Evaluating Deliberative Competence:
A Simple Method with an Application to Financial Choice

February 2021

Sandro Ambuehl
Department of Economics,
University of Zurich

B. Douglas Bernheim
Department of Economics,
Stanford University

Annamaria Lusardi
Global Financial Literacy Excellence
Center (GFLEC), The George Washington
University School of Business

Abstract:

We introduce a method for experimentally evaluating interventions designed to improve the quality of choices in settings where people imperfectly comprehend consequences. Among other virtues, our method yields an intuitive sufficient statistic for welfare that admits formal interpretations even when consumers suffer from biases outside the scope of analysis. We use it to study a financial education intervention, which we find improves the quality of decisions only when it incorporates practice and feedback, contrary to the implications of analyses based on conventional efficacy metrics. We trace the failures of conventional metrics to violations of assumptions that our method avoids.

GW | Business

www.gflec.org

Evaluating Deliberative Competence:

A Simple Method

with an Application to Financial Choice

Sandro Ambuehl, B. Douglas Bernheim, and Annamaria Lusardi[*]

February 18, 2021

Abstract

We introduce a method for experimentally evaluating interventions designed to improve the quality of choices in settings where people imperfectly comprehend consequences. Among other virtues, our method yields an intuitive sufficient statistic for welfare that admits formal interpretations even when consumers suffer from biases outside the scope of analysis. We use it to study a financial education intervention, which we find improves the quality of decisions only when it incorporates practice and feedback, contrary to the implications of analyses based on conventional efficacy metrics. We trace the failures of conventional metrics to violations of assumptions that our method avoids.

1 Introduction

Mounting evidence documents the prevalence of low-quality decision making in an assortment of policy-relevant domains, such as household finance (Beshears et al., 2018), health insurance (Loewenstein et al., 2013), and the consumption of durable goods (Grubb and Osborne, 2015; Allcott and Taubinsky, 2015). The field of Behavioral Public Economics aims to evaluate economic policies in light of such decision-making imperfections (for a review, see Bernheim and Taubinsky, 2018). One branch of the field studies standard economic policies that alter consumers' opportunities, such as the taxes and subsidies. Research within this first branch shows, for example, how governments can use these policy instruments to correct *internalities* (see for example Bernheim and Rangel, 2009; Gruber and Kőszegi, 2004; Farhi and Gabaix, 2020). A second branch focuses on *opportunity-neutral* interventions that influence decisions without changing the available options. Such interventions include (1) programs that seek, through education and training, to equip decision makers with the knowledge and skills necessary to process and interpret naturally occurring information, (2) efforts to "nudge" people toward better decisions through defaults, reminders, and other various alterations in the framing of their decision tasks (also known as *choice architecture*, Thaler and Sunstein, 2009) and (3) measures that influence decision making by altering the nature or frequency of personal interactions with professional advisers, family members, or peers.

In this paper, we develop and apply a new method for experimentally evaluating the impact of opportunity-neutral interventions on the *quality of decision making*. Following other recent research in Behavioral Public Economics, we assess decision-making quality based on the prevalence and severity of identifiable mistakes (see Bernheim and Taubinsky, 2018). In this literature, the notion of mistake involves a critical distinction between an option and the preference-relevant outcomes it induces. For example, people make choices over financial portfolios but care about the patterns of returns those portfolios generate, rather than the portfolios themselves. Likewise, people make choices among over-the-counter medications but care about the health consequences those products yield, rather than the medications themselves. We will refer to such choice options as *instruments* because they matter instrumentally rather than intrinsically. The distinction between an instrument I and its intrinsically valued consequences y is important because it allows for the possibility that people may not understand the mapping $Y(I)$ from instruments to consequences, a phenomenon known as *characterization failure*.[1]

In formulating our approach to assessing the quality of decision making, we address a critical unresolved conceptual concern that arises in virtually all existing applications of Behavioral Welfare Economics: When evaluating policies designed to treat a class of decision-making defects, how should one factor in second-

---

[1]One should think of the function Y as conditional on the information that is available to the consumer. To understand why, imagine that an asset is worth \$100 when the economy is good and \$50 when it is bad, where the two states occur with equal probability. If the consumer does not know the state, then Y maps the purchase of the asset to a 50-50 lottery between \$50 and \$100. But if the consumer knows the state is good, then Y maps the purchase of the asset to \$100.

best considerations arising from the potential existence of other defects, known and unknown? The fields of Psychology and Behavioral Economics have identified a wide assortment of choice anomalies commonly interpreted as cognitive biases, and evidence of novel biases emerges periodically. Focusing on a single bias while ignoring others potentially generates misleading conclusions concerning welfare. For example, as noted by Blumenthal and Volpp (2010), accurate information about the caloric content of food can exacerbate excessive consumption if people tend to overestimate calories.[2] A handful of papers have made some progress in this direction by modeling two biases at a time (e.g. Allcott et al., 2014, 2018), but the general methodological concern remains.

Our method builds on the analytic techniques surveyed by Bernheim and Taubinsky (2018). Evaluating welfare effects in settings with potential characterization failure requires one to identify a *welfare-relevant choice domain*, defined as the collection of decision settings for which such failures are absent, and to infer notions of "better" and "worse" from choices within that domain. Sometimes the analyst accomplishes this objective by creating two objectively identical decision problems, one with all the complexity of a naturally occurring decision, the other simplified to make the consequences transparent. The welfare-relevant domain is assumed to consist of the second type of decisions, which the analyst uses to evaluate choices of the first type.[3]

More specifically, our method involves assessing the severity of characterization failures through a controlled experiment in which we present subjects with *paired valuation tasks*. We assess reservation valuations for an instrument I when its description is naturalistic, in the sense that we identify it as I and leave the potentially complex problem of inferring $Y(I)$ to the consumer, and also when its description is limited to a simple and transparent explanation of its consequences $(y = Y(I))$. The magnitude of the difference between these valuations constitutes our measure of deliberative competence. Intuitively, if complex faming poses no barrier to comprehension of an option's consequences, then we should observe the same reservation valuation as with simple framing. Any discrepancy reflects a misunderstanding of the instrument's consequences when described in the complex frame. Interventions that mitigate characterization failures should reduce that discrepancy.

Our main theoretical results establish that our measure of deliberative competence is a *robust sufficient statistic* for the dollar-equivalent welfare loss a consumer suffers due to characterization failure when making her decision in the complex frame (see Chetty, 2009, for a discussion of the sufficient statistics approach). To draw formal conclusions about robustness, we introduce the notion of *idealized welfare analysis* which, we argue, provides an appropriate and useful strategy for attacking normative questions in settings with multiple (and in some cases unknown or poorly understood) sources of inefficiency. We prove that, under

---

[2]Likewise, according to Downs et al. (2009), "dietary information is likely to improve self-protective behavior only if existing biases encourage unhealthy eating, but the reverse is equally likely. When it comes to smoking, for example, there is evidence that smokers tend to overestimate the health risks, in which case providing risk information could undermine their motivation to quit."

[3]Bernheim and Rangel (2009); Bernheim (2016); Bernheim and Taubinsky (2018) outline the general strategy. For applications, see, e.g., Chetty et al. (2009); Bertrand and Morse (2011); Kling et al. (2012); Atanasov and Baker (2014); Abeler and Jäger (2015); Allcott and Taubinsky (2015); Kalayci and Serra-Garcia (2016); Bhargava et al. (2017b)

reasonably broad conditions, our method yields idealized welfare measures that are robust with respect to large classes of additional decision-making defects, both known and unknown, outside the scope of analysis. Thus, an intervention enhances welfare in this robust sense if an only if it improves measured deliberative competence.

The Deliberative Competence Method offers other notable advantages. The fact that it directly yields a sufficient statistic for the welfare effects of an opportunity-neutral intervention means that it entirely avoids the need to interpret choice patterns through the lens of a structural theory, and likewise the strong assumptions that routinely accompany such interpretations.[4] Significantly, it yields these measurements of welfare losses at the *individual level*, and consequently can easily accommodate population heterogeneity with respect to preferences and decision-making defects. As Taubinsky and Rees-Jones (2018) have shown, behavioral welfare analyses that neglect heterogeneity are potentially subject to severe measurement biases. Accommodating heterogeneity is particularly straightforward when using our method.

Our method also addresses another problematic possibility that arises in behavioral policy evaluation: an intervention that seeks to reduce characterization failure may also change behavior through *confounding framing effects*. For example, educating consumers about sales taxes to improve their understanding of tax-inclusive prices may also trigger aversive reactions to tax payments. In that case, the behavioral response to the intervention exaggerates the impact of improved comprehension, rendering welfare calculations potentially misleading. Our method yields a simple test for confounding framing effects, as well as a straightforward corrective adjustment to be used when such effects are present.

We demonstrate the practical value of our method through an application involving the quality of financial decision making and the efficacy of educational interventions intended to improve those decisions.[5] Specifically, we examine intertemporal allocation problems that require an understanding of compound interest. The educational interventions closely follow a popular text on investing (Malkiel and Ellis, 2013) and differ from each other only in that one provides practice with individualized feedback, while the other does not. We elicit reservation valuations for complexly framed financial instruments, I, such as the following: the proceeds from $5 invested for 36 days at an interest rate of 2 percent per day, compounded daily. We also elicit the same subjects' reservations valuations for equivalent simply framed financial instruments, $y = Y(I)$ (i.e., for the preceding example, roughly $20 received in 36 days), without informing them that these prospects are substantively equivalent. Our method instructs us to evaluate deliberate competence with respect to compound interest by calculating the discrepancy between a subject's reservation valuations for the two prospects.

---

[4]For a recent example, see Goldin and Reck (2020). For a review of structural behavioral economics, see DellaVigna (2018).

[5]While there is a substantial literature on the effects of financial education, most of it focuses on measuring changes in behavior rather than on the quality of decision making; see (Duflo and Saez, 2003; Bayer et al., 2009; Mandell, 2009; Cole and Shastry, 2010; Bertrand and Morse, 2011; Cole et al., 2011; Skimmyhorn, 2012, 2016; Servon and Kaestner, 2008; Collins, 2013; Drexler et al., 2014; Carlin et al., 2014; Heinberg et al., 2014; Goda et al., 2014; Luhrmann et al., 2018; Lusardi et al., 2015; Song, 2015; Bruhn et al., 2016; Urban et al., 2018; Sutter et al., 2020). Exceptions primarily include studies, discussed below, that evaluate financial literacy or draw inferences about decision quality from directional effects on behavior.

We also evaluate the interventions using more conventional metrics. Traditionally, assessments in the domain of household financial decision making take one of two forms (see Hastings et al., 2013; Lusardi and Mitchell, 2014; Beshears et al., 2018; Kaiser et al., 2020, for reviews). The first is to investigate whether an intervention changes behavior in a direction that counteracts a suspected bias. For example, proceeding from the assumption that some people do not save enough, some studies ask whether financial education leads them to save more (Bernheim et al., 2001; Bernheim and Garrett, 2003; Cole and Shastry, 2010; Bruhn et al., 2014, 2016).[6] Limitations of that method include the challenge of justifying the attribution of consumers' choices to biases rather than to their individual preferences (which may differ from the analyst's judgments), the associated possibility that interventions may influence choices through indoctrination, deference to authority, or social pressure, and the difficulty of determining whether interventions induce consumers to overcorrect. A second approach is to assess comprehension of the principles that govern relationships between actions and opportunities using batteries of exam-style questions. Assessments of financial literacy fall into this category (Council for Economic Education, 2006; Jump$tart Coalition for Personal Financial Literacy, 2006; Mandell, 2009; Mandell and Klein, 2009; Walstad et al., 2010; Carpena et al., 2011; Collins, 2013; Heinberg et al., 2014; Lusardi et al., 2015; Bruhn et al., 2014, 2016). A limitation of these assessments is that they cannot reveal whether people deploy their knowledge, either correctly or at all, in practical decisions. Neither of these methods measures the welfare losses from poor decision making directly, and they permit indirect inferences about the quality of choice only under strong assumptions, which we identify and test.[7]

Using our method, we find that the educational intervention improves the average quality of decision making when it includes practice and feedback, but not when it omits these components. As a practical matter, this finding suggests that incorporating practice and feedback can have a dramatic impact on the efficacy of financial education, even when the feedback is fully automated and relatively simple. Crucially, according to our theoretical analysis, one can interpret these findings as pertaining to the idealized welfare effects of the treatments, which means our conclusions are robust with respect to the possible existence of other biases (the leading possibilities being present bias and misconceptions about the likelihood of experimenter follow-through on future payments). In contrast, the conventional metrics incorrectly suggest that interventions with and without practice and feedback are both effective, and that their benefits are roughly the same. In particular, both interventions substantially increase subjects' valuations for compound-interest-bearing assets, apparently counteracting *exponential growth bias*, the well-documented tendency to underestimate

---

[6]Other directional biases include the tendency to choose low-deductible health insurance plans (Sydnor, 2010; Loewenstein et al., 2013), naïve diversification (Benartzi and Thaler, 2001), and investors' tendency to exit the market in downturns (Bennyhoff and Kinniry Jr, 2011).

[7]Bernheim and Taubinsky (2018) also discuss two other approaches to evaluating the quality of decision making. One is to examine the frequency of either dominated or dominant choices (Ernst et al., 2004; Calvet et al., 2007, 2009; Agarwal et al., 2009; Baltussen and Post, 2011; Choi et al., 2011; Aufenanger et al., 2016; Bhargava et al., 2017a). One limitation of this method is that it can remove personal preferences from the mix, deactivating mechanisms such as motivated reasoning that can degrade the quality of choice; a second is that a reduction in the frequency of dominated choices is neither necessary nor sufficient for a welfare improvement (except when the efficient frontier is degenerate); a third is that true dominance is difficult to establish in practical settings that potentially implicate preferences. Another approach is to measure the frequency of WARP or GARP violations (Afriat, 1972; Choi et al., 2014; Echenique et al., 2011). A limitation of this method is that it cannot detect consistent errors (e.g., mistaking apples for oranges). Nor can it accommodate the possibility that inconsistencies reflect the vagaries of preference construction.

exponential growth (Wagenaar and Sagaria, 1975; Eisenstein and Hoch, 2007; Stango and Zinman, 2009; Almenberg and Gerdes, 2012; Levy and Tasoff, 2016). Furthermore, both improve measured financial literacy concerning compound interest.

We trace the failure of the conventional metrics to violations of the assumptions on which those metrics depend. First, absent practice problems, the intervention increases subjective valuations of interest-bearing assets across the board, in spite of the fact that there is substantial heterogeneity in the extent of subjects' initial bias. As a consequence, it leads some subjects to overcorrect for their initial biases, and it yields greater bias among those who initially overestimate exponential growth.[8] Second, absent practice and feedback, the intervention's effect on test-performance is entirely disconnected from its effect on behavior, as we show through two treatments. For one treatment, we remove most of the substantive material from the intervention and maintain only its rhetorical elements (e.g. "Albert Einstein is said to have called compound interest the most powerful force in the universe"). The resulting intervention largely reproduces the behavioral effect of the full intervention, but has almost no effect on test performance. For the second treatment, we retain the substantive material and remove the rhetoric. That treatment roughly reproduces the effect of the original intervention on measured financial literacy, but largely fails to alter behavior.[9] Thus, in the absence of practice and feedback, the effects of the educational intervention on financial literacy arise primarily from the substantive elements of instruction, but its behavioral effects arise primarily from the rhetorical elements. These findings explain why the intervention can enhance measured financial literacy and shift behavior in the "desired" direction without actually improving the quality of decision making.

While our method produces sharp conclusions that depend on few assumptions, the fact that we diagnose decision quality based on choices in experimental settings raises potential issues. Experimental methods are widely used in research on the quality of decision making both in the lab and in the field (see, e.g. Bernheim and Taubinsky, 2018; Beshears et al., 2018, for reviews). They provide the ability both to control the environment tightly and to collect the type of detailed measurements required for rigorous inference about characterization failures. As a result, they usefully complement research based on naturally occurring data, which typically lacks these features. A reasonable question is whether, with larger real-life stakes, people might deliberate more carefully, and might be more likely to employ analytic tools or seek advice. In general, Enke et al. (2020) provide direct evidence that raising the stakes for decisions in experiments to levels that equal or exceed subjects' monthly wages does little to mitigate cognitive biases. Focusing on the specific experiment in this paper, a body of empirical evidence on large-stakes financial decisions suggests that real-life stakes often fail to induce more careful decision making. A survey of non-faculty staff at the University of Southern California revealed that 58 percent of respondents spent under one hour – less than

---

[8]Goda et al. (2015) and Levy and Tasoff (2017) also document considerable heterogeneity with respect to the perceived benefits of compounding. Relatedly, Harrison et al. (2020) finds that increased take-up of seemingly beneficial products (index insurance) can lead to welfare losses due to subject heterogeneity.

[9]Related dissociations have been observed in other contexts. Enke and Zimmermann (2015) shows that many people tend to neglect correlations in decision making, despite knowing how to account for them when prompted explicitly. Taubinsky and Rees-Jones (2018) find that many consumers underreact to excise taxes, even though they can properly compute tax-inclusive prices when prompted to do so.

the duration of a typical behavioral experiment – determining their contribution rates and asset allocations for their retirement savings (Benartzi and Thaler, 1999). Most people do not consult with anyone other than friends and family members, and make their financial decisions without assistance (Bernheim, 1998; Benartzi and Thaler, 1999; Lusardi and Mitchell, 2011).[10] Similarly, only a small fraction of the population takes the time to develop an explicit financial or retirement plan (Lusardi and Mitchell, 2011) and informal plans tend to be unsophisticated (Bernheim, 1994).[11] The existence of large default effects in retirement savings (Madrian and Shea, 2001; Bernheim et al., 2015) demonstrates that many people are inattentive, while others engage in simplistic reasoning, even when facing enormously consequential financial decisions. In any case, because our method remains applicable in settings with large stakes, it can facilitate formal explorations of the relationship between stakes and the quality of decision making.

While we focus here on a single application, our method is potentially useful in a much wider assortment of contexts. In other work, we are examining the impact of peer-to-peer communication on the quality of financial decision making (Ambuehl et al., 2018). One can use our method to assess the quality of other types of complex financial choices such as those involving insurance (Loewenstein et al. (2013); Johnson et al. (2013)), loans and leases, and contracts with multi-part tariffs (Grubb and Osborne, 2015), and to gauge the welfare effects of other types of interventions, such as the provision of professional advice. As we have already emphasized, our methods are also applicable outside of the financial domain, for example to the evaluation of efforts to improve the quality of choices involving health or durable goods, or to assess the welfare effects of a wide assortment of "nudges."

The remainder of the paper is organized as follows. Section 2 develops our method and Section 3 applies it to financial choices involving compound interest. Section 4 discusses the implications of our research and concludes.

2 The Deliberative Competence Method

In this section, we describe our method for measuring deliberative competence and explain its interpretation. Subsection 2.1 details the decision-making context and introduces the measurement method. Subsection 2.2 provides formal welfare-theoretic foundations. Subsection 2.3 establishes the robustness of the method to additional misperceptions and biases, known or unknown, outside the scope of analysis. Subsection 2.4 addresses the problem of potentially confounding framing effects and presents our solution. Finally, Subsection 2.5 provides formal foundations for aggregating our measure of deliberative competence over members of a population.

---

[10]The fraction of individuals reporting that they did not seek advice in our experiment (three quarters) matches experience in the field (Lusardi and Mitchell, 2011).

[11]According to Lusardi and Mitchell (2011), a mere 19% of respondents in the Health and Retirement Study had developed a plan for their retirement savings.

2.1 Measures of deliberative competence

Our method of measuring deliberative competence involves simple binary decisions: whether to purchase an instrument I at a price p. We envision a consumer who makes this decision in one of two settings which differ only with respect to the instrument's description. We call these descriptions *frames*. The *complex frame c* describes the instrument in naturalistic terms, and as a result typically requires the consumer to extrapolate consequences based on her understanding of the underlying processes. In contrast, the *simple frame s* involves a direct and transparent description of the consequences. We refer to the pair (I, f) (where $f = c, s$) as a *framed instrument*.

We assume the consumer acts as if she makes her decisions to maximize the utility function $V(m, i, f)$, where m represents any payment she receives (or makes if negative) aside from returns to the instrument, and i indicates whether the consumer chooses to purchase ($i = 1$) or forego ($i = 0$) the instrument. Critically, V is an *indirect* utility function: it subsumes her beliefs in frame f about the preference-relevant consequences she thinks will follow from owning or not owning the instrument (which is not by itself preference-relevant).[12] In the interest of generality, we leave those consequences unspecified for the moment. We assume only that V has a bounded range and that it is strictly increasing and continuously differentiable in its first argument, with $\frac{\partial V}{\partial m}$ bounded away from zero.

In frame f, the consumer purchases the instrument at price p iff:[13]

$$V(-p, 1, f) \geq V(0, 0, f)$$

We can therefore define her reservation valuation in frame f, call it r_f, as follows:

$$V(-r_f, 1, f) = V(0, 0, f) \tag{1}$$

Existence of r_f follows from our assumptions on V.

At the outset, we will assume that characterization failure can occur when the instrument is complexly framed, but does not generally occur when its description transparently states its consequences.[14] It follows that choices in the simple frame reveal welfare-relevant judgments (see Bernheim and Rangel (2009)). Other studies in Behavioral Public Economics make similar assumptions (see for example Chetty et al., 2009; Bertrand and Morse, 2011; Kling et al., 2012; Atanasov and Baker, 2014; Abeler and Jäger, 2015; Allcott

---

[12]Similarly, the consumer cares about the goods she could purchase with any payment she makes or receives rather than the payment itself.

[13]Technically, equality implies indifference.

[14]Under this assumption, the ideal policy is to re-engineer the choice architecture of real-world decision problems so that they are simply framed. However, there may be practical barriers to achieving this ideal. The market may provide firms with incentives for obfuscation that lead them to resist transparency, and that undermine the political viability of collective solutions (Gabaix and Laibson, 2006; Jin et al., 2015). In settings where an instrument's consequences depend on the decision maker's circumstances, it may also be impractical to provide everyone with an explanation that is both transparent and adequately individualized.

and Taubinsky, 2015; Kalayci and Serra-Garcia, 2016; Bhargava et al., 2017b). In Subsection 2.3, we examine the possibility that other failures also contaminate the simply framed choices.

Our method of assessing deliberative competence involves *paired valuation tasks*, which elicit r_s and r_c for, respectively, simply framed and complexly framed versions of the same instrument. Intuitively, if complex faming poses no barrier to comprehension of an option's consequences, then we should observe $r_c = r_s$. Any discrepancy must reflect a misunderstanding of the instrument's consequences when described in the complex frame.

Consistent with this intuition, we define the subject's *deliberative competence* with respect to the framed instrument (I, c) as the negative of the distance between r_s and r_c for some distance metric d.

Definition 1. *A subject's* deliberative competence *regarding framed instrument (I, c) is given by $D(I, c) = -d(r_s, r_c)$.*

Obviously there are many potential candidates for the distance metric. In the next subsection, we exhibit formal welfare-theoretic foundations for using notions of deliberative competence based on either of the following distance metrics: $d(x, y) = (x - y)^2$ or $d(x, y) = |x - y|$.

2.2 Interpretation as formal welfare measures

To conduct formal welfare analysis, one must specify an objective function. We consider two candidates, one of which justifies the distance metric $d(x, y) = (x - y)^2$, the other of which justifies $d(x, y) = |x - y|$. In both cases, we show that $d(x, y)$ is a money-metric measure of the welfare loss from characterization failure in the complex frame.

