)$, implying $y = m(z)(1 + r(z))^{t(z)}$. If consumers nevertheless employ some blend of simple and compound interest that depends on their financial sophistication, they might instead infer $y = \theta m(z)(1 + r(z))^{t(z)} + (1 - \theta)m(z)(1 + t(z)r(z))$. One could use the same framework to study financial competence with respect to other concepts, such as inflation (by employing an instrument that pays a nominal return in a currency that loses value at a known rate), leverage (by employing an instrument that blends an asset with a loan), or portfolio returns (by treating y as a vector and employing an instrument that blends assets earning different returns).

As noted above, the decisions we examine concern paired valuation tasks: for a given instrument z , we assess the consumer’s WTA for z and for an equivalent asset that simply promises a specified payment. We refer to these decisions as involving complex and simple framing, respectively. Within

⁶We use the term financial instrument to refer to any *indirect* description of income flows, either random or deterministic.

the context of our model, the consumer's WTA for instrument z , call it $x^V(z, \theta)$, is given by

$$V_1(x^V(z, \theta)) = \int V_2(y) dG_z(y, \theta) \quad (1)$$

Likewise, her WTA for an equivalent simplified claim, call it $x_0^V(z)$, is given by

$$V_1(x_0^V(z)) = \int V_2(y) dF_z(y) \quad (2)$$

If the consumer possesses and operationalizes a proper understanding of instruments, we should observe $x^V(z, \theta) = x_0^V(z)$. Thus, $x^V(z, \theta) - x_0^V(z)$, represents the valuation error resulting from the consumer's misunderstanding of the relationship between instruments and payoffs, according to the function V . As we demonstrate in Section 2.3, $|x^V(z, \theta) - x_0^V(z)|$ has a formal welfare interpretation: it measures the largest possible welfare loss the consumer can suffer when choosing between the instrument z and an immediate payoff of $\$d$. As we demonstrate, that conclusion is robust with respect to alternative assumptions about other flaws in the consumer's decision-making apparatus.

2.2 The key assumptions

Our formalism invokes two assumptions that merit acknowledgement, discussion, and empirical scrutiny. First, we assume that the financial instrument z is not an argument of V_1 or V_2 , and consequentially influences x^V only through G . In other words, the characteristics of financial instruments affect choices only insofar as they change anticipated future income. This assumption entails two mild restrictions: the packaging of claims on future or state-contingent income does not matter intrinsically to consumers, and it does not give rise to framing effects aside from its impact on the anticipated payoff.

Second, we assume that θ is not an argument of V_1 or V_2 , and consequently that it influences x^V only through G . In other words, the policy interventions under consideration affect choices only insofar as they change the anticipated future and/or state-contingent income flowing from the designated instruments; they do not change preferences over income flows. This assumption may be either reasonable or unreasonable depending on other features of the decision environment. In the context of our experiment, it is reasonable to assume that V_t is largely independent of θ for two reasons. First, V_t implicitly reflects the subject's solution to her overall intertemporal planning problem. It is unlikely that a subject would internalize newly acquired knowledge of compound interest into that solution instantaneously. Second, if the subject's planning horizon is short and r^M is small, even complete internalization of the aforementioned knowledge may have a modest effect on V_t .

Empirically, one can assess the validity of these assumptions by examining the following implications. If the first assumption holds, the perceived distribution of monetary outcomes, G_z , will

determine x^V . Because a financial instrument z describes the distribution of outcomes indirectly, the consumer must spend time trying to infer that distribution. Hence, subjects should take longer to make decisions in complexly framed valuation tasks than in their simply framed counterparts (inasmuch as only the former require assessment of the cash flow). In addition, because the description of the instrument in the simple frame is transparent, subjects should report deploying the principles governing the relationship between z and y in complexly framed tasks, but not in simply framed tasks. If the second assumption holds, then an intervention targeting those principles should affect valuations with complex framing, but not with simple framing. Additionally, any effects of such interventions on the time taken to make decisions should be confined to the complex frame. In the application of the present paper, the data support all these implications.

2.3 Welfare interpretation: The special case

We turn next to the welfare interpretation of $|x^V(z, \theta) - x_0^V(z)|$, our measure of financial competence. First, we explain our procedure for evaluating financial competence under the restrictive assumption that discrepancies between F_z and G_z are the *only* flaws in the consumer's decision making process. Specifically, we assume that the functions V_1 and V_2 do not incorporate any incorrect expectations about the mapping from current and future income flows to consumption, and are free from other evaluative biases. These assumptions permit us to derive welfare measures based on the functions V_1 and V_2 . If this assumption is violated, the use of V_1 and V_2 arguably involves an arbitrary and mistaken welfare standard. We examine that important possibility in section 2.4, and prove that our approach yields a useful welfare measure with considerable generality, even in the absence of specific information concerning the nature of other decision-making flaws.

Under the assumptions stated above, $|x^V(z, \theta) - x_0^V(z)|$ measures the largest possible welfare loss the consumer can suffer when choosing between the instrument z and an immediate payment of $\$d$.⁷ To understand this point, suppose first that $x^V(z, \theta) > x_0^V(z)$. If $d \geq x^V(z, \theta)$ or $d \leq x_0^V(z)$, there is no welfare loss, because the consumer would make the same choice regardless of which claim she considers.⁸ Mistakes occur when $x^V(z, \theta) > d > x_0^V(z)$. In this case, the consumer chooses the complex claim over $\$d$, even though she would willingly exchange the returns for $\$d$ if she fully anticipated the consequences of her choice. If she started out with her best option, $\$d$, she would be willing to

⁷Rigorous foundations for the welfare perspective taken in this subsection are found in Bernheim and Rangel (2009) and Bernheim (2016)(see also Bernheim (2009), and Bernheim and Rangel (2004)). Within that framework, one classifies a decision as a mistake when it involves *characterization failure*, and when there is some other option in the opportunity set that the decision maker would select over the mistakenly chosen one in settings where characterization failure does not occur. For the purpose of this section, we adopt the view that characterization failure is present in decision problems involving instruments (as evidenced by demonstrable failures to understand applicable financial principles), but not in the equivalent problems with transparent payments. In section 2.3, we address the possibility that characterization failure may occur in both settings.

⁸Technically, in the special case where $d \geq x^V(z, \theta)$, she would definitely choose the simple claim over d , and is willing to choose the complex claim over d .

give up $\$(d - x_0^V(z))$ to avoid swapping the cash for the income stream both claims promise. Hence, $\$(d - x_0^V(z))$ is the equivalent variation of the swap: it represents the dollar loss the consumer regards as equivalent to suffering the consequences of decision error.⁹ This loss is greatest when $d = x^V(z, \theta)$.

Next suppose that $x^V(z, \theta) < x_0^V(z)$. Reasoning as above, we see that mistakes occur only when $x^V(z, \theta) < d < x_0^V(z)$. In this case, the consumer chooses $\$d$ over the complex claim even though she would willingly exchange $\$d$ for that claim if she properly understood it. If she started out with her best option (the claim), she would require $\$(x_0^V(z) - d)$ as compensation for switching to $\$d$. Hence $\$(x_0^V(z) - d)$ is the compensating variation of the swap.¹⁰ Assuming income effects are negligible over the relevant range, compensating and equivalent variations coincide, and $\$(x_0^V(z) - d)$ then measures the dollar loss the consumer regards as equivalent to suffering the consequences of decision error.¹¹ This loss is again greatest when $d = x^V(z, \theta)$.¹²

Of course, the largest possible welfare loss generally overstates the actual loss. Another possibility is to compute the consumer's average or expected loss. Naturally, the expected loss depends on the process generating the consumer's opportunities. In the context of our experiment, the value of $\$d$ is drawn from a uniform distribution. The probability of incurring a loss is therefore proportional to $|x^V(z, \theta) - x_0^V(z)|$, and the expected loss conditional upon suffering one is $|x^V(z, \theta) - x_0^V(z)|/2$. Thus, the expected loss is proportional to $(x^V(z, \theta) - x_0^V(z))^2$. More generally, one can think of $\pi(x^V(z, \theta) - x_0^V(z))^2$ (where π is the density of the CDF governing the distribution of d at $d = x_0^V(z)$) as a second-order approximation of the expected welfare loss, much in the spirit of Harberger's (1964) well-known formula for the deadweight loss of a commodity tax.¹³

2.4 Welfare interpretation: The general case

We now turn to the important possibility that V_1 and V_2 involve an arbitrary and mistaken welfare standard, either because the consumer has an incorrect understanding of the mapping from current and future income flows to consumption, or because she suffers from evaluative biases. How one proceeds depends on the role one intends the policy of interest to play within a potentially multifaceted policy agenda.

⁹Formally, from equation (2), we have $V_1(d - (d - x_0^V(z))) = \int V_2(y)dF_z(y)$.

¹⁰Formally, from equation (2), we have $V_1(d + (x_0^V(z) - d)) = \int V_2(y)dF_z(y)$.

¹¹In other words, we assume that a consumer who is indifferent between the complex claim and $\$(d + r)$ immediately is also indifferent between a bundle consisting of the complex claim with a loss of $\$r$ immediately, and $\$d$ immediately (because the immediate income for both options is reduced by the same amount, $\$r$), which implies that $\$r$ is the equivalent variation associated with the switch from the complex claim to $\$d$ immediately.

¹²For the purpose of the application considered in this paper, the assumption of negligible income effects is reasonable. More generally, one can handle the case of non-negligible income effects by adjusting our valuation-elicitation procedure.

¹³Fix the value of $x_0^V(z)$, and assume that the CDF governing the distribution of d is twice continuously differentiable at $d = x_0^V(z)$. Taking a second-order Taylor series expansion of the expected welfare loss as a function of $x^V(z, \theta)$ in a neighborhood of $x_0^V(z)$, we obtain $\frac{\pi(x_0^V(z))}{2}(x^V(z, \theta) - x_0^V(z))^2$.

