The Welfare Economics of Reference Dependence

Daniel Reck, University of Maryland
Arthur Seibold, University of Mannheim

March 2024

Motivation

- Individuals often evaluate options relative to a reference point, especially seeking to avoid losses
  - Evidence from classic experiments (e.g. Kahneman & Tversky 1979; Kahneman, Knetsch, & Thaler 1990)
  - Field evidence: **labor supply** (Camerer et al. 1997, Fehr & Goette 2007, Crawford & Meng 2011), **responses to taxation** (Homonoff 2018, Rees-Jones 2018), **job search** (DellaVigna et al 2017), **retirement** (Seibold 2021; Lalive et al 2023)

    $\rightarrow$ reference dependence shapes responses to policy reforms
- **Open question:** How to evaluate the welfare effects of policy reforms in the presence of reference dependence?
  - Evaluating price instruments/taxes
  - Evaluating policies that influence reference points

Motivation

- Individuals often evaluate options relative to a reference point, especially seeking to avoid losses
  - Evidence from classic experiments (e.g. Kahneman & Tversky 1979; Kahneman, Knetsch, & Thaler 1990)
  - Field evidence: **labor supply** (Camerer et al. 1997, Fehr & Goette 2007, Crawford & Meng 2011), **responses to taxation** (Homonoff 2018, Rees-Jones 2018), **job search** (DellaVigna et al 2017), **retirement** (Seibold 2021; Lalive et al 2023)

    $\rightarrow$ reference dependence shapes responses to policy reforms
- **Open question:** How to evaluate the welfare effects of policy reforms in the presence of reference dependence?
  - Evaluating price instruments/taxes
  - Evaluating policies that influence reference points

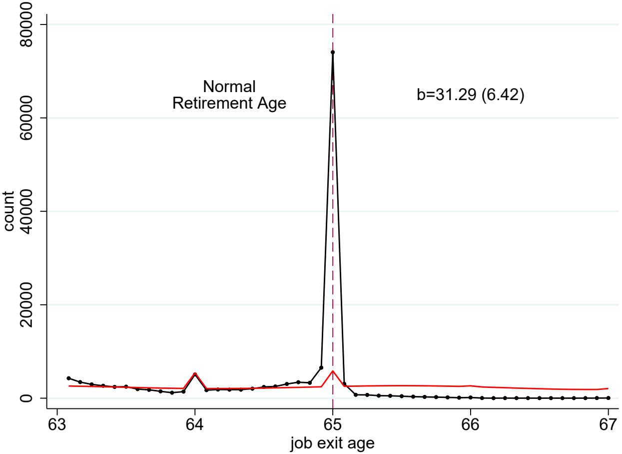
Challenges

1. Normative ambiguity: Is reference dependence a bias or a preference? (see e.g. O'Donoghue & Sprenger 2018)
    - **Our approach**: parametrize as normative judgment, identify map to welfare conclusions (Goldin & Reck 2022)
2. Positive ambiguity: many formulations of reference-dependent payoffs proposed in prior literature
    - Prior focus on tractability & identification, not welfare
    - **Our approach:** derive sufficient statistics
        - Reduced-form characterization of welfare under minimal conditions
        - Relate first-order determinants of welfare to parametric payoff formulations and empirical bunching designs

Challenges

1. Normative ambiguity: Is reference dependence a bias or a preference? (see e.g. O'Donoghue & Sprenger 2018)
    - **Our approach**: parametrize as normative judgment, identify map to welfare conclusions (Goldin & Reck 2022)
2. Positive ambiguity: many formulations of reference-dependent payoffs proposed in prior literature
    - Prior focus on tractability & identification, not welfare
    - **Our approach:** derive sufficient statistics
        - Reduced-form characterization of welfare under minimal conditions
        - Relate first-order determinants of welfare to parametric payoff formulations and empirical bunching designs

Empirical Application: Retirement Behavior



- Evaluate welfare effects of pension reforms: Normal Retirement Age as reference point + financial incentives

Preview of Results: Theory

- We decompose welfare effects of changes to reference points and prices into **direct effects** and **behavioral effects**
  - Normative judgments determine which effects matter
  - Payoff formulation determines the sign of the effects
- Propose flexible **reduced form** of reference-dependent payoffs capturing key features relevant for welfare
  - Encompasses wide range of formulations from prior literature
  - Two key parameters govern (i) strength and (ii) direction of loss aversion
- Show that reduced-form parameters are
  - **Sufficient statistics** for welfare (together with a price elasticity)
  - **Empirically identified** by bunching designs

Preview of Results: Theory

- We decompose welfare effects of changes to reference points and prices into **direct effects** and **behavioral effects**
  - Normative judgments determine which effects matter
  - Payoff formulation determines the sign of the effects
- Propose flexible **reduced form** of reference-dependent payoffs capturing key features relevant for welfare
  - Encompasses wide range of formulations from prior literature
  - Two key parameters govern (i) strength and (ii) direction of loss aversion
- Show that reduced-form parameters are
  - **Sufficient statistics** for welfare (together with a price elasticity)
  - **Empirically identified** by bunching designs

Preview of Results: Theory

- We decompose welfare effects of changes to reference points and prices into **direct effects** and **behavioral effects**
  - Normative judgments determine which effects matter
  - Payoff formulation determines the sign of the effects
- Propose flexible **reduced form** of reference-dependent payoffs capturing key features relevant for welfare
  - Encompasses wide range of formulations from prior literature
  - Two key parameters govern (i) strength and (ii) direction of loss aversion
- Show that reduced-form parameters are
  - **Sufficient statistics** for welfare (together with a price elasticity)
  - **Empirically identified** by bunching designs

Preview of Results: Empirical Application

Evaluate welfare effects of pension reforms using German administrative data

- Consider two types of reforms:
  - Shift Normal Retirement Age (NRA) $\Longrightarrow$ influence reference points
  - Change financial retirement incentives $\Longrightarrow$ price change
- Find positive welfare effects of increasing NRA (locally)
  - Crucial: bunching estimation suggests strong loss aversion over leisure $\Longrightarrow$ increasing NRA *lowers* reference points
  - Optimal NRA disciplined by potential consumption reference dependence
- Welfare effects of subsidizing later retirement ambiguous

Preview of Results: Empirical Application

Evaluate welfare effects of pension reforms using German administrative data

- Consider two types of reforms:
  - Shift Normal Retirement Age (NRA) $\Longrightarrow$ influence reference points
  - Change financial retirement incentives $\Longrightarrow$ price change
- Find positive welfare effects of increasing NRA (locally)
  - Crucial: bunching estimation suggests strong loss aversion over leisure $\Longrightarrow$ increasing NRA *lowers* reference points
  - Optimal NRA disciplined by potential consumption reference dependence
- Welfare effects of subsidizing later retirement ambiguous

Literature

1. **Behavioral welfare economics**: Chetty et al. (2009), Mullainathan et al. (2012), Allcott & Taubinsky (2015), Allcott et al. (2019), List et al. (2023)
    - Normative ambiguity: Bernheim & Rangel (2009), Goldin & Reck (2022)
2. **Reference-dependent preferences**: Kahneman & Tversky (1979), Tversky & Kahneman (1991), Köszegi & Rabin (2006, 2007), O'Donoghue & Sprenger (2018), Masatlioglu & Ellis (2022)
    - Field evidence: Camerer et al. (1997), DellaVigna et al. (2017), Rees-Jones (2018), Seibold (2021), Andersen et al. (2022), etc.

