Non-Standard Policy Instruments Part 2

Prof. B. Douglas Bernheim
Stanford University

NBER/Sloan Behavioral Public Economics Bootcamp
May 2022

Outline of Lecture

I. Financial education

II. Default options

I. Financial Education

I. Financial education

A. Motivation

- Evidence of pervasive financial illiteracy and “problematic” choice patterns raises questions about the quality of financial decision making
  - Surveys of the literature on financial literacy: Hastings et al., 2013, Lusardi and Mitchell, 2014
  - Examples of questionable decisions: inadequacy of saving (Bernheim 1993, Bernheim, Skinner, & Weinberg, 2001), low enrollment in pension plans that offer generous matches, naïve diversification, & tendency to invest heavily in employer’s stock (Benartzi & Thaler, 1999, 2001, 2007)
- Education seeks to improve the quality of decision making in two ways:
  - Provide factual information (standard welfare economics may or may suffice, depending on whether the information addresses biases)
  - Train people to use whatever information they receive more effectively in their decision making (“abstract” knowledge → deliberation skill, e.g. “thinking at the margin”)
- Main types of financial education: high school classes (often mandated), employer-based programs (>3/4 of pension plan sponsors provide seminars)

I. Financial education

B. Conventional methods of evaluating benefits

Method #1: Measure Average Treatment Effects (ATEs) on behavior, and ask whether they offset a known or presumed bias

- Literature
  - Begins with Bernheim, Garrett, & Maki, 2001 ("natural experiment" with high school curriculum mandates), Bernheim & Garrett, 2003, Duflo & Saez, 2003 (workplace interventions)
- Limitations of this method of evaluation:
  - How do we know the effect is offsetting a bias, rather than introducing one? Could involve indoctrination, deference to authority, social pressure.
  - These analyses ignore heterogeneity of biases and effects. If biases are heterogeneous and education works as intended, effects should be negatively correlated with biases (e.g., those saving too little/much should save more/less). Are they?

I. Financial education

B. Conventional methods of evaluating benefits

Method #2: Measure Average Treatment Effects (ATEs) on financial literacy

- Literature
  - Begins with Jump$tart, 2006, and Mandell, 2008 (correlational analyses of high school students' financial knowledge and whether they completed a class)
- Limitation of this method of evaluation: Conceptual knowledge may not translate into decisions
  - Correlational evidence relating knowledge to behavior: beginning with Hilgert, Hogarth, & Beverly, 2003
  - Various studies employ instruments, such as high school curriculum mandates (Lusardi & Mitchell, 2009), the presence (or opening) of universities (Christiansen, Joensen, & Rangvid, 2008), the financial experience of siblings & parents (Van Rooij, Lusardi, & Alessie, 2011), political attitudes (Bucher-Koenen & Lusardi, 2011). But in each case the instruments may proxy for other taste differences.

I. Financial education

B. Conventional methods of evaluating benefits

Kaiser, Lusardi, Menkhoff, & Urban (2021): meta-analysis of subsequent studies measuring the impact of financial education on behavior & financial literacy

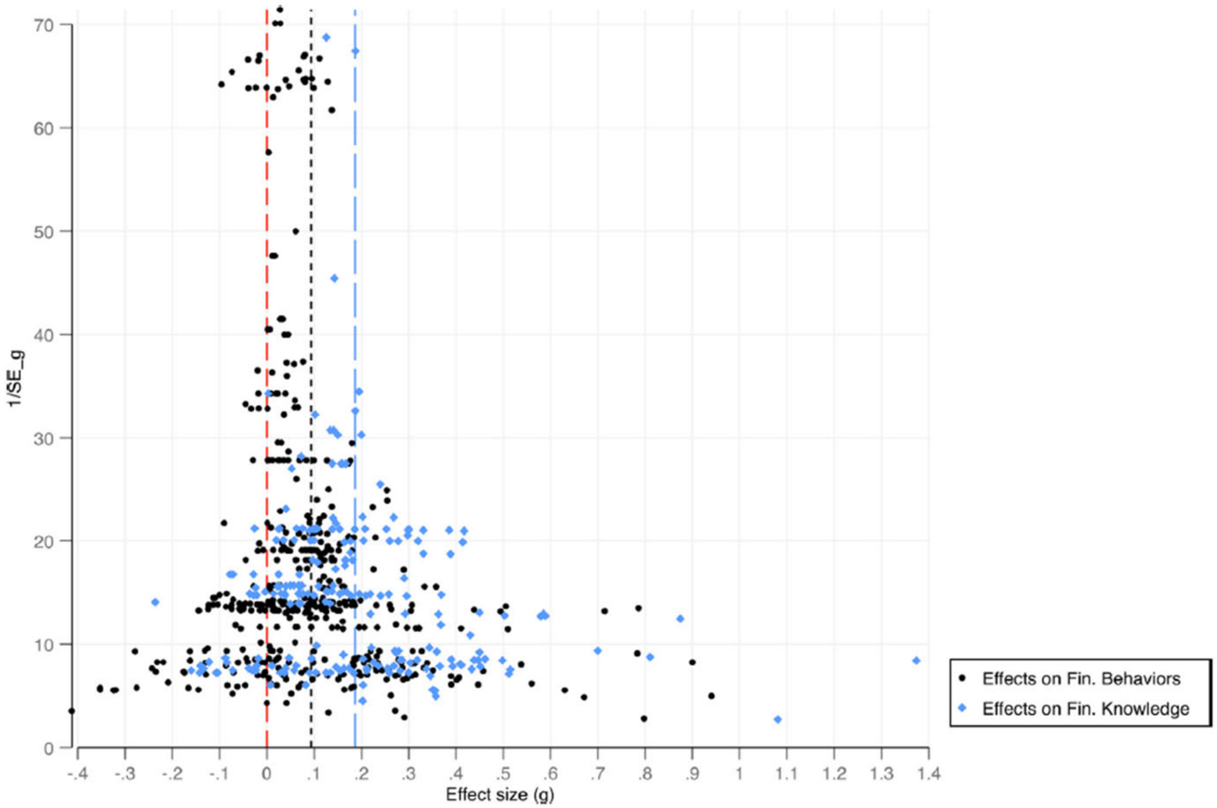


Fig. 2. Distribution of raw financial education treatment effects and their standard errors.
Effect size (g) is the bias corrected standardized mean difference (Hedges' g). 1/SE_g is its inverse standard error. The number of observations in the treatment effects on financial behaviors sample is 458 effect size estimates from 64 studies. The number of observations in the treatment effects on financial knowledge sample is 215 effect size estimates from 50 studies. Thirty-eight studies report treatment effects on both types of outcomes. The mean effect size on financial behaviors is 0.094 SD units, and the mean effect size on financial knowledge is 0.186 SD units.