We begin with a standard welfarist objective function. For this purpose, we assume the price p is distributed according to some CDF H with density h. The planner is concerned with the expected value of the individual's welfare:

$$W(r_f) = \int_0^{r_f} V(-p, 0, s)h(p)dp + \int_{r_f}^{\infty} V(0, 0, s)h(p)dp$$

We can then write the welfare loss of limited deliberative competence as:

$$\mathcal{L}_W(r_c, r_s) = W(r_s) - W(r_c) = \begin{cases} \int_{r_c}^{r_s} \left[V(-p, 0, s) - V(0, 0, s)\right] h(p)dp & \text{if } r_s \geq r_c \\ \int_{r_c}^{r_s} \left[V(0, 0, s) - V(-p, i, s)\right] h(p)dp & \text{if } r_c < r_s \end{cases} \tag{2}$$

To understand this formula, focus on the case of $r_s \geq r_c$. For $p > r_s$, the consumer does not purchase the instrument in either frame, so there is no welfare loss. Similarly, for $p < r_c$, she purchases it in both frames, so again the loss is zero. However, when $p \in [r_c, r_s]$, she purchases the instrument in the complex frame but not in the simple frame. Evaluating those outcomes based on $V(\cdot, \cdot, s)$ (because the Welfare-Relevant Domain consists of simply framed choices), we infer that the welfare loss is $V(-p, 0, s) - V(0, 0, s)$, which

we integrate to obtain the total loss. The case of $r_s < r_c$ is analogous, except that the consumer purchases the instrument in the simple frame but not in the complex frame for $p \in [r_s, r_c]$.

The welfarist approach assumes H, the probability distribution of p, is known. While it is known for any given experiment, it is hard to say which prices would be relevant to decision making about related instruments in real decisions. Treating the probability distribution as unknown, we can employ an alternative approach, widely used in Computer Science, that has recently found its way into Economic applications: evaluate an option based on the maximum possible loss it could induce.[15] The corresponding loss function is:

$$\mathcal{L}_M(r_c, r_s) = \begin{cases} \max_{p \in [r_c, r_s]} [V(-p, 1, s) - V(0, 0, d)] & \text{if } r_s \geq r_c \\ \max_{p \in [r_s, r_c]} [V(0, 0, s) - V(-p, 1, s)] & \text{if } r_c \geq r_s \end{cases} \tag{3}$$

The intuition for the form of this loss function is the same as for equation (2), except that we evaluate the maximum value of the loss rather than its integral.

Both $\mathcal{L}_W$ and $\mathcal{L}_M$ express the welfare loss in terms of "utils." To obtain money-metric measures of the loss, we can examine $\frac{1}{\nu_s}\mathcal{L}_W$ and $\frac{1}{\nu_s}\mathcal{L}_M$, where $\nu_s \equiv \frac{\partial}{\partial m} V(-r_s, 1, s)$ is the marginal utility of money for the scenario in which the consumer purchases the instrument in the simple frame at the price $p = r_s$. For notational convenience, we also define $\eta_s = h(r_s)$.

Having defined the social objectives, we can now state our first proposition, which provides formal welfare-theoretic foundations for our measures of deliberative competence under the assumption that simply framed choices are free of characterization failure. All proofs appear in Appendix A.

Proposition 1. *(i) Let $L_M(r, r_s) \equiv |r - r_s|$. To a first-order approximation around $r = r_s$,*

$$\frac{1}{\nu_s}\mathcal{L}_M(r, r_s) \approx L_M(r, r_s)$$

(ii) Let $L_W(r, r_s) \equiv (r - r_s)^2$. To a second-order approximation around $r = r_s$,

$$\frac{1}{\nu_s}\mathcal{L}_W(r, r_s) \approx \frac{\eta_s}{2} L_W(r, r_s)$$

We use a second-order approximation for $\frac{1}{\nu_s}\mathcal{L}_W(r, r_s)$ rather than a first-order approximation because the first-order term is identically equal to zero. To understand the intuition for Proposition 1, note that a consumer whose valuations differ across the frames overpays for the instrument by $p - r_s$ if $r_c \geq p \geq r_s$, misses out on net value of $r_s - p$ if $r_c \leq p \leq r_s$, and suffers no loss if p does not lie between r_c and r_s. Hence, the magnitude of any non-zero dollar-equivalent loss equals $|r_s - p|$. For part (i), regardless of the relative ranking of r_s and r_c, the dollar-equivalent loss reaches a maximum at $p = r_c$, and the maximized value is $|r_c - r_s|$. For part (ii), if the distribution of p is uniform with density η_s and a support that includes

---

[15]See, for example, Carroll (2019), who adopts this approach to study optimal mechanism design in settings where the distribution of agent characteristics is unknown. Other applications include decision theory (Machina and Siniscalchi, 2014) and public policy (Sunstein, 2020).

both r_c and r_s, the expected loss or expected foregone gain equals $\int_{\min\{r_c,r_s\}}^{\min\{r_c,r_s\}} \eta_s \, |r_s - p| \, dp = \frac{\eta_s}{2} \left(r_c - r_s \right)^2$.[16]
Our claim concerning the second-order approximation follows from the fact that any probability distribution with a continuous PDF is approximately uniform over a sufficiently narrow slice of its support.

Despite the fact that $\frac{1}{\nu_s}\mathcal{L}_W$ and $\frac{1}{\nu_s}\mathcal{L}_M$ reflect different approaches to conceptualizing the loss function, they admit similar welfare approximations, inasmuch as $\left(r_c - r_s \right)^2$ is simply the square of $|r_c - r_s|$. While $\frac{1}{\nu_s}\mathcal{L}_W$ reflects a more conventional approach, there are two reasons to prefer $\frac{1}{\nu_s}\mathcal{L}_M$, one practical and one conceptual. The practical advantage of $\frac{1}{\nu_s}\mathcal{L}_M$ is that $|r_c - r_s|$ is less sensitive to outliers than $\left(r_c - r_s \right)^2$. The conceptual advantage of $\frac{1}{\nu_s}\mathcal{L}_M$ is that, unlike $\frac{1}{\nu_s}\mathcal{L}_W$, it is interpretable as a dollar-value measure of welfare loss, rather than as an unspecified rescaling of a dollar-value measure, when the price distribution H (and in particular of the density η_s) is unknown.[17]

2.3 Welfare analysis with multiple biases

The fields of psychology and behavioral economics have identified a wide assortment of possible cognitive biases. This observation raises a central methodological question for methods that aim to measure and improve decision quality: when evaluating policies designed to treat one class of decision-making errors, how should one factor in other types of errors?

The most common approach, which we call *partial welfare analysis*, is to focus on remedies for a single bias under the simplifying assumption that the decision-making apparatus is otherwise faultless. Examples include O'Donoghue and Rabin (2006); Gruber and Kőszegi (2001, 2004); DellaVigna and Malmendier (2004). Unfortunately, welfare analyses that abstract from the pervasiveness and multiplicity of framing effects and biases may overlook critical second-best considerations (in the sense of Lipsey and Lancaster, 1956) that could overturn their implications.

A second strategy is to model the interactions among more than one source of bias. To our knowledge, existing applications focus on two biases at a time; see Allcott et al. (2014); Farhi and Gabaix (2020). While these analyses can highlight important second-best considerations, they remain susceptible to the same criticism as partial welfare analysis, in that they likewise assume the decision making apparatus is otherwise faultless. To overcome the concerns pertaining to biases outside the scope of analysis, one would have to ensure that the analysis is *comprehensive*. While the theoretical framework developed by Farhi and Gabaix (2020) accommodates comprehensive welfare analysis, they acknowledge that empirical implementation with multiple biases is a "momentous task" (p. 311). Unfortunately, the field of Behavioral Economics remains far from delivering a comprehensive empirical model of human behavior usable for such purposes.

---

[16]Formal justifications for the Harberger triangle (Harberger, 1964) involve a related argument.

[17]It is worth emphasizing that $\frac{1}{\nu_s}\mathcal{L}_W$ remains useful even when H and η_s are unknown. Specifically, in light of the fact that $L_W(r_c, r_s) \propto (r_c - r_s)^2$, we know that $\left(r_c - r_s \right)^2$ approximates $\frac{1}{\nu_s}\mathcal{L}_W$ up to a multiplicative scalar. It follows, for example, that if we are asked to evaluate competing policies that aim to equip consumers with tools for interpreting complexly framed instruments, we can rank their welfare benefits using the loss function $\left(r_c - r_s \right)^2$, and even gauge the percentage differences between the dollar-equivalents of those benefits. What we cannot do is assess the absolute levels of their welfare benefits. In contrast, $|r_c - r_s|$ approximates $\frac{1}{\nu_s}\mathcal{L}_M$ without a factor of proportionality.

Simultaneous consideration of multiple biases raises an additional issue: if the set of policy instruments considered is insufficiently broad, this strategy can produce misleading conclusions.To illustrate, suppose consumers initially underestimate the effects of interest compounding (exponential growth bias, see, e.g., Stango and Zinman, 2009; Almenberg and Gerdes, 2012; Levy and Tasoff, 2016). Imagine in addition that the government could eliminate this bias by adopting a financial education program, E. Ignoring the possibility that consumers suffer from other biases, the program is plainly beneficial. But what if consumers also suffer from "debt aversion," in the sense that they avoid using attractive credit instruments such as student loans even when the potential benefits of borrowing are substantial (Callender and Jackson, 2005; Rasmussen, 2006)? Considering all sources of inefficiency, the policy is likely harmful. Indeed, formal welfare analysis might favor an alternative "educational" intervention, D, that misleads consumers into underestimating compound interest even further. (As we noted in Section 1, policy prescriptions based on similar reasoning appear in the literature on nutrition and health.)

Yet this prescription hinges on a conceptual error: the analysis attempts to treat sources of inefficiency comprehensively, but does not treat policy options comprehensively.[18] Distorting policies that target consumers' understanding of compound interest in order to address concerns arising from debt aversion makes little sense if other policy tools are better suited for the latter purpose. Here, the optimal comprehensive policy might consist of E combined with measures that subsidize certain types of borrowing or remove artificial psychological barriers to indebtedness.

We propose an alternative strategy for addressing second-best considerations arising from the existence of additional biases, known and unknown, outside the scope of analysis. The essence of this strategy, which we call *idealized welfare analysis*, is to envision a compartmentalized approach to correcting decision making imperfections. The following analogy illustrates the main idea. Picture a group of engineers tasked with redesigning a poorly functioning satellite launch system. Because the task consists of many challenging components, the group splits them up by creating separate teams for propulsion, roll control, guidance, and so forth. Each team operates on the assumption that the others will perform their tasks properly. They do not attempt to compensate for pre-existing deficiencies in the other sub-systems. Combining the compartmentalized solutions yields a well-functioning system. Similarly, an economic study can contribute to the ultimate objective of improving human decision making by devising strategies for ensuring that a given decision-making "sub-system" performs optimally on the assumption that other studies will identify strategies for optimizing the performance of other "sub-systems." In the context of the preceding example, one can imagine an effort aimed at ensuring a proper understanding of compound interest (E), and a separate effort aimed at addressing debt aversion. By achieving all compartmentalized objectives, we arrive at a medley of local solutions that together achieve the overall optimum. Even if we do not yet understand

---

[18]One could also object to policy D based on concerns about the ethics of spreading misinformation, or about the government's long-term credibility. Those considerations are orthogonal to our current focus.

how to address all the compartmentalized objectives, this strategy allows us to make incremental progress toward a global solution.[19]

It is important to understand the essential distinction between partial welfare analysis and idealized welfare analysis. In the context of the binary choices described in Subsection 2.1, partial welfare analysis treats the simply framed choices one actually observes as welfare-optimal. Accordingly, the objective is to maximize $V(m, i, s)$. In sharp contrast, for idealized welfare analysis, the objective is to maximize some other function, $U(m, i, s)$, that rationalizes the simply framed choices the consumer would make *if all other decision-making defects were corrected.*

At first, it might appear that idealized welfare analysis is impractical. After all, observed choices reveal V, not U. Fortunately, however, there are surprisingly broad conditions under which V can serve as a legitimate stand-in for U, in the sense that it allows one to measure the (approximate) welfare effects of alternative policies up to a multiplicative scalar that depends on other biases. To put the matter differently, partial welfare analysis is actually robust with respect to the possible existence of a broad class of additional defects, known and unknown, if it employs our method (but not necessarily otherwise).

In the remainder of this section, we develop these ideas formally. We begin with an example and then explore its generality.

2.3.1 An example

A consumer cares about current consumption, m, and a vector of future outcomes, y. We assume she acts as if she maximizes the expectation of the utility function $u(m) + \gamma v(y)$, where $u(m) = \frac{m^{1-\sigma}-1}{1-\sigma}$.

The consumer starts with current income m_0 and must decide whether to purchase an instrument, I. For simplicity, she spends all current income immediately, so that $m = m_0 - p$ if she purchases I, and $m = m_0$ otherwise. Without the instrument, her future outcomes are y_0. The instrument changes these outcomes to y_s. With naturally occurring (complex) framing, she believes the outcomes will be y_c. We are concerned with measuring the welfare costs of this misperception and the associated benefits of interventions that potentially bring y_c more closely in line with y_s.

The potential existence of additional biases complicates our task. First, the consumer may discount the future excessively due to "present bias." Correcting this cognitive cognitive imperfection would reveal a discount factor $\delta = \frac{\gamma}{\beta}$ rather than γ (where $\beta < 1$ is the present-bias parameter). Second, even though the instrument is perfectly reliable, the consumer falsely believes it will have no impact with probability $1 - \pi$. We incorporate these particular biases into this example because they may be present in our experiment.

Because present bias, excessive skepticism concerning reliability, and discrepancies between y_s and y_c all impact the demand for the instrument, these biases plainly interact. However, from the perspective of idealized welfare analysis, we take the view that the best approach is to compartmentalize the three problems,

---

[19]Like all other approaches, idealized welfare analysis involves compromises. If there is no good way to address a source of inefficiency outside the scope of analysis, the approach will overlook potentially important second-best considerations.

devising separate remedies for each. When focusing on misperceptions of y_s, we imagine that other analyses will eventually provide (but have not yet provided) solutions for the other two biases.

From the preceding assumptions, we infer that

$$V(0, 0, s) = V(0, 0, c) = u(m_0) + \gamma v(y_0),$$

$$V(-p, 1, s) = u(m_0 - p) + \pi \gamma v(y_s) + (1 - \pi) \gamma v(y_0)$$

$$V(-p, 1, c) = u(m_0 - p) + \pi \gamma v(y_c) + (1 - \pi) \gamma v(y_0)$$

To determine the reservation price with simple framing, we can rewrite equation (1) as follows:

$$\frac{1}{1 - \sigma} \left(1 - \frac{r_s^V}{m_0} \right)^{1-\sigma} - 1 = m_0^{\sigma-1} \pi \gamma \left(v(y_0) - v(y_s) \right)$$

(We use the superscript "V" to indicate that a quantity, in this case the reservation price r_s, is evaluated based on the utility function V.) As long as r_s^V is a relatively small share of current resources m_0, we can use the first-order approximation, $\frac{1}{1-\sigma} \left(1 - \frac{r_s^V}{m_0} \right)^{1-\sigma} \approx 1 - \frac{r_s^V}{m_0}$, to obtain $r_s^V \approx \pi \gamma m_0^\sigma \left(v(y_0) - v(y_s) \right)$. Similar reasoning for the complex frame yields $r_c^V \approx \pi \gamma m_0^\sigma \left(v(y_0) - v(y_c) \right)$. Therefore, $L_M(r_c^V, r_s^V) \approx \pi \gamma m_0^\sigma \left| v(y_c) - v(y_s) \right|$.

If we were conducting partial welfare analysis, we would simply use $L_M(r_c^V, r_s^V)$ as our measure of loss. To conduct idealized welfare analysis, we need to measure the loss the consumer would incur if present bias and excessive skepticism were both corrected. In that case, a different utility function, U, would govern the consumer's choices. Under our assumptions,

$$U(0, 0, s) = U(0, 0, c) = u(m_0) + \delta v(y_0),$$

$$U(-p, 1, s) = u(m_0 - p) + \delta v(y_s)$$

$$U(-p, 1, c) = u(m_0 - p) + \delta v(y_c)$$

Reasoning as before, we infer that $L_M(r_c^U, r_s^U) \approx \delta m_0^\sigma \left| v(y_c) - v(y_s) \right|$.

The main challenge associated with conducting idealized welfare analysis is that we want to measure $L_M(r_c^U, r_s^U)$, but the data only allow us to measure $L_M(r_c^V, r_s^V)$ directly. Notice, however, that the preceding analysis implies

$$L_M(r_c^U, r_s^U) \approx K L_M(r_c^V, r_s^V)$$

where $K = \frac{\delta}{\pi \gamma} = \frac{1}{\pi \beta}$. The same property holds for $L_W(r_c^U, r_s^U)$ and $L_W(r_c^V, r_s^V)$, except that K^2 is the factor of proportionality. Several conclusions follow immediately.

First, to the extent we can measure the parameters governing the other biases, we can convert the partial welfare-analytic objective function into the idealized welfare-analytic objective function by making a simple

multiplicative adjustment. It is important to emphasize that this is a special property of the welfare measures we derive from paired valuation tasks. We have no reason to think that other welfare measures share this analytically attractive feature. Notice also that we obtain idealized welfare measures without needing to estimate the curvature of the underlying functions v and u.

Second, because the critical factor of proportionality is independent of both the misperception and the instrument, we can conduct useful welfare analysis *even when we do not have estimates of the parameters governing other biases.* Suppose our objective is to evaluate a collection of policy proposals designed to mitigate a particular type of characterization failure, such as a mistaken understanding of the payoffs that flow from instruments paying compound interest. Imagine in addition that we have identified a representative collection of those instruments, $\mathcal{I}$. For any member of this class, $I \in \mathcal{I}$, let y_{sI} denote the actual consequences of instrument I and $y_{cI\theta}$ denote the perceived consequences with complex framing under policy θ. Using our approximation, we know that, for all I and θ, we have $L_M\left(r_{cI\theta}^U, r_{sI}^U\right) \approx K L_M\left(r_{cI\theta}^V, r_{sI}^V\right)$ for the same value of K (where we have supplemented the subscripts of the reservation prices to indicate the instrument and policy). Suppose we are interested in aggregating $L_M\left(r_{cI\theta}^U, r_{sI}^U\right)$ over instruments to obtain an overall measure of the welfare impact, rather than an instrument-specific measure. If we think all members of $\mathcal{I}$ are equally representative, we might simply take the average of $L_M\left(r_{cI\theta}^U, r_{sI}^U\right)$. More generally, we could take a weighted average using weights λ_I, which could reflect judgments about each instrument's relevance for real-world choices.[20] Then for any two policies θ and θ', we have

$$\frac{\sum_{I \in \mathcal{I}} \lambda_I L_M\left(r_{cI\theta}^U, r_{sI}^U\right)}{\sum_{I \in \mathcal{I}} \lambda_I L_M\left(r_{cI\theta'}^U, r_{sI}^U\right)} \approx \frac{\sum_{I \in \mathcal{I}} \lambda_I L_M\left(r_{cI\theta}^V, r_{sI}^V\right)}{\sum_{I \in \mathcal{I}} \lambda_I L_M\left(r_{cI\theta'}^V, r_{sI}^V\right)} \tag{4}$$

An identical statement follows for L_W.[21] The preceding equation implies that we can rank these policies according to the idealized objective function $U\left(\cdot, \cdot\right)$, which is (currently) unknown and unobservable, by ranking them according to the observed choice-rationalizing objective function $V\left(\cdot, \cdot\right)$. We can even gauge the percentage differences between the dollar-equivalents of the benefits flowing from these policies.

Third, $L_M\left(r_{cI\theta}^V, r_{sI}^V\right)$ and $L_W\left(r_{cI\theta}^V, r_{sI}^V\right)$ remain valid measures of idealized welfare (up to an unknown multiplicative scalar) even when the analyst is unaware that the other biases exist. Consequently, our approach is robust with respect to the existence of certain types of unspecified biases, whether known and unknown, outside the scope of analysis. To be clear, we contend only that the welfare interpretation of our deliberative competence measure is usefully robust rather than universally valid. One of our objectives in the next subsection is to determine the theoretical limits of its robustness.

---

[20] As a default, we take all instruments within $\mathcal{I}$ to be equally relevant, and hence all λ_I to be identical.

[21] Because we are ultimately interested in approximating $\mathcal{L}_W$, the weights λ_I used for L_W would presumably vary across instruments to reflect variation in the value of the density parameter η_s. One can ignore this consideration if there is no reason to suspect systematic variation in η_s across instruments.

2.3.2 The general result

To generalize the conclusions reached in the preceding example, we supplement the parsimonious model of decision making of Section 2.1 with additional structure.

As in the example, we use y to denote the vector of consequences other than current consumption falling within the scope of the consumer's concerns. If she foregoes the instrument, she anticipates the consequences y_0. In frame f, she associates the instrument I with consequences y_f. Accordingly, she thinks the instrument changes her consequences by $y_f - y_0$. The characteristics of the instrument and the frame only impact the consumer's choices insofar as they determine these perceived consequences.

For our current purposes, it is convenient to write the choice-rationalizing utility function, V, as a function of m and y, allowing framing (s versus c) and expenditures on the instrument (p) to enter through these two channels. We will assume preferences are separable between m and y – in other words, that the utility function has the form $V(m, \phi(y))$. The reservation value for frame f given the objective function V is then given by the solution r_f^V to the following equation:

$$V\left(-r_f^V, \phi(y_f)\right) = V(0, \phi(y_0))$$

While separability is a restrictive assumption, this setting nevertheless subsumes a wide range of important possibilities. For instance, if we interpret m as current consumption and y as the vector of consumption for future periods, then standard formulations of intertemporal preferences, such as $V(m, \phi) = u(m) + \beta\phi$ and $\phi(y) = \sum_{t=1}^{T} \delta^t u(y_t)$ (quasihyperbolic discounting) satisfy the separability condition. Similarly, if we interpret m as current consumption and y as the vector of future consumption for various states of nature, then various representations of risk preferences, such as $V(m, \phi) = u(m) + \delta\phi$ and $\phi(y) = \sum_{s=1}^{T} \pi_s v(y_s)$, where π_s is the probability of state s (expected utility), also respect this restriction.[22] Alternatively, if we interpret m as cash and y as a vector of health consequences, then separability is satisfied as long as the elements of y contribute to perceived well-being through a composite health good, $\phi(y)$.

Next we assume that, if all other biases were corrected, the individual would make decisions according to the utility function $U(m, \phi(y))$. Notice that, in addition to imposing the same separability condition on U, we also assume that the subutility function ϕ is the same as for V. Consider the example of quasihyperbolic discounting. If one takes the view that a value of β below unity reflects present bias, then it may be natural to assume that $U(m, \phi) = u(m) + \phi$. Our assumption concerning U allows for this possibility. Next consider examples involving health applications. The thrust of the assumption is that biases do not affect the aggregation of perceived consequences into the composite health good; rather, they affect the anticipated consequences (y_c versus y_s) and the consumer's tradeoff between cash and health (V versus U). Under these

---

[22]Our assumption also accommodates other forms of risk preferences, such as probability weighting and cumulative probability weighting.

assumptions, the idealized reservation value for frame f is given by the solution r_f^U to the following equation:

$$U\left(-r_f^U, \phi\left(y_f\right)\right) = U\left(0, \phi\left(y_0\right)\right)$$

Proposition 1 tells us that $L_i\left(r_c^V, r_s^V\right)$ and $L_i\left(r_c^U, r_s^U\right)$ are approximations of dollar-equivalent welfare losses for V and U, respectively (where $i = M, W$ references the type of loss function). Our objective is to understand the relationship between $L_i\left(r_c^V, r_s^V\right)$ and $L_i\left(r_c^U, r_s^U\right)$.