   $\rightarrow$ **Our contribution**: first welfare analysis
3. **Retirement behavior**: Behaghel & Blau (2012), Brown (2013), Manoli & Weber (2016), Gelber et al. (2020), Gruber et al. (2022), Lalive et al. (2023)
    - Welfare and pension reforms: Haller (2022), Kolsrud et al. (2023)

   $\rightarrow$ **Our contribution**: incorporate reference dependence into welfare effects of pension reforms

Model: Setup

- Consumption good x, numeraire y, quasi-linear preferences, non-stochastic environment, price p, *reference point* r.

  ▸ Whence r?

$$\max_{x,y} \underbrace{u(x) + y}_{\textit{Intrinsic Utility}} + \underbrace{v(x, r)}_{\textit{Ref.-dep. payoff}}$$

subject to $px + y = z$

- **Welfare:** should reference-dependent payoffs be given normative weight? $\rightarrow$ parameter $\pi \in \{0, 1\}$.

$$w(p, r) = u(x(p, r)) + z - px(p, r) + \pi v(x(p, r), r)$$

  ▸ Revealed Preferences

Model: Setup

- Consumption good x, numeraire y, quasi-linear preferences, non-stochastic environment, price p, *reference point* r.

▸ Whence r?

$$\max_{x,y} \underbrace{u(x) + y}_{\textit{Intrinsic Utility}} + \underbrace{v(x, r)}_{\textit{Ref.-dep. payoff}}$$

$$\text{subject to } px + y = z$$

- **Welfare:** should reference-dependent payoffs be given normative weight? $\rightarrow$ parameter $\pi \in \{0, 1\}$.

$$w(p, r) = u(x(p, r)) + z - px(p, r) + \pi v(x(p, r), r)$$

▸ Revealed Preferences

Theoretical Results: Welfare and Reference Points

Formal Version

$$w = u(x) + z - px + \pi v(x, r)$$

General characterization: under minimal conditions on $v(x, r)$,

$$w_r = \underbrace{-(1-\pi) v_x x_r}_{\text{Behavioral Effect}} \underbrace{+\pi v_r,}_{\text{Direct Effect}}$$

- Which effect matters for welfare depends on π
- Assume no diminishing sensitivity
  - Behavioral & direct effects are **same-signed** $\rightarrow$ sign of w_r invariant to judgment π!
  - To determine sign, pinning down v_x is crucial $\leftrightarrow$ How does ref. dep. modify willingness to pay for x?

Note: Partial derivatives v_x, v_r do not exist where $x(p, r) = r$ (i.e. when bunching at reference point). We derive behavioral/direct effects characterization there too.

Theoretical Results: Welfare and Reference Points

Formal Version

$$w = u(x) + z - px + \pi v(x, r)$$

General characterization: under minimal conditions on $v(x, r)$,

$$w_r = \underbrace{-(1-\pi) v_x x_r}_{\text{Behavioral Effect}} \; \underbrace{+\pi v_r}_{\text{Direct Effect}},$$

- Which effect matters for welfare depends on π
- Assume no diminishing sensitivity
  - Behavioral & direct effects are **same-signed**
    $\rightarrow$ sign of w_r invariant to judgment π!
  - To determine sign, pinning down v_x is crucial
    $\leftrightarrow$ How does ref. dep. modify willingness to pay for x?

Note: Partial derivatives v_x, v_r do not exist where $x(p, r) = r$ (i.e. when bunching at reference point). We derive behavioral/direct effects characterization there too.

Theoretical Results: Welfare and Reference Points

Formal Version

$$w = u(x) + z - px + \pi v(x, r)$$

General characterization: under minimal conditions on $v(x, r)$,

$$w_r = \underbrace{-(1-\pi) v_x x_r}_{\text{Behavioral Effect}} \; \underbrace{+\pi v_r,}_{\text{Direct Effect}}$$

- Which effect matters for welfare depends on π
- Assume no diminishing sensitivity
  - Behavioral & direct effects are **same-signed**
    $\rightarrow$ sign of w_r invariant to judgment π!
  - To determine sign, pinning down v_x is crucial
    $\leftrightarrow$ How does ref. dep. modify willingness to pay for x?

Note: Partial derivatives v_x, v_r do not exist where $x(p, r) = r$ (i.e. when bunching at reference point). We derive behavioral/direct effects characterization there too.

Theoretical Results: Welfare and Prices

$$w = u(x) + z - px + \pi v(x, r)$$

General characterization:

$$w_p = \underbrace{-(1-\pi)v_x x_p}_{\text{Behavioral Effect}} \quad \underbrace{-x(p, r)}_{\text{Direct Effect (Roy)}}$$

- First-order behavioral effect only in the bias case $(\pi = 0)$
- Scope for corrective taxation pivots on normative judgment: *marginal internality* $= -(1-\pi)v_x$
- Again, v_x is key $\rightarrow$ next, turn to payoff formulations

Theoretical Results: Welfare and Prices

$$w = u(x) + z - px + \pi v(x, r)$$

General characterization:

$$w_p = \underbrace{-(1-\pi) v_x x_p}_{\text{Behavioral Effect}} \quad \underbrace{-x(p, r)}_{\text{Direct Effect (Roy)}}$$

- First-order behavioral effect only in the bias case $(\pi = 0)$
- Scope for corrective taxation pivots on normative judgment: *marginal internality* $= -(1-\pi) v_x$
- Again, v_x is key $\rightarrow$ next, turn to payoff formulations

Theoretical Results: Welfare and Prices

$$w = u(x) + z - px + \pi v(x, r)$$

General characterization:

$$w_p = \underbrace{-(1-\pi) v_x x_p}_{\text{Behavioral Effect}} \quad \underbrace{-x(p, r)}_{\text{Direct Effect (Roy)}}$$

- First-order behavioral effect only in the bias case $(\pi = 0)$
- Scope for corrective taxation pivots on normative judgment: *marginal internality* $= -(1-\pi) v_x$
- Again, v_x is key $\rightarrow$ next, turn to payoff formulations

Reduced-Form Reference-Dependent Payoffs

$$v(x,r) = \begin{cases} -\beta\Lambda(x-r) & x \geq r \\ (1-\beta)\Lambda(x-r) & x < r \end{cases}$$