I. Financial education

B. Conventional methods of evaluating benefits

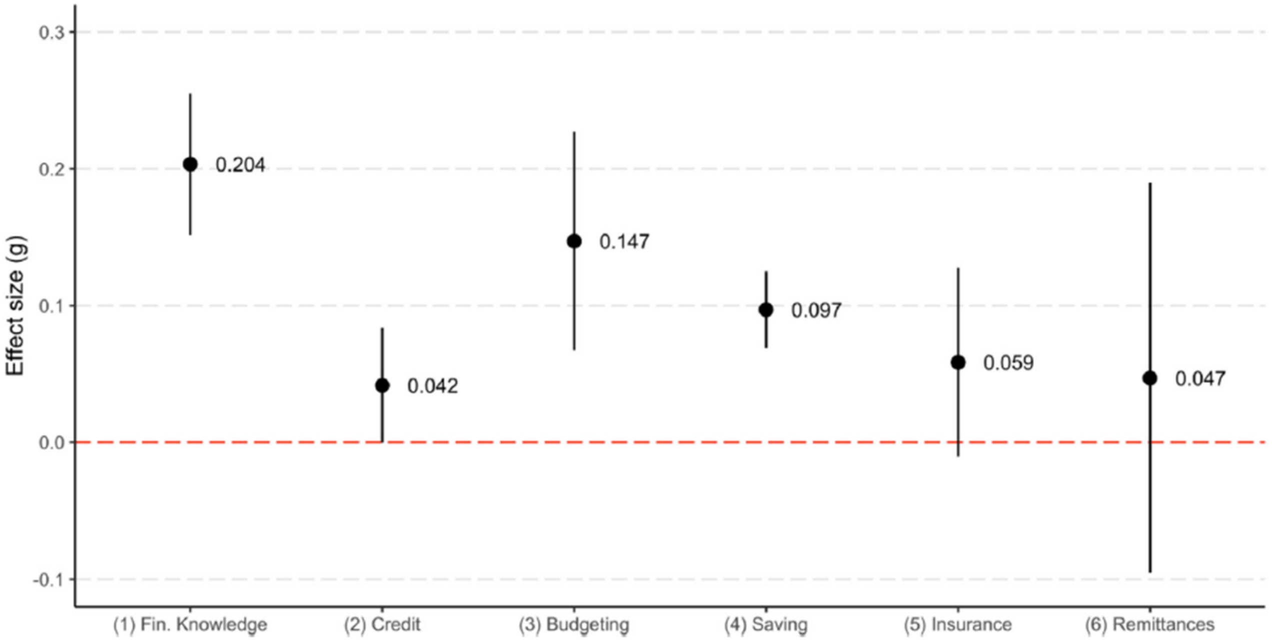


Fig. 4. Financial education treatment effects by outcome domain.
Results from robust variance estimation in meta-regression with dependent effect size estimates (RVE) (Hedges et al., 2010). The number of observations for the financial knowledge sample (1) is 215 effect size estimates within 50 studies. The number of observations for the credit behavior sample (2) is 115 within 22 studies. The number of effect size estimates for the budgeting behavior sample (3) is 55 within 23 studies. The number of observations in the saving and investing behavior (4) sample is 253 effect size estimates within 54 studies. The number of observations in the insurance behavior sample (5) is 18 effect sizes within six studies. The number of observations on remittance behavior (6) is 17 effect size estimates reported within six studies. Dots show the point estimates, and the solid lines indicate the 95% confidence interval.

- A comforting message (?): behavior is changing in the right direction for the right reason

I. Financial education

C. Deploying the tools of Behavioral Public Economics

Ambuehl, Bernheim, and Lusardi (2022)

- Objectives:

  1. Uses an experiment to examine the reliability of the conventional evaluative metrics

  2. Proposes and implements an alternative evaluative metric based on the principles of Behavioral Public Economics.

- Focus is on comprehension of compound interest

  - Foundational concept in finance

  - Suitable for an experiment: relatively easy to explain in a brief intervention

  - People tend to suffer from a known bias (*exponential growth bias*): Wagenaar and Sagaria, 1975, Eisenstein & Hoch, 2007, Stango & Zinman, 2009, Almenberg & Gerdes, 2012, Levy & Tasoff, 2016

  - Experiment assesses test performance to evaluate effects on literacy, and valuations (e.g., WTP for a $10 investment in an asset that pays 2% interest per day, compounded daily, for 36 days) to determine directional effects on behavior

I. Financial education

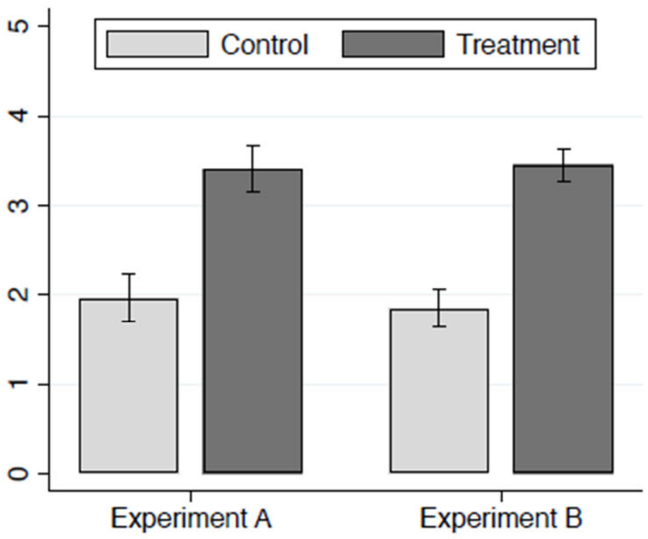
C. Deploying the tools of Behavioral Public Economics

- Structure of the experiment
  - Stage 1: Educational intervention
  - Stage 2: Valuation decisions
  - Stage 3: Exam-style questions (to assess ability to compute compound interest)
- Details of educational interventions
  - Treatment: A narrated video of a section on compound interest from a popular investment guide (Malkiel and Ellis). Includes substance (explanation of compound interest plus the “Rule of 72”), and motivational rhetoric.
  - Control: A narrated video based on another section of the same investment guide covering an unrelated topic.
- Versions of experiment
  - Experiment A: The educational intervention does not include practice and feedback
  - Experiment B: The educational intervention does include practice and feedback

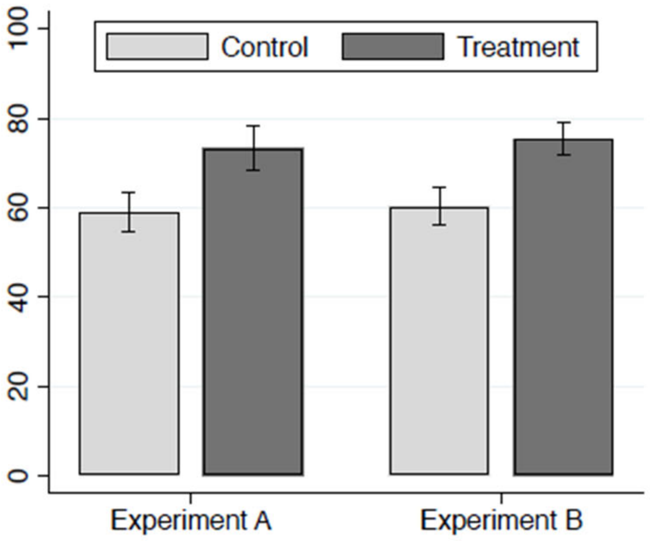
I. Financial education

C. Deploying the tools of Behavioral Public Economics

- According to conventional measures, the intervention is highly successful, and equally successful regardless of whether it includes practice and feedback:



A. Test scores



B. Valuations in complex frame

- Both interventions appear to have the “right effects” for the “right reasons”