As in the previous subsection, we make use of an approximation. Define

$$y_f^\alpha \equiv \alpha y_f + (1 - \alpha) y_0$$

This formulation allows us to rescale the consequences of the instrument as perceived in the simple and complex frames up and down by the factor α. For notational clarity, we will write the reservation valuations as $r_f^V(y_f^\alpha)$ and $r_f^U(y_f^\alpha)$ to reflect their dependence on the perceived consequences. We will also write the associated loss measures as $L_i^V\left(y_c^\alpha, y_s^\alpha\right)$ and $L_i^U\left(y_s^\alpha, y_s^\alpha\right)$ for $i = M, W$ (where we obtain these functions by substituting the formulas for the reservation values).

The following result shows, in effect, that to a first-order approximation around $\alpha = 0$, L_i^U and L_i^V are related by a factor of proportionality that does not depend on either the instrument I or the misperception of its consequences $y_c - y_s$.

Proposition 2. *There exists a strictly positive constant K, such that for all values of y_s and y_c, we have*

$$\lim_{\alpha \to 0} \left(\frac{L_M^U\left(y_c^\alpha, y_s^\alpha\right)}{L_M^V\left(y_c^\alpha, y_s^\alpha\right)} \right) = K$$

and

$$\lim_{\alpha \to 0} \left(\frac{L_W^U\left(y_c^\alpha, y_s^\alpha\right)}{L_W^V\left(y_c^\alpha, y_s^\alpha\right)} \right) = K^2$$

Using the proposition to approximate values for $\alpha = 1$, we see that $L_M^U\left(y_c, y_s\right) \approx K L_M^V\left(y_c, y_s\right)$ and $L_W^U\left(y_c, y_s\right) \approx K^2 L_M^V\left(y_c, y_s\right)$, as in our example. It follows that the three implications discussed at the end of the last subsection generalize: first, if we can measure other biases, we can in principle recover K (using the formula in the appendix) and convert the partial welfare-analytic objective function into an idealized welfare-analytic objective function by making a simple multiplicative adjustment;[23] second, even when we do not have estimates of the parameters governing other biases, we can usefully assess policies impacting misperceptions according to the idealized objective function $U\left(\cdot, \cdot\right)$, which is (currently) unknown and unobservable, by evaluating them according to the observed choice-rationalizing objective function $V\left(\cdot, \cdot\right)$; third, the approach is robust with respect to the existence of certain types of unspecified biases outside the scope of analysis.

---

[23] According to the formula, K represents the marginal rate of substitution between m and ϕ for U and V, evaluated at the status quo.

A feature of Proposition 2 is that all consequences of the instrument become small in the limit. Accordingly, in settings where ϕ represents expected utility (i.e., $V(m, \phi) = u(m) + \delta\phi$ and $\phi(y) = \sum_{s=1}^{T} \pi_s v(y_s)$), the first-order approximation involves an extrapolation from a limit in which risk and risk aversion vanish. For applications in which risk aversion is thought to play a central role, that property may raise concerns. Note, however, that for this special case, we can proceed exactly as in the example of Subsection 2.3.1, using a first-order approximation only for the terms $u\left(m_0 - r_f^U\right)$ and $u\left(m_0 - r_f^V\right)$. While this alternative approximation involves an extrapolation from a limit in which the reservation price is small, there is no requirement that the limit eradicates either risk or risk aversion.

2.4 Confounding framing effects

So far, we have assumed that neither V nor U depends directly on the policies we seek to compare (e.g., θ and θ') – in other words, that policies do not create confounding framing effects. This assumption is responsible for the invariance of the multiplicative constant, K, across policies, which these comparisons exploit. Part of the assumption is empirically testable: if θ has no direct effect on V, then simply framed valuations should be insensitive to the intervention. In any given application, we may find on the contrary that an intervention changes valuations not only in the complex frame, but also in the simple frame. For example, educating consumers about compound interest may induce them to attend more carefully to all future consequences, whether described in terms of compound returns or otherwise. Such choices would falsify our assumption by revealing that V depends on θ.

How should an analyst proceed if the data reveal these types of confounding framing effects?[24] We will continue to assume that U, the idealized utility function, does not depend directly on the intervention θ. This assumption implies that the direct dependence of V on θ reflects framing effects associated with biases that are as yet undiagnosed.[25] In that case, a simple adjustment to our measure of deliberative competence preserves all of the implications that follow from Proposition 2.

Formally, we extend our model as follows. Proceeding as in Section 2.3.1, we use θ to index a collection of policies that potentially address characterization failure. In addition to allowing for the dependence of perceived complexly framed outcomes on θ, we also posit the existence of generalized choice-rationalizing utility functions, $V(m, \phi(y), \theta)$, that also depend directly on θ. This direct dependence constitutes the additional and potentially confounding framing effects. We assume that the bias-free choice-rationalizing utility function, which we continue to write as $U(m, \phi(y))$, is not directly affected by θ. We use the same notation as before, except that we include θ as an argument of r_f^V and L_i^V to accommodate the confounding framing effects, and we add subscripts where appropriate for instruments and policies.

---

[24] To be clear, we attribute any dependence of V or U on θ to framing effects, rather than to the possibility that the intervention is itself part of the consumption bundle (which would imply that it affects valuations by serving as a complement to or substitute for m or y).

[25] To the extent the direct dependence of V on θ does not reflect undiagnosed biases, then θ must also create framing effects (aside from those involving comprehension of y_s) for idealized choices. In that case, welfare analysis would require the full apparatus of Bernheim and Rangel, 2009.

The following proposition identifies the adjustment to measured Deliberative Competence required to restore comparability across policies:

Proposition 3. *For each y_s, there exists a constant $\kappa(y_s)$ such that for all y_c and θ we have*

$$\lim_{\alpha \to 0} \left(\frac{L_M^U (y_{cI\theta}^\alpha, y_{sI}^\alpha)}{L_M^V (y_{cI\theta}^\alpha, y_{sI}^\alpha, \theta)} - \frac{\alpha\kappa(y_s)}{r_s^V (y_{sI}^\alpha, \theta)} \right) = 0$$

and

$$\lim_{\alpha \to 0} \left(\frac{L_W^U (y_{cI\theta}^\alpha, y_{sI}^\alpha)}{L_W^V (y_{cI\theta}^\alpha, y_{sI}^\alpha, \theta)} - \left(\frac{\alpha\kappa(y_s)}{r_s^V (y_{sI}^\alpha, \theta)} \right)^2 \right) = 0$$

Furthermore, when y_s is a scalar, $\kappa(y_s) = (y_s - y_0)\kappa^*$ for some constant κ^*.

Using the proposition to approximate values for $\alpha = 1$, we see that

$$L_M^U (y_{cI\theta}, y_{sI}) \approx \frac{\kappa(y_s)}{r_s^V (y_{sl}, \theta)} L_M^V (y_{cI\theta}, y_{sI}, \theta)$$

For any particular instrument I and any two policies θ and θ', we then have

$$\frac{L_M \left(r_{cI\theta}^U, r_{sI}^U \right)}{L_M \left(r_{cI\theta'}^U, r_{sI}^U \right)} \approx \frac{L_M^V (y_{cI\theta}, y_{sI}, \theta) / r_s^V (y_{sl}, \theta)}{L_M^V (y_{cI\theta'}, y_{sI}, \theta') / r_s^V (y_{sl}, \theta')},$$

This expression tells us that, for any instrument, we can rank the policies according to the idealized objective function $U(\cdot, \cdot)$ by ranking them according to *adjusted* welfare measures based on the observed choice-rationalizing objective function $V(\cdot, \cdot)$ (where the adjustment simply involves rescaling by simply framed valuations). Once again, we can even gauge the percentage differences between the dollar-equivalents of the benefits flowing from these policies. A limitation of this extension is that the dependence of κ on y_s complicates aggregation of welfare measures over instruments. However, in cases where y is a scalar (as in our experiment), the proposition yields the following approximate values for $\alpha = 1$:

$$L_M^U (y_{cI\theta}, y_{sI}) \approx \kappa^* \frac{y_s - y_0}{r_s^V (y_{sl}, \theta)} L_M^V (y_{cI\theta}, y_{sI}, \theta)$$

Supposing in particular that we are interested in aggregating $L_M^U (y_{cI\theta}, y_{sI})$ over instruments using weights λ_I (as discussed in Section 2.3.1), then for any two policies θ and θ', we have

$$\frac{\sum_{I \in \Im} \lambda_I L_M \left(r_{cI\theta}^U, r_{sI}^U \right)}{\sum_{I \in \Im} \lambda_I L_M \left(r_{cI\theta'}^U, r_{sI}^U \right)} \approx \frac{\sum_{I \in \Im} \lambda_I \frac{y_{sI} - y_0}{r_s^V (\alpha, y_{sI}, \theta)} L_M^V (y_{cI\theta}, y_{sI}, \theta)}{\sum_{I \in \Im} \lambda_I \frac{y_{sI} - y_0}{r_s^V (y_{sI}, \theta')} L_M^V (y_{cI\theta'}, y_{sI}, \theta')} \tag{5}$$

This formula is the same as expression (4), except that we multiply each measure of L_M^V by the associated ratio of the future payment to the observed simply framed valuation. Identical statements follow for L_W^V and L_W^U, except that the adjustment term is squared.

2.5 Aggregation

So far we have focused on the measurement of each individual's deliberative competence. Welfare analyses of public policies ultimately entail aggregation. The standard practice in the optimal taxation literature (see, for example, Mirrlees, 1971) is to assume that aggregate social welfare is given by some function $W(U_1, ..., U_J)$, where U_j is the utility of individual j and J is the population size. For a marginal change in some policy variable θ we have:

$$\frac{dW}{d\theta} = \sum_{j=1}^{J} \frac{\partial W}{\partial U_j} \frac{dU_j}{d\theta} = \sum_{j=1}^{J} g_j \left(\frac{\frac{dU_j}{d\theta}}{\frac{dU_j}{dm}} \right), \tag{6}$$

where $g_j = \frac{\partial W}{\partial U_j} \frac{dU_j}{dm}$ is the marginal social value of a dollar given to individual j, and $\frac{dU_j}{d\theta} / \frac{dU_j}{dm}$ is the dollar-equivalent marginal effect of the policy on j's welfare. Unless the analyst wishes to explicitly incorporate distributional concerns, the marginal social value of a dollar, g_i, is the same for everyone. In that case, equation (6) implies that the analyst simply adds up dollar-equivalent welfare measures. For the minmax welfare criterion, one can proceed similarly, focusing on the worst possible outcome for each individual.[26]

Focusing first on L_M, we see that aggregation presents no difficulties when we assume away decision-making defects outside the scope of the analysis ($U = V$). When U and V differ, aggregation becomes more challenging, because the unobserved constant of proportionality between the approximations, K, can vary from one person to another.

For each individual j, Proposition 2 tells us that, as an approximation, $L_M^V(j) \approx L_M^U(j) K(j)$. It follows that

$$\overline{L}_M^V = \overline{L}_M^U \overline{K} + \text{cov}\left(L_W^U, K \right)$$

where we use bars to denote the mean value of each variable. Therefore, by computing the mean of L_M^V, we obtain the mean of L_M^U up to a multiplicative constant, with an adjustment term involving the covariance between the ideal measure of deliberative competence and the factor of proportionality from Proposition 2.

Our objective is to compare policies according to their average idealized welfare losses, $\overline{L}_M^U$. Hence, average observed losses, $\overline{L}_M^V$, are a valid gauge of the proportional benefits of policy if decision making defects within the scope of the analysis are uncorrelated with defects outside the scope of the analysis – in other words, if $\text{cov}\left(L_M^U, K \right) = 0$. This assumption strikes us as rather strong, particularly in light of evidence that biases tend to be correlated across domains (Dean and Ortoleva, 2019; Stango and Zinman, 2019; Chapman et al., 2018).

If we aim for an ordinal rather than cardinal comparison of policies, weaker assumptions will suffice. The rankings of two policies according to $\overline{L}_M^V$ and $\overline{L}_M^U$ will be identical as long as $\text{cov}\left(L_M^U, K \right)$ ranks them in the same order as $\overline{L}_M^U$. This condition is obviously satisfied whenever $\text{cov}\left(L_M^U, K \right)$ is policy-invariant. To obtain a weaker condition, we can rewrite the covariance as the product of the correlation coefficient between

---

[26]This practice assumes that different consumers may end up paying different prices, perhaps because they make their purchase decisions at different points in time. The worst possible social outcome is then the one that delivers the worst possible outcome for every consumer.

L_M^U and K, the standard deviation of L_M^U, and the standard deviation of K. To the extent the correlation between L_M^U and K reflects the relation between underlying susceptibilities, it will be insensitive to changes in policy. Policies will then impact $\text{cov}\left(L_M^U, K\right)$ only through the standard deviation of L_M^U. With L_M^U bounded below by zero, it is natural to assume that its standard error is non-decreasing in its mean. But then the addition of the adjustment term preserves the policy ranking, as desired.

The following special case illustrates the preceding point. Suppose consumers are characterized by two binary parameters, τ_L and τ_K. With respect to biases within the scope of the analysis, they are either rational ($\tau_L = R$) with $L_M^U(R) = 0$, or behavioral ($\tau_L = B$) with $L_M^U(B) > 0$. Likewise, with respect to biases outside the scope of analysis, they are either rational ($\tau_K = r$) with $K(r) = 1$, or behavioral ($\tau_K = b$) with $K(b) \neq 1$. Thus there are four types of consumers, each of which appears in the population with some arbitrary frequency. It is then straightforward to show that

$$\overline{L}_M^V = \overline{L}_M^U E\left(K \mid \tau_L = B\right)$$

In other words, $\overline{L}_M^V$ and $\overline{L}_M^U$ are related by a multiplicative factor, $E\left(K \mid \tau_L = B\right)$. Furthermore, as long as the policies do not change consumers' underlying types, $E\left(K \mid \tau_L = B\right)$ does not vary. Accordingly, $\overline{L}_M^V$ and $\overline{L}_M^U$ rank policies in the same order, and indeed $\overline{L}_M^V$ correctly measures dollar-equivalent welfare correctly up to a multiplicative scalar. [27]

Precisely the same set of issues arise in the context of our other deliberative competence measure, L_W, plus one additional consideration: the density parameter η_s may vary from person to person. However, there is no particular reason to think η_s is correlated with either L_W^U or with K. To the extent these correlations are zero, heterogeneity in η_s does not impact the proportionality between $\overline{L}_W^U$ and $\overline{L}_W^V$.

3 Application to financial decision making

As an application of our method of Deliberative Competence, we investigate the efficacy of two educational interventions aimed at improving intertemporal decision making by enhancing comprehension of compound interest, one of the most fundamental concepts in finance, and a core topic in the vast majority of courses and books on the principles of financial decision making. To assess these types of interventions, the literature has primarily examined either performance on exam-style questions designed to gauge comprehension of pertinent

---

[27]To the extent concerns about the heterogeneity of K remain, two options are available. First, one can deploy diagnostic tests to determine whether a problem is present and to gauge its magnitude. For example, one could measure a collection of standard biases (other than those impacting the consumer's understanding of the instrument I) and ask whether the rankings of policies vary meaningfully between high-bias and low-bias subgroups. Under the plausible hypothesis that the other measured biases are related to $K(i)$ (again because biases tend to be correlated across domains), invariance of the ranking would rule out problematic patterns of correlation between K and L_M^U. Second, for specific biases outside the scope of the analysis, it may be possible to measure $K(i)$ and adjust $L_M^V(i)$ accordingly. For the example given in Subsection 2.3.1, $K(i)$ is simply $(\beta(i)\pi(i))^{-1}$. Consequently, measuring $\beta(i)$ and $\pi(i)$ permits one to calculate $\frac{L_M^V(i)}{\beta(i)\pi(i)}$.

concepts (financial literacy) or directional effects on behavior. Accordingly, we compare the conclusions that follow from our method and from these two conventional approaches.[28]

With respect to financial literacy, we focus on comprehension of compound interest rather than a broad range of financial topics. Improved performance on batteries of exam-style questions implies better financial decisions if (A1) subjects employ that knowledge when making decisions, and (A2) changes in financial choices are only due to the additional knowledge, and do not involve mechanisms such as propaganda or brow-beating.[29] With respect to directional effects on behavior, the existence of exponential growth bias creates a presumption that people undervalue assets that pay compound interest. A directional shift in average behavior that opposes this presumed bias indicates better financial decision making if (B1) the shift is not so large that it results in overshooting, and (B2) heterogeneity across decision makers is sufficiently limited, so that the behavior of the average subject is an accurate proxy for the behavior of each individual. We directly test these assumptions and find that all are violated. Our measure of Deliberative Competence, in contrast, does not rely on these assumptions.

We structure this section as follows. Subsection 3.1 describes our experimental design. Subsection 3.2 details its implementation and describes preliminary analyses. Subsections 3.3 and 3.4 contrast evaluations of these interventions based on our measure of deliberative competence and the conventional outcome metrics. Subsection 3.5 then documents violations of the assumptions underlying the use of the conventional metrics.

3.1 Design

Our investigation involves two experiments. Both consist of three stages. First, subjects participate in a randomly assigned financial education intervention. Second, subjects complete paired valuation tasks. Third, subjects answer a battery of exam-style questions on compound interest. Additional detail concerning each stage follows.

Stage 1: Education intervention. Using material from a popular investment guide, *The Elements of Investing: Easy Lessons for Every Investor* by Malkiel and Ellis (2013), we produced an instructional video on compound interest that covers the topic through a narrated slide presentation.[30] The narration is

---

[28]Other methods for assessing the quality of decision making are not easily adapted to the types of decisions that provide the focus of our study. Consistency-based methods are unlikely to detect problems arising from a consistent misunderstanding of compound interest (see, e.g. Stango and Zinman, 2009; Levy and Tasoff, 2016), and dominance-based methods are inapplicable when the best choice depends on intertemporal preferences or perceptions of reliability. Furthermore, within this domain, the existence of numerous open questions concerning the nature of intertemporal preferences provides a reason to prefer our "sufficient statistic" approach (or some other direct welfare proxy) over structural methods. See (Ericson and Laibson, 2018) for a discussion of open questions concerning intertemporal preferences. Some important puzzles pertain specifically to choices with monetary payments (Cohen et al., 2016; Andreoni et al., 2018).

[29]This approach also assumes that the effects of knowledge on the quality of decision making are monotonic, which may not be correct if "a little knowledge is a dangerous thing."

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

verbatim from the text (with a few minor adjustments), while the slides summarize key points. Stylistically, the video resembles those offered through the educational internet platform *www.khanacademy.org.*[31]

For Experiment A, the video constitutes the entire Treatment. For Experiment B, we split the video into three parts, each of which is followed by practice problems with automated, individualized feedback. If a subject provides an answer that is mistaken but consistent with the calculation of simple interest, the intervention mentions the likely source of the error, explains why the answer is mistaken, and suggests how to get started with a calculation of the correct response.[32] There are six practice questions in total. For the first four, subjects who do not answer correctly in three attempts move on to the next part of the Treatment. For the last two questions, subjects still receive feedback, but they must select the unique correct answer from 13 options before they can continue.[33] We reproduce the complete practice stage in Appendix E.

Subjects in the Control conditions for each experiment participate in interventions that are stylistically similar and that require comparable amounts of cognitive effort as the respective Treatment interventions. In Experiment A, subjects watch a video based on a section about index funds from the same investment guide. Experiment B employs a Control condition concerning portfolio diversification that also contains practice problems with personalized feedback, based on Malkiel and McCue (1985), supplemented with material from Malkiel and Ellis (2013). Neither control intervention mentions compound interest or the time value of money.

In order to isolate the features of the intervention that affect test scores and behavior, Experiment A employs two additional treatments that dissect the main Treatment into its constituent parts. The complete educational module begins with a simple explanation of compound interest illustrated through an iterative calculation.[34] The remainder of the text consists of two components:

---

[31]While the intervention is brief, it is important to bear in mind that financial education in the workplace is also brief. A meta-analysis by Fernandes et al. (2014) finds that the average financial education program involves only 9.7 hours of instruction. That time is 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.

[32]The following is an example of our individualized feedback: "Now you try: $100 is invested at 9% for 32 years, compounded yearly. How much will be in the account after these 32 years?" The subject selects from the options $100, $200, $388, $400, $600, $800, $1200, $1600. The feedback the subject receives depends on her choice. If the subject selects $100, she sees: "You selected $100. That's not quite right. You start out with $100. Then you get 9% interest each year! Hence, after 32 years, you will have MORE than $100!" If the subject selects $200, she reads: "You selected $200. That's not quite right. You probably remember from the example above that at 9%, an investment doubles in 8 years. Thus, the $100 double to $200 after 8 years. These $200 then double to $400 in the next 8 years. (That is, until year 16.) In the next 8 years, from year 16 to year 24, these $400 double to $800! Then, in the next 8 years, from year 24 to year 32, it will double again. Please give it another try." If the subject selects $388: "You selected $388. That's not quite right. You probably got this because you thought you'd get 32 times the interest of 9% on your $100, which is $9. But, starting from the second year, you also get interest on the interest you earned! Here's how. [Explanation omitted] Please watch the video again, so you'll understand how compound interest works." Similar feedback screens follow the remaining options.

[33]A single subject dropped out of the study when attempting to answer these questions.

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

(i) An explanation of a simple, memorable, and potentially valuable heuristic, the rule of 72, along with five illustrative applications.[35] 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) Rhetorical material. 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, and characterizes compounding as a "miracle."

Subjects in the *Substance-Only* treatment view a video covering all of the substantive material, but omitting exhortations and atmospheric quotes.[36] In contrast, subjects in the *Rhetoric-Only* treatment view a video containing all of the rhetorical material, as well as the introductory explanation of compound interest, but omitting all material on the rule of 72.

Stage 2: Valuation tasks Subjects perform 10 paired valuation tasks. Each task elicits an equivalent current dollar value for a reward r to be received in either 36 or 72 days. With *simple framing*, we describe the reward as follows: "We will pay you \$$r$ in t days." With *complex framing*, we describe the same reward in terms of a return on an initial investment, as follows: "We will invest \$$a$ at an interest rate of X% per day. Interest is compounded daily. We will pay you the proceeds in t days." Subjects make two sets of choices pertaining to each future reward, one with simple framing, the other with complex framing.[37] For each frame f, we elicit each subject j's immediate dollar equivalent of a payment R received in t days, using the iterated multiple price list method with a resolution of \$0.20 (Andersen et al., 2006). Figure 1 presents an example of the first decision list of a round for a complexly framed prospect.[38]

We randomize the order of the valuation tasks at the subject level. Subjects are not told that some of the tasks are substantively equivalent, and they typically do *not* perform equivalent simply and complexly framed tasks consecutively.

---

[35]We 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.

[36]In 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?"

[37]We chose the parameters of the tasks so that the complexly framed version yielded roughly 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.

[38]Throughout, we use the midpoint of the pertinent interval for analysis. For further details on the iterated multiple price lists, see Appendix B.