We begin this section with a discussion of conceptual issues, which leads us to formulate a notion of *idealized welfare analysis*. Relying on that conceptual framework, we then prove our main result. Finally, we detail the difficulties with the natural alternative to idealized welfare analysis.

2.4.1 Idealized welfare analysis

The following concrete example helps to clarify the issues. Suppose a consumer initially overestimates the benefits of compound interest, and in addition suffers from severe present bias,¹⁴ so that, on balance, she saves too little. Imagine our objective is to evaluate the welfare effects of a financial education program θ^T that provides the consumer with perfect knowledge of interest rate principles. Considering all sources of inefficiency, we would conclude that the policy is likely harmful. Indeed, we might end up recommending an alternative ‘educational’ intervention θ^D that misleads consumers into exaggerating the benefits of compound interest.

Prescribing the policy θ^D is potentially objectionable, even aside from concerns about the ethics of spreading misinformation, and about the government’s long-term credibility. Arguably, the prescription follows from a conceptual error: the analysis attempts to treat sources of inefficiency comprehensively, but does not treat policy options comprehensively. Distorting policies that target consumers’ understanding of compound interest in order to address concerns arising from present bias makes little sense if other policy tools are better suited for the latter purpose. For instance, an optimal comprehensive policy might consist of θ^T combined with measures that create appropriate commitment opportunities.

We see two potential solutions to the problem described in the preceding paragraph. One is to insist on treating all sources of inefficiency and policy responses comprehensively. Unfortunately, this strategy is impractical. As a general matter, economists compartmentalize policy analyses, focusing on one (or a few) policies at a time, because a fully comprehensive treatment encompassing all potentially interacting policies and their motivating concerns would be intractable.

The second alternative is to compartmentalize policies and the concerns that motivate them in parallel. For a concrete illustration, we return to our previous example. Suppose we see financial education as addressing limited comprehension of compound interest, and the creation of commitment opportunities as addressing present bias. A compartmentalized evaluation of financial education would focus on welfare effects involving comprehension, and would treat concerns about present bias as if they will be (but are not yet) fully resolved through appropriate commitments. Likewise, a compartmentalized evaluation of commitment opportunities would focus on welfare effects involving present bias, and would treat concerns about comprehension as if they will be (but are not yet) fully

¹⁴For the purpose of this example, we assume that present focus constitutes a mistake, as is often assumed.

resolved through appropriate education. We refer to this approach as *idealized welfare analysis* to indicate that it treats sources of inefficiency outside the scope of the analysis as if other policies will provide ideal resolutions.

Idealized welfare analysis allows one to solve policy problems one by one and still achieve an overall optimum in a single pass, provided each solution fully resolves the associated problem. In our illustration, it would correctly identify the optimum as consisting of θ^T combined with appropriate commitment opportunities. That said, because the compartmentalization of policy analysis abstracts from interactions, it necessarily involves conceptual compromises. In particular, if the best solutions to some concerns are imperfect, then idealized welfare effects do not capture potentially significant second-best considerations. Because the latter effects are complex, difficult to measure, and highly sensitive to assumptions, we contend that it is useful to begin policy evaluations with measures of idealized welfare effects, and then to make adjustments for second-best considerations from this baseline. We see most applied welfare analyses as implicitly (and informally) proceeding in this spirit, inasmuch as normative conclusions are typically derived from models that depict isolated market failures or decision-making flaws.

2.4.2 The main result

At first, it might appear that idealized welfare analysis requires a deep understanding of all decision-making flaws and their solutions, because it references judgments made in an idealized setting, rather than actual decisions. On the contrary, we prove below that one can approximate idealized welfare effects using the same data on WTAs described in Section 2.1, even if one has no information concerning the existence or nature of expectational and evaluative flaws embedded in the functions V_1 and V_2 .

Returning to the setting of Section 2.1, we imagine that, if all decision-making flaws aside from the discrepancy between F_z and G_z were resolved, the individual would make decisions according to indirect utility functions $U_1(x)$ and $U_2(y)$, which by assumption are free from other biases and expectational errors. Were we in that world, the individual's WTA for instrument z , call it $x^U(z, \theta)$, would be given by the equation

$$U_1(x^U(z, \theta)) = \int U_2(y) dG_z(y, \theta) \quad (3)$$

Given a proper understanding of instruments, her WTA would be $x_0^U(z)$, defined as follows

$$U_1(x_0^U(z)) = \int U_2(y) dF_z(y) \quad (4)$$

Thus, in parallel to our analysis of V_1 and V_2 , her use of G_z rather than F_z would lead to a valuation error of

$$x^U(z, \theta) - x_0^U(z)$$

One might think that the measurement of $x^U(z, \theta) - x_0^U(z)$ would be highly challenging, because it appears to require knowledge of unobserved functions, U_1 and U_2 . Fortunately, that is not the case. As it turns out, there is a close mathematical relationship between $x^V(z, \theta) - x_0^V(z)$ which we can observe, and $x^U(z, \theta) - x_0^U(z)$, which we seek to measure.

We establish this point by introducing a scaling parameter, α , representing ‘shares’ of z . Setting a given value of α rescales any actual or anticipated cash flow from y to αy .¹⁵ We also make the following technical assumption: U_t and V_t are continuously differentiable, as well as unbounded above and below. With these restrictions, we can prove the following result:

Theorem 1. *There exists a strictly positive constant K such that, for all θ and nondegenerate instruments z ,*¹⁶

$$\lim_{\alpha \rightarrow 0} \left[\frac{x^U(z, \theta, \alpha) - x_0^U(z, \alpha)}{x^V(z, \theta, \alpha) - x_0^V(z, \alpha)} \right] = K$$

The theorem tells us that, to a first order approximation, $x^V(z, \theta) - x_0^V(z)$ identifies $x^U(z, \theta) - x_0^U(z)$ up to a multiplicative scalar. Consequently, $x^V(z, \theta) - x_0^V(z)$ provides a useful approximation of the idealized welfare effect: Because K is positive, it always has the right sign. Because K is independent of θ , it ranks policies in the correct order and provides a valid gauge of their proportional costs or benefits. And because K is independent of z , it is strictly comparable across different instruments. It is worth emphasizing that the theorem holds regardless of U_1 and U_2 , and allows us to conduct welfare analysis using these functions to evaluate outcomes, even though we do not have sufficient information to identify them. See Appendix A.1 for a formal proof. As explained in Appendix A.2, subject to some qualifications, the result extends to non-expected utility.

To illustrate the robustness of our local approximation, we consider a parametric example that admits a global solution.

Example. $V_1(x) = U_1(x) = x$, $V_2(y) = \delta u(y)$, and $U_2(y) = \beta \delta u(y)$. One can think of this example as representing a case in which the individual is not only mistaken about F_z , but is also present-biased ($\beta < 1$). Then $x_0^U(z) = \delta u(f(z))$, $x^U(z, \theta) = \delta u(g(z, \theta))$, $x_0^V(z) = \beta \delta u(f(z))$, and $x^V(z, \theta) =$

¹⁵We are assuming here that people understand scaling, e.g., that doubling the number of shares of an instrument doubles all income flows. Even in light of the evidence on limited numeracy, this assumption strikes us as relatively innocuous.

¹⁶A degenerate instrument is one that yields a payment of zero with certainty.

$\beta\alpha u(g(z, \theta))$. In this case, the local approximation given in the theorem is globally valid: for any z , we have

$$\frac{x^U(z, \theta) - x_0^U(z)}{x^V(z, \theta) - x_0^V(z)} = \frac{\delta(u(g(z, \theta)) - u(f(z)))}{\rho(u(g(z, \theta)) - u(f(z)))} = \frac{1}{\beta} \equiv K$$

An additional conclusion follows from our example: if some other analysis yields an estimate of β , one can recover the level of $x^U(z, \theta) - x_0^U(z)$ exactly, rather than up to a factor of proportionality—simply rescale $x^V(z, \theta) - x_0^V(z)$ by the multiplicative factor $1/\beta$. A close reading of the proof of the theorem reveals that the factor of proportionality, K , always equals the ratio of the marginal rates of substitution between current and future income according to the U_t functions and the V_t functions, and consequently that this observation does not depend on the assumption of quasilinearity.

2.4.3 The alternative to idealized welfare analysis

The alternative to idealized welfare analysis is to evaluate the welfare effects of an isolated policy comprehensively in light of all decision making flaws. Detailed knowledge of those flaws becomes an absolute necessity, which renders such analysis highly challenging and susceptible to controversy. Returning to the setting of Section 2.1, we posit the existence of yet another pair of indirect utility functions, W_1 and W_2 , also defined on x and y , that correctly account for the manner in which the individual *actually* disposes of resources, and that is also free from evaluation bias. The “correct” evaluation of instrument z is then given by the equation

$$W_1(x_0^W(z)) = \int W_2(y)dF_z(y)$$

As a result, the consumer’s misunderstanding leads to a valuation error of

$$x^V(z, \theta) - x_0^W(z)$$

This alternative approach encounters numerous conceptual and practical difficulties. First, as explained at the outset of this section, concerns best addressed through other means can artificially distort policy prescriptions. Second, welfare measures and policy prescriptions become sensitive to changes in other policies affecting the severity of decision-making flaws that the policy in question does not seek to address. Resolving the n^{th} problem changes the solutions to the first $n - 1$ problems; thus the process of seeking solutions can cycle, and there is no guarantee it converges. Finally, implementation may be impractical, in that the approach requires a complete model of biases, expectational errors, and contingent plans. Going this route, one cannot make meaningful progress unless one’s understanding of decision making is complete, a condition that is impossible to satisfy (at least without considerable controversy) in practice. There is no counterpart to our theorem for idealized welfare analysis.