I. Financial education

C. Deploying the tools of Behavioral Public Economics

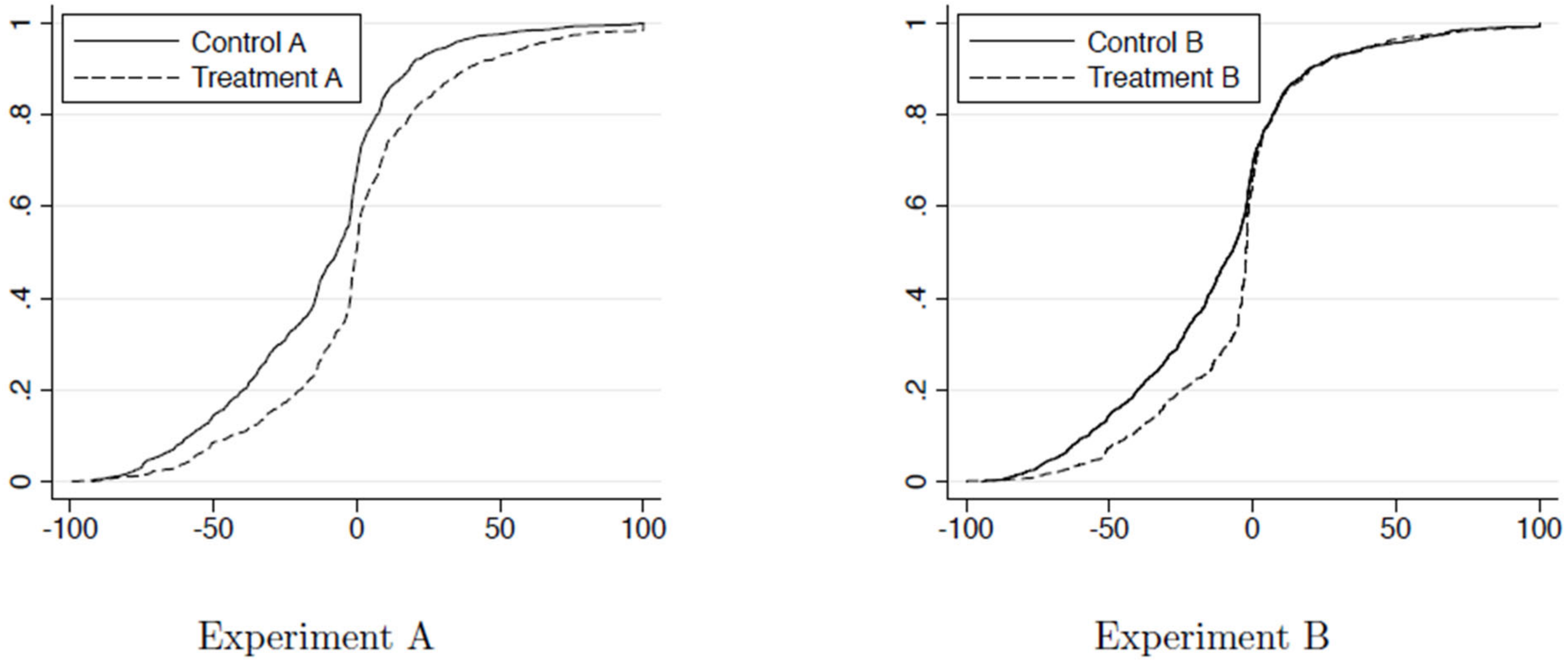
- But the experiment also assesses *money metric biases*
  - General strategy: study two objectively equivalent decision problems, one with naturally occurring complexity, the other simplified to make the consequences transparent, and define the WRD to include only the second.
  - Specific strategy uses *paired valuation tasks:* in addition to assessing (i) WTP for a $10 investment in an asset that pays 2% interest per day, compounded daily, for 36 days, also assess (ii) WTP for $20 in 36 days.
  - The money-metric bias is the difference in valuations

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- Intervention A (without practice and feedback) actually fails because the impact on the money-metric bias is uncorrelated with the initial bias.
- Intervention B (with practice and feedback) actually succeeds because the bias and the effect are negatively correlated

B. CDF of valuation differences



I. Financial education

C. Deploying the tools of Behavioral Public Economics

- To understand the difference between the conventional results and the results for money-metric bias, we also field:
  - A substance-only treatment (no motivational rhetoric)
  - A motivational-rhetoric treatment (no Rule of 72)
- Findings:
  - Effects on tested financial literacy come almost entirely from the substantive elements of instruction
  - Without practice and feedback, effects on valuations come almost entirely from the motivational rhetoric

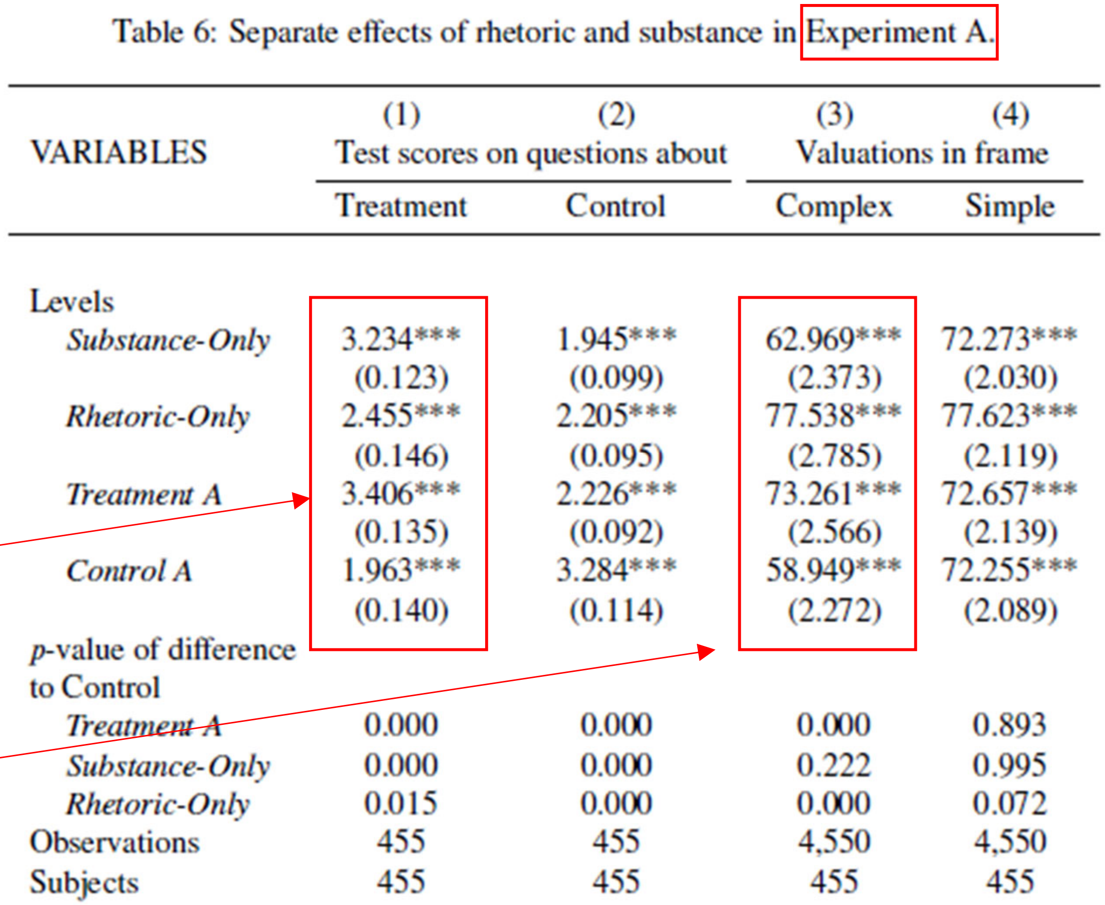


Table 6: Separate effects of rhetoric and substance in Experiment A.