23

	You will get the specified dollar amount within two days from today	We will invest $\$a$ in an account with $X\%$ interest per day. Interest is compounded daily. We will pay you the proceeds in t days.
$20	◯	◯
$18	◯	◯
$16	◯	◯
. . .	. . .	. . .
$2	◯	◯
$0	◯	◯

Figure 1: Elicitation of valuations. The figure displays the first decision list of a round with complex framing. Payment amounts on the left include all even dollar amounts between and including $0 and $20. For simply framed choices, the text on the upper right is replaced by "We will pay you R in t days."

Table 1 lists the parameters t, X, 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.[39] Our study is thus biased in favor of finding beneficial behavioral effects of the Treatment interventions.

Future Reward R	Investment Amount a	Daily Interest Rate X	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.

Stage 3: Exam-style questions. The final stage of each experiment is an incentivized test consisting of the five questions about compound interest listed in Table 2, as well as five questions about the material

---

[39]We use two different time frames so subjects face a greater variety of decision problems, and hence are less likely to consider successive problems highly similar.

covered in the video shown to the control group.[40] subjects are aware that performance is incentivized prior to participating in the education intervention.

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: Exam-style questions. Questions were presented in random order and intermingled with the five questions concerning material covered in the respective control interventions.

Payment   All subjects know from the outset that they will be paid according to one randomly selected decision from the entire experiment with 75% probability, and according to their performance on the test with 25% probability. In the latter case, they earn $1 for each of the ten questions answered correctly. The instructions emphasize that subjects will be paid within at most two days of the promised payment date. The instructions repeatedly highlight that subjects "should make every decision as if it is the one that counts, because it might be!" In addition to the incentive payment, each subject receives a completion payment of $10.

3.2   Implementation and preliminary analysis

We conducted both experiments through the online labor market Amazon Mechanical Turk (AMT).[41] Our primary reason for employing this subject pool is that the typical member has a poor understanding of compound interest.[42] Also, this group resembles the target populations for many financial education programs in terms of demographic characteristics such as age and income. We ran Experiment A in eight sessions with a total of 504 subjects during April and May 2014 and Experiment B in two sessions with a total of 401

---

[40]The test questions for the material in the control interventions are in Appendix Table B.2. We randomize the order of all ten test questions at the subject level.

[41]See Horton et al. (2011); Mason and Suri (2012); Peysakhovich et al. (2014); Buhrmester et al. (2018) on the use of AMT in social science research.

[42]We opted for this subject pool after pilot experiments at Stanford and at The Ohio State University indicated that the vast majority of student subjects at these universities correctly apply compound interest calculations.

subjects in October 2018, all on weekday mornings. Each subject participated in exactly one treatment of exactly one experiment.[43] We restricted participation to subjects who reside in the US and are at least 18 years of age. We took several precautionary measures to ensure that subjects were able to view the videos and that they would pay attention to them. These measures are detailed in Appendix B, as are additional implementation details.

Each experiment begins with a two-and-a-half minute video recording of one of the authors (Bernheim) vouching that we will pay subjects exactly the amount we promise within at most two days of the promised date,[44] and ends with a number of non-incentivized questions about subjects' decision making.[45] We intentionally place no restriction on the use of resources such as calculators, the internet, or personal advice when making decisions, as subjects always have those options when making real-world decisions. Roughly a quarter of our subjects in Experiment A and less than eight percent in Experiment B report using such resources when completing the incentivized test, a fraction that does not vary meaningfully across treatments. Subjects could complete all valuation tasks at their own pace. Subjects in Experiment A took 62 minutes on average (s.d. 22 minutes) to complete the study, while those in experiment B took 75 minutes (s.d. 25 minutes). The difference in completion times is attributable to the longer interventions in experiment B.

On average, subjects earned \$22.86 in Experiment A and \$23.14 in Experiment B. In both experiments, earnings include a fixed \$10 participation fee and range 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; Hara et al., 2018).

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. Hence, we informed subjects that "most people begin a decision list by preferring the option on the left and then switch to the option on the right." In practice, 7.7% of subjects (39 of 504) in Experiment A and 13.2% of subjects (53 of 401) in Experiment B switched two or more times in at least one price list. This number does not significantly differ across treatments ($p = 0.85$ and $p = 0.18$ for Experiments A and B, respectively). In laboratory studies of risky choices by undergraduate subjects (such as Holt and Laury, 2002), the corresponding figure typically falls in the range of 10 to 15%. Following the usual convention (see, for example, Harrison, Lau, Rutström and Sullivan, 2005), we focus attention on the 803 subjects who respected monotonicity.

Demographics. While our conclusions about the validity of methods for assessing decision quality do not depend on the demographic composition of our sample, we note that our participants are slightly more financially literate than other pools of U.S. subjects (see Lusardi and Mitchell, 2009, and Lusardi, 2011). There are also differences between the samples used for Experiments A and B, possibly due to changes in

---

[43]We excluded all individuals who have participated in any other study of ours on financial decision making.

[44]The video invites subjects to click a link to the authors' homepages so they can verify the authenticity of the video. Before participating in the main stages of the experiment, subjects also complete an unincentivized questionnaire concerning demographics, as well as a standard battery of five questions designed to assess financial literacy (Lusardi and Mitchell, 2009). We reproduce the five questions in Appendix Table B.1.

[45]Appendix Table C.4 details the questions and responses by experiment and treatment.

the composition of the AMT subject pool.[46] Subjects in experiment A are on average five years younger, the fraction of women is 6.6 percentage points lower, and their median income is 30% lower. Appendix C.1 lists all demographic details.[47] Due to the differences between the samples, we analyze each experiment separately, without making direct statistical comparisons. Appendix C.2 complements our main analysis by combining data across experiments. It shows that the relative performance of the two compound-interest interventions are driven by their differences rather than by divergences between the subject pools.

Randomization and attrition Randomization into treatments was successful. Of the 68 F-tests we perform to assess the differences in demographic characteristics across treatments and experiments (one for each characteristic and each experiment), two are significant at the 5%-level, and five more are significant at the 10% level (see Appendix C.1). These figures are well within the expected range.

Because we conduct the experiment over the internet, attrition is a possibility. In Experiment A, we find it to be negligible and unrelated to the treatments. Only four subjects who reached the stage at which they may have viewed a treatment video failed to complete the study (compared to 504 who completed it).[48] In Experiment B, two subjects in the Control condition and nine subjects in the Treatment condition began but did not finish the intervention. While these magnitudes are small compared to the sample size of 401 subjects, they differ statistically at the 5%-level.

3.3 Assessment based on conventional outcome measures

In this section, we analyze the effects of the Treatment interventions in each experiment using the two conventional metrics, exam-style questions that test comprehension, and directional effects on behavior.[49] We defer the analysis of the Substance-Only and Rhetoric-Only treatments of Experiment A to Section 3.5.

Effects on tested knowledge Panel A of Figure 2 shows how the treatments affect subjects' average scores on the five test questions pertaining to compound interest. In the Control condition of Experiment A, the average subject answers just under two of five, or 39%, of the questions correctly. The Treatment intervention increases the average score by roughly 29 percentage points (1.4 additional correct answers), to 68%.[50] The numbers in Experiment B are remarkably similar, with 37% correct responses in the Control condition and 69% in the Treatment condition.

---

[46]In addition, for experiment B, we required subjects to have at least 500 completed Human Intelligence Tasks (HITs) and at least a 98% approval rating.

[47]On average, our samples are somewhat poorer, better educated, and more likely to live in larger households than the average US citizen. While our samples mirror the general population with respect to the prevalence of full-time employment, the fraction of respondents who describe themselves as working part-time is roughly twice as high. Perhaps because we recruited our subjects through the internet, our sample also over-represents young adults, whites, urban residents, and people who have never been married.

[48]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.

[49]All results reported in this section are robust with respect to various statistical controls and alternative specifications. For details, see Appendix C.

[50]See Appendix C.3 for the effects on individual test questions.

A. Test scores

B. Valuations in complex frame
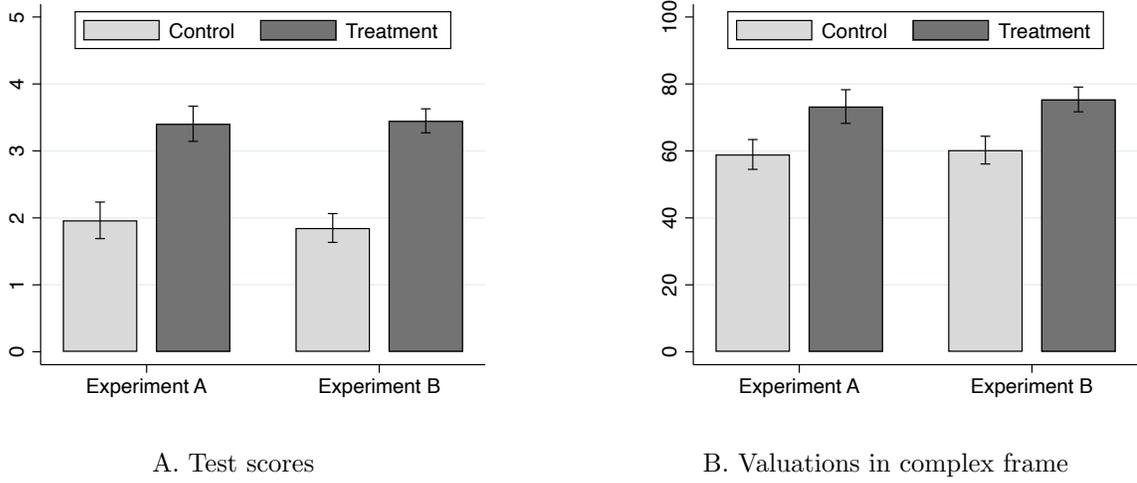
Figure 2: Assessment by conventional methods. Panel A: Fraction of correct answers on the five exam-style questions about compound interest. Panel B: Mean valuation of complexly framed future payments, $r_c^{i,R,t}$, rescaled to range from 0 to 100 (percentage points) rather than 0 to 1. Whiskers indicate 95% confidence intervals. Panel B includes multiple observations per subject; standard errors are clustered on the subject level.

	(1)	(2)	(3)	(4)
	Test score compounding		Test score control module	
	(out of 5)		(out of 5)	
Experiment	A	B	A	B
Levels				
Control	1.963***	1.849***	3.284***	3.078***
	(0.140)	(0.110)	(0.114)	(0.101)
Treatment	3.406***	3.450***	2.226***	2.704***
	(0.135)	(0.092)	(0.092)	(0.098)
Difference	1.442***	1.601***	-1.058***	-0.374***
	(0.194)	(0.143)	(0.146)	(0.141)
Observations	215	348	215	348

Table 3: Performance in exam-style test. Each column displays the coefficients of a separate OLS regression. Standard errors in parentheses. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

Columns 1 and 2 of Table 3 summarize these results and present tests of equality between the Treatment and Control conditions using OLS regressions. The treatment effects in both experiments are highly statistically significant ($p < 0.01$). Columns 3 and 4 establish that the improvements in test performance are not due to effects of the Treatment on general motivation or attention. If the effects were spurious, we would find comparable effects for subjects' scores on the five test questions concerning the Control interventions. On the contrary, in each experiment, subjects in the Treatment condition perform substantially worse on questions about the respective Control intervention than subjects in the Control condition. The differences of 1.05 and 0.37 questions in Experiments A and B, respectively, are again highly statistically significant ($p < 0.01$). The variation in effect size across experiments is attributable to the fact that the two experiments employ different Control conditions and different test questions about those conditions.

Valuations in complexly framed tasks

	(1)	(2)	(3)	(4)	(5)	(6)
Delay in days	both	both	72	72	36	36
Experiment	A	B	A	B	A	B
Levels						
Control	58.949***	60.244***	56.401***	58.181***	61.496***	62.308***
	(2.275)	(2.115)	(2.336)	(2.134)	(2.374)	(2.169)
Treatment	73.261***	75.378***	70.629***	71.747***	75.893***	79.010***
	(2.569)	(1.886)	(2.718)	(1.904)	(2.621)	(2.001)
Difference	14.312***	15.134***	14.227***	13.566***	14.398***	16.702***
	(3.431)	(2.834)	(3.583)	(2.860)	(3.536)	(2.951)
Observations	2,150	3,480	1,075	1,740	1,075	1,740
Subjects	215	348	215	348	215	348

Table 4: Valuations in the complex frame, $r_c^{i,R,t}$, in percentage points. Each column displays the coefficients of a separate OLS regression. Standard errors in parentheses, clustered by subject. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

Directional behavioral effects   As we have noted, it is well-established that people, on average, tend to underestimate the power of compound interest. Consequently, following the approach sometimes adopted in the literature, one would deem an intervention potentially welfare-improving if it causes subjects to state higher valuations for investments that involve compound interest. To make these comparisons, we express each valuation as a percentage of the associated future reward. Formally, letting $\tilde{r}_f^{i,R,t}$ denote subject i's valuation for the future reward R received with delay t when presented in frame $f \in \{s, c\}$, we define the normalized valuation as $r_f^{i,R,t} = \tilde{r}_f^{i,R,t}/R$.

Panel B of Figure 2 displays normalized valuations for complexly framed prospects by experiment and condition. Averaged across reward amounts and timeframes, subjects in the Control condition in both experiments regard the interest-bearing investment as equivalent to an immediate payment of about 60% of the future reward amount. Subjects in the respective Treatment conditions state valuations that are about 15

percentage points higher on average. This pattern is precisely what one would expect if the Treatment in each experiment successfully counteracts exponential growth bias. Given the magnitude of the bias documented in the existing literature (Stango and Zinman, 2009), the size of the average treatment effect does not appear to raise concerns about systematic overcorrection. Following the standard approach in the literature, we would thus conclude that the Treatment interventions substantially increase the quality of financial decision making, and do so similarly across the two experiments.[51]

We formalize these findings by regressing normalized valuations in the complex frame, $r_c^{i,r,t}$, on treatment indicators, clustering standard errors by subject. Column 1 of Table 4 shows the results for Experiment A. In the Control condition, subjects value an investment that compounds to one future dollar at 59 cents on average. The Treatment condition raises average valuations by a substantial and statistically significant increment, to 73 cents ($p < 0.01$). The estimates for Experiment B, displayed in column 2, are remarkably similar. Subjects in the Control condition value a future dollar at 60 cents on average, and the Treatment condition increases this valuation by a substantial and statistically significant increment, to 75 cents ($p < 0.01$). Columns 3 - 6 of Table 4 replicate this analysis separately for decisions with 36-day delays and those with 72-day delays, respectively. The same conclusions follow in each instance.

Taken at face value, results based on both conventional methods for assessing the quality of decision making imply that the two versions of the educational intervention used in our experiments are not only highly successful, but are also equally effective. They both increase performance on tests of knowledge, and they both significantly increase average valuations for complexly framed tasks, as one would expect if an intervention successfully counteracts exponential bias. In each case, the effects are comparable. The apparent implication of these findings, at least for this setting, is that instruction is highly beneficial, but that providing people with opportunities to practice their knowledge and receive feedback yields no incremental benefit.

3.4   Assessment based on Deliberative Competence

Next we assess the effects of the two Treatments using the measure of Deliberative Competence developed in Section 2. Following definition 1, we measure subject i's Deliberative Competence for the valuation pair involving future reward R and delay t as the negative of the welfare loss from deficient decision making,

$$DC_{i,R,t} = - \left| r_c^{i,R,t} - r_s^{i,R,t} \right|.$$

The sign convention used here aids interpretation because it associates larger numerical values with higher levels of deliberative competence. For the reasons discussed in Section 2.2, we use the absolute value of the

---

[51]Some prominent studies do not directly observe behavior, but rely on self-reported data about behavior (e.g. Bruhn et al., 2016). Our experiments also elicit such data. At the end of the study, subjects report the number of decision problems for which they explicitly calculated future values, whether they have used the rule of 72 in complexly framed problems, whether they have used it in simply framed problems, and whether they have used external help to make their decisions. As we find in our analysis on directional behavioral effects, our analysis of self-reports suggests that both interventions improve decisions and that they are similarly effective. See Appendix C.4 for details.

difference rather than its square as our measure of welfare loss. Appendix C.5 shows that our empirical results based on $DC_W^{i,R,t} = -\left(r_c^{i,R,t} - r_s^{i,R,t}\right)^2$ are generally similar.

Panel A of Figure 3 shows the levels of Deliberative Competence across Control and Treatment groups in Experiments A and B, respectively, averaged across all valuation pairs. In each experiment, Deliberative Competence in the Control condition is approximately -25% (-24.4% in Experiment A and -25.7% in Experiment B). According to Proposition 1, this figure – 25 cents per dollar – represents the maximal welfare loss the average subject could suffer due to her misperception of the complexly framed rewards, under the assumption that choices in the simple frame are free of other decision-making defects.

A key finding is that treatment effects differ dramatically across Experiments A and B. For Experiment A, the Treatment has no discernible effect on Deliberative Competence. In stark contrast, for Experiment B, the Treatment increases Deliberative Competence by a substantial and statistically significant increment, from -25.7% to -18.6%. In other words, the Treatment in Experiment B eliminates 27.6 percent of the welfare loss associated with characterization failure in the complex frame $(1 - 0.186/0.257 = 0.276)$.[52] As a practical matter, this finding suggests that incorporating practice and feedback can have a dramatic impact on the efficacy of financial education, even when the feedback is fully automated and relatively simple.

Crucially, according to the analysis of Section 2.3, one can interpret the preceding finding as pertaining to the idealized welfare effects of the treatments, which means our conclusions are robust with respect to the possible existence of other biases (Proposition 2). This robustness is important, because the steep discounting of future payments observed with simple framing suggests that other biases are likely present. The leading possibilities include present bias and misconceptions about the likelihood of experimenter follow-through on future payments. Recall that we addressed both of these possibilities explicitly in the example of Section 2.3.1.

At first glance, the dramatic difference in treatment effects across Experiments A and B appears to conflict with the evidence on directional behavioral changes from the previous subsection—if two interventions change mean valuations in complexly framed problems by essentially the same amounts, how is it possible that one substantially lowers the average distance between valuations in complexly and simply framed tasks, whereas the other fails to affect that average distance? Panel B of Figure 3 provides the answer. It displays the empirical cumulative distribution function of the valuation difference $r_c^{i,r,t} - r_s^{i,r,t}$ for the Treatment and Control conditions in each experiment. In each Control condition, roughly 60% of subjects are afflicted by exponential growth bias $(r_c^{i,r,t} < r_s^{i,r,t})$, while a nontrivial fraction of subjects are well-calibrated, or overestimate compound interest $(r_c^{i,r,t} \geq r_s^{i,r,t})$.[53] An effective intervention would increase valuations in complexly framed tasks for subjects who underestimate compound interest, and would *decrease* them for subjects who overestimate compound interest. As a result, the CDF of the valuation difference $r_c^{i,r,t} - r_s^{i,r,t}$ would become more tightly centered around zero. In contrast to this hypothetical ideal scenario,

---

[52]These estimates are based on equal weighting of observed Deliberative Competence across subjects and decision problems. We discuss the implications of equal weighting of subjects toward the end of this section.

[53]Appendix C.6 tests and refutes the hypothesis that the overvaluation of complexly framed opportunities is solely attributable to noisy choice. A fraction of subjects in Goda et al. (2015) and in Levy and Tasoff (2016) also overestimate compound interest.

31

A. Deliberative Competence

Experiment A

Experiment B

B. CDF of valuation differences


Experiment A
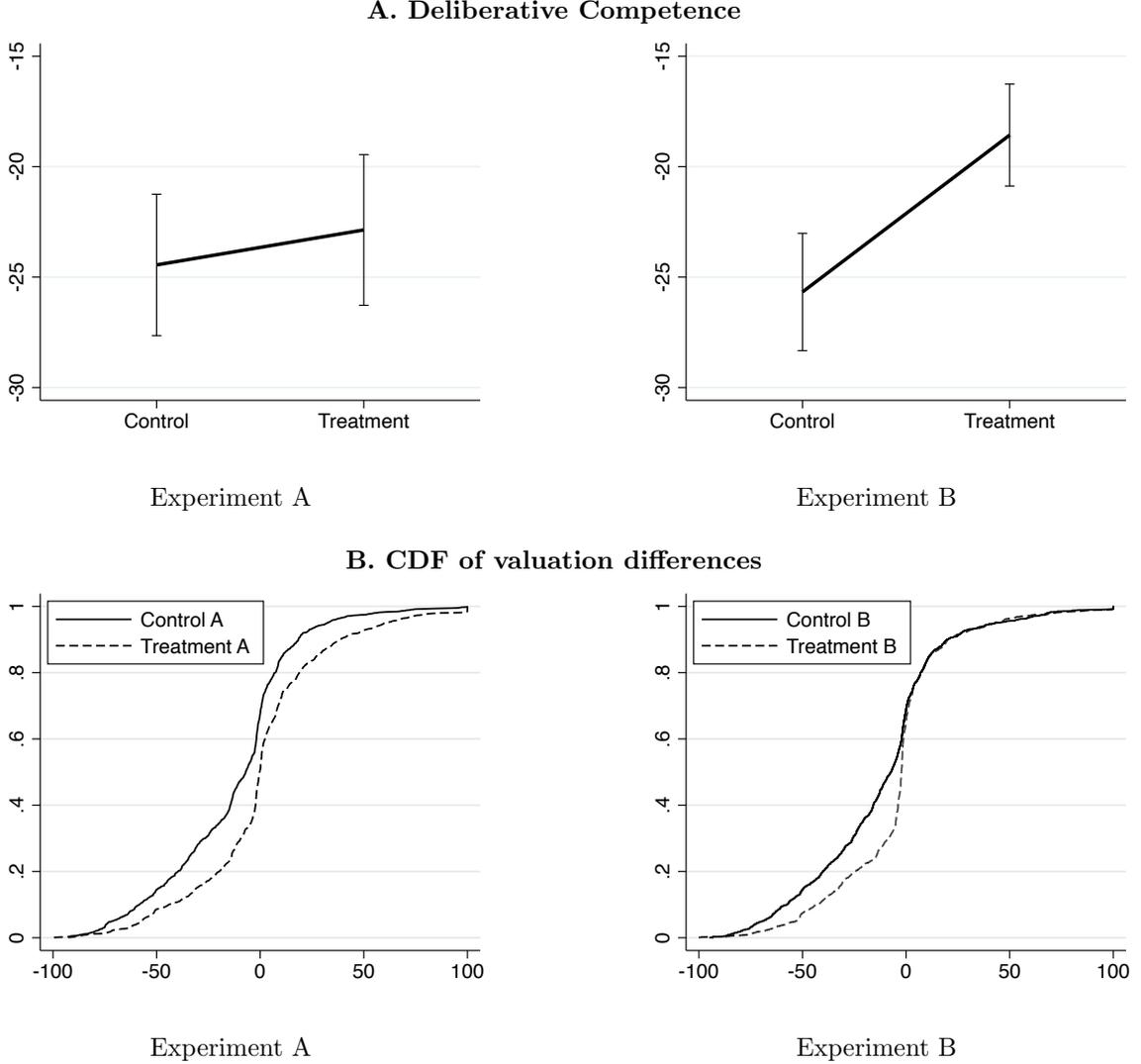
Experiment B

Figure 3: Assessment based on Deliberative Competence. Panels A and B show mean Deliberative Competence, $DC^{i,R,t} = -|r_c^{i,R,t} - r_s^{i,R,t}|$, in the Treatment and Control conditions of experiments A and B, respectively, with $r_f^{i,R,t}$ rescaled to range from 0 to 100 (percentage points) rather than 0 to 1. Whiskers represent 95% confidence intervals with standard errors clustered on the subject level. Panels C and D show cumulative distribution functions for individual-level valuation differences between the complex and simple frame, $r_c^{i,R,t} - r_s^{i,R,t}$, by treatment. Mass to the left of zero indicates underestimation of compound interest (exponential growth bias). Mass to the right of zero indicates overestimation.

the Treatment intervention in Experiment A shifts the entire CDF to the right. It increases valuations for complexly framed tasks across the board, irrespective of whether a subject initially underestimates compound interest, estimates it correctly, or overestimates it. This indiscriminate effect helps some subjects but hurts others, leading to a null effect on decision quality, on average. In sharp contrast, the effects of the Treatment intervention in Experiment B are sensitive to subjects' initial bias. The Treatment increases valuations of complexly framed tasks only for subjects who would have underestimated compound interest, and not for those who would have estimated compound interest correctly, or who would have overestimated it. Accordingly, the Treatment in Experiment B yields a strictly positive net increase in decision quality.