- $\Lambda > 0$ captures the *magnitude* of loss aversion
- $\beta \in [0,1]$ captures the *direction* of loss aversion (over x vs. y), and other potential factors (e.g. payoffs over gains)
- Encompasses formulations from prior literature (incl. Tversky & Kahneman 1991; Köszegi & Rabin 2006; Crawford & Meng 2011, DellaVigna et al. 2017, Rees-Jones 2018, Thakral & Tô 2021, Seibold 2021, Andersen et al. 2022) Examples Details
- But avoids imposing strong ex ante structure on welfare
  - e.g. *Simple Loss Aversion* requires $\beta = 0 \implies v_x \geq 0$

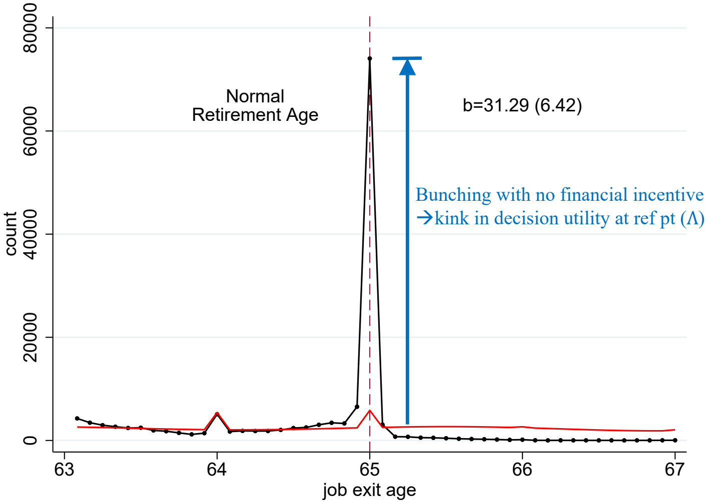
Reduced-Form Reference-Dependent Payoffs

$$v(x,r) = \begin{cases} -\beta\Lambda(x-r) & x \geq r \\ (1-\beta)\Lambda(x-r) & x < r \end{cases}$$

- $\Lambda > 0$ captures the *magnitude* of loss aversion
- $\beta \in [0,1]$ captures the *direction* of loss aversion (over x vs. y), and other potential factors (e.g. payoffs over gains)
- Encompasses formulations from prior literature

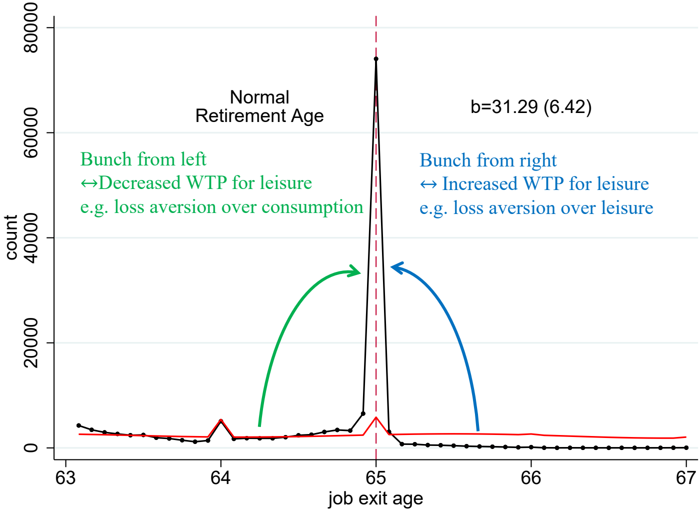
  (incl. Tversky & Kahneman 1991; Köszegi & Rabin 2006; Crawford & Meng 2011, DellaVigna et al. 2017, Rees-Jones 2018, Thakral & Tô 2021, Seibold 2021, Andersen et al. 2022) Examples Details
- But avoids imposing strong ex ante structure on welfare
  - e.g. *Simple Loss Aversion* requires $\beta = 0 \implies v_x \geq 0$

Reduced-Form Intuition: Rationalizing Bunching



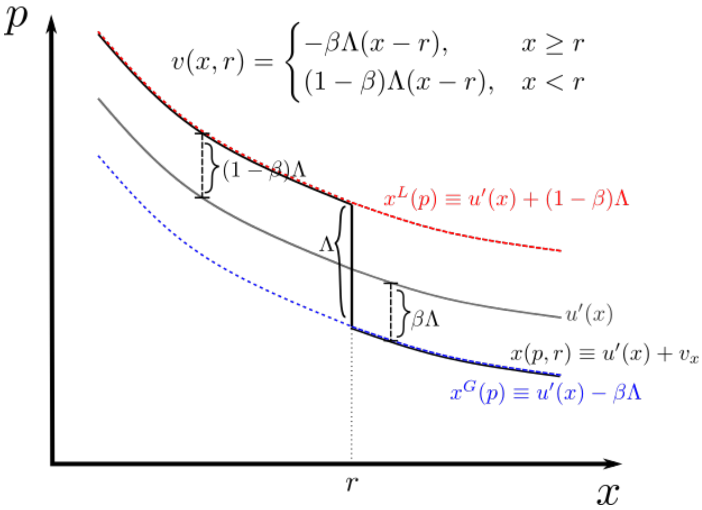
Magnitude of bunching responses governed by Λ

Reduced-Form Intuition: Rationalizing Bunching



Direction of bunching responses governed by β Illustration

Demand with Reduced-Form Payoff Formulation



Welfare effects of interest correspond to areas in graph ▸ Illustration

Social Welfare: Sufficient Statistics Formulas

Assume Utilitarian social welfare, index individuals by i. Groups G, L, R with $x_i(p,r)$ above, below and equal to r.

Social welfare effect of a change in the reference point Δr:

$$\Delta W \approx \Delta r \pi \left\{ \underbrace{E[\beta_i \Lambda_i | G] P[G]}_{\text{Direct effect for } G} - \underbrace{E[(1-\beta_i)\Lambda_i | L] P[L]}_{\text{Direct effect for } L} \right\} + \Delta r \underbrace{E\left[\Lambda_i \left(\beta_i - \frac{1}{2} \right) \middle| R \right] P[R]}_{\text{Direct=Behavioral effect for } R}.$$

Social welfare effect of a price change Δp:

$$\Delta W \approx (1-\pi)\frac{\Delta p}{p} \left\{ E\left[(1-\beta_i)\Lambda_i \varepsilon_i x_i \right| L \right] P[L] - E\left[\beta_i \Lambda_i \varepsilon_i x_i \right| G \left] P[G] \right\} - \Delta p E[x_i]$$

Social Welfare: Sufficient Statistics Formulas

Assume Utilitarian social welfare, index individuals by i. Groups G, L, R with $x_i(p, r)$ above, below and equal to r.