VARIABLES	(1)	(2)	(3)	(4)
	Test scores on questions about		Valuations in frame	
	Treatment	Control	Complex	Simple
Levels				
Substance-Only	3.234***	1.945***	62.969***	72.273***
	(0.123)	(0.099)	(2.373)	(2.030)
Rhetoric-Only	2.455***	2.205***	77.538***	77.623***
	(0.146)	(0.095)	(2.785)	(2.119)
Treatment A	3.406***	2.226***	73.261***	72.657***
	(0.135)	(0.092)	(2.566)	(2.139)
Control A	1.963***	3.284***	58.949***	72.255***
	(0.140)	(0.114)	(2.272)	(2.089)
p-value of difference to Control				
Treatment A	0.000	0.000	0.000	0.893
Substance-Only	0.000	0.000	0.222	0.995
Rhetoric-Only	0.015	0.000	0.000	0.072
Observations	455	455	4,550	4,550
Subjects	455	455	455	455

Intervention A is not having the "right effects" for the "right reasons"

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- In place of conventional outcome metrics, the paper proposes using the absolute value or the square of money metric bias (*deliberative competence*)

- These measures can be rationalized as the dollar-equivalent welfare loss a consumer suffers due to characterization failure when making her decision in the complex frame.

  - Imagine the consumer deciding whether to buy similar financial instruments in settings where the price will be realized from some distribution

  - Absolute money metric bias gives the largest possible loss (i.e., the most costly mistake the individual can make)

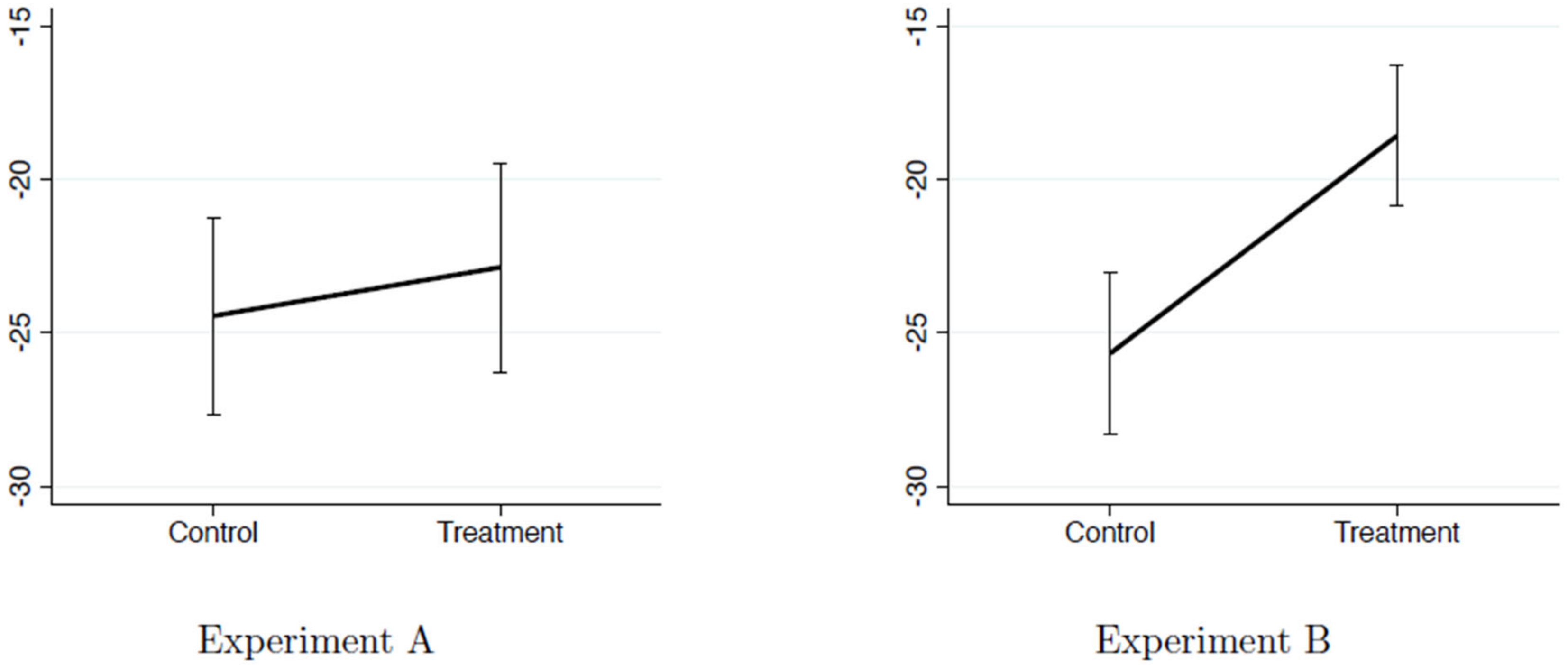
  - Squared money metric bias approximates the expected loss

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- According to such measures, the intervention unambiguously improves the quality of financial decision making only if it includes practice and feedback:

A. Financial competence



I. Financial education

C. Deploying the tools of Behavioral Public Economics

- Using money metric bias as an evaluative metric raises a general concern: even if simple framing removes the type of characterization failure it is designed to address, other valuation biases may remain. Whether simply framed choices provide valid normative benchmarks is then unclear.

- Example: suppose people not only underestimate compound interest, but are also time-inconsistent
  - The assumption is plausible: decisions in the experiment for simple frames reflect very high discounting
  - Maybe a good policy would distort consumers' understanding of compound interest in a way that offsets such biases

- What do we do about other distortions, which may not even be known?

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- This is an old question that arose long ago outside of behavioral economics: how should we evaluate a policy that ameliorates or aggravates a distortion when there are other distortions elsewhere in the economy?
- Lipsey and Lancaster (1956) argued that the right way to handle these types of issues is to analyze all distortions and all remedies simultaneously (*comprehensive second-best welfare analysis*).
- Due to feasibility concerns, the overwhelming preference of economists is to address distortions and their solutions one or two at a time (i.e., solve the overall policy puzzle piece by piece).
  - Meade (1955): proposed compartmentalizing by evaluating one policy targeting one distortion, but account for all other distortions (*narrow second-best welfare analysis*).
  - Concerns about feasibility remain, and there’s a conceptual problem: why take another distortion into account if another policy will correct it? E.g., why distort consumers’ understanding of compound interest to counter time inconsistency if we can combine effective education with commitment opportunities?