To formalize these findings, we regress $DC_{i,r,t}$ on a treatment indicator and cluster standard errors by subject, separately for each experiment. Panel A of Table 5 displays the results. Column 1 shows that the Treatment intervention in Experiment A increases Deliberative Competence by a small and statistically insignificant 1.5 percentage points. Column 2 exhibits the corresponding regression for Experiment B. It shows that while Control subjects display a level of Deliberative Competence similar to those in Experiment A, the Treatment intervention substantially raises Deliberative Competence from -25.7% to -18.6%. The increase of 7.1 percentage points is highly statistically significant ($p < 0.01$). Columns 3 and 4 replicate this analysis separately using the set of investments with 36-day delays, whereas Columns 5 and 6 do so for investments with 72-day delays. The same conclusions follow in both cases.

Valuations in the simple frame We have defined Deliberative Competence based on the difference between valuations in the complex and simple frame. While the main welfare interpretation of our measure assumes the intervention affects measured competence only by changing valuations with complex framing (Propositions 1 and 2), in principle changes could occur in either frame. When valuations also change in the simple frame, Proposition 3 provides the appropriate adjustment factor.

Table 6 presents regressions describing reservation valuations in the simple frame, separately for each timeframe. Columns 1 and 3 show that, in Experiment A, there are no treatment effects on valuations in the simple frame. In each condition, subjects value a dollar received with a 72-day delay at roughly 70 cents, and a dollar received with a 32-day delay at roughly 75 cents.[54] In contrast, a treatment effect is present in Experiment B. The valuations of subjects in the Control condition are similar to those elicited from subjects in Experiment A, as columns 2 and 4 show. Average simply framed valuations in the Treatment condition, however, are higher by 8.2 and 7.4 cents for the 72-day and 36-day timeframes, respectively.

Accordingly, we deploy proposition 3 to adjust for the change in simply framed valuations in experiment B. To do so, we divide Deliberative Competence by the simply framed valuation separately for the Treatment observations and the Control observations.[55] Observe that this correction changes the units of measurement;

---

[54]These magnitudes are typical for studies that elicit time preferences over short horizons (Frederick et al., 2002). Not only do subjects discount the future heavily, but the discount function decreases much less steeply for longer delays than for short delay. Part of the explanation may involve perceptions of experimenter reliability.

[55]We apply a version of equation (5) that accounts for noise in the elicitation of valuations in the simple frame. Specifically, we calculate the mean valuation for simply framed choices for each timeframe and use the resulting average as the correction factor. Moreover, by our normalization, $y_{sI} - y_0 = 1$ for all instruments I.

<div align="center">**A. Deliberative Competence**</div>

	(1)	(2)	(3)	(4)	(5)	(6)
Delay in days	both	both	72	72	36	36
Experiment	A	B	A	B	A	B
Levels						
Control	-24.448***	-25.673***	-24.076***	-25.945***	-24.820***	-25.401***
	(1.635)	(1.357)	(1.745)	(1.428)	(1.684)	(1.393)
Treatment	-22.864***	-18.569***	-22.959***	-18.441***	-22.769***	-18.697***
	(1.742)	(1.177)	(1.801)	(1.208)	(1.888)	(1.253)
Difference	1.584	7.104***	1.117	7.504***	2.050	6.704***
	(2.389)	(1.796)	(2.508)	(1.870)	(2.530)	(1.874)
Observations	2,150	3,480	1,075	1,740	1,075	1,740
Subjects	215	348	215	348	215	348

B. Deliberative Competence corrected for changes in simply framed valuations

	(1)	(2)	(3)	(4)	(5)	(6)
Delay in days	both	both	72	72	36	36
Experiment	A	B	A	B	A	B
Levels						
Control	-34.639***	-37.541***	-36.246***	-39.326***	-33.032***	-35.756***
	(2.118)	(1.906)	(2.538)	(2.127)	(2.055)	(2.001)
Treatment	-37.204***	-25.491***	-38.363***	-26.293***	-36.044***	-24.689***
	(3.792)	(1.775)	(3.886)	(1.865)	(4.206)	(1.882)
Difference	-2.565	12.050***	-2.117	13.032***	-3.013	11.067***
	(4.343)	(2.605)	(4.641)	(2.829)	(4.682)	(2.747)
Observations	2,150	3,470	1,075	1,735	1,075	1,735
Subjects	215	347	215	347	215	347

Table 5: Deliberative Competence. Each column displays the coefficients of a separate OLS regression of Deliberative Competence, on treatment indicators. Standard errors in parentheses, clustered by subject. Panel A: Deliberative Competence measured as $DC^{i,R,t} = -|r_c^{i,R,t} - r_s^{i,R,t}|$, with $r_c^{i,R,t}$ rescaled to range from 0 to 100 (percentage points) rather than 0 to 1. Panel B: Deliberative Competence is corrected for the change in simply framed valuations; see Section 2.4. The reason for the smaller number of observations in Panel B in Experiment B is one subject who consistently made choices consistent with a valuation of zero in the simple frame. As the correction consists in dividing by simply framed valuations, this subject is excluded from that analysis. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

Valuations in simply framed tasks

	(1)	(2)	(3)	(4)
Delay in Days	72		36	
Experiment	A	B	A	B
Levels				
Control	68.637***	69.698***	75.874***	75.343***
	(2.246)	(1.766)	(2.015)	(1.628)
Treatment	69.513***	77.908***	75.801***	82.791***
	(2.265)	(1.707)	(2.076)	(1.451)
Difference	0.876	8.210***	-0.073	7.448***
	(3.190)	(2.456)	(2.893)	(2.181)
Observations	1,075	1,740	1,075	1,740
Subjects	215	348	215	348

Table 6: Valuations in simply framed tasks, $r_s^{i,R,t}$, rescaled to range from 0 to 100 (percentage points) rather than 0 to 1. Each column displays the coefficients of a separate OLS regression. Standard errors in parentheses, clustered by subject. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

the valuation difference in Panel A is stated in subjects' present value of the prospects, whereas the division by simply framed valuations in Panel B leads to magnitudes stated in subjects' future value of the prospects.[56] Panel B of Table 6 reproduces Panel A with the corrected measure of financial competence, clustering standard errors on the subject level. Columns 1 and 2 average across both timeframes; the remaining columns display estimates separately within each timeframe. We continue to observe large and statistically highly significant improvements in deliberative competence in Experiment B. Treatment effects in Experiment A remain insubstantial and statistically indistinguishable from zero.

Heterogeneous treatment effects In Section 2.5, we showed that, when aggregating our measure of deliberative competence over multiple consumers, the weight implicitly attributed to the welfare of any particular consumer can depend on the extent to which they suffer from biases outside the scope of analysis. In the context of our experiment, this observation implies, for example, that we may give greater or smaller weight to subjects based on their present bias or the accuracy of their beliefs about experimenter reliability. While we have identified theoretical conditions under which this possibility would not impact our results, the issue is also amenable to empirical investigation.

We focus on biases that would express themselves through subjects' valuations in the simple frame, as is the case with present bias and with false beliefs about the reliability of future rewards. Accordingly, we sort subjects into quartiles based on their mean valuations of simply framed prospects. To account for the fact that the Treatment intervention in Experiment B affected those valuations, we classify subjects separately for each treatment. For each experiment, we then regress measured Deliberative Competence on

---

[56]In principle, we could have conducted the experiments eliciting future values rather than present values directly, and thus abstract from preferences entirely. The involvement of preferences, however, may interact with subjects' understanding of the compounding principles (see, e.g., Kunda, 1990), and hence is a potential driver underlying our results.

a Treatment dummy, an indicator for each quartile of simply framed valuations, and an interaction between those indicators and the Treatment dummy. We cluster standard errors on the subject level.

Experiment A

VARIABLE	Deliberative Competence			
	Quartile simply framed valuation			
	1	2	3	4
Levels				
Control	-17.007***	-21.693***	-26.589***	-32.216***
	(2.037)	(2.596)	(3.410)	(3.893)
Treatment	-24.967***	-22.345***	-22.855***	-21.367***
	(4.484)	(2.779)	(2.559)	(3.772)
Effect	-7.961	-0.652	3.734	10.849**
	(4.925)	(3.803)	(4.263)	(5.421)
Observations	2,150			
Subjects	215			

Experiment B

VARIABLE	Deliberative Competence			
	Quartile simply framed valuation			
	1	2	3	4
Levels				
Control	-17.734***	-25.920***	-25.706***	-33.156***
	(1.893)	(2.073)	(2.414)	(3.605)
Treatment	-17.818***	-21.239***	-17.215***	-18.018***
	(2.155)	(2.392)	(2.254)	(2.547)
Effect	-0.085	4.681	8.492**	15.138***
	(2.869)	(3.165)	(3.302)	(4.414)
Observations	3,480			
Subjects	348			

Table 7: Effect of treatments on Deliberative Competence, $DC^{i,R,t} = -|r_c^{i,R,t} - r_s^{i,R,t}|$, by quartiles of simply framed valuations, rescaled to range from 0 to 100 (percentage points) rather than 0 to 1. Each panel presents the output of a single OLS regression. Standard errors in parentheses, clustered by subject. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

Table 7 displays the results. Within each experiment, we find that the Treatment has larger beneficial effects on subjects who have higher mean valuations in the simple frame. Within each quartile, the Treatment effect in Experiment B exceeds that in Experiment A by several percentage points. For Experiment B, the benefit of the intervention is roughly zero for subjects in the top quartile, positive for all other quartiles, and statistically significant for the bottom two quartiles. For Experiment A, the benefit is substantially negative (though not statistically significant at conventional levels) for those in the highest quartile, slightly negative for those in the second quartile, and positive for the lowest two quartiles, but statistically significant only for the lowest quartile. We conclude that the Treatment intervention in Experiment B has a (weakly) positive effect on Deliberative Competence regardless of the welfare weights assigned to different subjects

based on biases outside the scope of the analysis, whereas the Treatment intervention in Experiment A has mixed effects, clearly benefiting only those in the lowest quartile. Because the Treatment intervention in Experiment B has a more positive effect within each quartile than the intervention in Experiment A, our conclusions concerning the benefits of practice and individualized feedback are not driven by an unintended relationship between welfare weights and biases that impact simply framed valuations.

3.5 Testing the assumptions underlying conventional approaches

As we have seen in the preceding subsections, unlike Deliberative Competence, conventional measures of decision-making quality imply that the educational interventions used in Experiments A and B are equally effective, and consequently that the opportunities for practice and feedback offered in Experiment B have little or no incremental value. In this section, we conduct additional analyses that shed light on the reasons for these divergent findings.

The analysis of the previous section has already established that assumptions (B1) and (B2), which would permit one to draw inferences from directional changes in behavior, do not hold in Experiment A. Due to heterogeneity in subjects' misperceptions of compound interest (which violates assumption B2) as well as the intervention's indiscriminate impact, it produces either overcorrection or a magnification of biases for a substantial fraction of subjects (which violates assumption B1).

To examine assumptions (A1) and (A2), which one needs to rationalize inferences drawn from performance on exam-style questions, we study the effects of the Rhetoric-only and Substance-only interventions in Experiment A. First we examine their effects on assessed comprehension by regressing test scores on each of the four treatments in Experiment A, clustering standard errors at the subject level. Column 1 of Table 8 displays the results. The Substance-Only intervention increases performance on test questions concerning compound interest by a similar amount as the full intervention, while the Rhetoric-Only intervention has a much smaller effect.[57] To verify that these results are not due to an effect of the Substance-Only intervention on general motivation, Column 2 uses performance on test questions about the Control intervention as the dependent variable. Indeed, subjects in the Control intervention perform significantly better on these questions than subjects in any other treatment. Accordingly, increases in performance on our assessment of comprehension arise through the expected mechanism (substantive instruction).

Next we investigate the connection between the changes in behavior observed for those exposed to the full intervention in Experiment A and its effects on comprehension. Assumption (A1) maintains that behavior responds to knowledge. Because we have just demonstrated that the Substance-Only treatment improves knowledge, this assumption implies that it ought to change valuations in complexly framed decision problems. Assumption (A2) maintains that behavior only responds to features of the intervention that enhance comprehension. Because we have just demonstrated that the Rhetoric-Only treatment has only a modest

---

[57]The fact that the effect of that intervention is positive may be attributable to the inclusion of an example that illustrates the calculation of compound interest.

VARIABLES	(1) Test scores on questions about		(3) Valuations in frame	(4)	(5) Deliberative Competence
	Treatment	Control	Complex	Simple	
Levels					
Substance-Only	3.234***	1.945***	62.969***	72.273***	-22.012***
	(0.123)	(0.099)	(2.373)	(2.030)	(1.433)
Rhetoric-Only	2.455***	2.205***	77.538***	77.623***	-19.797***
	(0.146)	(0.095)	(2.785)	(2.119)	(1.406)
Treatment A	3.406***	2.226***	73.261***	72.657***	-22.864***
	(0.135)	(0.092)	(2.566)	(2.139)	(1.741)
Control A	1.963***	3.284***	58.949***	72.255***	-24.448***
	(0.140)	(0.114)	(2.272)	(2.089)	(1.633)
p-value of difference to Control					
Treatment	0.000	0.000	0.000	0.893	0.507
Substance-Only	0.000	0.000	0.222	0.995	0.263
Rhetoric-Only	0.015	0.000	0.000	0.072	0.031
Observations	455	455	4,550	4,550	4,550
Subjects	455	455	455	455	455

Table 8: Separate effects of rhetoric and substance in Experiment A. Each column displays the coefficients of a separate OLS regression. Standard errors in parentheses, clustered by subject. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

effect on knowledge, this assumption implies that it ought to change valuations in complexly framed decisions to a much smaller degree than the full intervention.

In fact, we find precisely the opposite. Column 3 displays the coefficients of a regression of complexly framed valuations on treatment indicators. The Substance-Only intervention does not have a statistically discernible effect on these valuations. In contrast, the Rhetoric-Only intervention yields effects comparable to those of the full intervention. Moreover, there are no corresponding changes in simply framed valuations, as Column 4 shows. Accordingly, the Substance-Only intervention leaves Deliberative Competence unchanged (column 5). While the Rhetoric-Only intervention produces a small improvement in Deliberative Competence, this effect does not appear to flow from improved comprehension; rather, the motivational rhetoric provides a nudge that proves fortuitously beneficial to a small degree. We conclude that neither assumption (A1) nor assumption (A2) hold in Experiment A. Behavioral effects in that experiment are wholly disconnected from effects on comprehension. These findings invalidate inferences about the quality of decision making based on assessments of comprehension in Experiment A.

4 Conclusion

In this paper, we have introduced a new method for evaluating the quality of decision making experimentally. It measures the extent to which a consumer's assessments of value coincide with those she would make if

she possessed an accurate understanding of her opportunity set. It yields a sufficient statistic that admits formal and intuitive welfare interpretations, even when (1) consumers suffer from additional decision making biases, known or unknown, outside the scope of analysis, and (2) the interventions of interest potentially create confounding framing effects. It yields measurements of welfare losses at the individual level, and consequently can easily accommodate population heterogeneity with respect to preferences and decision-making defects. It is designed to accommodate settings in which optimal choices depend on preferences (so that dominance-based methods do not apply), and in which misunderstandings of opportunities are governed by consistent principles (so that consistency-based methods are inappropriate). In comparison to structural approaches, it is far simpler and avoids the need for strong assumptions.

We have demonstrated empirically that the Deliberate Competence Method can lead to important insights one might otherwise miss. Specifically, we have studied the effects of two financial education interventions that seek to foster an understanding of compound interest. The interventions are identical except that one provides opportunities for practice and automated, individual feedback while the other does not. Our analysis of Deliberative Competence shows that the intervention improves the average quality of decision making only when it includes those opportunities; without them, it harms decision quality as much as it helps. It follows that, as a practical matter, incorporating opportunities for practice and feedback into financial education can substantially improve programmatic efficacy, even when the feedback is automated and relatively simple. In stark contrast, two conventional evaluation metrics—performance on exam-style questions and directional behavioral changes that counteract a suspected bias—mistakenly imply that both interventions are effective, and to similar degrees, so that the incremental benefits of practice and feedback are essentially non-existent.

We have shown that the implicit assumptions required to draw inference about the quality of choice from the conventional methods do not apply in our setting. First, even though there is a pronounced behavioral bias for the average subject, there is also substantial heterogeneity. As a consequence, without practice and feedback, the intervention leads substantial fractions of subjects either to overcorrect or to magnify their preexisting biases. For these subjects, the effect is harmful rather than beneficial. Second, the behavioral effects of the same intervention are wholly disconnected from its effects on comprehension. Behavioral changes are exclusively attributable to the rhetorical elements of the intervention. In contrast, the substantive material improves test performance, but leaves behavior unchanged. Our investigation raises the possibility that similar violations of the assumptions that rationalize the use of conventional metrics as measures of decision quality are present in other assessments of interventions that target financial decision making.

More generally, our method is potentially useful for evaluating programs that seek to improve decision making through education and training, that aim to evaluate person-to-person influences, including advice (both professional and casual) and mimicry, and that attempt to "nudge" people toward better decisions by manipulating framing.

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.

Allcott, Hunt and Dmitry Taubinsky, "Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market," *American Economic Review*, 2015, *105* (8), 2501–38.

_ , Benjamin Lockwood, and Dmitry Taubinsky, "Ramsey Strikes Back: Optimal Commodity Tax and Redistribution in the Presence of Salience Effects," in "AEA Papers and Proceedings," Vol. 108 2018, pp. 88–92.

_ , Sendhil Mullainathan, and Dmitry Taubinsky, "Energy policy with externalities and internalities," *Journal of Public Economics*, 2014, *112*, 72–88.

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

Ambuehl, Sandro, B Douglas Bernheim, Fulya Ersoy, and Donna Harris, "Peer Advice on Financial Decisions: A case of the blind leading the blind?," *NBER Working Paper*, 2018, *w25034*.

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, Christina Gravert, Michael A Kuhn, Silvia Saccardo, and Yang Yang, "Arbitrage Or Narrow Bracketing? On Using Money to Measure Intertemporal Preferences," *NBER working paper*, 2018.

Atanasov, Pavel and Tom Baker, "Putting health back into health insurance choice," *Medical Care Research and Review*, 2014, *71* (4), 337–355.

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

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.

Benartzi, Shlomo and Richard H Thaler, "Risk aversion or myopia? Choices in repeated gambles and retirement investments," *Management science*, 1999, *45* (3), 364–381.

_ and Richard H. Thaler, "Naive diversification strategies in defined contribution saving plans.," *American Economic Review*, 2001, pp. 79–98.

Bennyhoff, Donald G and Francis M Kinniry Jr, "Advisor's Alpha," *Vanguard research paper (December). https://advisors. vanguard. com/iwe/pdf/ICRAA. pdf*, 2011.

Bernheim, B. Douglas, "Personal Saving, Information, and Economic Literacy: New Directions for Public Policy," in "Tax Policy for Economic Growth in the 1990s" American Council for Capital Formation Washington, DC 1994, pp. 53–78.

_ , "Financial Illiteracy, Education, and Retirement Saving," in Olivia S. Mitchell and Sylvester J. Schieber, eds., *Living with Defined Contribution Pensions. Remaking Responsibility for Retirement.*, University of Pennsylvania Press, 1998, chapter 3.

_ , "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, "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.

Bernheim, B Douglas and Dmitry Taubinsky, "Behavioral public economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 381–516.

Bernheim, B. Douglas, Daniel M. Garrett, 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.

_ , **Igor Popov, and Andrey Fradkin**, "The Welfare Economics of Default Options in 401(k) Plans," *American Economic Review*, 2015, *105* (9), 2798–837.

Bertrand, Marianne and Adair Morse, "Information Disclosure, Cognitive Biases, and Payday Borrowing," *The Journal of Finance*, 2011, *66* (6), 1865–993.

Beshears, John, James J Choi, David Laibson, and Brigitte C Madrian, "Behavioral household finance," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 177–276.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor, "Choose to lose: Health plan choices from a menu with dominated option," *The Quarterly Journal of Economics*, 2017, *132* (3), 1319–1372.

_ , _ , **and Shlomo Benartzi**, "The costs of poor health (plan choices) & prescriptions for reform," *Behavioral Science & Policy*, 2017, *3* (1), 1–12.

Blumenthal, Karen and Kevin G Volpp, "Enhancing the effectiveness of food labeling in restaurants," *JAMA*, 2010, *303* (6), 553–554.

Bruhn, Miriam, Gabriel Lara Ibarra, and David McKenzie, "The minimal impact of a large-scale financial education program in Mexico City," *Journal of Development Economics*, 2014, *108*, 184–189.

_ , **Luciana de Souza Leão, Arianna Legovini, Rogelio Marchetti, and Bilal Zia**, "The impact of high school financial education: Evidence from a large-scale evaluation in Brazil," *American Economic Journal: Applied Economics*, 2016, *8* (4), 256–95.

Buhrmester, Michael D, Sanaz Talaifar, and Samuel D Gosling, "An Evaluation of Amazon's Mechanical Turk, Its Rapid Rise, and Its Effective Use," *Perspectives on Psychological Science*, 2018, *13* (2), 149–154.

Callender, Claire and Jonathan Jackson, "Does the fear of debt deter students from higher education?," *Journal of social policy*, 2005, *34* (4), 509–540.

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*.

Carroll, Gabriel, "Robustness in mechanism design and contracting," *Annual Review of Economics*, 2019, *11*, 139–166.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer, "Econographics," *NBER Working Paper*, 2018.

Chetty, Raj, "Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods," *Annu. Rev. Econ.*, 2009, *1* (1), 451–488.

_ , **Adam Looney, and Kory Kroft**, "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–177.

Choi, James J., David Laibson, and Brigitte C. Madrian, "$100 Bills on the Sidewalk: Suboptimal 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.

Cohen, Jonathan D, Keith Marzilli Ericson, David Laibson, and John Myles White, "Measuring Time Preferences," *NBER working paper*, 2016, *22455*.

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.

Dean, Mark and Pietro Ortoleva, "The empirical relationship between nonstandard economic behaviors," *Proceedings of the National Academy of Sciences*, 2019, *116* (33), 16262–16267.

DellaVigna, Stefano, "Structural behavioral economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 613–723.

_ **and Ulrike Malmendier**, "Contract design and self-control: Theory and evidence," *The Quarterly Journal of Economics*, 2004, *119* (2), 353–402.