Social welfare effect of a change in the reference point Δr:

$$\Delta W \approx \Delta r \pi \left\{ E[\beta_i \Lambda_i | G] P[G] - E[(1-\beta_i)\Lambda_i | L] P[L] \right\} + \Delta r E\left[\Lambda_i \left(\beta_i - \frac{1}{2} \right) \middle| R \right] P[R].$$

Social welfare effect of a price change Δp:

$$\Delta W \approx (1-\pi)\frac{\Delta p}{p} \left\{ \underbrace{E\left[(1-\beta_i)\Lambda_i \varepsilon_i x_i \right| L\right] P[L]}_{\text{internality*response}>0 \text{ in } L} - \underbrace{E\left[\beta_i \Lambda_i \varepsilon_i x_i \right| G\left] P[G]}_{\text{internality*response}<0 \text{ in } G} \right\} - \underbrace{\Delta p E[x_i]}_{\text{Direct effect for any } \pi}$$

Sufficient Statistics and Empirical Identification

Key Result 1: **Sufficient Statistics** for Welfare

- Sufficient statistics for welfare effects are $E[\Lambda_i]$, $E[\beta_i]$ and π (assuming mutual independence)
- Plus price elasticity $E[\varepsilon_i]$ for Δp

Key Result 2: **Empirical Identification** from Bunching

- Bunching at reference point identifies $E[\Lambda_i]$
  - See also Rees-Jones (2018), Seibold (2021)
- Share of bunching from the left identifies $E[\beta_i]$
  - "Counterfactual density" captures *intrinsic WTP*, left bunching share captures how ref. dep. modifies WTP (v_x)

Sufficient Statistics and Empirical Identification

Key Result 1: **Sufficient Statistics** for Welfare

- Sufficient statistics for welfare effects are $E[\Lambda_i]$, $E[\beta_i]$ and π (assuming mutual independence)
- Plus price elasticity $E[\varepsilon_i]$ for Δp

Key Result 2: **Empirical Identification** from Bunching

- Bunching at reference point identifies $E[\Lambda_i]$
  - See also Rees-Jones (2018), Seibold (2021)
- Share of bunching from the left identifies $E[\beta_i]$
  - "Counterfactual density" captures *intrinsic WTP*, left bunching share captures how ref. dep. modifies WTP (v_x)

Empirical Application: Retirement Behavior

- Seibold (2021): reference dependence explains bunching responses to Normal Retirement Age (NRA) in Germany
  - NRA: salient threshold, framed as "normal time to retire"
- Simulate effects of two policies
  1. Increasing the NRA from 65 to 66 $\rightarrow$ shifts reference points
     - Strong effect on average retirement age: +4.5 months
  2. Increasing financial incentives for late retirement (Delayed Retirement Credit, DRC) $\rightarrow$ changes price (of leisure)
     - DRC increase from 6% to 10.4% per year yields same effect on average retirement age as NRA reform
- Goal: estimate (money-metric) welfare effects of these reforms
- Use high-quality administrative data on German retirees

Empirical Application: Retirement Behavior

- Seibold (2021): reference dependence explains bunching responses to Normal Retirement Age (NRA) in Germany
  - NRA: salient threshold, framed as "normal time to retire"
- Simulate effects of two policies
  1. Increasing the NRA from 65 to 66 $\rightarrow$ shifts reference points
     - Strong effect on average retirement age: +4.5 months
  2. Increasing financial incentives for late retirement (Delayed Retirement Credit, DRC) $\rightarrow$ changes price (of leisure)
     - DRC increase from 6% to 10.4% per year yields same effect on average retirement age as NRA reform
- Goal: estimate (money-metric) welfare effects of these reforms
- Use high-quality administrative data on German retirees

Empirical Application: Retirement Behavior

- Seibold (2021): reference dependence explains bunching responses to Normal Retirement Age (NRA) in Germany
  - NRA: salient threshold, framed as "normal time to retire"
- Simulate effects of two policies
  1. Increasing the NRA from 65 to 66 $\rightarrow$ shifts reference points
     - Strong effect on average retirement age: $+4.5$ months
  2. Increasing financial incentives for late retirement (Delayed Retirement Credit, DRC) $\rightarrow$ changes price (of leisure)
     - DRC increase from 6% to 10.4% per year yields same effect on average retirement age as NRA reform
- Goal: estimate (money-metric) welfare effects of these reforms
- Use high-quality administrative data on German retirees

Direction of Loss Aversion in the Empirical Application

- Challenge: point-identifying β via counterfactual density requires strong assumptions (Blomquist et al. 2021)
- We begin with a specification assuming Simple Loss Aversion over leisure $(\beta = 0)$
  - Empirically, loss aversion over leisure appears *a priori* dominant
    Illustration
- Then we relax this restriction, allow for $\beta \geq 0$. Here: loss aversion over consumption (Behaghel-Blau 2012)
  1. Point-identify direction of loss aversion (β) under additional assumptions $\rightarrow$ similar qualitative results
  2. Partially identify possibilities consistent with observed bunching $\rightarrow$ for most plausible combinations, similar qualitative results

Direction of Loss Aversion in the Empirical Application

- Challenge: point-identifying β via counterfactual density requires strong assumptions (Blomquist et al. 2021)
- We begin with a specification assuming Simple Loss Aversion over leisure $(\beta = 0)$
  - Empirically, loss aversion over leisure appears *a priori* dominant
    ▸ Illustration
- Then we relax this restriction, allow for $\beta \geq 0$. Here: loss aversion over consumption (Behaghel-Blau 2012)
  1. Point-identify direction of loss aversion (β) under additional assumptions $\rightarrow$ similar qualitative results
  2. Partially identify possibilities consistent with observed bunching $\rightarrow$ for most plausible combinations, similar qualitative results

Empirical Specification

Baseline Model with Simple Loss Aversion over Lifetime Leisure ($\beta = 0$):

$$U_i(C, R) = C - \frac{n_i}{1 + \frac{1}{\varepsilon}} \left(\frac{R}{n_i} \right)^{1+\frac{1}{\varepsilon}} - \begin{cases} 0 & R < \hat{R} \\ \widetilde{\Lambda}(R - \hat{R}) & R \geq \hat{R} \end{cases}$$

R: retirement age, $\hat{R}$: reference pt, C: consumption (NPV at 65).