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- The dominant approach in Behavioral Public Economics: analyze a small number (usually just one) of biases and associated policies while assuming (implicitly) that the consumer's decision-making apparatus is otherwise flawless (*myopic welfare analysis*)
  - Tractable, but there is no reason to think that this "piecemeal" approach yields desirable solutions (Lipsey and Lancaster, 1956)
- An alternative approach to compartmentalization: evaluate policies designed to address a single bias (or small collection) under the assumption that effective remedies for other biases *will be forthcoming* (*idealized welfare analysis*)
  - Avoids the Lipsey-Lancaster critique of myopic welfare analysis
  - Logically permits a compartmentalized approach to resolving biases (solving problems one at a time)

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- A natural concern: Idealized welfare analysis may be no simpler than comprehensive welfare analysis, because it requires us to figure out how the consumer would behave if all other biases were corrected
- The paper shows that, under a separability condition, deliberative competence (calculated myopically) nevertheless approximates the idealized welfare effect up to an unknown scalar
  - Despite the unknown scale, we can rank policies according to their effectiveness, gauge the percentage differences between the dollar-equivalents of the associated benefits, and aggregate over decision problems. Also robust with respect to normative ambiguity.
- Illustration:
  - A financial instrument z yields a future payoff $f(z)$, which the consumer perceives as $g(z, \theta)$, where θ is an educational policy.
  - For simplicity, the consumer expects to spend income when it is received, and evaluates outcomes according to the utility function $c_1 + \gamma u(c_2)$, where c_1 is current consumption and c_2 is future consumption
  - WTP for z is $\gamma u(f(z))$, but it should be $\gamma u\big(g(z, \theta)\big)$. The measured WTP error is $E_M = \gamma u\big(f(z)\big) - \gamma u\big(g(z, \theta)\big)$

I. Financial education

C. Deploying the tools of Behavioral Public Economics

- Illustration (continued):
  - Suppose that, possibly unknown to the analyst, the consumer discounts the future excessively due to present bias, and that the appropriate normative standard is $c_1 + \delta u(c_2)$.
  - Assuming correction of the present bias, the idealized WTP error is $E_I = \delta u\big(f(z)\big) - \delta u\big(g(z,\theta)\big)$
  - Notice that $E_I = kE_M$, where $k = \delta/\gamma$ is a multiplicative constant.
  - Therefore, myopic welfare analysis yields the correct sign for the welfare effect, correctly ranks policies from the perspective of idealized welfare analysis, and properly measures their proportional costs and benefits
  - For arbitrary decision utility functions and objective functions satisfying a separability condition, the same finding holds to a first-order approximation (i.e., for “small” securities)

II. Default effects

II. Optimal defaults

A. Empirical findings on default effects

- Every decision has a default option – i.e., the option that prevails if no choice is made
- In some settings, the choice of the default option is potentially consequential
  - Contribution rates and portfolio allocation in pension plans
  - Health insurance plan choices
  - Organ donation elections on drivers' license applications

II. Optimal defaults

A. Empirical findings on default effects

- Literature begins with Madrian and Shea, 2001
  - Fortune 500 company changed the default contribution rate for it's 401(k) plan from zero to 3% as of 4/1/1998, with default investment in a money market fund.
  - The paper examines three groups: those hired during the year before the change ("Window"), those hired earlier ("Old"), and those hired in the year after the change ("New"). Verifies that they are similar demographically (including age when hired).
  - Compares elections for the "Window" group on 6/30/1998 and the "New" group on 6/30/1999 (so that members of each group have 3 to 15 months to tenure)

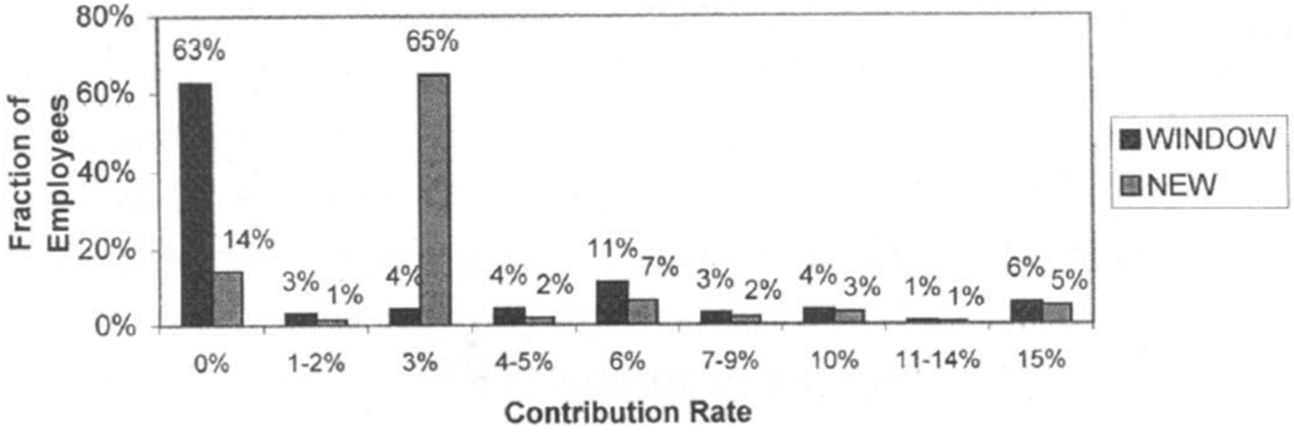


FIGURE IIc
Distribution of 401(k) Contribution Rates for the WINDOW and NEW Cohorts Including Nonparticipation

II. Optimal defaults

A. Empirical findings on default effects

- Rules out possibility that the difference reflects the change to immediate eligibility (by comparing contribution rates of Old and Window groups at comparable levels of tenure)
- Also finds enormous effects on portfolio allocation