Downs, Julie S, George Loewenstein, and Jessica Wisdom, "Strategies for promoting healthier food choices," *American Economic Review*, 2009, *99* (2), 159–64.

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.

_ , **Uri Gneezy, Brian J Hall, David Martin, Vadim Nelidov, Theo Offerman, and Jeroen van de Ven**, "Cognitive Biases: Mistakes or Missing Stakes?," *Unpublished Manuscript*, 2020.

Ericson, Keith Marzilli and David Laibson, "Intertemporal Choice," *NBER working paper*, 2018.

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

Farhi, Emmanuel and Xavier Gabaix, "Optimal taxation with behavioral agents," *American Economic Review*, 2020, *110* (1), 298–336.

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.

Gabaix, Xavier and David Laibson, "Shrouded attributes, consumer myopia, and information suppression in competitive markets," *The Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Goda, 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*.

Goldin, Jacob and Daniel Reck, "Revealed-Preference Analysis with Framing Effects," *Journal of Political Economy*, 2020, *128* (7), 2759–2795.

Grubb, Michael D and Matthew Osborne, "Cellular service demand: Biased beliefs, learning, and bill shock," *American Economic Review*, 2015, *105* (1), 234–71.

Gruber, Jonathan and Botond Kőszegi, "Is addiction "rational"? Theory and evidence," *The Quarterly Journal of Economics*, 2001, *116* (4), 1261–1303.

_ **and** _ , "Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes," *Journal of Public Economics*, 2004, *88* (9-10), 1959–1987.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P Bigham, "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk," in "Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems" ACM 2018, p. 449.

Harberger, Arnold C, "The measurement of waste," *The American Economic Review*, 1964, *54* (3), 58–76.

Harrison, Glenn, Karlijn Morsink, and Mark Schneider, "Do No Harm? The Welfare Consequences of Behavioral Interventions," *Unpublished*, 2020.

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., 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.

Jin, Ginger Zhe, Michael Luca, and Daniel Martin, "Is no news (perceived as) bad news? An experimental investigation of information disclosure," Technical Report, National Bureau of Economic Research 2015.

Johnson, Eric J, Ran Hassin, Tom Baker, Allison T Bajger, and Galen Treuer, "Can consumers make affordable care affordable? The value of choice architecture," *PloS one*, 2013, *8* (12), e81521.

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

Kaiser, Tim, Annamaria Lusardi, Lukas Menkhoff, and Carly J Urban, "Financial Education Affects Financial Knowledge and Downstream Behaviors," *NBER Working Paper*, 2020.

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

Kling, Jeffrey R, Sendhil Mullainathan, Eldar Shafir, Lee C Vermeulen, and Marian V Wrobel, "Comparison friction: Experimental evidence from Medicare drug plans," *The Quarterly Journal of Economics*, 2012, *127* (1), 199–235.

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.

Lipsey, Richard G and Kelvin Lancaster, "The general theory of second best," *The review of economic studies*, 1956, *24* (1), 11–32.

Loewenstein, George, Joelle Y Friedman, Barbara McGill, Sarah Ahmad, Suzanne Linck, Stacey Sinkula, John Beshears, James J Choi, Jonathan Kolstad, David Laibson et al., "Consumers' misunderstanding of health insurance," *Journal of Health Economics*, 2013, *32* (5), 850–862.

Luhrmann, Melanie, Marta Serra-Garcia, and Joachim Winter, "The impact of financial education on adolescents' intertemporal choices," *American Economic Journal: Economic Policy*, 2018, *10* (3), 309–32.

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.

Machina, Mark J and Marciano Siniscalchi, "Ambiguity and ambiguity aversion," in "Handbook of the Economics of Risk and Uncertainty," Vol. 1, Elsevier, 2014, pp. 729–807.

Madrian, Brigitte C and Dennis F Shea, "The power of suggestion: Inertia in 401 (k) participation and savings behavior," *The Quarterly journal of economics*, 2001, *116* (4), 1149–1187.

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

Malkiel, Burton Gordon and Kerin McCue, *A random walk down Wall Street*, Norton New York, 1985.

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.

Mirrlees, James A, "An exploration in the theory of optimum income taxation," *The Review of Economic Studies*, 1971, *38* (2), 175–208.

O'Donoghue, Ted and Matthew Rabin, "Optimal sin taxes," *Journal of Public Economics*, 2006, *90* (10-11), 1825–1849.

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*.

Rasmussen, Christopher James, "Effective cost-sharing models in higher education: Insights from low-income students in Australian Universities," *Higher Education*, 2006, *51* (1), 1–25.

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: 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.

_ **and** _ , "We are all behavioral, more or less: Measuring and using consumer-level behavioral sufficient statistics," *NBER Working Paper*, 2019.

Sunstein, Cass R, "Maximin," *Yale Journal on Regulation*, 2020.

Sutter, Matthias, Michael Weyland, Anna Untertrifaller, and Manuel Froitzheim, "Teaching financial literacy has an impact on risk and time preferences—Insights from a field experiment," *Unpublished*, 2020.

Sydnor, Justin, "(Over) insuring modest risks," *American Economic Journal: Applied Economics*, 2010, *2* (4), 177–99.

Taubinsky, Dmitry and Alex Rees-Jones, "Attention variation and welfare: theory and evidence from a tax salience experiment," *The Review of Economic Studies*, 2018, *85* (4), 2462–2496.

Thaler, Richard H and Cass R Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, 2009.

Urban, Carly, Maximilian Schmeiser, J Michael Collins, and Alexandra Brown, "The effects of high school personal financial education policies on financial behavior," *Economics of Education Review*, 2018.

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.

ONLINE-APPENDIX
NOT FOR PUBLICATION

Sandro Ambuehl, B. Douglas Bernheim, Annamaria Lusardi

Table of Contents

A Proofs

A.1 Proof of Proposition 1

Proof. **Part (i)**

Suppose $r_s \geq r_c$. Because V is strictly decreasing in its first argument, we have

$$\mathcal{L}_M (r_c, r_s) = \max_{p \in [r_c, r_s]} [V(-p, 1, s) - V(0, 0, s)] = V(-r_c, 1, s) - V(0, 0, s),$$

which equals 0 for $r_c = r_s$ by the definition of r_s. It follows that

$$\frac{d\mathcal{L}_M}{dr_c} = -V_m(-r_c, 1, s),$$

and accordingly that a first-order approximation around $r = r_s$ (for $r < r_s$) is given by

$$\frac{1}{\nu_s} \mathcal{L}_M (r, r_s) \approx r_s - r$$

Now suppose $r_c \geq r_s$. Again appealing to the strict monotonicity of V in p, we have

$$\mathcal{L}_M (r_c, r_s) = \max_{p \in [r_s, r_c]} [V(0, 0, s) - V(-p, 1, s)] = V(0, 0, s) - V(-r_c, 1, s)$$

It follows that

$$\frac{d\mathcal{L}_M}{dr_c} = V_m(r_c, 1, s),$$

and accordingly that a first-order approximation around $r = r_s$ (for $r > r_s$) is given by

$$\frac{1}{\nu_s} \mathcal{L}_M (r, r_s) \approx r - r_s$$

Thus we have

$$\frac{1}{\nu_s} \mathcal{L}_M (r, r_s) \approx |r - r_s|$$

Part (ii)

Suppose $r_s \geq r_c$. Fixing r_s and treating $\mathcal{L}_W$ as a function of r_c, we have

$$\frac{d\mathcal{L}_W}{dr_c} = -[V(-r_c, 1, s) - V(0, 0, s)] h(r_c)$$

Notice that, by the definition of r_s,

$$\left. \frac{d\mathcal{L}_W}{dr_c} \right|_{r_c = r_s} = -[V(-r_s, 1, s) - V(0, 0, s)] h(r_s) = 0,$$

so the second-order approximation does not include a first-order term. Next we have

$$\frac{d^2\mathcal{L}_W}{dr_c^2} = V_m\left(-r_c, 1, s\right) h(r_c) - \left[V\left(-r_c, 1, s\right) - V\left(0, 0, s\right)\right] h_p(r_c)$$

from which it follows that

$$\left.\frac{1}{\nu_s} \frac{d^2\mathcal{L}_W}{dr_c^2}\right|_{r_c=r_s} = -\eta_s$$

The claim follows immediately.

The argument for $r_c \leq r_s$ is analogous.

□

A.2   Proof of Proposition 2

Proof. We begin by assessing the following limit:

$$\lim_{\alpha \to 0} \left(\frac{r_c^U\left(y_c^\alpha\right) - r_s^U\left(y_s^\alpha\right)}{r_c^V\left(y_c^\alpha\right) - r_s^V\left(y_s^\alpha\right)}\right)$$

Because the numerator and denominator both converge to zero, we apply L'Hopital's rule:

$$\lim_{\alpha \to 0} \left(\frac{r_c^U\left(y_c^\alpha\right) - r_s^U\left(y_s^\alpha\right)}{r_c^V\left(y_c^\alpha\right) - r_s^V\left(y_s^\alpha\right)}\right) = \frac{\lim_{\alpha \to 0} \frac{\partial r_c^U(y_c^\alpha)}{\partial \alpha} - \lim_{\alpha \to 0} \frac{\partial r_s^U(y_s^\alpha)}{\partial \alpha}}{\lim_{\alpha \to 0} \frac{\partial r_c^V(y_c^\alpha)}{\partial \alpha} - \lim_{\alpha \to 0} \frac{\partial r_s^V(y_s^\alpha)}{\partial \alpha}}$$

Recall that $r_c^V\left(y_c^\alpha\right)$ is defined by the following equation:

$$V\left(-r_c^V\left(y_c^\alpha\right), \phi\left(\alpha y_c + (1-\alpha)\, y_0\right)\right) = V\left(0, \phi\left(y_0\right)\right)$$

We differentiate implicitly to obtain:

$$\frac{\partial r_c^V\left(y_c^\alpha\right)}{\partial \alpha} = \frac{V_\phi\left(-r_c^V\left(y_c^\alpha\right), \phi\left(\alpha y_c + (1-\alpha)\, y_0\right)\right)}{V_m\left(-r_c^V\left(y_c^\alpha\right), \phi\left(\alpha y_c + (1-\alpha)\, y_0\right)\right)} \left(\nabla\phi\left(\alpha y_c + (1-\alpha)\, y_0\right) \cdot \left(y_c - y_0\right)\right)$$

Using the fact that $\lim_{\alpha \to 0} r_c^V\left(y_c^\alpha\right) = 0$, it follows that

$$\lim_{\alpha \to 0} \frac{\partial r_c^V\left(y_c^\alpha\right)}{\partial \alpha} = \frac{V_\phi\left(0, \phi\left(y_0\right)\right)}{V_m\left(0, \phi\left(y_0\right)\right)} \left(\nabla\phi\left(y_0\right) \cdot \left(y_c - y_0\right)\right)$$

An analogous calculation reveals that

$$\lim_{\alpha \to 0} \frac{\partial r_s^V\left(y_s^\alpha\right)}{\partial \alpha} = \frac{V_\phi\left(0, \phi\left(y_0\right)\right)}{V_m\left(0, \phi\left(y_0\right)\right)} \left(\nabla\phi\left(y_0\right) \cdot \left(y_s - y_0\right)\right)$$

Therefore,

$$\lim_{\alpha \to 0} \frac{\partial r_c^V (y_c^\alpha)}{\partial \alpha} - \lim_{\alpha \to 0} \frac{\partial r_s^V (y_s^\alpha)}{\partial \alpha} = \frac{V_\phi (0, \phi (y_0))}{V_m (0, \phi (y_0))} (\nabla \phi (y_0) \cdot (y_c - y_s))$$

Repeating these calculations for U, we obtain

$$\lim_{\alpha \to 0} \frac{\partial r_c^U (y_c^\alpha)}{\partial \alpha} - \lim_{\alpha \to 0} \frac{\partial r_s^U (y_s^\alpha)}{\partial \alpha} = \frac{U_\phi (0, \phi (y_0))}{U_m (0, \phi (y_0))} (\nabla \phi (y_0) \cdot (y_c - y_s))$$

Accordingly, we have

$$\lim_{\alpha \to 0} \left(\frac{r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha)}{r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha)} \right) = \frac{U_\phi (0, \phi (y_0)) V_m (0, \phi (y_0))}{U_m (0, \phi (y_0)) V_\phi (0, \phi (y_0))} \equiv K > 0$$

Notice that K does not depend on y_s or y_c (even though the limits of the numerator and denominator do).

For $i = M$, we have

$$\lim_{\alpha \to 0} \frac{\left| r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha) \right|}{\left| r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha) \right|} = \lim_{\alpha \to 0} \left| \frac{r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha)}{r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha)} \right| = \left| \lim_{\alpha \to 0} \frac{r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha)}{r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha)} \right| = K$$

For $i = W$, we have

$$\lim_{\alpha \to 0} \frac{\eta_s \left(r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha) \right)^2}{\eta_s \left(r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha) \right)^2} = \lim_{\alpha \to 0} \left(\frac{r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha)}{r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha)} \right)^2 = \left(\lim_{\alpha \to 0} \frac{r_c^U (y_c^\alpha) - r_s^U (y_s^\alpha)}{r_c^V (y_c^\alpha) - r_s^V (y_s^\alpha)} \right)^2 = K^2$$

$\square$

A.3   Proof of Proposition 3

Proof. Proceeding as in the proof of Proposition 2, we see that

$$\lim_{\alpha \to 0} \left(\frac{L_M^U (y_{cI\theta}^\alpha, y_{sI}^\alpha)}{L_M^V (y_{cI\theta}^\alpha, y_{sI}^\alpha, \theta)} \right) = \frac{U_\phi (0, \phi (y_0)) V_m (0, \phi (y_0), \theta)}{U_m (0, \phi (y_0)) V_\phi (0, \phi (y_0), \theta)} \tag{7}$$

Moreover, a step in the derivation of the preceding formula establishes that:

$$\lim_{\alpha \to 0} \frac{\partial r_s^V (y_s^\alpha, \theta)}{\partial \alpha} = \frac{V_\phi (0, \phi (y_0), \theta)}{V_m (0, \phi (y_0), \theta)} (\nabla \phi (y_0) \cdot (y_s - y_0)) \tag{8}$$

Because $r_s^V (y_s^0) = 0$, we can write $\lim_{\alpha \to 0} \frac{\partial r_s^V (y_s^\alpha, \theta)}{\partial \alpha} = \lim_{\alpha \to 0} \frac{r_s^V (y_s^\alpha, \theta)}{\alpha}$. Defining $\kappa(y_s) \equiv \frac{U_\phi (0, \phi(y_0))}{U_m (0, \phi(y_0))} \nabla \phi (y_0) \cdot (y_s - y_0)$, equation (8) then implies

$$\lim_{\alpha \to 0} \frac{\alpha \kappa(y_s)}{r_s^V (y_{sI}^\alpha, \theta)} = \frac{U_\phi (0, \phi (y_0)) V_m (0, \phi (y_0), \theta)}{U_m (0, \phi (y_0)) V_\phi (0, \phi (y_0), \theta)} \tag{9}$$

Using equations (8) and (9), the rest of the steps are essentially the same as for Proposition 2.

For settings in which y is a scalar, we have $\kappa(y_s) \equiv \kappa^* (y_s - y_0)$ for $\kappa^* = \frac{U_\phi (0, \phi(y_0))}{U_m (0, \phi(y_0))} \phi' (y_0)$.

$\square$

3

B   Experiment details

In this section we detail the implementation of the experiment. Screenshots of the instructions and the experimental interface are in Appendix D.

Amazon Mechanical Turk   Workers log on to AMT through an interface that displays a list of *Human Intelligence Tasks* (HITs), each with a title, an estimated duration, and an estimated remuneration rate. Other HITs include taking surveys, categorizing images, writing product descriptions, and identifying performers on music recordings.

To ensure that subjects were *technically* able to view the videos, we told them at the outset of the study that access to youtube.com was required. We also asked them to reproduce the last word spoken in the welcome video, and the last word of the title slide of whichever treatment video they viewed. Subjects who were not able to complete these tasks correctly were not allowed to continue with the study. The videos were embedded in the survey so that subjects could not find the other treatment videos used in this study.

We ensured that each subject participated in our study only once using the unique identifying numbers assigned by AMT.[58] A subject can only receive payment for participation in the study if she correctly provides this information, and hence has no incentive for misrepresentation.

Initial Financial Literacy   Before participating in the main stages of the experiment, subjects completed the unincentivized financial literacy test in Table B.1. This test of financial literacy originated with Lusardi and Mitchell (2009) and van Rooij et al. (2011), and has been used in many other studies (Lusardi and Mitchell, 2014).

Attention to the Video   Before subjects watched the treatment video, we informed them that, with 25% probability, their earnings would be entirely determined by their performance on a test,[59] and that 'to be able to answer the questions in the test, you need to both understand and know the contents of the video.' We also explained that the video could help them make better decisions both during the experiment and in real life, inasmuch as it was made by 'internationally recognized academic experts on financial decision making.' Finally, we disabled the *continue* button for the duration of the video.

Iterated Multiple Price List   Each line of each price list was a binary choice between the future reward and a specified dollar amount to be received no more than two days after completion of the experiment. For the first price list, the immediate payment varied from $0 to $20 in increments of $2. For the second price list, it varied from x to $(x + 1.8)$ in increments of $0.20, where $x + 2$ is the smallest amount chosen over the future reward in the first list. (See appendix D for screenshots of the computer interface.) If a subjects'

---

[58]Nonetheless, one subject managed to participate in our study twice. Both times, this subject exhibited multiple switching points, and hence is excluded from all analyses.

[59]Hastings et al. (2013) criticize most existing studies that use such test scores as outcome measures on the grounds that the tests are unincentivized. One of the few exceptions is Levy and Tasoff (2016).

payment was determined according to a price list, the randomization over lines proceeded as follows. A line was randomly selected from the first price list. If that line did not correspond to x (defined above), it was implemented. Otherwise, a random line from the second price list was selected, and the decision for that line was implemented. With this procedure, truthful revelation of preferences is optimal.

Questionnaire Questions concerning decision strategies employed the following wording. Use of the rule of 72 in complexly framed problems: "Sometimes in this experiment, you were given a choice such as 'We will invest $10 in an account with 1% interest per week. Interest is compounded weekly. We will pay you the proceeds in 72 days.' When deciding about this choice, did you use the rule of 72?"[60] Use of the rule of 72 in simply framed problems: "Sometimes in this experiment, you were given a choice such as 'We will pay you $20 in 36 days.' When deciding about such a choice, did you use the rule of 72?" In both cases, subjects answered either "Yes", "No", or "I don't know the rule of 72." Number of problems for which the future reward was calculated explicitly: "In total, you were given 10 rounds in which one of the options was something like 'we will invest $... in an account with ...% interest per day. Interest is compounded daily. We will pay you the proceeds in... days.' Out of these 10 rounds, how many times did you explicitly calculate the money amount that this investment would yield within the specified time?" Subjects responded by selecting an integer between 0 and 10. Use of external help on the test: "When you completed the test about the video on financial investing, did you use external resources (such as other websites, books, etc.) to find the right answers?" Subjects answered either "Yes" or "No."

We also asked subjects how much attention they had paid to their choices, how much attention they had paid to the video, whether they had any suggestions about the study, and whether they had experienced any technical difficulties. The overwhelming majority of subjects reported the highest level of attention in answer to both questions—a finding we interpret with caution.

---

[60]The survey question incorrectly described the interest rate as pertaining to a week rather than a day. We believe the meaning of the question was nevertheless clear despite this typo.

FL1. Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow?
More than $102 (92.86%), Exactly $102 (3.37%), Less than $102 (1.98%), Do not know (1.79%)

FL2. Suppose you had $100 in a savings account and the interest rate is 20 percent per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?
More than $200 (72.62%), Exactly $200 (22.62%), Less than $200 (2.98%), Do not know (1.79%)

FL3. Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, how much would you be able to buy with the money in this account?
More than today (8.33%), exactly the same (6.94%), less than today (1.15%), do not know (3.57%)

FL4. Assume a friend inherits $10,000 today and his sibling inherits $10,000 3 years from now. Who is richer because of the inheritance?
My friend (55.36%), his sibling (9.13%), they are equally rich (29.37%), do not know (6.15%)

FL5. Suppose that in the year 2015, your income has doubled and prices of all goods have doubled too. In 2015, how much will you be able to buy with your income?
More than today (4.76%), the same (89.29%), less than today (4.76%), do not know (1.19%)

Table B.1: Financial Literacy questionnaire. This questionnaire was administered to subjects at the beginning of the survey. Numbers in brackets indicate the percentage of subjects who chose a given answer.

Test questions in Control interventions Panels A and B of Table B.2 display the test questions about the Control intervention in Experiments A and B, respectively. The two experiments involve different sets of questions about their respective Control interventions because the Control interventions differ. We decided to use a different Control intervention for Experiment B because the Control intervention in Experiment A is largely descriptive, and hence is not well-suited to incorporating practice questions with individualized feedback.

Experiment A

Which of the following quotes is attributed to Benjamin Franklin?
Compound interest is the most powerful force in the universe; Youth is wasted on the young; Money makes money. And the money that money makes, makes money;

Which quote is attributed to the author Upton Sinclair?
Only liars manage always to be out of the market during bad times and in during good times; It is difficult to get a man to understand something when his salary depends upon his not understanding it; There are three classes of people who do not believe that markets work: the Cubans, the North Koreans, and active managers; Nobody knows more than the market.

What percentage of mutual funds tends to be outperformed by the market (S&P 500 Index) each year?
between 10 and 30% between 30 and 50% between 50 and 70% between 70 and 90%

What is an "indexing" investment strategy?
Buying index funds, which hold assets that have been indexed as particularly profitable by financial experts; Buying index funds, which hold stocks of companies that provide information about the stock market as a whole (stock market indices); Buying index funds, which hold the market portfolio; Buying index funds, which hold optimally diversified, custom tailored portfolios.

Professional investors as a whole are responsible for what percentage of stock market trading?
30%; 50%; 70%; 90%.

Experiment B

In order to limit your risk, you might invest in which of the following pairs of stocks?
Microsoft and Google; General Motors and Chrysler; Coca-Cola and Pepsi; General Motors and Microsoft; Facebook and Twitter.

We would expect the degree of relation between the returns of Coca-Cola stock and the returns of Pepsi stock to be closest to ____? [-1 means perfect negative relation and +1 means perfect positive relation]:
-0.7; -0.3; 0; 0.3; 0.7.

Considering a long time period (for example 10 or 20 years), which asset normally gives the highest return?
Savings accounts; Corporate bonds; Government bonds; T-Bills; Stocks.

Normally, which asset displays the highest fluctuations over time?
Savings accounts; Corporate bonds; Government bonds; T-Bills; Stocks.

A degree of relation of ____between two assets will NOT help reduce your risk.
1; 0.5; 0; -0.5; -1.

Table B.2: Test questions concerning the Control interventions. Questions were displayed in individually randomized order.

C   Additional Data Analysis

C.1   Demographics

Table C.1 presents detailed demographics of our subject pool by treatment, as well as their initial financial literacy.[61] Column 5 lists data for the representative US citizen. Demographic variables are taken from the

---

[61]These statistics only include subjects who did not exhibit multiple switching points in any of the price lists.

2010 US Census. Employment variables are for April 2014, and come from the Bureau of Labor Statistics. Financial literacy scores are from Lusardi (2011), and from the 2012 FED bulletin for stock holdings.[62] (Representative data on financial literacy only exist for questions FL1 and FL3.) For empty cells, no representative data are available. Column 6 reports, for each variable, the p-value of an F-test for differences across treatments. The number of significant differences is well within the range we would expect given the number of tests performed.