2.5 Additional remarks concerning the current application

Given the focus of the current paper, some remarks on the form of the indirect utility functions are in order. If y represents a single future payment received with certainty, then with perfect capital markets and a textbook consumer, we would have $V_2(y) = V_1\left(\frac{y}{(1+r^M)^t}\right)$, and valuation would be a math problem, involving no expression of preference. For standard experimental tasks, t is measured in days, so $r^M \approx 0$, which implies (as a good approximation) that the consumer would evaluate future payments according to the function $V_1(y)$. Experimental evidence overwhelmingly rejects this hypothesis. Subjects discount future payments at rates far exceeding any reasonable estimate of r^M (see, e.g. Frederick et al. (2002)), and often display a strict preference for interior allocations when confronted with linear tradeoffs between current and future payments (Andreoni and Sprenger, 2012a). We are aware of three possible explanations. First, subjects may expect to consume income when they receive it, and discount the pleasure derived from future consumption at a high rate, possibly as a consequence of a behavioral bias. (For example, many classify present focus as a bias.) Second, subjects may entertain doubts about the reliability of future payments. For example, an otherwise “textbook” subject who expects the experimenter to default on future payments with probability π , and who anticipates spending incremental cash no sooner than period t , might evaluate y according to the function $V_2(y) = (1 - \pi)V_1\left(\frac{y}{(1+r^M)^t}\right)$. Third, subjects may violate textbook assumptions concerning decision making, for example by evaluating alternatives according to their income flows rather than consumption. In what follows, we take no stand concerning the correct explanation, and, as our theorem shows, we do not need to.¹⁷ For our purposes, the important point is that, from the typical consumer’s perspective, the comparison of x and y involves subjective considerations, and is not simply a matter of computing present values based on market returns.

2.6 Comparisons to Other Approaches

Economists have developed and deployed different methods of evaluating the quality of financial decision making for different purposes. With respect to the current application, our approach offers important advantages.

The most common alternative is to evaluate the prevalence of dominated choices; see Ernst et al. (2004), Calvet et al. (2007, 2009), Agarwal et al. (2009), Baltussen and Post (2011), Choi et al. (2011), and Aufenanger et al. (2016). The essence of this approach is to select diagnostic tasks that remove personal preferences from the mix. In effect, each decision boils down to solving a math problem that has one and only one correct answer. Consequently, the approach amounts to administering an

¹⁷It is worth noting, however, that our experimental tasks involve tradeoffs between payoffs received after a short delay on the order of a few days, and payoffs received after longer delays. Consequently, even if subjects consume income when they receive it, conventional $\beta\delta$ discounting cannot account for their high discount rates.

incentivized test of financial literacy. Conversely, every incentivized financial literacy test, including the one we administer as part of this experiment, consists of decision tasks in which a single choice – the correct answer – is the dominant option.

In contrast, the vast majority of real-world financial decisions are not simply math problems: the ‘right’ choice almost always depends on preferences. Thus, a central issue when evaluating financial education interventions is whether people operationalize pertinent knowledge and concepts when preferences remain in the mix. They may not.¹⁸ Posing a problem that has no objectively correct answer may reduce the resemblance to textbook examples, making the applicable principles harder to recognize. People may be less likely to deploy mathematical tools when mathematics potentially govern only one amongst several aspects of evaluation. Consideration of preferences may also activate specialized heuristics or psychological mechanisms, such as motivated reasoning (Kunda, 1990), that sweep relevant principles into the background, even if they are invoked. An important advantage of our approach is that, unlike the dominance agenda, it permits us to evaluate the quality of decision making rigorously even when preferences remain in the mix.

In principle, by deploying structural methods involving explicit models of preferences and choices, one could achieve the same advantage. Unfortunately, in any given application, that approach may necessitate much stronger assumptions than many analysts are willing to make or accept. To our knowledge, Song (2015) is the only existing empirical study that employs this approach in the context of financial education. He uses a life-cycle consumption model to evaluate the welfare effects of changes in retirement contributions resulting from an educational intervention targeting compound interest. His analysis hinges on the accuracy with which a particular life-cycle model, calibrated with data drawn from other choice domains, describes lifetime opportunities, unobserved future choices, and ‘true’ preferences.¹⁹ By focusing on consistency within paired valuation tasks, our approach avoids the need to endorse a particular structural model and allows us to proceed under much weaker assumptions.

There are, of course, other notions of internal consistency such as WARP and GARP, and these have also been used to assess the quality of financial decision making (Choi, Kariv, Müller and Silverman, 2014). These tools complement our approach because they allow one to examine the consistency of non-equivalent choices made in a fixed decision frame, rather than the consistency of

¹⁸This disconnect has been observed in other contexts. Enke and Zimmermann (2015) show that many people tend to neglect correlations even in simple settings, despite knowing how to account for them. Taubinsky and Rees-Jones (2016) find that many consumers underreact to excise taxes, even though they can properly compute tax-inclusive prices. Likewise, in the current context, consider the contrast between the conclusions we reach when evaluating our financial education intervention based on an incentivized test of financial literacy (effectively choice problems with dominant options) and our measures of financial competence.

¹⁹He concludes that the intervention improved welfare on average even though its effect on behavior was indiscriminate. 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. Another interpretation is that the life-cycle model poorly captured actual objectives.

equivalent choices across different frames. However, for the following reasons, measures of within-frame consistency are less well-suited to the task of assessing financial education interventions than our approach. First, they are not designed to detect the types of decision making failures that primarily concern us. A consumer who misunderstands a financial concept in a consistent manner will nevertheless respect such axioms. For example, one who incorrectly believes that bundle i will ultimately lead to a better consumption bundle than bundle j , perhaps because she uses the simple interest formula to assess compound interest, will choose i over j , and will never choose j when i is available; therefore, her choices among bundles will satisfy WARP. Second, financial education does not target conformance with WARP directly, and non-conformance may result from a variety of considerations that are unrelated to the consumer's understanding of specific financial principles (such as incompleteness of underlying preferences). In contrast, our approach allows one to design the paired valuation tasks so that the targeted principles govern their equivalence. Third, our approach more readily yields measures of non-conformance that are interpretable as welfare losses.²⁰

One important point of differentiation among the various studies mentioned above is that some evaluate real-world decisions while others examine choices in experimental tasks. The current application of our method falls into the latter category.²¹ While the use of experimental data permits us to proceed with fewer assumptions and facilitates sharper conclusions, there is also a cost, in that it is less obvious whether the conclusions generalize to the choices that actually matter. We will defer our discussion of this concern to Section 7, which broadly addresses questions of generalizability.

3 Experimental Design

We now deploy our method to examine the effects of a financial education intervention on the quality of decision making.

Our experiment involves a web-based financial education intervention narrowly focused on the concept of compound interest. We chose this topic for a number of reasons. First, it is associated with a well-documented behavioral bias that an intervention, if effective, would counteract. Second, it is a fundamental concept in financial decision making and most financial education courses cover

²⁰To be clear, some measures of non-conformance with GARP, such as the Afriat (1972) critical cost efficiency index, do have efficiency interpretations; see, e.g., Choi et al. (2014) for a related application. Moreover, Echenique et al. (2011) provide a measure of non-conformance that is interpretable as the maximal amount of money one can extract from a decision maker with specific violations of GARP.

²¹One can also use our method to assess real-world choices, but the implementation is more challenging. Admittedly, it may be easier to find naturally occurring opportunities to study the frequency of dominated choices. That said, dominance is typically hard to establish in the field, because the complexity of the real world invites many possible rationalizations for ostensibly poor choices. As an example, consider the use of payday loans by consumers with unused credit card balances. While agreeing that this practice is generally ill-advised, we question whether one can legitimately categorize it as dominated, as some have claimed (see Ernst et al., 2004). In principle, it could be rational for a consumer to preserve some of the instant liquidity credit cards offer for emergencies requiring immediate outlays.

it. Third, its narrowness, and the corresponding brevity of treatments in standard investment guides and employer-sponsored financial education programs, make it suitable for an intervention of limited duration.

The experiment consisted of three stages. First, subjects watched one of four educational videos, selected at random. Second, they completed incentivized 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 Online Appendix B.

Education intervention. We used a video based on 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.²² The remainder of the text consists of two components:

- (i) An explanation of a simple, memorable, and potentially valuable heuristic, the rule of 72, along with five illustrative applications.²³ 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).
- (ii) Motivational material (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. These anecdotes do not include any computations, and hence are not helpful for understanding the mechanics of compound interest. It also explicitly exhorts readers to behave frugally, asserting that "the power of

²²The 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."

²³We used this particular investment guide in part because it teaches a useful quantitative heuristic. Some investment guides and educational interventions cover this topic without offering useful quantitative tools.

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 2×2 *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.²⁴ 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.²⁵ 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 www.khanacademy.org. 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 the Online Appendix B.

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.²⁶ 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.,

²⁴In cases where it was impossible to remove sentences containing rhetorical material, we substituted neutral language. 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?”

²⁵We chose this approach because existing research indicates that financial education videos are generally more effective than written text (Lusardi et al., 2015).

²⁶We chose the parameters of the tasks so that the complexly framed version yielded the same future payment as the simply framed version according to the rule of 72. Since that rule is an approximation, future values actually differ by small amounts between the two frames.

2006).²⁷ 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.

Table 1 lists the parameters t , r , a , and R used for the paired valuation tasks. We chose time horizons of 36 and 72 days to simplify applications of the rule of 72.²⁸ 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 chose values for the remaining parameters 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.

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, or personal advice when making decisions, as subjects always have those options when making real-world decisions.²⁹ As detailed below, only a quarter of our subjects report using such resources when completing the incentivized test, a fraction that does not vary meaningfully across treatments. That pattern mirrors findings concerning real financial decisions (Lusardi and Mitchell, 2011).

Knowledge test and self-reports. We also gathered data to evaluate the educational intervention according to conventional metrics. Many studies have used 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., 2015; Walstad, Rebeck and MacDonald, 2010; Council for Economic Education, 2006; Collins, 2013). Accordingly, we administered an incentivized test consisting of the five questions about compound interest listed in Table 2, as well as five questions about the material covered in the video shown to the control group.³⁰

Previous studies have also examined self-reported decision strategies (for instance Heinberg et al., 2014; Lührmann, Serra-Garcia and Winter, 2015b; Carlin, Jiang and Spiller, 2014). In the final stage

²⁷Throughout, we set $V_{j,r,t}^f$ equal to the midpoint of the pertinent interval. For further details, see Online Appendix B.