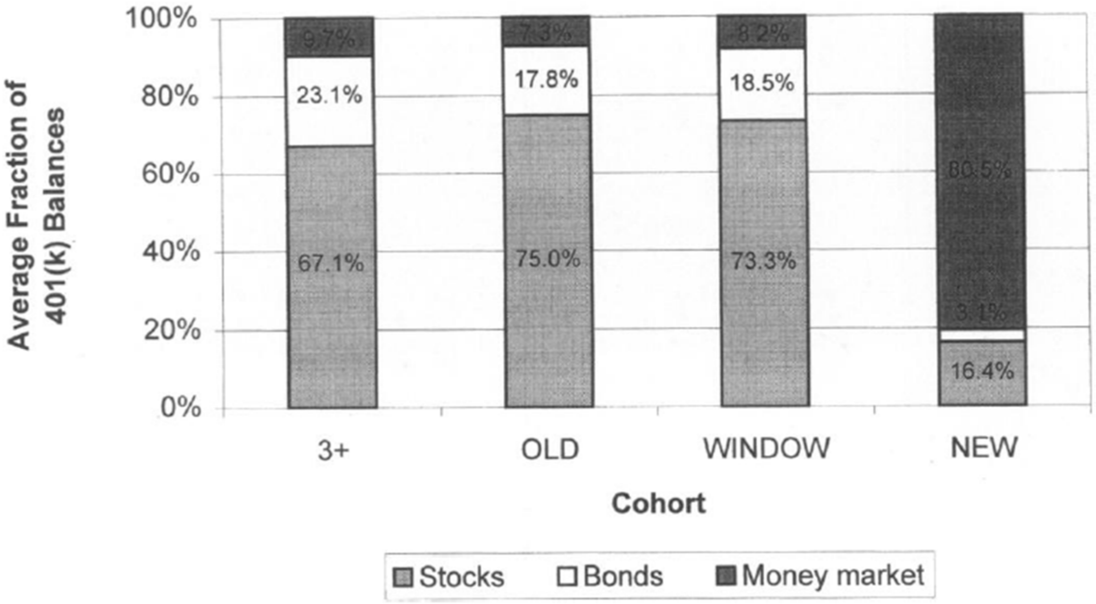


Figure III
401(k) Asset Allocation by Cohort

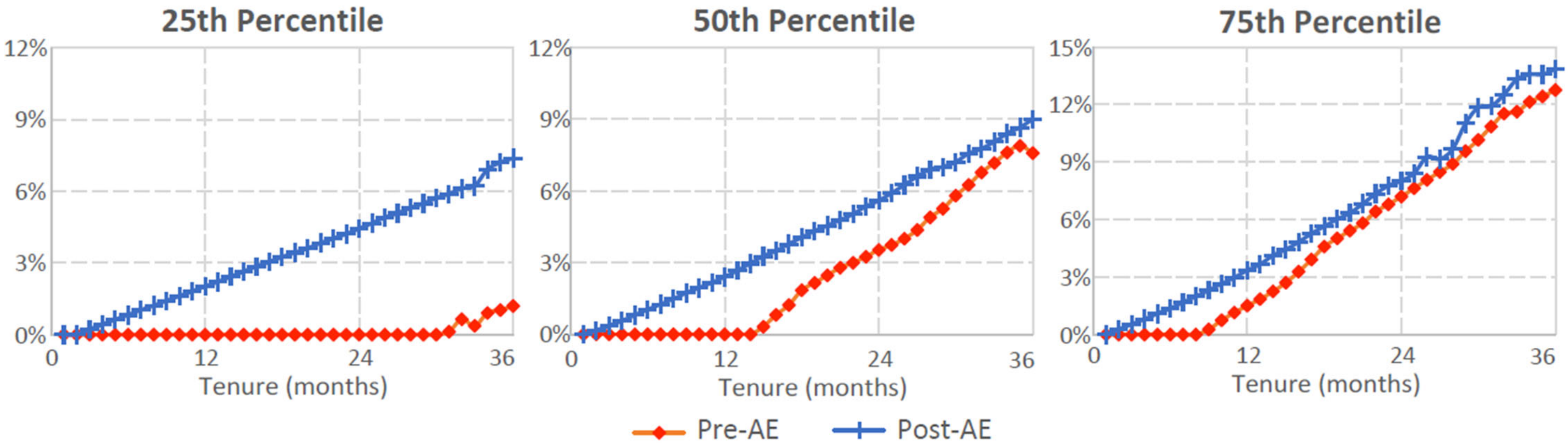
II. Optimal defaults

A. Empirical findings on default effects

- Findings corroborated in many subsequent studies: see Beshears, Choi, Laibson, and Madrian, 2018 (*Handbook* chapter) for a review

- An important qualification (Choukhmane, 2021): there appears to be substantial catch-up of contributions after 12 months (below), and at subsequent jobs

Figure 1: Cumulative employee 401(k) contributions to annual salary ratios at the 25th, 50th and 75th percentiles.



Notes: The graphed series correspond to cumulative employee 401(k) contributions at each month of tenure divided by individuals' annual salary. The Pre-AE (Post-AE) cohort corresponds to employees hired in the 12 months before (after) the introduction of an auto-enrollment policy at 3% of salary for new hires. The sample at each tenure level corresponds to workers still employed by the firm at that time. Data source: administrative 401(k) records from 34 U.S. firms.

II. Optimal defaults

B. Theory

- Ideas concerning optimal default rates:
  - Set the default to maximize contributions, because people “don’t save enough” (Thaler and Sunstein, 2008). But why? 401(k) elections have no immediate consequences…
  - Set the default to minimize the frequency of opt-out (Thaler and Sunstein, 2003), because those who stick with the default must find it acceptable (“ex post validation”)
  - Force all employees to make active decisions, so they express their preferences (Carroll et al., 2009). But it is obviously not in the interests of someone who wants to avoid the costs of making a contribution election.
  - Notice the tension between the second approach (which minimizes opt-out), and the third approach (which maximizes opt-out)
- Determining the best policy requires an understanding of default effects. Examples of possible theories:
  - Opt-out is costly (but implied as-if opt-out costs are implausibly); time inconsistency (sophisticated or naïve) and procrastination; inattention; anchoring

II. Optimal defaults

B. Theory

- Formal analyses of optimal defaults include:
  - Carroll, Choi, Laibson, Madrian, & Metrick, 2009 (time inconsistency)
  - Bernheim, Fradkin, & Popov, 2015 (empirical implementation of time inconsistency, inattention, anchoring)
  - Goldin & Reck, 2019 (a class of theories equivalent to time inconsistency)
  - Choukhmane, 2019 (empirical implementation of time inconsistency)
  - Bernheim & Mueller-Gastell, 2021 (same as Goldin-Reck)
- Much of the literature focuses on identifying conditions that justify the opt-out minimization criterion, or its opposite, opt-out maximization
- We will follow Bernheim & Mueller-Gastell, which illuminates the problem's mathematical structure and arrives at the most general results

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta V(x, x^*, \rho) - \eta\lambda I(x \neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \to 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta\boxed{V(x,x^*,\rho)}-\eta\lambda I(x\neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x=x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda\rightarrow 0$
  - $I(x\neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta\equiv(\beta,\eta,\rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta V(x, \boxed{x^*}, \rho) - \eta\lambda I(x \neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \to 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta V(x, x^*, \boxed{\rho}) - \eta\lambda I(x \neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \rightarrow 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta V(x, x^*, \rho) - \boxed{\eta\lambda} I(x \neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \to 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta V(x, x^*, \rho) - \eta \boxed{\lambda} I(x \neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \rightarrow 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\beta V(x, x^*, \rho) - \eta\lambda \boxed{I(x \neq D)}$$

- V is future utility from the choice of x. It’s interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker’s ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \to 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\boxed{\beta}V(x, x^*, \rho) - \eta\lambda I(x \neq D)$$

- V is future utility from the choice of x. It's interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker's ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \to 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- Facing a default of D, the worker chooses the contribution rate x to maximize

$$\boxed{\beta}V(x,\boxed{x^*,\rho}) - \boxed{\eta}\lambda I(x \neq D)$$

- V is future utility from the choice of x. It’s interpretable as a state evaluation function, so the formulation implicitly encompasses dynamics
  - x^* is the worker’s ideal point (V maximized at $x = x^*$)
  - ρ impacts the shape of V
- $\eta\lambda$ is the opt-out cost
  - The purpose of λ is to allow us consider the limiting case of $\lambda \to 0$
  - $I(x \neq D)$ just indicates whether the worker opts out
- β is the weight attached to future consequences
- We allow for arbitrary heterogeneity in $\theta \equiv (\beta, \eta, \rho)$ and x^*

II. Optimal defaults

B. Theory

- A worker who opts out chooses x^*. Opt-out therefore occurs when

$$\Delta(D, x^*, \rho) \equiv V(x^*, x^*, \rho) - V(D, x^*, \rho) \geq \frac{\eta\lambda}{\beta}$$

- Employer:
    - Must set a single default, D, without conditioning on θ or x^*.
    - Is assumed to be utilitarian who believes $\beta - 1$ is a bias

II. Optimal defaults

B. Theory

- A worker who opts out chooses x^*. Opt-out therefore occurs when

$$\Delta(D, x^*, \rho) \equiv V(x^*, x^*, \rho) - V(D, x^*, \rho) \geq \frac{\eta\lambda}{\beta}$$