- **Crucial:** reference dependence in terms of retirement age $\equiv$ loss aversion over lifetime leisure
  - $R \geq \hat{R}$ is the *loss domain* for leisure
  - Increase NRA $\equiv$ decrease reference point
- We estimate parameters via bunching and simulate behavior, welfare under various policy scenarios

Empirical Specification

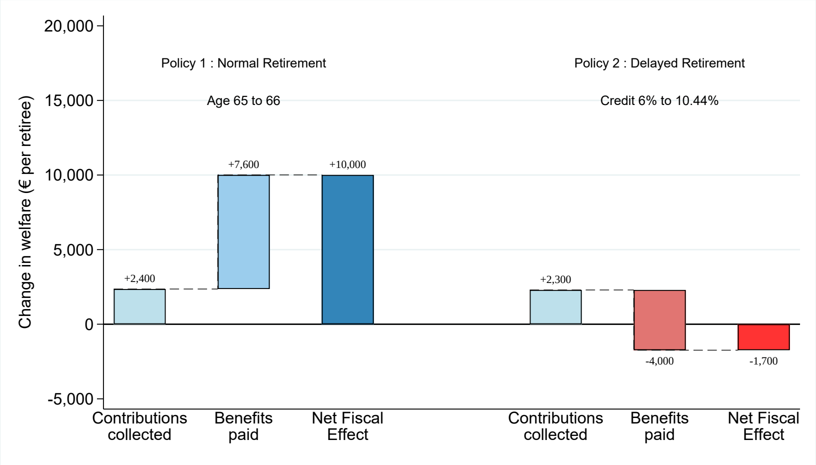
Baseline Model with Simple Loss Aversion over Lifetime Leisure ($\beta = 0$):

$$U_i(C, R) = C - \frac{n_i}{1 + \frac{1}{\varepsilon}} \left(\frac{R}{n_i} \right)^{1 + \frac{1}{\varepsilon}} - \begin{cases} 0 & R < \hat{R} \\ \widetilde{\Lambda}(R - \hat{R}) & R \geq \hat{R} \end{cases}$$

R: retirement age, $\hat{R}$: reference pt, C: consumption (NPV at 65).

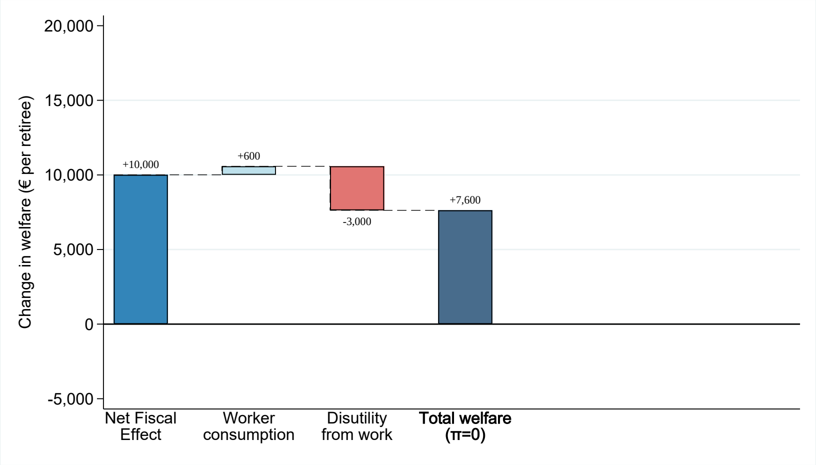
- **Crucial:** reference dependence in terms of retirement age $\equiv$ loss aversion over lifetime leisure
  - $R \geq \hat{R}$ is the *loss domain* for leisure
  - Increase NRA $\equiv$ decrease reference point
- We estimate parameters via bunching and simulate behavior, welfare under various policy scenarios

Simulated Reforms: Fiscal Effects



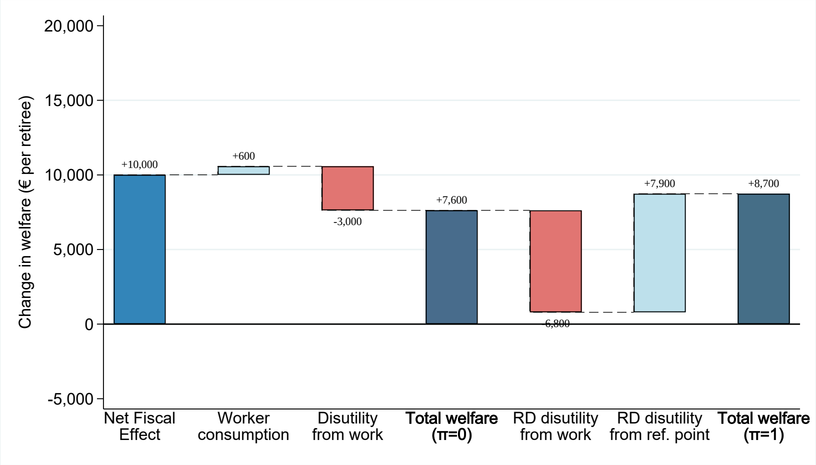
Fiscal externalities already favor increasing the NRA.

Increasing the Normal Retirement Age



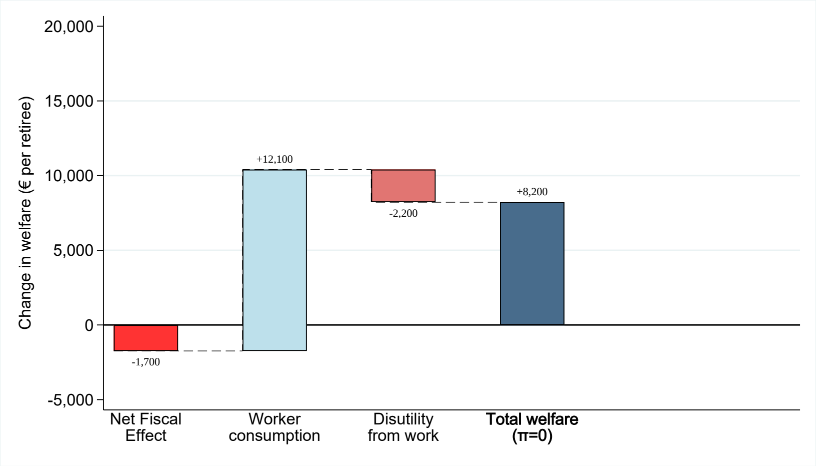
$\pi = 0$: **Reducing consumption of leisure improves welfare (behavioral effect).**

Increasing the Normal Retirement Age



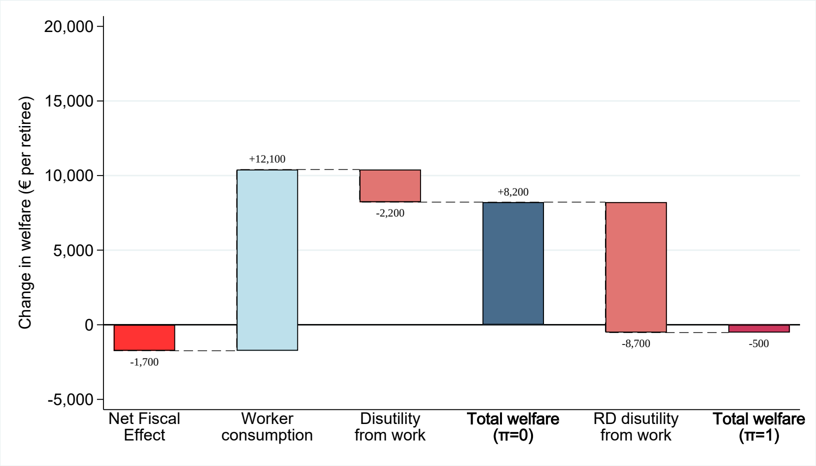
$\pi = 1$: Reduced leisure offset by ref. dep. payoff.
But raising NRA shrinks losses in leisure (direct effect).