²⁸We 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.

²⁹This feature differentiates our study from most of the literature on the effects of financial education (Hastings, Madrian and Skimmyhorn, 2013). An exception is Levy and Tasoff (2016) who also conduct an internet-based study.

³⁰The test questions for the material in the control video are available upon request. We randomized the order of all ten test questions at the subject level. Subjects knew 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.

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

Table 1: Decision problems. *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.

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.³¹

4 Implementation and Preliminary Analysis

We conducted our experiment through the online labor market Amazon Mechanical Turk (AMT).³² 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 (Horton, Rand and Zeckhauser, 2011; Mason and Suri, 2012; Peysakhovich, Nowak and Rand, 2014).

We ran eight sessions with a total of 504 subjects during April and May 2014, all on weekday mornings. 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-

³¹The questionnaire also addressed a small number of additional issues.

³²An advantage of conducting the experiment online is that it mirrors many real-world financial decisions, which have steadily migrated to internet platforms.

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.

Table 2: Test questions. Questions were presented in random order and intermingled with the questions concerning material covered in the Control video.

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.³³ 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.³⁴

The average length of a session was 62 minutes (s.d. 22 minutes). Attrition was negligible and unrelated to the treatments.³⁵ 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 (Mason and Suri, 2012).

³³The video invited subjects to click a link to the author's homepage so they could verify the authenticity of the video. It also provided a link to the homepage of a graduate-student co-author (Ambuehl) in case they felt uncomfortable contacting and inconveniencing a professor.

³⁴This test of financial literacy originated with Lusardi and Mitchell (2009) and van Rooij, Lusardi and Alessie (2011), and has been used in many other studies (Lusardi and Mitchell, 2014). We reproduce the five questions in the Online Appendix Table B.1. It is standard practice to administer this test without incentivization.

³⁵Only 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.

Multiple switching. Any subject with coherent preferences will switch her 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 laboratory studies of risky choices by undergraduate subjects (such as Holt and Laury, 2002), the comparable figure typically falls in the range of 10 to 15%. Following the usual convention (see, for example, Harrison et al., 2005), we focus attention on the 455 subjects who respected monotonicity .

Demographics. While our subjects are not highly representative of the US population, neither are they highly unusual. On average, our sample is 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 subjects through the internet, our sample over-represents males, young adults, whites, urban residents, and people who have never been married. The level of financial literacy slightly exceeds that found in other studies of US subjects (see Lusardi and Mitchell, 2009, and Lusardi, 2011). Online Appendix C.1 provides additional details.

Randomization into treatments was successful. Of the 34 F -tests we performed to assess the 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. Online Appendix C.1 also includes these tests.

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. Third, we obtain similar results when subjects who exhibited either unusually noisy or unresponsive behavior – the likely hallmarks of inattention – are dropped from the sample; see Online Appendix D.2. Finally, we take a small

degree of reassurance from the fact that, when completing the exit survey, the overwhelming majority of subjects reported paying the highest level of attention to the video and to their choices.

Baseline discounting and exponential growth bias. The extent to which subject j discounts a reward r in a decision task with time horizon t and frame $f \in \{\text{simple, complex}\}$ is given by

$$\delta_{j,r,t}^f = \frac{V_{j,r,t}^f}{r} \quad (5)$$

We also refer to this quantity as the subject's *normalized valuation*. Focusing on the Control condition, the average normalized valuations with simple framing (that is, discount factors) are 0.767 and 0.706 for tasks with 36 and 72 day horizons, respectively.³⁶ There is also significant exponential growth bias: normalized valuations with complex framing are lower than with simple framing by an average of 13.3 percent of the promised reward.

5 Conventional Outcome Measures

As noted in Sections 1 and 3, studies that evaluate financial education interventions frequently focus on financial literacy, self-reported decision strategies, and directional changes in behavior. In this section, we show that our intervention appears to be successful according to these conventional outcome measures.³⁷ In section 6 we then show that these measures fail to detect crucial deficiencies.

Column 1 of table 3 shows the effects of our 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, or 39%, of the questions correctly. The Full intervention increases the average score dramatically, by roughly 1.4 additional correct answers, or equivalently by 29 percentage points, 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 average score improves by only 0.5 additional correct answers, or equivalently 10 percentage points.³⁸ Thus, according to standard measures, the interventions that include substantive material are highly effective at promoting financial literacy.³⁹

³⁶Thus our typical subject discounts future payments rather heavily. A longer horizon results in greater discounting, but the relative magnitudes of these rates across horizons are inconsistent with exponential discounting. These patterns are common in studies that elicit time preferences over short horizons (Frederick, Loewenstein and O'Donoghue, 2002). They do not, however, reflect conventional present-bias, because our subjects expect to receive all payments with a one-to-two-day lag.

³⁷All results reported in this section are robust with respect to various statistical controls and alternative specifications. For details, see Online Appendix D.

³⁸The fact that there is still a gain is not surprising given that Rhetoric-Only treatment, unlike the Control treatment, includes a simple explanation of compound interest, illustrated through an iterative calculation.

³⁹See Online Appendix D.1 for the effects on individual test questions.

VARIABLES	(1) Test score compounding	(2) Test score control	(3) External help	(4) Uses rule in complex framing	(5) Uses rule in simple framing	(6) Explicit calculation	(7) $100 \times \delta_{j,r,c}^c$
Level in Control	1.963*** (0.139)	3.284*** (0.103)	0.220*** (0.042)	0.128*** (0.040)	0.092** (0.039)	6.404*** (0.354)	58.95*** (2.272)
<i>Treatment effects</i>							
Full	1.442*** (0.197)	-1.058*** (0.146)	-0.013 (0.059)	0.579*** (0.056)	0.172*** (0.055)	1.738*** (0.504)	14.31*** (3.427)
Substance-Only	1.271*** (0.189)	-1.339*** (0.140)	0.061 (0.057)	0.637*** (0.054)	0.260*** (0.053)	1.737*** (0.482)	4.021 (3.285)
Rhetoric-Only	0.492** (0.195)	-1.079*** (0.144)	0.066 (0.058)	0.104* (0.056)	0.060 (0.054)	0.418 (0.497)	18.59*** (3.595)
$P(\beta_{Substance} = \beta_{Rhetoric})$	0.000	0.062	0.937	0.000	0.000	0.006	0.000
$P(\beta_{Full} = \beta_{Rhetoric})$	0.000	0.885	0.184	0.000	0.040	0.009	0.259
$P(\beta_{Substance} = \beta_{Full})$	0.368	0.047	0.196	0.285	0.099	0.999	0.003
$P(\text{joint insignificance})$	0.000	0.000	0.400	0.000	0.000	0.000	0.000
Observations	455	455	455	455	455	455	4,550
Number of subjects	455	455	455	455	455	455	455

Table 3: Conventional outcome measures. The dependent variable in columns 1 - 7 are, respectively, the mean number of test questions answered correctly (1 to 5), the self-reported answer to whether the subjects used external help in the test, the answer to the question whether the rule of 72 was used in the complexly framed problems, the answer to the question whether the rule of 72 was used in the simply framed problems, the self-reported number of complexly framed problems (out of 10) for which the subject explicitly calculated the future reward, and the normalized valuation for complexly framed decision tasks. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

These improvements in performance on test questions pertaining to compound interest are not due to effects of the Full, Substance-Only, and Rhetoric-Only videos on general motivation. If they were, we would find comparable effects for subjects' scores on the five test questions pertaining to topics covered in the Control video. On the contrary, as shown in Column 2, the Control video increases the average score on this portion of the test by more than one additional correct answer (over 20%) relative to all three treatments. We conclude that subjects learn the substantive material contained in whichever video they view.

A natural concern is that education may simply displace the use of reference materials or reliance on knowledgeable friends. Such displacement could dampen the effects of the interventions on test scores and choices. Column 3 shows that the various educational interventions do not affect the (self-reported) extent to which subjects employ external help.

Subjects report operationalizing the knowledge they acquire from the substantive interventions, as Column 4 shows. Only 13% of subjects in the Control report using the rule of 72 when making complexly framed choices. In sharp contrast, the corresponding figure exceeds 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.⁴⁰ As shown in Column 5, 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, this finding is not entirely unexpected.⁴¹

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, Column 6 shows that the Full and Substance-Only interventions significantly increase the average number of complexly framed decision tasks for which subjects report 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 is much smaller and statistically insignificant. Thus, the educational interventions do not simply increase (self-reported) reliance on the rule of 72 by migrating subjects from other methods of explicit calculation.

Next we turn to the effects of financial education on behavior. Many studies draw informal inferences concerning the success of these types of interventions by asking whether they directionally

⁴⁰There 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.

⁴¹Subjects 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.

counteract presumed biases. 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, it is well-established that people on average underestimate the power of compound interest, a phenomenon known as *exponential growth bias* (see the references cited in Section 1). Consequently, following the approach adopted in the literature, one would deem an intervention potentially welfare-improving if it leads subjects to value investments involving compound interest (our complexly framed rewards) more highly.

Column 7 shows the effects of the various treatment videos on normalized valuations for complexly framed tasks. According to the table, the Full video increases valuations for complexly framed choices by a 14.31 percentage points relative to the Control video, and the effect is highly significant. Furthermore, given the magnitude of the exponential growth bias documented in the existing literature, the size of the average treatment effect raises no concerns about systematic overcorrection.⁴²

Taken at face value, the preceding results suggest that the Full intervention has the right effects for the right reasons. It successfully increases performance on an incentivized knowledge test and, as one would hope, this increase results from the substantive elements of the intervention rather than from motivational rhetoric. Moreover, subjects report operationalizing their newly obtained knowledge in their decisions, and there is no indication that the use of new quantitative tools crowds out reliance on other resources or computational methods. Finally, valuations in complexly framed tasks change in a direction that counteracts a known bias (which we have verified for this sample), and the change does not appear to be excessive on average. 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.