- Employer:
  - Must set a single default, D, without conditioning on θ or x^*.
  - Is assumed to be utilitarian who believes $\beta - 1$ is a bias

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$-\eta\lambda + Pr_{x^*}\left(\Delta(D,x^*,\rho) \leq \frac{\eta\lambda}{\beta}\middle|\theta\right)\left[\eta\lambda - E_{x^*}\left(\Delta(D,x^*,\rho)\middle|\theta,\Delta(D,x^*,\rho) \leq \frac{\eta\lambda}{\beta}\right)\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_D E_\theta\left[\omega(\theta)Pr\left(\Delta(D,x^*,\rho) \leq \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$\boxed{-\eta\lambda} + Pr_{x^*}\left(\Delta(D,x^*,\rho) \leq \frac{\eta\lambda}{\beta}\middle|\theta\right)\left[\eta\lambda - E_{x^*}\left(\Delta(D,x^*,\rho)\middle|\theta, \Delta(D,x^*,\rho) \leq \frac{\eta\lambda}{\beta}\right)\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_D E_\theta\left[\omega(\theta) Pr\left(\Delta(D,x^*,\rho) \leq \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$-\eta\lambda + \boxed{Pr_{x^*}\left(\Delta(D, x^*, \rho) \leq \frac{\eta\lambda}{\beta}\middle|\theta\right)}\left[\eta\lambda - E_{x^*}\left(\Delta(D, x^*, \rho)\middle|\theta, \Delta(D, x^*, \rho) \leq \frac{\eta\lambda}{\beta}\right)\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_D E_\theta\left[\omega(\theta)Pr\left(\Delta(D, x^*, \rho) \leq \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$-\eta\lambda + Pr_{x^*}\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\left[\eta\lambda - \boxed{E_{x^*}\left(\Delta(D,x^*,\rho)\middle|\theta,\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\right)}\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_D E_\theta\left[\omega(\theta)Pr\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$-\eta\lambda + Pr_{x^*}\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\left[\boxed{\eta\lambda - E_{x^*}\left(\Delta(D,x^*,\rho)\middle|\theta,\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\right)}\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_D E_\theta\left[\omega(\theta)Pr\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$\boxed{-\eta\lambda} + Pr_{x^*}\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\left[\eta\lambda - E_{x^*}\left(\Delta(D,x^*,\rho)\middle|\theta,\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\right)\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_{D} E_{\theta}\left[\omega(\theta) Pr\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- Aggregate welfare (relative to the first-best) is the expectation over θ of

$$-\eta\lambda + Pr_{x^*}\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\left[\eta\lambda - \boxed{E_{x^*}\left(\Delta(D,x^*,\rho)\middle|\theta,\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\right)}\right]$$

- The first term does not depend on D, and is therefore irrelevant.
- Observation: If it were the case that the $E_{x^*}(\cdot)$ term didn't depend on D, then the bracketed term would be some function $\omega(\theta)$, in which case welfare maximization would be equivalent to weighted opt-out minimization:

$$\max_D E_\theta\left[\omega(\theta)Pr\left(\Delta(D,x^*,\rho) \le \frac{\eta\lambda}{\beta}\middle|\theta\right)\right]$$

II. Optimal defaults

B. Theory

- **Claim:** for small λ,

$$\lambda^{-1} E_{x^*}\left(\Delta(D, x^*, \rho) \middle| \theta, \Delta(D, x^*, \rho) \le \frac{\eta\lambda}{\beta}\right) \approx \frac{\eta}{3\beta}$$

- Assuming the claim is correct, then

$$\lambda^{-1}\left[\eta\lambda - E_{x^*}\left(\Delta(D, x^*, \rho) \middle| \theta, \Delta(D, x^*, \rho) \le \frac{\eta\lambda}{\beta}\right)\right] \approx \eta\left(1 - \frac{1}{3\beta}\right) \equiv \omega(\beta, \eta)$$

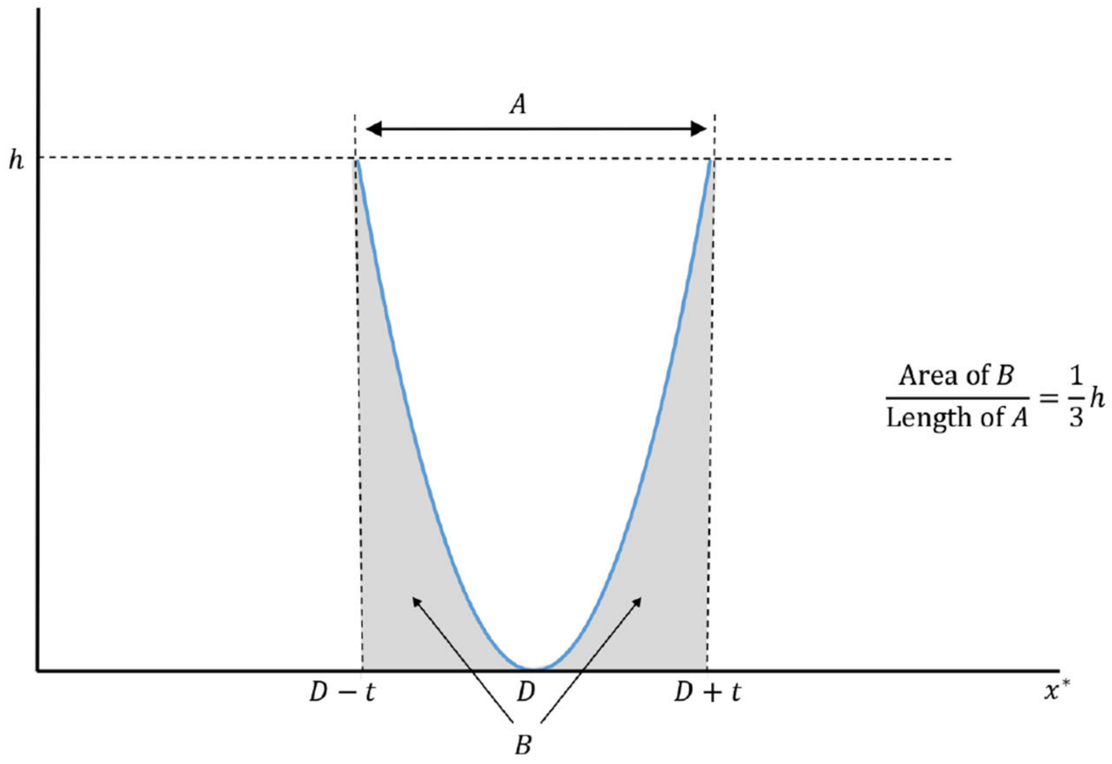
- So welfare maximization becomes (approximately) equivalent to opt-out minimization with weights $\omega(\beta, \eta)$

II. Optimal defaults

B. Theory

$$\lambda^{-1} E_{x^*}\left(\Delta(D, x^*, \rho)\middle|\theta, \Delta(D, x^*, \rho) \le \frac{\eta\lambda}{\beta}\right)$$

- To a 2nd-order approx., $\Delta(D, x^*, \rho)$ is a parabola with a minimized value of 0 at $x^* = D$
- For small λ, density is approx. constant over the opt-in interval
- $E_{x^*}(\cdot)$ is (approx.) the area under the parabola (B in the figure) divided by the width (A in the figure), which always equals $\frac{1}{3}h$
- Here, $h = \frac{\eta\lambda}{\beta}$, so $\lambda^{-1}E_{x^*}(\cdot)$ is approximately $\frac{\eta}{3\beta}$