Increasing the Delayed Retirement Credit



$\pi = 0$: higher DRC corrects over-consumption of leisure (behavioral effect).

Increasing the Delayed Retirement Credit

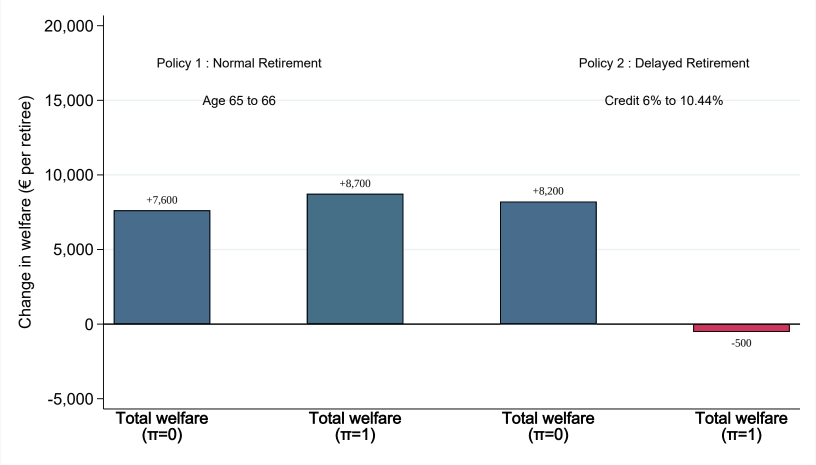


$\pi = 1$: no behavioral welfare effect.

Higher DRC is a distortionary tax on leisure.

Total Welfare Effects

▸ Extended Simulations ▸ NRA-Benefit Linkage

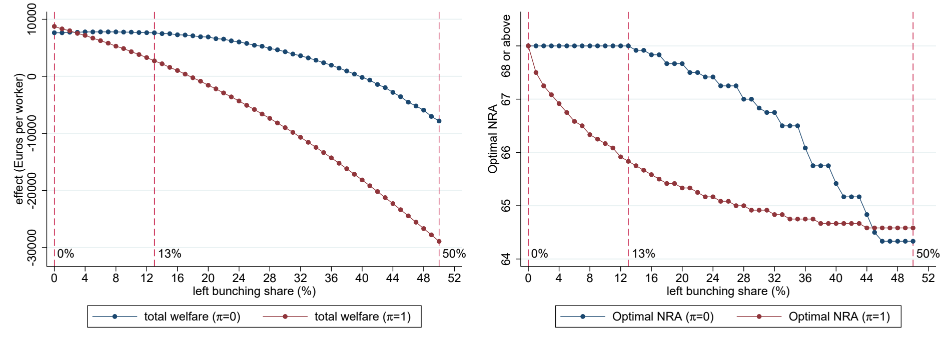


Increasing the NRA has positive welfare effects regardless of π. Effects of financial incentives (DRC) highly ambiguous.

Welfare under Two-Dimensional Loss Aversion ($\beta > 0$)

(a) Welfare Effect of Increasing NRA

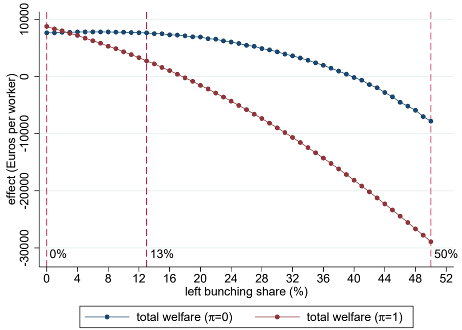
(b) Optimal NRA



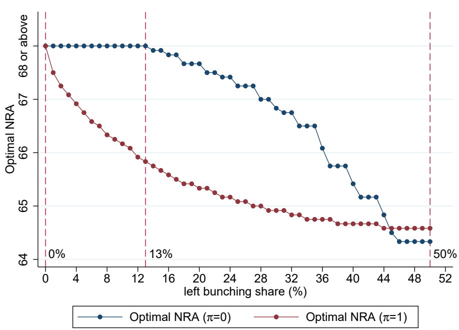
- We estimate $\approx 13\%$ bunching from the left. ▸ Graph
- With larger β, increasing NRA
  - implies more sub-optimally late retirement ($\pi = 0$) *OR* mounting consumption losses ($\pi = 1$)
  - makes it costlier to increase NRA, optimal NRA is lower

Welfare under Two-Dimensional Loss Aversion ($\beta > 0$)

(a) Welfare Effect of Increasing NRA



(b) Optimal NRA



- We estimate $\approx$ 13% bunching from the left. ▸ Graph
- With larger β, increasing NRA
  - implies more sub-optimally late retirement ($\pi = 0$) *OR* mounting consumption losses ($\pi = 1$)
  - makes it costlier to increase NRA, optimal NRA is lower

Conclusion

- We characterize welfare effects of policies under reference dependence:
  - General characterization: behavioral effects vs. direct effects
  - Sign of effects depends on form of payoffs; which effects matter depends normative judgements
- We apply the insights to pension design:
  - Loss aversion over *leisure* empirically dominant $\implies$ increasing NRA improves welfare (locally)
  - Optimal NRA increase disciplined by loss aversion over consumption (and potentially other factors)
  - Welfare effects of financial retirement incentives highly ambiguous

▸ Further Questions

Conclusion

- We characterize welfare effects of policies under reference dependence:
  - General characterization: behavioral effects vs. direct effects
  - Sign of effects depends on form of payoffs; which effects matter depends normative judgements
- We apply the insights to pension design:
  - Loss aversion over *leisure* empirically dominant
    $\Longrightarrow$ increasing NRA improves welfare (locally)
  - Optimal NRA increase disciplined by loss aversion over consumption (and potentially other factors)
  - Welfare effects of financial retirement incentives highly ambiguous

▸ Further Questions

THANK YOU!

Questions/Comments:

dreck@umd.edu
seibold@uni-mannheim.de

APPENDIX SLIDES

Is the reference point a policy parameter? Back

- We assume individuals evaluate options relative to an exogenous reference point r that can be influenced by policy
- The literature is unsettled on the origins of reference points
  - Salient options (Rosch 1975); status quo (Kahneman et al 1990); goals (Heath et al. 1999), beliefs/expectations (Köszegi and Rabin 2006, 2007), past experiences (Thakral and Tô 2020, DellaVigna et al. 2017)
- Growing evidence suggests policy can shift reference points in some settings, *at least locally*
  - Normal Retirement Age (Seibold 2021, Lalive et al 2023 Gruber et al 2020); Tax withholding rules (Rees-Jones 2018); Framing of Pigouvian incentives as taxes/subsidies (Homonoff 2018). Related experimental results in e.g. Kahneman et al (1990).
- Think of a generic policy reform dP:

$$\frac{dW}{dP} = \frac{\partial W}{\partial r}\frac{\partial r}{\partial P} + \frac{\partial W}{\partial P}$$

- We characterize $\frac{\partial W}{\partial r}$ in the theory, confront questions about $\frac{\partial r}{\partial P}$, $\frac{\partial W}{\partial P}$ in our empirical context.