As we will see, the results presented in the next section paint a much different picture, which demonstrates the value of formally assessing the quality of financial decision making, as we do in our work. A closer examination of the regression in the final column of Table 3 alerts us to the source of the problem: the estimated effect on valuations in complexly framed tasks for the Substance-Only treatment (4.02 percentage points) is statistically indistinguishable from zero and significantly smaller than that of the Full treatment (14.31 percentage points, $p = 0.003$). In contrast, the estimated effect for the Rhetoric-Only treatment (18.59 percentage points) is actually larger than that of the Full treatment, and we do not reject equality ($p = 0.259$). Accordingly, despite demonstrable effects of

⁴²Stango and Zinman (2009) posit that subjects assess future value (FV) based on the magnitude of an initial investment (I) and the interest rate (i) according the formula $FV = I \times (1 + i)^{\theta t}$. They estimate this equation for each member of their subject pool. The median estimate of θ is 0.8 (see their footnote 24). Given the tasks in our experiment, a subject with $\theta = 0.8$ underestimates future values on average by a factor of 0.71. Assuming that current valuation varies proportionately with the magnitude of the future receipt, the elimination of exponential growth bias would increase the average current valuation by 40.1% (because $1/0.71 = 1.401$). In contrast, the Full treatment increases the mean valuations for complexly framed tasks by $14.31/58.95 = 24.3\%$. Thus it appears from this calculation that the Full treatment did not cause subjects to overcorrect on average.

substantive instruction on comprehension as well as subjects' statements concerning their proclivities to operationalize substantive knowledge in their decisions, *the behavioral effects of the Full treatment are traceable almost entirely to motivational rhetoric rather than substance.*

6 The Quality of Decision Making

Next we assess the effects of our educational interventions on the quality of decision making using the approach developed in Section 2. We demonstrate that, despite the generally encouraging results of the previous section, the Full intervention fails to make subjects better off on average. Additional results link this finding to the role of motivational rhetoric.

VARIABLES	(1) 100 $\times \delta_{j,r,c}^s$	(2) 100 $\times d_{j,r,c}$	(3) 100 $\times C_e$	(4) 100 $\times C_m$
Level in Control	72.26*** (2.089)	-13.31*** (2.221)	11.69*** (1.232)	24.45*** (1.633)
<i>Treatment effects</i>				
Full	0.402 (2.99)	13.91*** (3.332)	0.155 (2.035)	-1.584 (2.386)
Substance-Only	0.018 (2.913)	4.002 (2.961)	-1.461 (1.669)	-2.436 (2.172)
Rhetoric-Only	5.368* (2.975)	13.22*** (2.952)	-2.546 (1.700)	-4.651** (2.155)
$P(\beta_{Substance} = \beta_{Rhetoric})$	0.069	0.001	0.505	0.270
$P(\beta_{Full} = \beta_{Rhetoric})$	0.010	0.827	0.177	0.171
$P(\beta_{Substance} = \beta_{Full})$	0.897	0.002	0.413	0.706
$P(\text{joint insignificance})$	0.202	0.000	0.390	0.178
Observations	4,550	4,550	4,550	4,550
Number of subjects	455	455	455	455

Table 4: Results pertaining to the quality of decision making. $\delta_{j,r,t}^f$ is subject j 's normalized valuation for reward r to be received at time t when presented in frame f . $d_{j,r,c} = \delta_{j,r,t}^c - \delta_{j,r,t}^s$ is the framing distortion. If subject j underestimates compound interest, $d_{j,r,c} < 0$. Subject j 's expected and maximal welfare losses from characterization failure are proportional to $C_e = (d_{j,r,c})^2$ and $C_m = |d_{j,r,c}|$, respectively. Standard errors clustered by subject. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Simply framed valuations. We start by verifying that subjects' valuations for simply framed opportunities are largely invariant with respect to the educational interventions. As explained in section 2, the welfare interpretation of our financial competence measures presupposes this stability property. Column 1 of Table 4 shows that, for normalized valuations in simply framed tasks, the estimated effects of the Full and Substance-Only interventions are close to zero and statistically

insignificant. While the corresponding effect of the Rhetoric-Only condition is somewhat larger, it is less than one-third the size of its counterpart in the regression for complexly framed valuations (column 7 of Table 3), and it lacks statistical significance at the 5% level.

Framing distortions. Next we investigate the effect of our educational interventions on the degree of exponential growth bias. To this end, we define the *framing distortion* as the difference between the normalized complex and simple valuations for the same task: $d_{j,r,c} = \delta_{j,r,t}^c - \delta_{j,r,t}^s$. (Note that we do *not* take the absolute value.) An individual who underestimates (overestimates) the power of compound interest will exhibit $d_{j,r,t} < 0$ ($d_{j,r,t} > 0$). As we mentioned in Section 4, our subjects exhibit substantial exponential growth bias in the Control condition. Indeed, column 2 of table 4 shows subjects' normalized valuations are lower with complex framing than with simple framing by, on average, 13.31 percentage points (measured relative to the promised reward).

In light of the fact that our Full intervention increases valuations for complexly framed tasks but not for simply framed tasks, one should not be surprised to learn that it reduces the magnitude of the average framing distortion. Even so, the extent of the reduction is striking. According to column 2, the average value of $d_{j,r,c}$ falls by 13.91 percentage points, leaving a gap of only 0.6 percentage points (s.e. = 2.48), thereby effectively eliminating exponential growth bias on average. While this result is in line with the encouraging findings of the previous section, we emphasize that the effect flows almost entirely from motivational rhetoric rather than the substantive elements of instruction. In particular, the estimated effect on the mean framing distortion 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, $p = 0.002$). 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 ($p = 0.827$).

Because the elimination of the average framing distortion results from motivational rhetoric rather than substance, one suspects that these averages may mask many inappropriate subject-level responses. To investigate this possibility, we examine the cumulative distribution of $d_{j,r,t}$ for each treatment; see Figure 1. While subjects usually exhibit exponential growth bias in the Control treatment ($d_{j,r,t} < 0$ in roughly 65% of tasks), they also *overestimate* compound interest with reasonably high frequency ($d_{j,r,t} > 0$ in roughly 35% of tasks). Moreover, much of the observed variation in the framing distortion reflects individual-level heterogeneity rather than task-specific noise.⁴³

⁴³The Cronbach- α statistics for $d_{j,r,t}$ show that subjects who underestimate (overestimate) compound interest in some decisions tend do so in all decisions, and by comparable amounts. The values of the statistic are 0.92, 0.92, 0.94, and 0.95 for the Control, Full, Substance-Only and Rhetoric-Only treatments, respectively. These values compare favorably with the standard benchmark of 0.8, indicating a high level of individual consistency.

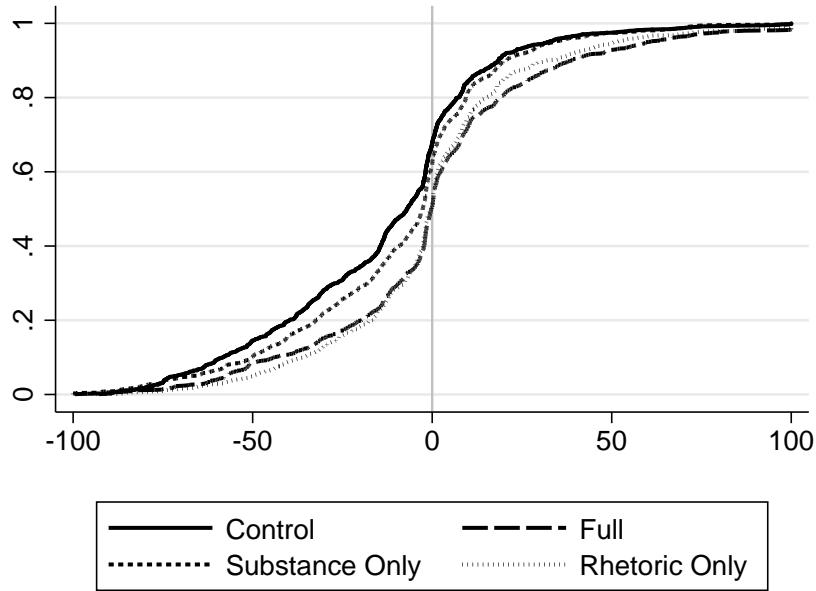


Figure 1: C.D.F. of framing distortion, by treatment. For better visibility, the graph is truncated at -100 and at 100.

An effective intervention would bring valuations for equivalent complexly and simply framed problems more closely in line. For subjects who underestimate compound interest, it would increase valuations complexly framed tasks. For subjects who overestimate compound interest, it would *decrease* these valuations. As a result, the two CDFs would *cross* at $d = 0$, and the distribution for the intervention would be more tightly concentrated around zero.

Instead, *the Full intervention shifts the entire CDF to the right*. In other words, it generally increases valuations for complexly framed tasks irrespective of whether the subject initially underestimates or overestimates compound interest. This indiscriminate effect helps in some instances but hurts in others.⁴⁴ The Rhetoric-Only treatment yields a similar shift in the CDF, while the Substance-Only treatment has a much smaller effect. These findings are consistent with the hypothesis that behavioral effects of the Full intervention primarily reflect motivational elements of instruction rather than substantive elements, and that consequently they bear no systematic relation to the appropriate response.

Financial competence and welfare. In light of the preceding findings, the effects of the Full intervention on welfare are unclear. On the one hand, it significantly enhances financial literacy,

⁴⁴We are not alone in finding that some people overestimate compound interest; see, for example, Goda et al. (2015) and Levy and Tasoff (2016).

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. On the other hand, its behavioral effects are driven almost entirely by its motivational elements rather than its substantive elements, and as a result its impact is largely indiscriminate (that is, unrelated to the initial framing distortion).