As reported in section 3.2, our sample is poorer, better educated, and more likely to live in larger households than the average US citizen. While the incidence of full-time employment in our sample mirrors that of the general population, the fraction of respondents who classify themselves as employed part-time is double that of the general population. Our subjects are also disproportionately male and white, younger, slightly more urban, and more likely to have never been married than the representative US citizen.

C.2   Main results controlling for demographics

Table C.2 presents our main results in a regression that includes data from both experiments and controls for demographics. Demographic controls consist of all variables listed in Table C.1, except for the summary statistics "FL1-FL3 all correct" and "FL1-FL5 all correct". For brevity, we pool across the timeframes.

In each case we see that coefficient estimates are barely changed in comparison to the estimates in the main text, which do not control for demographic characteristics. We conclude that the differences between experiments A and B reflect differences in the interventions rather than differences in subject characteristics.

---

[62] J. Bricker, A. B. Kennickell K. B. Moore, and J. Sabelhaus, 2012, *Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances*, Federal Reserve Bulletin, 98(2).

[63] In our survey, household income is interval coded. The values stated are the midpoints of the median intervals.

[64] Percentage of civilian noninstitutional population that is full-time employed.

[65] Percentage of civilian noninstitutional population that is part-time employed.

[66] Our questionnaire included the option "Prefer not to say". The three subjects who chose this response are not accounted for in this table.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Experiment A				Experiment B		
Treatment	Treat.	Cont.	Subst. only	Rhet. only	p-value	Treat.	Cont.	p-value	US
FL1	0.93	0.92	0.92	0.95	0.83	0.90	0.92	0.59	0.65
FL2	0.81	0.73	0.73	0.71	0.32	0.75	0.80	0.29	-
FL3	0.82	0.82	0.83	0.85	0.93	0.88	0.85	0.38	0.64
FL4	0.58	0.64	0.50	0.59	0.17	0.53	0.59	0.31	-
FL5	0.96	0.90	0.87	0.91	0.09*	0.91	0.88	0.30	-
FL1 - FL3 all correct	0.71	0.63	0.62	0.62	0.44	0.68	0.70	0.72	-
FL1 - FL5 all correct	0.47	0.45	0.34	0.40	0.20	0.40	0.42	0.60	-
Male	0.57	0.57	0.61	0.50	0.40	0.47	0.53	0.28	0.49
Age (median)	28	32	29	29	0.09*	36	34	0.11	37.2
Household Income (median)[63]	45,000	35,000	45,000	35,000	0.06**	45,000	57,500	0.15	53,046
Race									
African-american	0.08	0.06	0.08	0.04	0.68	0.05	0.10	0.10*	0.13
Asian	0.08	0.11	0.12	0.05	0.22	0.06	0.06	0.90	0.05
Caucasian	0.81	0.72	0.72	0.77	0.34	0.82	0.79	0.41	0.63
Hispanic	0.03	0.07	0.03	0.10	0.06**	0.04	0.04	0.88	0.17
Other	0.01	0.04	0.05	0.04	0.44	0.02	0.01	0.37	0.02
Education									
Less than high school	0.01	0.00	0.00	0.00	0.35	0.00	0.00	1.00	0.14
High school	0.13	0.12	0.15	0.14	0.92	0.14	0.13	0.71	0.31
Vocational / technical	0.08	0.08	0.08	0.03	0.29	0.07	0.06	0.56	0.09
Some college	0.35	0.37	0.33	0.44	0.34	0.31	0.34	0.67	0.19
College	0.39	0.37	0.38	0.34	0.90	0.37	0.37	0.97	0.18
Graduate degree	0.05	0.06	0.07	0.05	0.88	0.11	0.11	0.88	0.09
Employment									
Full time employed	0.50	0.50	0.48	0.43	0.70	0.61	0.68	0.16	0.48[64]
Part time employed	0.21	0.23	0.26	0.27	0.72	0.18	0.16	0.60	0.11[65]
Marital Status[66]									
Married	0.28	0.30	0.32	0.29	0.94	0.46	0.47	0.80	0.27
Widowed	0.00	0.00	0.00	0.00	1.00	0.01	0.01	0.53	0.56
Divorced	0.07	0.05	0.04	0.04	0.80	0.04	0.06	0.37	0.06
Never married	0.64	0.65	0.64	0.64	1.00	0.49	0.46	0.69	0.10
Urban / Rural									
Urban and suburban	0.17	0.17	0.11	0.17	0.48	0.25	0.17	0.06*	0.81
Rural	0.83	0.83	0.89	0.83	0.48	0.75	0.83	0.06*	0.19
Household size									
1	0.18	0.13	0.11	0.19	0.26	0.10	0.13	0.33	0.22
2	0.22	0.24	0.25	0.24	0.95	0.27	0.23	0.42	0.36
3	0.14	0.19	0.17	0.22	0.46	0.18	0.22	0.35	0.17
4 or more	0.46	0.44	0.47	0.35	0.23	0.45	0.41	0.50	0.26
Owns stocks	0.16	0.23	0.20	0.23	0.54	0.43	0.39	0.44	0.15
N	109	106	128	112	-	169	179	-	-

Table C.1: Demographics and financial literacy. The sample includes all subjects who completed the study and did not exhibit any multiple switching points. Column 5 presents the p-values of an F-test for joint equality of the coefficients listed in columns 1 - 4. Column 8 presents the p-value of a t-test for joint equality of the coefficients in columns 6 - 7. Column 9 lists comparison values for the representative US citizen whenever available. See text for data sources.

VARIABLES	(1) (2) Test scores on module		(3) (4) Valuation in frame		(5) (6) Deliberative Competence	
	Treatment	Control	Complex	Simple		
Correction for changes in valuations in simple frame					No	Yes
Difference to Control B						
Treatment B	1.640***	-0.375***	15.067***	7.242***	7.901***	12.883***
	(0.139)	(0.124)	(2.790)	(2.164)	(1.773)	(2.563)
Treatment A	1.528***	-0.660***	11.928***	1.522	1.133	-1.283
	(0.197)	(0.158)	(3.969)	(3.189)	(2.521)	(4.652)
Control A	0.117	0.405**	-1.504	1.121	0.832	3.151
	(0.201)	(0.174)	(3.740)	(3.059)	(2.430)	(3.309)
Substance-Only	1.400***	-0.886***	3.406	2.158	2.525	-0.418
	(0.190)	(0.164)	(3.789)	(3.086)	(2.258)	(5.042)
Rhetoric-Only	0.608***	-0.651***	16.741***	6.611**	5.247**	7.755**
	(0.210)	(0.165)	(4.082)	(3.148)	(2.259)	(3.386)
Demographic controls	Yes	Yes	Yes	Yes	Yes	Yes
p-value						
Control A = Treatment A	0.000	0.000	0.000	0.890	0.894	0.277
Observations	803	803	8,030	8,030	8,030	8,020
Subjects	803	803	803	803	803	802

Table C.2: Main results controlling for demographics in joint analysis of both experiments. Each column corresponds to a separate regression. Column 6 omits the subject for whom $r_S^{i,R,t} = 0$ for all instruments. Standard errors in parentheses, clustered on the subject level. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

C.3 Effects on individual test questions

We analyze the effect of the treatments on answers to individual test questions in table C.3. The test questions differ by how closely they follow the material in the education intervention, and by how easily they are answered without knowledge of the rule of 72.

Q1 is the only question for which the answer was explicitly given in the education video (including in the Substance-Only treatment but not in the Rhetoric-Only treatment). The video also discussed an example that is similar, but not identical, to Q2.[67]

The remaining questions require more flexible thinking. Q3 and Q4 can easily be answered with the rule of 72. Knowledge of this rule, however, is not necessary to answer these questions correctly. Q3 can be answered by iteratively multiplying a starting value with 1.07, and counting the number of iterations required for the amount to increase to the desired value. Likewise, Q4 can be answered by calculating the factor by which an investment grows within 8 years at 9 percent interest (either iteratively, or using the compound interest formula), and then dividing 500 by this number. Q5 is a standard compound interest calculation, and parallels the calculations that need to be made in the complexly framed decision problems.

---

[67]The example is: "To double your money in 10 years, what rate of return do you need? The answer: 10 times $X = 72$, so $X = 7.2$ percent."

Table C.3 displays the treatment effects on the success rates for each of these questions. Baseline rates of correct answers are highly similar across the two experiments. Moreover, in both experiments, the significant effect of the Full and Substance-Only treatments on the total score derive from questions Q1, Q2, and Q5. The fact that performance in Q5 increased in these treatments is reassuring, as it demonstrates that the increase in test scores is at least partly due to subjects' increased ability to analyze previously unseen problems properly. Moreover, while treatment effects are similar across the experiments for questions Q1 to Q4, the treatment effect on Q5 in Experiments B is more than double that in Experiment A, tentatively hinting at our finding that our intervention in Experiment B is more effective than that in Experiment A.

Question	Q1	Q2	Q3	Q4	Q5
	Experiment A				
Treatment effects					
Treatment	0.566***	0.619***	0.062	0.021	0.174***
	(0.054)	(0.053)	(0.068)	(0.068)	(0.067)
Substance-Only	0.584***	0.592***	-0.037	0.023	0.109*
	(0.051)	(0.053)	(0.065)	(0.065)	(0.065)
Rhetoric-Only	0.072	0.191***	0.067	0.114*	0.050
	(0.065)	(0.061)	(0.067)	(0.067)	(0.067)
Level in Control	0.330***	0.220***	0.514***	0.422***	0.477***
	(0.045)	(0.040)	(0.048)	(0.047)	(0.048)
Observations	455	455	455	455	455
	Experiment B				
Treatment effect	0.559***	0.696***	0.045	-0.075	0.375***
	(0.042)	(0.038)	(0.054)	(0.052)	(0.048)
Level in Control	0.346***	0.168***	0.464***	0.436***	0.436***
	(0.036)	(0.028)	(0.037)	(0.037)	(0.037)
Observations	348	348	348	348	348

Table C.3: Fraction of correct responses on individual questions in the test about the Treatment intervention. Standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

C.4 Self-reported behavior

The study ends with a brief non-incentivized questionnaire. We ask 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 elicit the number of complexly framed valuation tasks for which subjects explicitly calculated the future value of the investment, and ask whether they obtained help when taking the test on compound interest. The questionnaire also addresses a small number of additional issues.

Subjects in the Control condition report similar numbers of decisions for which they engaged in explicit calculations in each of the experiments (6.4 and 6.7 in Experiments A and B, respectively). The Treatment condition significantly increases that number, by 1.7 problems in Experiment A (column 1, $p < 0.01$) and

by 2.6 problems in Experiment B (column 2, $p < 0.01$). The treatment effect on the fraction of subjects reporting to have used the rule of 72 in their decision making in complexly framed decisions does not differ substantially across experiments (57.9% and 60.5% in Experiments A and B, respectively). There is, however, a difference in levels. Only 12.8% of subjects in the Control condition of Experiment A report using the rule, whereas 31.8% of subjects in the Control condition of Experiment B do so.

As expected, the fraction of subjects reporting to have used the rule of 72 for simply framed problems is substantially smaller; averaging 9.2% and 22.3% in the Control conditions of experiments A and B, respectively. In both experiments the Treatment condition increases the frequency of such reports, but does so almost twice as much in Experiment A (by 17.2 percentage points) than in Experiment B (by 9.6 percentage points). Finally, when asked about the use of external help with the test questions at the end of the experiment, we do not find treatment effects in either experiment, although the fraction of subjects reporting the use of such help exceeds 20% in Experiment A, whereas it is lower than 8% in experiment B.

Unlike performance on test scores and directional behavioral changes, these self-reported behaviors suggest that the effects of the Treatment interventions differ across the experiments, though that interpretation is complicated by the fact that baseline levels differ across the experiments. Like the conventional measures, however, data on self-reported behavior suggest that the Treatment interventions are effective in either experiment.

Self-report	(1) (2) Engages in explicit calculation		(3) (4) Uses of rule of 72 in complex frame		(5) (6) Uses of rule of 72 in simple frame		(7) (8) External help with test	
Experiment	A	B	A	B	A	B	A	B
Levels								
Control	6.404***	6.693***	0.128***	0.318***	0.092***	0.223***	0.220***	0.078***
	(0.377)	(0.277)	(0.032)	(0.035)	(0.028)	(0.031)	(0.040)	(0.020)
Treatment	8.142***	9.331***	0.708***	0.923***	0.264***	0.320***	0.208***	0.059***
	(0.342)	(0.209)	(0.044)	(0.021)	(0.043)	(0.036)	(0.040)	(0.018)
Difference	1.738***	2.639***	0.579***	0.605***	0.172***	0.096**	-0.013	-0.019
	(0.509)	(0.347)	(0.055)	(0.041)	(0.051)	(0.048)	(0.056)	(0.027)
Observations	215	348	215	348	215	348	215	348

Table C.4: Self-reported behavior. Each column displays the coefficients of a separate OLS regression. Standard errors in parentheses. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

C.5 Deliberative Competence based on expected welfare loss

Here, we present our main empirical results on Deliberative Competence using the measure $DC_W^{i,R,t} = -\left(r_c^{i,R,t} - r_s^{i,R,t}\right)^2$ which approximates the average rather than the maximal loss from characterization failure.

Table C.5 replicates Table 5 with this alternative measure. Panel A shows that the intervention in Experiment A leaves $DC_W^{i,R,t}$ nearly unchanged on average whereas the intervention in Experiment B leads

to a substantial and statistically highly significant increase both in the pooled sample and separately within each timeframe.

Panel B applies Proposition 3 to correct for changes in valuations in the simple frame.[68] Again, we find that the intervention in Experiment A, if anything, harms subjects, whereas the intervention in Experiment B significantly increases their welfare. The relative magnitude of these effects differs from those in Table 5. Here, we find that the harm caused by the intervention in experiment A is of a similar magnitude as the benefits caused by the intervention in Experiment B, whereas in Panel B of Table 5 the benefits of the intervention in Experiment B exceed the magnitude of the harm in Experiment A severalfold. One reason for this divergence is the stronger sensitivity of $DC_W^{i,R,t}$ to large valuation differences, $r_C^{i,R,t} - r_S^{i,R,t}$.

---

[68]Proceeding as in Table 5, we apply a version of equation (5) that accounts for noise in the elicitation of valuations in the simple frame. Specifically, we calculate the mean valuation for simply framed choices for each timeframe and use the square of the resulting average as the correction factor. Moreover, by our normalization, $y_{sI} - y_0 = 1$ for all instruments I.

A. Deliberative Competence

	(1) both	(2) both	(3) 72	(4) 72	(5) 36	(6) 36
Delay in days	both	both	72	72	36	36
Experiment	A	B	A	B	A	B

Levels

	(1)	(2)	(3)	(4)	(5)	(6)
Control	-11.693***	-12.713***	-11.564***	-12.972***	-11.822***	-12.453***
	(1.233)	(1.069)	(1.377)	(1.136)	(1.247)	(1.098)
Treatment	-11.848***	-8.646***	-11.703***	-8.176***	-11.993***	-9.116***
	(1.622)	(0.906)	(1.588)	(0.895)	(1.966)	(0.996)
Difference	-0.155	4.067***	-0.139	4.797***	-0.171	3.337**
	(2.038)	(1.401)	(2.101)	(1.447)	(2.327)	(1.483)
Observations	2,150	3,480	1,075	1,740	1,075	1,740
Subjects	215	348	215	348	215	348

B. Deliberative Competence corrected for changes in simply framed valuations

	(1) both	(2) both	(3) 72	(4) 72	(5) 36	(6) 36
Delay in days	both	both	72	72	36	36
Experiment	A	B	A	B	A	B

Levels

	(1)	(2)	(3)	(4)	(5)	(6)
Control	-0.232***	-0.292***	-0.261***	-0.307***	-0.203***	-0.276***
	(0.027)	(0.034)	(0.041)	(0.036)	(0.023)	(0.044)
Treatment	-0.416***	-0.174***	-0.397***	-0.177***	-0.436***	-0.171***
	(0.102)	(0.025)	(0.094)	(0.025)	(0.136)	(0.028)
Difference	-0.184*	0.118***	-0.136	0.130***	-0.233*	0.105**
	(0.105)	(0.042)	(0.103)	(0.044)	(0.138)	(0.052)
Observations	2,150	3,470	1,075	1,735	1,075	1,735
Subjects	215	347	215	347	215	347

Table C.5: Deliberative Competence. Each column displays the coefficients of a separate OLS regression of Deliberative Competence, $DC_W^{i,R,t} = -\left(r_c^{i,R,t} - r_s^{i,R,t}\right)^2$, on treatment indicators. Standard errors in parentheses, clustered by subject. The reason for the smaller number of observations in Panel B in Experiment B is one subject who consistently made choices consistent with a valuation of zero in the simple frame. As the correction consists in dividing by simply framed valuations, this subject is excluded from that analysis. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

Table C.6 replicates Table 7 using using the measure $DC_W^{i,R,t} = -\left(r_c^{i,R,t} - r_s^{i,R,t}\right)^2$. As in Table 7, we find that the intervention in Experiment A harms subjects in the lowest quartile of valuations in the simple frame ($p < 0.1$), and has beneficial effects for subjects in the highest quartile ($p < 0.01$). Also paralleling the result in Table 7, the intervention in Experiment B does not harm subjects in any quartile, but has substantially positive effects for subjects in the second-highest ($p < 0.1$) and highest quartiles ($p < 0.01$). We conclude that our inferences regarding practice and feedback are not driven by an unintended relationship between

welfare weights and biases that impact simply framed valuations, also when Deliberative Competence is measured by $DC_W^{i,R,t} = -\left(r_c^{i,R,t} - r_s^{i,R,t}\right)^2$.

	Experiment A			
VARIABLE	Deliberative Competence			
	Quartile simply framed valuation			
	1	2	3	4
Levels				
Control	-6.135***	-9.003***	-13.230***	-18.164***
	(1.406)	(2.201)	(2.620)	(2.735)
Treatment	-15.605***	-9.982***	-10.279***	-11.605***
	(5.189)	(2.100)	(1.728)	(2.792)
Effect	-9.470*	-0.979	2.951	6.558*
	(5.376)	(3.042)	(3.139)	(3.908)
Observations	2,150			
Subjects	215			

	Experiment B			
VARIABLE	Deliberative Competence			
	Quartile simply framed valuation			
	1	2	3	4
Levels				
Control	-6.755***	-11.928***	-12.102***	-19.933***
	(1.338)	(1.910)	(1.698)	(2.813)
Treatment	-7.078***	-10.033***	-7.906***	-9.546***
	(1.831)	(1.748)	(1.502)	(2.074)
Effect	-0.322	1.895	4.196*	10.387***
	(2.268)	(2.590)	(2.267)	(3.495)
Observations	3,480			
Subjects	348			

Table C.6: Effect of treatments on Deliberative Competence, $DC_W^{i,R,t} = -\left(r_c^{i,R,t} - r_s^{i,R,t}\right)^2$, by quartiles of simply framed valuations. Each panel presents the output of a single OLS regression. Standard errors in parentheses, clustered by subject. ***$p < 0.01$,***$p < 0.05$,*$p < 0.1$.

C.6 Valuation difference compared to noise in the simple frame

Here we investigate the possibility that the measured overestimation of compound interest exhibited by a fraction of our subjects is solely attributable to elicitation noise. To test the hypothesis, we estimate the amount of noise in simply framed decisions. We then compare the frequency of excessive valuations we would observe based on that amount of noise alone to the frequency of excessive valuations we actually observe in the complex frame. Specifically, we calculate, for each subject and each timeframe $t \in \{36, 72\}$, the values $\Delta_1^t = r_S^{j,20,t} - r_S^{j,18,t}$, $\Delta_2^t = r_S^{j,18,t} - r_S^{j,20,t}$ $\Delta_3^t = r_S^{j,16,t} - r_S^{j,14,t}$, and $\Delta_4^t = r_S^{j,14,t} - r_S^{j,16,t}$ (recall that valuations $r_S^{i,R,t}$ involve the normalization of the future value to \$1). Figure C.1 superimposes the CDF of Δ_k^t (pooled across $k \in \{1,..,4\}$ and $t \in \{36, 72\}$) on the CDF of the valuation difference from Figure 3. If excessive

	Experiment A	Experiment B
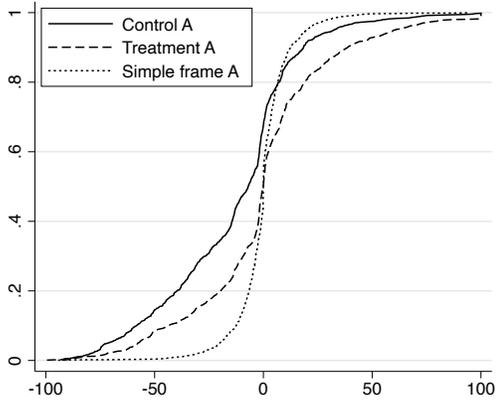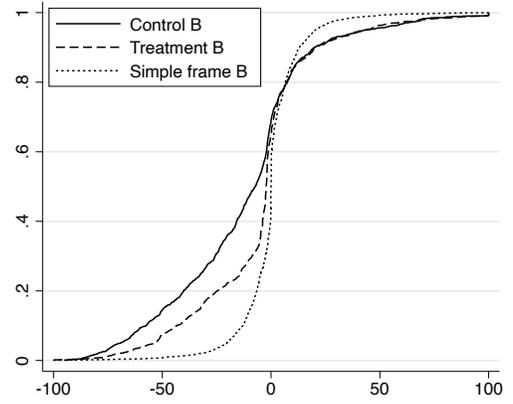
Experiment A   Experiment B

Figure C.1: Replication of CDFs from Figure 3 with a measure of the distribution of noise in the simple frame superimposed.

valuations in the complex frame were solely attributable to elicitation noise, the CDF of Δ_k^t should coincide with the CDF of $r_C^{i,R,t}$ to the left of zero. By contrast, we find substantial differences between these curves in each experiment, and especially for the treatment in Experiment A. Accordingly, elicitation noise alone cannot explain the overestimation of compound interest in either condition of either experiment.

D Instructions

This is a research study run by the department of economics at Stanford University.

IMPORTANT

This study may take up to ONE AND A HALF HOURS to complete. Please start this study only if you do have that much time in a single session.

If you do not complete the study, or if the HIT times out on you, we will not be able to pay you. (The HIT is set to time out in 3 hours.)

You will earn $10 just for completing this study. In addition, you will receive up to $20, depending on the decisions you make in this study.

Do not start this study if you do not have access to youtube.com. Some browsers will block embedded videos. Please make sure your browser will display them, as you may otherwise not be able to complete this study.

Click here to start the study: https://stanforduniversity.qualtrics.com/SE/?SID=SV_0GPXNo1f9TX5YlR

Provide the survey code here:

WELCOME

This is a research study run by the department of economics at Stanford University.

IMPORTANT

This study may take up to ONE AND A HALF HOURS to complete. Please start this study only if you do have that much time in a single session.

If you do not complete the study, or if the HIT times out on you, we will not be able to pay you. (The HIT is set to time out in 3 hours.)

You will earn $10 just for completing this study. In addition, you will receive up to $20, depending on the decisions you make in this study.

Do not start this study if you do not have access to youtube.com. Some browsers will block embedded videos. Please make sure your browser will display them.

By clicking the button below, you consent to participating in this research study.