Revealed Preference Foundations

Back

$$w(p, r) = u(x(p, r)) + z - px(p, r) + \pi v(x(p, r), r)$$

- Under $\pi = 1$, observed revealed preferences correspond to welfare
- Under $\pi = 0$, welfare coincides with intrinsic utility
  - Assume existence of a counterfactual frame in which individual maximizes intrinsic utility
  - Revealed preferences in this frame identify welfare (as in e.g. Chetty et al. 2009)
- Welfare criterion of Bernheim-Rangel (2009) $\iff$ Option A preferred to B for any $\pi \in \{0, 1\}$
- Quasi-linearity $\implies$ money-metric welfare, comparable under $\pi = 0$ and $\pi = 1$

Formulating Reference-Dependent Payoffs

Back

General form of reference-dependent payoffs:

$$v(x, r) = \nu(\mu(x) - \mu(r))$$

Assumptions:

- **A1**: $\mu(.)$ 2x-differentiable everywhere w/$\mu' > 0, \mu'' \leq 0$;
  $\nu(z)$ continuous everywhere & 2x-differentiable for any $z \neq 0$;
  $\nu(0) = 0$ (gain-loss payoff);
  $\nu'_-(0) > \nu'_+(0)$ (*loss aversion*).

- **A2**:
  1. $\nu(z)$ is monotone over $(-\infty, 0)$ and over $(0, \infty)$ (*domain-specific monotonicity*)
  2. $\nu''(z) = 0$ for any $z \neq 0$ (*No Diminishing Sensitivity*)

- These assumptions capture most payoff formulations proposed in prior literature, except diminishing sensitivity, see Appendix.

Formulating Reference-Dependent Payoffs ▸ Back

General form of reference-dependent payoffs:

$$v(x, r) = \nu(\mu(x) - \mu(r))$$

Assumptions:

- **A1**: $\mu(.)$ 2x-differentiable everywhere w/ $\mu' > 0, \mu'' \leq 0$;
  $\nu(z)$ continuous everywhere & 2x-differentiable for any $z \neq 0$;
  $\nu(0) = 0$ (gain-loss payoff);
  $\nu'_-(0) > \nu'_+(0)$ (*loss aversion*).

- **A2**:
  1. $\nu(z)$ is monotone over $(-\infty, 0)$ and over $(0, \infty)$ (*domain-specific monotonicity*)
  2. $\nu''(z) = 0$ for any $z \neq 0$ (*No Diminishing Sensitivity*)

- These assumptions capture most payoff formulations proposed in prior literature, except diminishing sensitivity, see Appendix.

Welfare Effect of Changing the Reference Point ▸ Back

For given (p, r) we find three cases for $x(p, r)$:

- $x(p, r) > r$: Gain domain (G); $x(p, r) < r$: Loss domain (L)
- $x(p, r) = r$: Reference domain (R)

Under A1, we find

$$(p, r) \notin R \implies w_r = \underbrace{-(1 - \pi) v_x x_r}_{\text{Behavioral Effect}} + \underbrace{\pi v_r}_{\text{Direct Effect}},$$

Partial derivatives (v_x, v_r) do not exist in R domain but we can find a similar characterization:

$$v^R(x, r) \equiv (1 - \pi) U(x, z - px) + \pi U(r, z - pr)$$

$$(p, r) \in R \implies w(p, r) = v^R(x(p, r), r)$$

$$\implies w_r = \underbrace{(1 - \pi) v_x^R x_r}_{\text{Behavioral Effect}} + \underbrace{\pi v_r^R}_{\text{Direct Effect}} = u'(r) - p.$$

Welfare Effect of Changing the Reference Point ▸ Back

For given (p, r) we find three cases for $x(p, r)$:

- $x(p, r) > r$: Gain domain (G); $x(p, r) < r$: Loss domain (L)
- $x(p, r) = r$: Reference domain (R)

Under A1, we find

$$(p, r) \notin R \implies w_r = \underbrace{-(1-\pi)v_x x_r}_{\text{Behavioral Effect}} + \underbrace{\pi v_r,}_{\text{Direct Effect}}$$

Partial derivatives (v_x, v_r) do not exist in R domain but we can find a similar characterization:

$$v^R(x, r) \equiv (1-\pi)U(x, z-px) + \pi U(r, z-pr)$$

$$(p, r) \in R \implies w(p, r) = v^R(x(p, r), r)$$

$$\implies w_r = \underbrace{(1-\pi)v_x^R x_r.}_{\text{Behavioral Effect}} + \underbrace{\pi v_r^R}_{\text{Direct Effect}} \quad = u'(r) - p.$$

Welfare Effect of Changing the Reference Point ▸ Back

For given (p, r) we find three cases for $x(p, r)$:

- $x(p, r) > r$: Gain domain (G); $x(p, r) < r$: Loss domain (L)
- $x(p, r) = r$: Reference domain (R)

Under A1, we find

$$(p, r) \notin R \implies w_r = \underbrace{-(1-\pi)v_x x_r}_{\text{Behavioral Effect}} + \underbrace{\pi v_r,}_{\text{Direct Effect}}$$

Partial derivatives (v_x, v_r) do not exist in R domain but we can find a similar characterization:

$$v^R(x, r) \equiv (1-\pi)U(x, z-px) + \pi U(r, z-pr)$$
$$(p, r) \in R \implies w(p, r) = v^R(x(p, r), r)$$
$$\implies w_r = \underbrace{(1-\pi)v_x^R x_r.}_{\text{Behavioral Effect}} + \underbrace{\pi v_r^R}_{\text{Direct Effect}} = u'(r) - p.$$

Signing Individual Welfare Effects of Δr Back

Proposition: Under A1 and A2, at least one of the following obtains:

- (*Everywhere Increasing*): $v_x \geq 0$ for all $x \neq r$, and $w_r(p, r) \leq 0$ *almost everywhere*
- (*Everywhere Decreasing*): $v_x \leq 0$ for all $x \neq r$, and $w_r(p, r) \geq 0$ *almost everywhere*
- (*Single-Peaked*) $v_x \geq 0$ for $x < r$ and $v_x \leq 0$ for $x > r$, and for the unique reference point r^* s.t. $u'(r^*) = p$, $w_r \geq 0$ for $r \leq r^*$ and $w_r \geq 0$ for $r \geq r^*$.

These conditions do not refer to π: sign of w_r invariant to normative judgments!

"Almost everywhere:" w_r might not exist at the boundary of R, which is measure zero.