To determine whether an intervention improves or reduces welfare, we examine its effects on financial competence, measured as either $C_e = (\delta_{j,r,t}^s - \delta_{j,r,t}^c)^2$ or $C_m = |\delta_{j,r,t}^s - \delta_{j,r,t}^c|$. These measures are always non-negative, and higher values imply lower competence. Columns 1 and 2 of Table 4 show how our interventions affect them. There is no evidence that the Full treatment benefits subjects by improving their financial competence on average. The point estimates for the effects of the Substance-Only and Rhetoric-Only treatments on C_e and C_m are a bit larger in magnitude, but only one of the four is statistically significant (the Rhetoric-Only treatment effect on C_m).

It is natural to wonder whether our findings concerning financial competence are attributable to a mismatch between the difficulty of the valuation problems and the depth of the material covered in the Full intervention. To investigate that possibility, we reexamine the effects of the various interventions on welfare, differentiating between tasks according to the difficulty of applying the rule of 72. 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-estimate the basic specification from Table 4 separately for valuation tasks with a single doubling, two to four doublings, and 2.5 doublings. Results appear in columns 3 - 5, respectively, of table 5.

If the ease of applying the rule of 72 improves the success of interventions that teach it, we should see systematic differences in the *relative* welfare effects of the substantive and Rhetoric-Only interventions across these three categories of valuation tasks.⁴⁵ Thus, in table 5, 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 to columns 4 and 5, thereby increasing the difficulty of applying the rule. In fact, no such pattern is observed. 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 ($p > 0.10$).

⁴⁵Notice that our focus here is on the relationship between *relative* welfare effects and the difficulty of applying the rule of 72. For any given treatment, the *absolute* welfare effects may vary with that degree of difficulty for other reasons. For example, difficulty is associated with the number of doublings, which in turn is associated with initial degree of exponential growth bias. Mechanically, any fixed increase in valuation is more likely to be welfare enhancing when the initial bias is greater.

for all pairwise comparisons).⁴⁶ Thus, one cannot attribute the poor performance of our substantive interventions in terms of welfare to the difficulty of applying the rule of 72 in our valuation tasks.

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

Table 5: Problem difficulty and response times. Columns 1 - 3 show the effect on average welfare for complexly framed decision tasks that differ according to the number of times the investment doubles over its life. Columns 4 and 5 show the effect of the treatments on mean response times for the complexly and simply framed problems, respectively. Standard errors clustered by subject. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Decision times. An examination of decision times corroborates some the assumptions made in Section 2 as well as various inferences we have drawn from our analysis of valuations. We derive this corroboration from the regressions in the final two columns of Table 5, which pertain to decision times in complexly and simply framed valuation tasks, $\tau_{j,r,t}^c$ and $\tau_{j,r,t}^s$, respectively. Several notable conclusions follow from these regressions.

First, on average, valuation tasks with complex framing take subjects nearly three times as long to complete than those with simple framing (59 seconds versus 22 seconds, $p < 0.001$). This finding is consistent with our premise that simply framed tasks are transparent and easily evaluated, while complexly framed tasks require additional cognitive effort, likely because subjects try to “translate” from complex to simple framing.

Second, decision times are sensitive to the educational interventions for complexly framed tasks, but not for simply framed tasks. This pattern is expected given the sensitivity of complexly framed

⁴⁶We 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.

valuations and the insensitivity of simply framed valuations to the same interventions, as well as the relative frequencies with which subjects report using information from the videos when performing the two types of tasks. Together, these findings provide a solid foundation for our assumption that the interventions change the way subjects think about and assess opportunities that are complexly framed, but not ones that are simply framed. As we hypothesized, education appears to alter the “translation” from complex to simple framing.

Third, the pattern of effects for complexly framed tasks corroborates our inferences about the role of motivational rhetoric. The Substance-Only intervention has the largest effect, increasing average decision times by 19.5 seconds relative to the Control, or roughly 40%. Thus, this intervention appears to alter the way subjects think about complexly framed opportunities, even though it does not produce much of a systematic shift in *average* valuations. (Below, we show that it *does* change the manner in which subjects value these opportunities, but the effect is directionally haphazard.) In contrast, the impact of the Full treatment is small and statistically insignificant. Thus, the provision of substantive information appears to induce greater effort and deliberation, but the addition of simplistic rhetorical assertions concerning the power of compound interest seem to negate that effect, perhaps because they point to a less cognitively demanding heuristic.

Reliance on simple interest calculations. According to previous research, many people project investment values based on linear rather than exponential growth – in other words, according to simple interest (Eisenstein and Hoch, 2007; Mckenzie and Liersch, 2011). As we show next, *all* of our interventions – including the Substance-Only video, for which the effect on average valuations is minimal – render this misconception less common. Thus, the problem is not one of intellectual stubbornness. Rather, the interventions apparently migrate subjects to other similarly inappropriate methods of making choices.

We estimate the frequency with which subjects employ simple interest calculations 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:

$$\frac{V_{j,r,t}^c}{V_{j,r,t}^s} = \sum_{\tau \in T} \left[\beta_0^\tau + \beta_1^\tau \frac{FV_{r,t}^{SI}}{FV_{r,t}^{CI}} \right] \mathbb{I}_j(\tau) + \epsilon_{j,r,t} \quad (6)$$

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 τ .⁴⁷ In this specification, β_1^τ 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 β_1^τ as the fraction of decisions that are consistent with simple rather than compound interest calculations in treatment τ . In the extreme, if all subjects correctly calculate future value, we would find $\beta_0^\tau = 1$ and $\beta_1^\tau = 0$, and if all subjects use the simple interest formula, we would find $\beta_0^\tau = 0$ and $\beta_1^\tau = 1$.

We estimate model (6) pooling data for all of our subjects, as well as separately for subjects with high and low financial literacy, as measured by the three questions concerning the time value of money that were included in the unincentivized financial literacy test administered at the start of the experiment. In each case, we pool data across all valuation tasks.⁴⁸ Here we use median regression because the distribution of the dependent variable is highly skewed due to the presence of observations with values of $V_{j,r,t}^s$ close to zero.

Results appear in Table 6. According to our basic specification, roughly 30% of the Control group's complexly framed decisions are made using the simple interest formula. That method appears to be far more prevalent among those with low financial literacy (49%) than among those with high financial literacy (20%). The Substance-Only treatment reduces reliance on simple interest calculations to roughly 9% overall (29% and 6% for those with low and high financial literacy, respectively). 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.⁴⁹ 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

⁴⁷Note that the dependent variable, $\frac{V_{j,r,t}^c}{V_{j,r,t}^s}$, 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 $\tilde{\delta}$.

⁴⁸In particular, our regressions employ data for valuation tasks with both 36 and 72 day horizons. As discussed elsewhere 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 model (6) 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.

⁴⁹We fail to reject the hypothesis that $\beta_0^\tau + \beta_1^\tau = 1$ in all cases with $p > 0.3$.

formula, they do *not* succeed in fostering the correct calculation of compound interest in the context of decisions that implicate preferences.

VARIABLE	$V_{j,r,t}^c/V_{j,r,t}^s$		
	(1) all	(2) high FL	(3) low FL
$\beta_1^{Control}$	0.304*** (0.100)	0.197** (0.091)	0.489** (0.212)
β_1^{Full}	0.009 (0.031)	0.0133 (0.027)	-0.004 (0.134)
$\beta_1^{Substance}$	0.088** (0.038)	0.060* (0.035)	0.294* (0.178)
$\beta_1^{Rhetoric}$	0.023 (0.033)	0.000 (0.032)	0.082 (0.108)
$\beta_0^{Control}$	0.721*** (0.085)	0.814*** (0.080)	0.527** (0.208)
β_0^{Full}	0.993*** (0.023)	0.994*** (0.019)	0.984*** (0.102)
$\beta_0^{Substance}$	0.906*** (0.030)	0.930*** (0.026)	0.730*** (0.149)
$\beta_0^{Rhetoric}$	0.983*** (0.020)	1.000*** (0.017)	0.926*** (0.082)
$P(\beta_1^{Control} = \beta_1^{Full})$	0.005	0.053	0.050
$P(\beta_1^{Control} = \beta_1^{Substance})$	0.044	0.160	0.481
$P(\beta_1^{Control} = \beta_1^{Rhetoric})$	0.008	0.041	0.088
Observations	4,550	2,920	1,630
Subjects	455	292	163

Table 6: Use of simple interest formula. High and low financial literacy (FL) are measured by the three questions concerning the time value of money that were included in the unincentivized test administered at the start of the experiment. Estimated using median regression. Standard errors clustered by subject. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Robustness Online Appendix D details a wide range of robustness analyses. First we show that our findings are not sensitive to the inclusion of various demographic control variables. This is unsurprising given that our samples are reasonably large and well-randomized. Second, we demonstrate that our main results are not attributable either to special features of particular experimental tasks such as the time horizon, or to special subgroups of subjects defined by initial levels of financial literacy, degree of responsiveness to variation in experimental stimuli, or degree to which a subject's implied rate of time preference is stable across simply framed tasks. Third, we adapt our analysis to allow

for the possibility that subjects’ valuations may be “fuzzy.” Here we employ two distinct analytic strategies. One is to assume that “true” valuations are well-defined, and that the fuzziness reflects noisy elicitation, which could in principle mask improvements in welfare. The other strategy is to proceed according to the Bernheim-Rangel welfare framework, treating fuzzy valuations as implying normative ambiguity. Both strategies leave our qualitative conclusions unchanged.

7 Generalizability

Naturally, one must exercise caution when generalizing from any study that focuses on a single financial education intervention. Certainly, our analysis does not justify a broad inference that financial education programs fail to improve welfare. Different interventions may have different effects. Indeed, even the same intervention may produce dissimilar outcomes in different populations.