Questions, Concerns, or Complaints: If you have any questions, concerns or complaints about this research study, its procedures, risks and benefits, you should ask the Protocol Director, Sandro Ambuehl, sambuehl@stanford.edu

Independent contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at (650)-723-2480 or toll free at 1-866-680-2906. You can also write to the Stanford IRB, Stanford University, Stanford, CA 94305-5401

>>

[Some browsers will ask you whether you want to display this content. Please click "display all content".]

WelcomeMovie

▶ 🔊 0:19 / 2:32   ℹ CC ⚙ ⛶

[There should be a video here. If if does not load, please click here]

Links to researchers' personal homepages

Professor B. Douglas Bernheim

Sandro Ambuehl

To continue, please enter the LAST word that Doug Bernheim said in this video. A continue button will appear after the duration of the video.

>>

Before we start this study, we would like to ask you a few questions about yourself. Please answer these questions truthfully. Your answers will not affect your payment from this experiment.

What is your gender?

○ male

○ female

What is your age?

[▲▼]

What is your ethinicity?

○ African-American

○ Asian

○ Caucasian

○ Hispanic

○ Other

Please indicate the highest level of education you completed.

○ Elementary School

○ Middle School

○ High School or equivalent

○ Vocational/Technical School (2 year)

○ Some College

○ College Graduate (4 year)

○ Master's Degree (MS)

○ Doctoral Degree (PhD)

○ Professional Degree (MD, JD, etc.)

What is your current marital status?

○ Divorced

○ Living with another

○ Married

○ Separated

○ Single

○ Widowed

○ Prefer not to say

Which of the following best describes the area you live in?

- ⭕ Urban
- ⭕ Suburban
- ⭕ Rural

Please indicate your current household income in U.S. dollars

- ⭕ Under $10,000
- ⭕ $10,000 - $19,999
- ⭕ $20,000 - $29,999
- ⭕ $30,000 - $39,999
- ⭕ $40,000 - $49,999
- ⭕ $50,000 - $74,999
- ⭕ $75,000 - $99,999
- ⭕ $100,000 - $150,000
- ⭕ Over $150,000
- ⭕ Prefer not to say

Please choose the option that describes your situation best

- ⭕ I am unemployed
- ⭕ I am employed part-time
- ⭕ I am employed full-time

How many people other than you live in your household?

[▲▼]

Do you own stocks or bonds?

- ⭕ Yes
- ⭕ No

[>>]

Please answer the following questions as well as you can. Your answers to these questions will not affect your payment from this study.

Imagine that the interest rate on your savings account was 1 percent per year and inflation was 2 percent per year. After 1 year, how much would you be able to buy with the money in this account?

○ More than today

○ Exactly the same

○ Less than today

○ Do not know

Suppose you had $100 in a savings account and the interest rate is 20 percent per year and you never withdraw money or interest payments. After 5 years, how much would you have on this account in total?

○ More than $200

○ Exactly $200

○ Less than $200

○ Do not know

Assume a friend inherits $10,000 today and his sibling inherits $10,000 3 years from now. Who is richer because of the inheritance?

○ My friend

○ His sibling

○ They are equally rich

○ Do not know

Suppose you had $100 in a savings account and the interest rate was 2 percent per year. After 5 years, how much do you think you would have in the account if you left the money to grow?

○ More than $102

○ Exactly $102

○ Less than $102

○ Do not know

Suppose that in the year 2015, your income has doubled and prices of all goods have doubled too. In 2015, how much will you be able to buy with your income?

○ More than today

○ The same

○ Less than today

○ Do not know

>>

You will now watch a

12-MINUTE VIDEO ABOUT FINANCIAL INVESTING.

Please follow this video carefully.

Please watch the ENTIRE video.

(a "continue" button will appear after 12 minutes.)

Doing so will be useful to you for three reasons:

1.   TEST with PAYMENT FOR CORRECT ANSWERS.
Your earnings from this experiment may be entirely determined by a test on this video. The final part of this experiment is a test about the contents of this video. There is a one in four chance that your earnings from this experiment are wholly determined by your performance in this test. The test has 10 questions. For each question you answer correctly, you will receive $1 within at most two days from today. For each question you answer incorrectly, you will receive $0. To be able to answer the questions in the test, you need to both *understand* and *know* the contents of the video. You may scroll back to watch parts of the video multiple times if you wish.

2.   REMAINDER OF THIS STUDY.
The video may help you with your decisions in the remainder of this experiment. In each remaining part of this experiment, you will make financial investment decisions. There is a three in four chance that one of these decisions wholly determines your earnings from this experiment.

3.   REAL LIFE
The video may help you with your decisions in real life.
This video was made by internationally recognized academic experts on financial decision making (Burton G. Malkiel, Charles D. Ellis, and B. Douglas Bernheim). This video may help you make financial decisions in your life in general.

>>

PLEASE FOLLOW THIS VIDEO CAREFULLY

PLEASE WATCH THE ENTIRE VIDEO

[Some browsers will ask you whether you want to display this content. Please click "display all content".]

[There should be a video here. If it does not load, please click here.]

To continue, enter the FOURTH word of the FIRST slide of this video. A continue button will appear after the duration of the video.

>>

PLEASE READ THESE INSTRUCTIONS CAREFULLY

The remainder of this experiment consists of 20 rounds of decision making.

Your payment may be determined entirely by ONE RANDOMLY CHOSEN decision you make in this part of the experiment.

This will happen with a three in four chance. Otherwise, your payment is determined by your performance in the test about the video you just watched.

Hence, you should make every decision as if it is the one that counts, because it might be!

>>

PLEASE READ THESE INSTRUCTIONS CAREFULLY

In each round, you will be presented with two lists. The first list will be like the following:

	you will get the specified dollar amount within two days from today	Option X
$20	○	○
$18	○	○
$16	○	○
$14	○	○
$12	○	○
$10	○	○
$8	○	○
$6	○	○
$4	○	○
$2	○	○
$0	○	○

Option X will vary from round to round. For instance, option X may be "get $15 in 8 weeks".

YOUR TASK:
Decide, on each line, whether you prefer the option on the left, or the option on the right.

Most people begin a decision list by preferring the option on the left, and then switch to the option on the right, for instance like this:

	you will get the specified dollar amount within two days from today	Option X
$20	●	○
$18	●	○
$16	●	○
$14	●	○
$12	●	○
$10	●	○
$8	●	○
$6	○	●
$4	○	●
$2	○	●
$0	○	●

After you have filled in the first list, you will be shown the second list. This list will have *different payment amounts*, for instance like this:

	you will get the specified dollar amount within two days from today	Option X
$ 7.80	○	○
$ 7.60	○	○
$ 7.40	○	○
$ 7.20	○	○
$ 7	○	○
$ 6.80	○	○
$ 6.60	○	○
$ 6.40	○	○
$ 6.20	○	○
$ 6	○	○

Again, your task is to decide, on each line, whether you prefer the option on the left, or the option on the right.

Read this paragraph if you want to know how the options on the second list are determined.

The options on the second list are determined by the point at which you switched from the left option to the right option in the first list. The second list will display payment amounts that lie between the two amounts at which you switched in the first list. In the above example, you switched between the amounts $6 and $8. Hence, the second list shows amounts between $6 and $8.

>>

PLEASE READ THESE INSTRUCTIONS CAREFULLY

Our payment procedure is designed such that it is in your best interest to choose, on each line of each decision list, the option you genuinely prefer.

Here's why: You'll get exactly what you chose, for one randomly drawn decision.

Read this paragraph if you want to know more details.

Question: When will I be paid according to the first decision list, and when will I be paid according to the second decision list in a round?

Answer: Suppose you filled in the *first* decision list of a round as follows:

	you will get the specified dollar amount within two days from today	Option X
$20	●	○
$18	●	○
$16	●	○
$14	●	○
$12	●	○
$10	●	○
$8	●	○
$6	○	●
$4	○	●
$2	○	●
$0	○	●

If the line randomly selected on the *first* list is NOT the line corresponding to $6, you will be paid according to the *first* decision list. Otherwise, you will be paid according to the *second* decision list.

That is, you are paid according to the FIRST decision list whenever the line randomly selected on that list is NOT the first line at which you chose the option on the right. Otherwise, you are paid according to the SECOND decision list.

>>

YOU WILL NOW MAKE YOUR DECISIONS

It is in your best interest to choose as you genuinely prefer. Please think about your choices carefully.

There are no right or wrong choices!

>>

Please choose, on each line, the option you genuinely prefer.

If you pick the option on the LEFT,

you will get the specified dollar amount within two days from today.

If you pick the option on the RIGHT,

we will invest $4.50 in an account with 2% interest per day. Interest is compounded daily. We will pay you the proceeds in 72 days.

You may switch from left to right at most once.

This is the
first
decision list for these options.

	you will get the specified dollar amount within two days from today	we will invest $4.50 in an account with 2% interest per day. Interest is compounded daily. We will pay you the proceeds in 72 days.
$20	○	○
$18	○	○
$16	○	○
$14	○	○
$12	○	○
$10	○	○
$8	○	○
$6	○	○
$4	○	○
$2	○	○
$0	○	○

>>

Please choose, on each line, the option you genuinely prefer.

If you pick the option on the LEFT,

you will get the specified dollar amount within two days from today.

If you pick the option on the RIGHT,

we will invest $4.50 in an account with 2% interest per day. Interest is compounded daily. We will pay you the proceeds in 72 days.

You may switch from left to right at most once.

**This is the
second
decision list for these options.**

	you will get the specified dollar amount within two days from today	we will invest $4.50 in an account with 2% interest per day. Interest is compounded daily. We will pay you the proceeds in 72 days.
$ 9.8	○	○
$ 9.6	○	○
$ 9.4	○	○
$ 9.2	○	○
$ 9	○	○
$ 8.8	○	○
$ 8.6	○	○
$ 8.4	○	○
$ 8.2	○	○
$ 8	○	○

>>

TEST

You will now participate in a test about the video you have watched at the beginning of the experiment. The test has 10 questions.

There is a one in four chance that your earnings from this study are entirely determined by your performance in this test.

IF you are randomly chosen to be paid according to this test, THEN: For each question you answer correctly, you will earn $1. For each question you answer incorrectly, you will earn $0. You will be paid within at most two days from today.

>>

What is an "indexing" investment strategy?

○ Buying index funds, which hold assets that have been indexed as particularly profitable by financial experts

○ Buying index funds, which hold stocks of companies that provide information about the stock market as a whole (stock market indices)

○ Buying index funds, which hold the market portfolio

○ Buying index funds, which hold optimally diversified, custom tailored portfolios

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

○ $210

○ $220

○ $230

○ $240

○ $250

○ $260

○ $270

○ $280

○ $290

○ $300

○ $310

○ $320

○ $330

○ $340

○ $350

○ $360

○ $370

○
　$380

○ $390

○ $400

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

If an investment grows at 8 percent per year (interest is compounded yearly), by how much has it grown after 4 years?

○ by 30%

○ by 31%

○ by 32%

○ by 33%

○ by 34%

○ by 35%

○ by 36%

○ by 37%

○ by 38%

○ by 39%

○ by 40%

Which of the following quotes is attributed to Benjamin Franklin?

○ Compound interest is the most powerful force in the universe

○ Youth is wasted on the young

○ Money makes money. And the money that money makes, makes money

What percentage of mutual funds tends to be outperformed by the market (S&P 500 Index) each year?

○ between 10 and 30%

○ between 30 and 50%

○ between 50 and 70%

○ between 70 and 90%

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

○ about 10 years

○ about 15 years

○ about 20 years

○ about 25 years

○ about 30 years

○ about 25 years

○ about 30 years

○ about 35 years

○ about 40 years

Which quote is attributed to the author Upton Sinclair

○ Only liars manage always to be out of the market during bad times and in during good times.

○ It is difficult to get a man to understand something when his salary depends upon his not understanding it.

○ There are three classes of people who do not believe that markets work: the Cubans, the North Koreans, and active managers.

○ Nobody knows more than the market

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%

Professional investors as a whole are responsible for what percentage of stock market trading?

○ 30%

○ 50%

○ 70%

○ 90%

[>>]

Please answer the following questions truthfully. Your answers to these questions DO NOT AFFECT YOUR PAYMENT for this study.

How much attention did you pay to your choices?

○ I paid quite a bit of attention for all of my choices.

○ For some choices I paid attention, for others I didn't pay much attention

○ I clicked through most of the choices without paying much attention.

At the beginning of the experiment, we asked you to watch a video about financial investing. Please indicate which of the following describes your situation best

○ I watched the entire video, and paid close attention

○ I watched the entire video, but sometimes didn't pay attention

○ I skipped parts of the video, because I already knew the material

○ I skipped parts of the video, because it was boring (but I did not already know the material)

○ I did not watch the video.

Sometimes in this experiment, you were given a choice such as "We will invest $10 in an account with 1% interest per day. Interest is compounded weekly. We will pay you the proceeds in 72 days." When deciding about this choice, did you use the rule of 72?

○ Yes

○ No

○ I don't know the rule of 72

Sometimes in this experiment, you were given a choice such as "We will pay you $20 in 36 days." When deciding about such a choice, did you use the rule of 72?

○ Yes

○ No

○ I don't know the rule of 72

In total, you were given 10 rounds in which one of the options was something like "we will invest $... in an account with ...% interest per week. Interest is compounded weekly. We will pay you the proceeds in … days". Out of these 10 rounds, how many times did you explicitly calculate the money amount that this investment would yield within the specified time?

[▲▼]

When you completed the test about the video on financial investing, did you use external resources (such as other websites, books, etc.) to find the right answers?

○ Yes

○ No

Do you have any suggestions for us about this experiment?

[   ]

Did you experience any technical difficulties with this study?

[   ]

[>>]

E Practice problems with personalized feedback

Link for the Education Intervention Part 1 Here: https://www.youtube.com/v/7r8XtqhNlIA

Part 1 Practice Question

If you invest $100 at 2% (compounded yearly), how much will be in your account after 36 years?

- ○ $102
- ○ $172
- ○ $200
- ○ $202
- ○ $300
- ○ $302
- ○ $400
- ○ $402

BACK   NEXT

If the answer is correct in the first trial:

Great job! Watch the next part of the video to hone your skills even more.

BACK   NEXT

Hmm, that's not quite right.

Please try again. The rule of 72 will help!

If you invest $100 at 2% (compounded yearly), how much will be in your account after 36 years?

- ○ $102
- ○ $172
- ○ $200
- ○ $202
- ○ $300
- ○ $302
- ○ $400
- ○ $402

BACK   NEXT

Nice! You got it this time!

Please watch the next part of the video to hone your skills even more!

BACK   NEXT

If the answer is incorrect in the second trial:

Hmm, that's still not quite right.

Watch the next part of the video, so you see how you can get a good idea about how compound interest works.

PLEASE FOLLOW THIS VIDEO CAREFULLY

PLEASE WATCH THE ENTIRE VIDEO

The next button will appear automatically when the video ends (after time equal to the duration of the video passes.).

You can stop the video by clicking on it once and make it full screen by clicking on it twice.

If you reload the page, you will again need to wait for the next button. So please do not close your web browser or reload the page unless it is necessary.

[Some browsers will ask you whether you want to display this content. Please click "display all content".]

Link for the Education Intervention Part 2 Here: https://www.youtube.com/v/75z3Rh6GMqw

Now you try:

$100 is invested at 9% for 32 years, compounded yearly. How much will be in the account after these 32 years?

○ $100
○ $200
○ $388
○ $400
○ $600
○ $800
○ $1200
○ $1600

BACK   NEXT

If the answer is correct in the first trial:

This is correct!

Please click next to move on the next part of the video.

BACK   NEXT

41

If the answer is incorrect in the first trial:

Subjects see one of the following explanations depending on their previous answer and they re-attempt the question.

If the subject selected answer $100:

You selected $100. That's not quite right.

You start out with $100. Then you get 9% interest each year! Hence after 32 years, you will have MORE than $100!

Please watch the video again to understand how much you will have.

BACK   NEXT

You selected $200. That's not quite right.

You probably remembered from the example above that at 9%, an investment doubles in 8 years.

Thus, the $100 double to $200 after 8 years.

These $200 then double to $400 in the next 8 years. (That is, until year 16).

In the next 8 years, from year 16 to year 24, these $400 double to $800!

Then, in the next 8 years, from year 24 to year 32, it will double again.

Please give it another try.

BACK   NEXT

You selected $388. That's not quite right.

You probably got this because you thought you'd get 32 times the interest of 9% on your $100, which is $9.

But, starting from the second year, you also get interest on the interest you earned!

Here's how: You do start out with $100. In the first year, you get 9% interest. That's $9. You start the second year with $109 in your account. Your interest in the second year is 9% of $109, which is MORE than $9. In fact, you'll get 9% of $109, which is $9.80.

Please watch the video again, so you'll understand how compound interest works.

BACK   NEXT

If the subject selected answer $400:

You selected $400. That's not quite right.

In this question, the $100 are invested for 32 years, not just for 16 years, as in the example above.

Please give it another try.

<div style="display:flex; justify-content:space-between;">
BACK   NEXT
</div>

You selected $600. That's not quite right.

You probably remembered from the example above that at 9%, an investment doubles in 8 years.

Thus, the $100 double to $200 after 8 years.

These $200 then double to $400 in the next 8 years. (That is, until year 16).

In the next 8 years, from year 16 to year 24, these $400 double to $800!

Then, in the next 8 years, from year 24 to year 32, it will double again.

Please give it another try.

BACK

NEXT

You selected $800. That's not quite right.

You probably remembered from the example above that at 9%, an investment doubles in 8 years.

Thus, the $100 double to $200 after 8 years.

These $200 then double to $400 in the next 8 years. (That is, until year 16).

In the next 8 years, from year 16 to year 24, these $400 double to $800!

Then, in the next 8 years, from year 24 to year 32, it will double again.

Please give it another try.

BACK   NEXT

You selected $1200. That's not quite right.

You probably remembered from the example above that at 9%, an investment doubles in 8 years.

Thus, the $100 double to $200 after 8 years.

These $200 then double to $400 in the next 8 years. (That is, until year 16).

In the next 8 years, from year 16 to year 24, these $400 double to $800!

Then, in the next 8 years, from year 24 to year 32, it will double again.

Please give it another try.

BACK   NEXT

Re-attempt the question:

Please try again:

$100 is invested at 9% for 32 years, compounded yearly. How much will be in the account after these 32 years?

○ $100
○ $200
○ $388
○ $400
○ $600
○ $800
○ $1200
○ $1600

BACK   NEXT

If the answer is correct in the second time:

This is correct!

Please click next to move on the next part of the video.

BACK   NEXT

If the answer is incorrect in the second trial:

Hmm, that's still not quite right.

But let's move to the next part of the video.

BACK   NEXT

The next button will appear automatically when the video ends (after time equal to the duration of the video passes.).

You can stop the video by clicking on it once and make it full screen by clicking on it twice.

If you reload the page, you will again need to wait for the next button. So please do not close your web browser or reload the page unless it is necessary.

Link for the Education Intervention Part 3 Here: $https://www.youtube.com/v/2NDOLWqXCQI$

Thanks for watching this video!

We'll now ask you to solve a bunch of problems on your own. We'll first walk you through in steps, and then it's up to you to find the right steps.

These questions are still a part of the education. They don't count for money, but you need to get them right so you can continue with the survey.

BACK

NEXT

Question 1(a)

You invest $50 at 8%. Eventually, we want to know how much will be in your account after 27 years. But we'll get there in three easy steps.

1. How long does it take for the money to double at 8%?
2. How many times does it double in 27 years?
3. Hence, how much will be in the account after it doubles that many times?

So let's start with the first one of these.

How many years does it take for this investment to double?

<div style="border:1px solid #ccc; padding:8px; width:200px;">   ▲▼</div>

BACK   NEXT

Your answer isn't quite correct. Remember: The rule of 72 says

percentage interest rate X number of years it takes for the investment to double = 72

Please try again: You invest $50 at 8%. How many years does it take for this investment to double?

[▲▼]

That's still not quite correct.

Here's how you can do it correctly:

The rule of 72 says that

percentage interest rate *times* the number of years it takes for the investment to double = 72

or, in mathematical notation,

X × Y = 72

In this problem, the percentage interest rate is 8%. Hence you just need to know: 8 times *what* equals 72?

That's how long it takes for the investment to double!

Enter your answer below.

[⬍]

BACK   NEXT

Great, you've got it!

Question 1(b)

We're still looking at that $50 invested at 8%. As you've figured out, at 8%, the investment doubles in 9 years.

Remember the three steps?

1. How long does it take for the money to double at 8%?
2. How many times does it double in 27 years?
3. Hence, how much will be in the account after it doubles that many times?

We now tackle the second step:

How many times does this investment double over the course of 27 years?

- ○ once
- ○ twice
- ○ three times
- ○ four times
- ○ five times
- ○ six times
- ○ seven times
- ○ eight times
- ○ nine times
- ○ ten times

Unfortunately, that's not quite right.

As you've figured out, the investment doubles in 9 years. It doubles in every 9 years over the course of 27 years!

Hence, 9 times *what* equals 27?

The answer to this question tells you how many times the investment doubles!

Please choose one of the answers below.

- ○ once
- ○ twice
- ○ three times
- ○ four times
- ○ five times
- ○ six times
- ○ seven times
- ○ eight times
- ○ nine times
- ○ ten times

BACK   NEXT

Nice job!

Now to the last one of the three steps.

1. How long does it take for the money to double at 8%?
2. How many times does it double in 27 years?
3. Hence, how much will be in the account after it doubles that many times?

You figured out that over the course of 27 years, your $50, invested at 8% double three times.

Hence, how much will be in your account after 27 years?

○ $50
○ $100
○ $150
○ $200
○ $250
○ $300
○ $350
○ $400
○ $450
○ $500
○ $600
○ $700
○ $800

Oops, that is not quite right.

In the first 9 years, your investment doubles by $50 and is then worth $100. In the second 9 years, these entire $100 double again. So after the second 9 years (that is after 18 years), you have $200.

So, how much will you have after 27 years?

○ $50
○ $100
○ $150
○ $200
○ $250
○ $300
○ $350
○ $400
○ $450
○ $500
○ $600
○ $700
○ $800

BACK   NEXT

Awesome job!

Now it's up to you to go through the steps in the right order.

Let's try this example:

Question 2

You invest $100 at 6%. How much will be in your account after 24 years?

- ○ $100
- ○ $106
- ○ $148
- ○ $200
- ○ $288
- ○ $300
- ○ $306
- ○ $400
- ○ $406
- ○ $500
- ○ $506
- ○ $600
- ○ $606

Oops, that's not quite right. Remember the three steps for using the rule of 72:

1. How long does it take for the money to double?
2. How many times will it double over the years?
3. Hence, how much will be in the account after it doubles that many times?

Give it another shot:

Question 2

You invest $100 at 6%. How much will be in your account after 24 years?

- ○ $100
- ○ $106
- ○ $148
- ○ $200
- ○ $288
- ○ $300
- ○ $306
- ○ $400
- ○ $406
- ○ $500
- ○ $506
- ○ $600
- ○ $606

If the answer to Question 2 is correct in the first trial or later trials:

Great! Thanks for paying attention to this education module. Before moving forward, we would like to ask you a question about the education module. Your answer to this question will not affect your payments from this study.

BACK

NEXT

This and other Global Financial Literacy Excellence Center
publications are available online at www.gflec.org

GFLEC
GLOBAL FINANCIAL LITERACY
EXCELLENCE CENTER