A
h
$\frac{\text{Area of } B}{\text{Length of } A} = \frac{1}{3}h$
$D - t$
D
$D + t$
x^*
B

II. *Optimal defaults*

B. *Theory*

- The preceding observation points to the main result: weighted opt-out minimization with weights $\omega(\beta, \eta) = \eta\left(1 - \frac{1}{3\beta}\right)$ is approximately optimal

- Simulations show that the approximation is good even for substantial opt-out costs

- Suppose there's no heterogeneity in β. Then with sufficient bias ($\beta < \frac{1}{3}$), the sign flips, and the objective becomes weighted opt-out *maximization* rather than minimization

- It also follows that, as long as β exceeds $\frac{1}{3}$, normative ambiguity (concerning β) does not impact the optimum (but it can have an enormous effect if β is less than $\frac{1}{3}$)

- For cases with bunching and finite menus, the same result holds, except that the weights are simply $\omega(\beta, \eta) = \eta$.

II. Optimal defaults

B. Theory

- A corollary of the main result: unweighted opt-out minimization (which is easier to implement) is approximately optimal when it coincides in the limit with weighted opt-out minimization
- Sufficient conditions:
  1. The characteristics β and η are independent of x^* and ρ
  2. Limited mass at “small” values of β
- One can relax condition 2 if the firm can impose optimal fees for passive choice

II. Optimal defaults

C. Empirical implementation

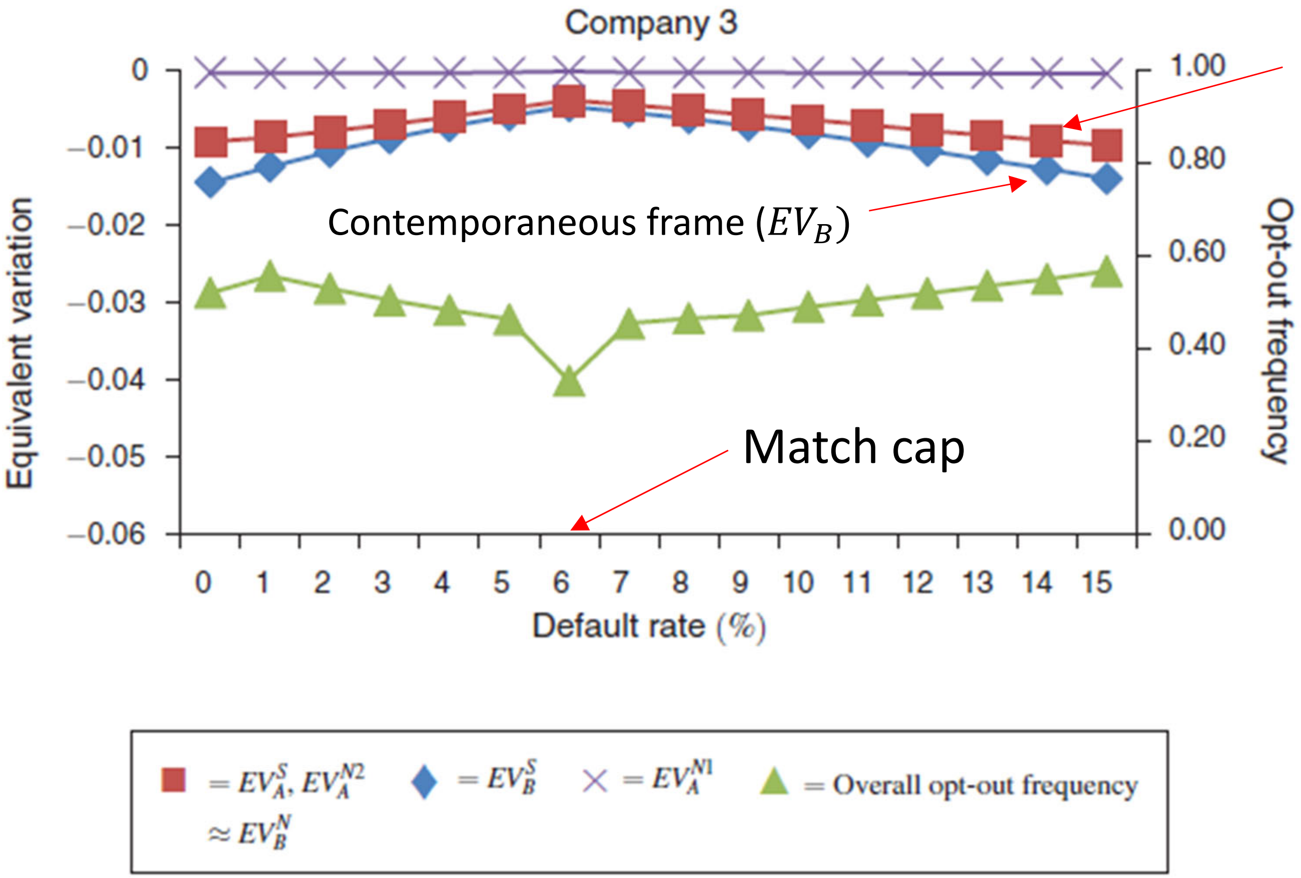
- Bernheim, Fradkin, and Popov (2015) study this issue empirically
  - Consider multiple theories of default effects (opt-out costs, sophisticated and naïve time inconsistency, inattention, anchoring)
  - Calibrate models to the available data
  - In each case, entertain the entire range of possible assumptions about the Welfare-Relevant Domain
- Findings:
  - As-if opt-out costs must be very high (thousands of dollars) to rationalize magnitude of default effects (Note: Choukhmane, 2021, finds opt-out costs of ~$250 rationalize choices in a fully dynamic model due to catch-up)
  - The data favor anchoring (because of size of as-if opt-out costs, and the absence of a "trough" in the distribution of choices near the default)
  - Generally, firms should optimize default contribution rates by setting them at points of accumulation in the distribution of ideal choices – boundaries (0 or max), kink points (match cap). Usually coincides empirically with opt-out minimization.
  - For three of the four theories, alternative assumptions about the WRD have surprisingly little effect on the optima or the implied welfare effects

II. Optimal defaults

C. Empirical implementation

For the case of sophisticated time-inconsistency (assuming costs of output are $25 to $30 from the forward-looking perspective):

Note: EV is measured as a percent of income



In-advance frame (EV_A)

Contemporaneous frame (EV_B)

Match cap

Generally,

II. Optimal defaults

C. Empirical implementation

- How could the normative ambiguity be so small?
  - Calculated welfare is higher when welfare is evaluated from the perspective of in-advance rather than contemporaneous choices because we effectively exclude almost all of the as-if opt-out costs from consideration
  - If the as-if opt-out costs are $2,500, then the difference should be about 5% of a $50,000 income, not 0.5%
- Explanation
  - There is a distribution of as-if opt-out costs
  - When calculating welfare from the perspective of the in-advance frame, we exclude almost all of the opt-out costs *for those who opt out*
  - Those with high opt-out costs don’t opt out, so changing the frame of evaluation doesn’t change the welfare calculation for them
  - Changing the frame of evaluation only changes the welfare calculation for those who do opt out, and they tend to have relatively low opt-out costs