Even so, our analysis does have important general implications. First, it highlights the potential pitfalls of educational interventions that are brief and laden with motivational rhetoric. In these and other respects, the intervention we examine is typical of the programs offered to millions of workers through their employers.⁵⁰ Unfortunately, as we have seen, behavior may respond primarily to motivational rhetoric even when people appear to understand and internalize the substantive elements of instruction. By making the material engaging and memorable, educators may also render its behavioral effects indiscriminate, and consequently of questionable value. This conclusion directly challenges received wisdom and argues for a reexamination of the principles governing the design of adult financial education interventions.

Second, we have shown that one cannot count on conventional outcome measures to reliably detect these deficiencies. The intervention we consider improves measured financial literacy, increases the self-reported use of the desired decision strategy without reducing reliance on advice or other analytic methods, and on average counteracts a known decision bias. Even so, it fails to enhance the average quality of decision making in simple choice tasks that are designed to permit easy application of the targeted tools. Our analysis therefore underscores the importance of conducting explicit and rigorous examinations of welfare when evaluating particular educational interventions. Moreover, our notion of financial competence leads to practical welfare measures that address this need.

It is of course appropriate to ask whether the welfare effects measured through our method are generalizable beyond the diagnostic tasks from which they are derived. A skeptic might raise one of three issues.

First, improved performance in simple diagnostic tasks does not necessarily translate into better real-world decision making in the complex contexts educational interventions target, such as saving

⁵⁰Another representative feature is that the intervention lacks opportunities for practice and feedback.

for retirement. But if an intervention does not even improve the quality of decision making in tasks to which the pertinent concepts are easily applied, the notion that it will do so in more complex real-world settings, except by chance, is far-fetched.⁵¹ Moreover, if one finds improvements in the simplest diagnostic tasks, one can then deploy the same methods in a sequence of increasingly complicated tasks that mimic additional features of real-world problems.

Second, a failure to improve performance in diagnostic tasks might not generalize to real-world contexts with higher stakes. Plainly, the diagnostic stakes must be large enough to motivate subjects. In the context of our current experiment this is certainly the case, in that we recruit our subjects through an online labor market, and offer payments that are substantial in comparison to the wages for which they normally work. One cannot plausibly attribute the absence of welfare gains to insufficient stakes, inasmuch as the intervention significantly improves performance on a incentivized test of financial literacy even though the stakes are also small. Once stakes are large enough to motivate serious effort, making them too large could actually be problematic, in that it might undermine the accuracy of the first-order approximation given in Theorem 1.

Third, a failure to improve performance in diagnostic tasks might not generalize to real-world contexts wherein people may be more likely to employ analytic tools or seek advice. We note, however, that roughly three-quarters of the US population reports making real financial decisions without assistance (Lusardi and Mitchell, 2011). Additionally, by design, we did not preclude subjects from seeking help. Indeed, because we conducted our experiment online, subjects had access to all resources available through the internet, and were given ample time to use them. Notably, the fraction of individuals reporting that they did not seek advice in our experiment (three quarters) matches experience in the field.

8 Conclusion

In this paper, we introduced a new method for measuring the quality of financial decision making built around the concept of *financial competence*. We used this notion to document the potential pitfalls of the types of brief rhetoric-laden interventions that are commonly used for adult financial education. We also demonstrated that conventional methods of evaluation do not reliably detect these deficiencies, thereby establishing the importance of including assessments of financial competence in evaluations of educational interventions.

⁵¹In principle, the loss function for decisions made in the field could be asymmetric, for instance with underestimation of compound interest more damaging than overestimation. In that case, the fact that we have designed diagnostic tasks with symmetric loss functions could cause us to underestimate the benefits of a measure that causes an indiscriminate increase in the valuations ascribed to interest-bearing assets. However, any improvement in welfare is then entirely fortuitous, and not the result of enhanced decision-making skill.

We say that consumers are financially competent with respect to specific financial principles if they make equivalent choices from equivalent opportunity sets whenever an understanding of those principles would enable them to verify the equivalencies. To assess financial competence, we compare a consumer's decisions across equivalent complexly framed and simply framed valuation tasks. As a method of evaluating the quality of financial decision making, this new approach offers a number of significant advantages over conventional metrics: it is non-paternalistic, it imposes modest information requirements, it is simple, intuitive, and easily implemented. We prove our measure admits a formal welfare interpretation even when consumers suffer from additional decision-making flaws, known and unknown, outside the scope of analysis.

The financial education intervention we study resembles typical employer-sponsored programs with respect to its brevity and emphasis on heuristic and motivational messages; subject to the constraints of brevity, it is ostensibly well-designed. Indeed, we find that it significantly improves measured financial literacy, and subjects report that they operationalize their improved knowledge when making choices. The intervention even eliminates exponential growth bias on average. However, financial competence does *not* improve. Further investigation reveals the explanation: behavior responds primarily to motivational rhetoric even when people appear to understand and internalize the substantive elements of instruction. While the rhetorical components make the material engaging and memorable, they also render its behavioral effects indiscriminate, and consequently of limited value.

Our main findings have potentially important implications for public policy. Most importantly, our analysis shows that it is important to evaluate the success of an intervention by assessing financial competence using the methods we have developed, rather than by administering simple tests of financial knowledge.

Potential strategies for addressing deficiencies in financial competence fall into three broad categories. The first is to devise educational methods that more effectively lead people to put pertinent knowledge into practice when making decisions, and to do so correctly. Given that brevity appears to be a design constraint for adult financial education, it is important to determine whether efficacy and brevity are compatible. In light of our analysis, we recommend exploring program designs that replace motivational rhetoric and simple prescriptive dicta with practical exercises that illustrate the application of the pertinent principles and that create opportunities for providing participants with practice and feedback. Rhetorical prods may be useful for the purpose of marketing educational programs and boosting participation, but counterproductive when incorporated into pedagogy.

A second strategy is to deploy educational programs targeted at populations known to manifest particular biases in order to create countervailing biases. In effect, this amounts to accomplishing the right objective for the wrong reason. To illustrate, in the current study, we have found that

the most beneficial intervention is actually the one with the least substance and the most rhetorical motivation. 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. This “targeted de-biasing” strategy is likely to prove challenging, however, because it seems likely that any success in balancing countervailing biases will be highly context-specific. It is also contrary to the principles of “idealized welfare analysis,” in that it ignores the availability of additional policy tools for addressing biases that lie outside the scope of the educational intervention.

A third strategy is to simplify the framing of naturally occurring decision problems, either by developing and deploying better tools for visualizing opportunities and consequences, or by requiring suppliers of financial products to characterize them in simple terms. In principle this is a promising approach, but its effective implementation 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 adult educational interventions in the workplace and other settings. Research on pedagogical design will, however, at least initially require extensive study of more narrowly focused interventions in the laboratory. Indeed, 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 have reservations concerning the call in Hastings, Madrian and Skimmyhorn (2013) for studies of “large scale interventions.” The effective design of such interventions likely requires a much more comprehensive micro-level understanding of financial education than we currently possess. An initial focus on narrow 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.

References

Abeler, Johannes and Simon Jäger, “Complex Tax Incentives,” *American Economic Journal: Economic Policy*, 2015, 7 (3), 1–28.

Afriat, Sidney N., “Efficiency Estimation of Production Functions,” *International Economic Review*, 1972, 13 (3), 568–98.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson, “The Age of Reason: Financial Decisions over the Life Cycle and Implications for Regulation,” *Brookings Papers on Economic Activity*, 2009, Fall, 51–101.

Almenberg, Johan and Christer Gerdes, “Exponential Growth Bias and Financial Literacy,” *Applied Economics Letters*, 2012, 19 (17), 1693–696.

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutstrom, “Elicitation using Multiple Price List Formats,” *Experimental Economics*, 2006, 9, 383–405.

Andreoni, James and Charles Sprenger, “Estimating Time Preferences from Convex Budgets,” *American Economic Review*, 2012, 102 (7), 3333–356.

— and —, “Risk preferences are not time preferences,” *The American Economic Review*, 2012, 102 (7), 3357–3376.

Aufenanger, Tobias, Friedemann Richter, and Matthias Wrede, “Measuring Decision-Making Ability in the Evaluation of Financial Literacy Education Programs,” *Unpublished Manuscript*, 2016.

Austin, Rob and Winfield Evens, “2013 Trends & Experience in Defined Contribution Plans,” *Aon Hewitt*, 2013.

Baltussen, Guido and Gerrit T. Post, “Irrational Diversification: An Examination of Individual Portfolio Choice,” *Journal of Financial and Quantitative Analysis*, 2011, 5, 1463–491.

Bayer, Patrick J., B. Douglas Bernheim, and John Karl Scholz, “The Effects of Financial Education in the Workplace: Evidence from a Survey of Employers,” *Economic Inquiry*, 2009, 47 (4), 605–24.

Bernheim, B. Douglas, “Behavioral Welfare Economics,” *Journal of the European Economic Association*, 2009, 7 (2-3), 267–319.

—, “The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics,” *Journal of Benefit-Cost Analysis*, 2016, 7 (1), 12–68.

- and Antonio Rangel, “Addiction and Cue-Triggered Decision Processes,” *American Economic Review*, 2004, 94 (5), 1558–590.
- and —, “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics,” *Quarterly Journal of Economics*, 2009, 124 (1), 51–104.
- and Daniel M. Garrett, “The Effects of Financial Education in the Workplace: Evidence from a Survey of Households,” *Journal of Public Economics*, 2003, 87, 1487–519.
- , —, and Dean M. Maki, “Education and Saving: The Long-Term Effects of High School Financial Curriculum Mandates,” *Journal of Public Economics*, 2001, 80, 435–65.
- Bertrand, Marianne and Adair Morse**, “Information Disclosure, Cognitive Biases, and Payday Borrowing,” *The Journal of Finance*, 2011, 66 (6), 1865–993.
- Brown, Alexandra, J Michael Collins, Maximilian Schmeiser, and Carly Urban**, “State Mandated Financial Education and the Credit Behavior of Young Adults,” *Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board, Washington, D.C., Finance and Economics Discussion Series*, 2014, 2014-68.
- Calvet, Laurent E., John Y. Campbell, and Paolo Sodini**, “Down or Out: Assessing the Welfare Costs of Household Investment Mistakes,” *Journal of Political Economy*, 2007, 115 (5), 707–47.
- , —, and —, “Measuring the Financial Sophistication of Households,” *American Economic Review*, 2009, 99 (2), 393–98.
- Carlin, Bruce I., Li Jiang, and Stephen A. Spiller**, “Learning Millennial-Style,” *NBER Working Paper*, 2014, 20268.
- Carpena, Fenella, Shawn Cole, Jeremy Shapiro, and Bilal Zia**, “Unpacking the Causal Chain of Financial Literacy,” *The World Bank Policy Research Working Paper*, 2011, 5798.
- Choi, James J., David Laibson, and Brigitte C. Madrian**, “\$100 Bills on the Sidewalk: Sub-optimal Investment in 401(k) Plans,” *Review of Economics and Statistics*, 2011, 93 (3), 748–63.
- Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman**, “Who is (More) Rational?,” *American Economic Review*, 2014, 104 (6), 1518–550.
- Cole, Shawn and Gauri Kartini Shastry**, “Is High School the Right Time to Teach Self-control? The Effect of Financial Education and Mathematics Courses on Savings Behavior,” *Unpublished Manuscript*, 2010.

– , **Thomas Sampson, and Bilal Zia**, “Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets?,” *The Journal of Finance*, 2011, 66 (6), 1933–967.

Collins, J.M., “The Impacts of Mandatory Financial Education: Evidence from a Randomized Field Study,” *Journal of Economic Behavior & Organization*, 2013, 95, 146–58.

Council for Economic Education, “Financing Your Future (DVD),” <http://financingyourfuture.councilforeconed.org/> 2006.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, “Keeping It Simple: Financial Literacy and Rules of Thumb,” *American Economic Journal: Applied Economics*, 2014, 6 (2), 1–31.

Duflo, Esther and Emmanuel Saez, “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment,” *Quarterly Journal of Economics*, 2003, 118 (3), 815–42.

Echenique, Federico, Sangmok Lee, and Matthew Shum, “The Money Pump as a Measure of Revealed Preference Violations,” *Journal of Political Economy*, 2011, 119 (6), 1201–223.

Eisenstein, Eric M. and Stephen J. Hoch, “Intuitive Compounding: Framing, Temporal Perspective, and Expertise,” *Unpublished Manuscript*, Dec 2007.

Enke, Benjamin and Florian Zimmermann, “Correlation Neglect in Belief Formation,” *Unpublished Manuscript*, 2015.

Ernst, Keith, John Farris, and Uriah King, “Quantifying the Economic Cost of Predatory Payday Lending,” Technical Report, Center for Responsible Lending 2004.

Fernandes, Daniel, John G Lynch Jr., and Richard G Netemeyer, “Financial Literacy, Financial Education, and Downstream Financial Behaviors,” *Management Science*, 2014, 60 (8), 1861–883.

Frederick, Shane, George Loewenstein, and Ted O’Donoghue, “Time Discounting and Time Preference: A Critical Review,” *Journal of Economic Literature*, 2002, 40 (2), 351–401.

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

– , **Matthew R Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff**, “The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings,” *NBER working paper*, 2015, 21482.

Harrison, Glenn W, Morten Igel Lau, E Elisabet Rutström, and Melonie B Sullivan, “Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues,” *Field Experiments in Economics*, 2005, 10, 125–218.

Hastings, Justine S. and Lydia Tejeda-Ashton, “Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence from Mexico,” *NBER Working Paper*, 2008, 14538.

– , **Brigitte C. Madrian, and William L. Skimmyhorn**, “Financial Literacy, Financial Education, and Economic Outcomes,” *Annual Review of Economics*, 2013, 5, 347–73.

Heinberg, Aileen, Angela Hung, Arie Kapteyn, Annamaria Lusardi, Anya Savikhin Samek, and Joanne Yoong, “Five Steps to Planning Success: Experimental Evidence from U.S. Households,” *Oxford Review of Economic Policy*, 2014, 30 (4), 697–724.

Holt, Charles A. and Susan K. Laury, “Risk Aversion and Incentive Effects,” *American Economic Review*, 2002, 92 (5), 1644–655.

Horton, John J., David G. Rand, and Richard J. Zeckhauser, “The Online Laboratory: Conducting Experiments in a Real Labor Market,” *Experimental Economics*, 2011, 14, 399–425.

Jump\$tart Coalition for Personal Financial Literacy, “Financial Literacy Shows Slight Improvement among Nation’s High School Students,” *Press Release*, 2006, Washington, D.C.

Kahneman, Daniel and Amos Tversky, “Prospect theory: An analysis of decision under risk,” *Econometrica: Journal of the econometric society*, 1979, pp. 263–291.

Kalayc, Kenan and Marta Serra-Garcia, “Complexity and Biases,” *Experimental Economics*, 2016, 19 (1), 31–50.

Kline, Paul, *Handbook of Psychological Testing*, 2 ed., London and New York: Routledge, 1999.

Kunda, Ziva, “The Case for Motivated Reasoning,” *Psychological bulletin*, 1990, 108 (3), 480–98.

Levy, Matthew and Joshua Tasoff, “Exponential Growth Bias and Lifecycle Consumption,” *Journal of the European Economic Association*, 2016, 14 (3), 545–83.

Levy, Matthew R. and Joshua Tasoff, “Exponential-Growth Bias and Overconfidence,” *Journal of Economic Psychology*, 2017, 58, 1–14.

Lührmann, Melanie, Marta Serra-Garcia, and Joachim Winter, “The Impact of Financial Education on Adolescents’ Intertemporal Choices,” *IFS Working Paper*, 2015, *W14/18*.

—, —, and —, “Teaching Teenagers in Finance: Does It Work?,” *Journal of Banking & Finance*, 2015, *54*, 160–74.

Lusardi, Annamaria, “Americans’ Financial Capability,” *NBER Working Paper*, 2011, *17103*.

— and **Olivia Mitchell**, “Financial Literacy and Planning: Implications for Retirement Well-being,” in Annamaria Lusardi and Olivia S Mitchell, eds., *Financial Literacy. Implications for Retirement Security and the Financial Marketplace*, Oxford University Press, 2011, pp. 17–39.

— and —, “The Economic Importance of Financial Literacy: Theory and Evidence,” *Journal of Economic Literature*, 2014, *52* (1), 1–44.

— and **Olivia S Mitchell**, “How Ordinary Consumers Make Complex Economic Decisions: Financial Literacy and Retirement,” *NBER Working Paper*, 2009, *15350*.

—, **Anya Samek, Arie Kapteyn, Lewis Glinert, Angela Hung, and Aileen Heinberg**, “Visual Tools and Narratives: New Ways to Improve Financial Literacy,” *Journal of Pension Economics and Finance*, 2015, pp. 1–27.

Malkiel, Burt G. and Charles D. Ellis, *The Elements of Investing: Easy Lessons for Every Investor*, New Jersey: Wiley, 2013.

Mandell, Lewis, “The Financial Literacy of Young American Adults: Results of the 2008 National Jump\$tart Coalition Survey of High School Seniors and College Students,” *Jump\$tart Coalition*, 2009, *Washington, D.C.*

— and **Linda Schmid Klein**, “The Impact of Financial Literacy Education on Subsequent Financial Behavior,” *Journal of Financial Counseling and Planning*, 2009, *20* (1), 15–24.

Mason, Winter and Siddarth Suri, “Conducting Behavioral Research on Amazon’s Mechanical Turk,” *Behavior Research Methods*, 2012, *44* (1), 1–23.

McKenzie, Craig R. M. and Michael J. Liersch, “Misunderstanding Savings Growth: Implications for Retirement Savings Behavior,” *Journal of Marketing Research*, 2011, *48*, 1–13.

Peysakhovich, Alexander, Martin A. Nowak, and David G. Rand, “Humans Display a ‘Cooperative Phenotype’ That Is Domain General and Temporally Stable,” *Nature Communications*, 2014, *5*.

Servon, Lisa J. and Robert Kaestner, “Consumer Financial Literacy and the Impact of Online Banking on the Financial Behavior of Lower-Income Bank Customers,” *Journal of Consumer Affairs*, 2008, 42 (2), 271–305.

Skimmyhorn, William, “Essays in Behavioral Household Finance.” PhD dissertation, Harvard Kennedy School, Cambridge, MA 2012.

— , “Assessing Financial Education: Promising Evidence From Boot Camp,” *American Economic Journal: Economic Policy*, 2016, 8 (2), 322–43.

Song, Changcheng, “Financial Illiteracy and Pension Contributions: A Field Experiment on Compound Interest in China,” *Unpublished Manuscript*, March 2015.

Stango, Victor and Jonathan Zinman, “Exponential Growth Bias and Household Finance,” *The Journal of Finance*, 2009, 64 (6), 2807–849.

Taubinsky, Dmitry and Alex Rees-Jones, “Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment,” *NBER Working Paper*, 2016, 22545.

Tversky, Amos and Daniel Kahneman, “Advances in prospect theory: Cumulative representation of uncertainty,” *Journal of Risk and Uncertainty*, 1992, 5 (4), 297–323.

van Rooij, Maarten, Annamaria Lusardi, and Rob Alessie, “Financial Literacy and Stock Market Participation,” *Journal of Financial Economics*, 2011, 101 (2), 449–72.

Wagenaar, William M. and Sabato D. Sagaria, “Misperception of Exponential Growth,” *Perception and Psychophysics*, 1975, 18 (6), 416–22.

Walstad, William B., Ken Rebeck, and Richard A. MacDonald, “The Effects of Financial Education on the Financial Knowledge of High School Students,” *Journal of Consumer Affairs*, 2010, 44 (2), 336–57.