Signing Individual Welfare Effects of Δr ▸ Back

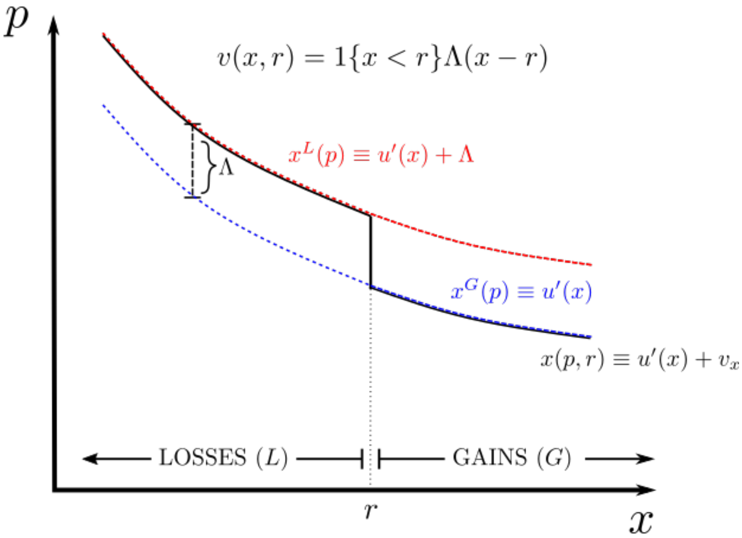
Proposition: Under A1 and A2, at least one of the following obtains:

- (*Everywhere Increasing*): $v_x \geq 0$ for all $x \neq r$, and $w_r(p, r) \leq 0$ *almost everywhere*
- (*Everywhere Decreasing*): $v_x \leq 0$ for all $x \neq r$, and $w_r(p, r) \geq 0$ *almost everywhere*
- (*Single-Peaked*) $v_x \geq 0$ for $x < r$ and $v_x \leq 0$ for $x > r$, and for the unique reference point r^* s.t. $u'(r^*) = p$, $w_r \geq 0$ for $r \leq r^*$ and $w_r \geq 0$ for $r \geq r^*$.

These conditions do not refer to π: sign of w_r invariant to normative judgments!

"Almost everywhere:" w_r might not exist at the boundary of R, which is measure zero.

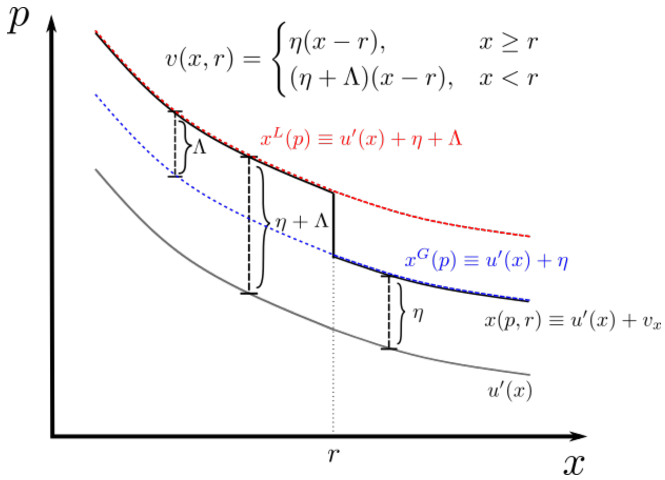
Example 1: Simple Loss Aversion ▸ Back



$v_x \geq 0$ everywhere; individually optimal r is any $r \in (-\infty, r^*]$, where $u'(r^*) = p$.

Ex 2: Loss Aversion Plus Gain Utility (Tversky & Kahneman 1991)

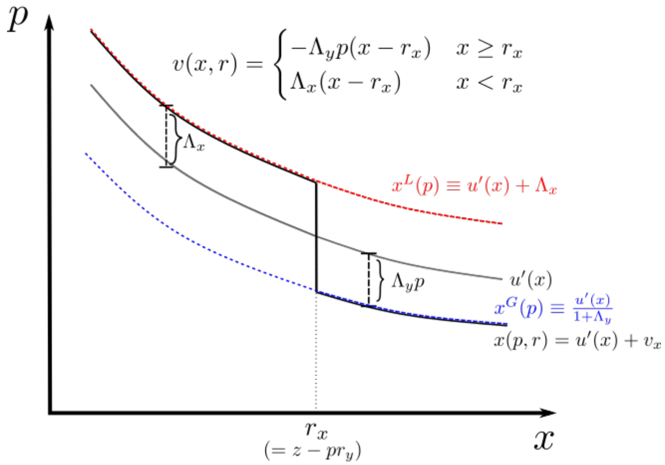
Back



$v_x > 0$ everywhere; individually optimal r is $(-\infty, r^*]$ for $\pi = 0$ and $-\infty$ for $\pi = 1$.

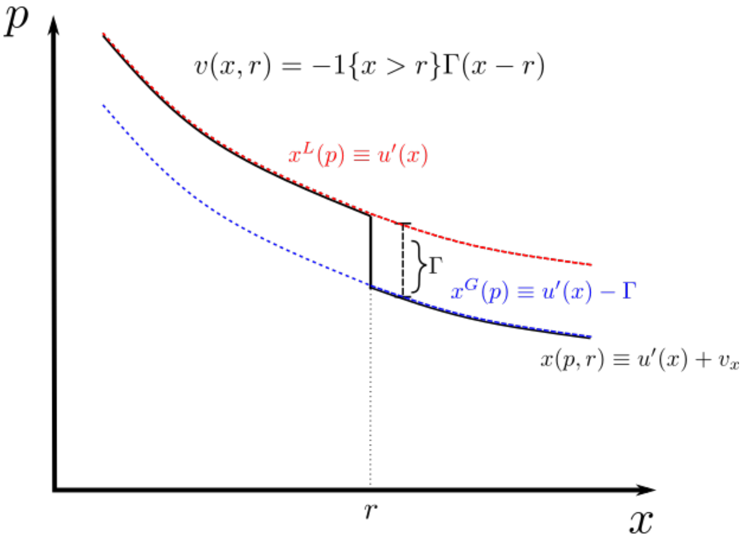
Ex 3: 2-Dimensional Loss Aversion, r on Budget Constraint

Back



v is *single-peaked* at r^*; welfare is peaked at intrinsic optimum r^*.

Ex 4: Gain Discounting ▸ Back



Resembles SLA over y; $v_x \leq 0$ everywhere. Individually optimal r is $r \in [r^*, \infty)$.

All Formulations (in Paper Appendix) Back

Description	(1) Reference-Dependent Payoff	(2) Assumptions A1 & A2	(3) Case
Simple Loss Aversion	$1\{x < r\}\Lambda(x-r)$	Yes	everywhere increasing + single-peaked
Loss Aversion with Gain Utility	$(\eta + 1\{x < r\}\Lambda)(x-r)$	Yes	everywhere increasing
Utils Formulation (Köszegi-Rabin)	$(\eta + 1\{x < r\}\Lambda)(u(x)-u(r))$	Yes	everywhere increasing
Gain Discounting	$1\{x > r\}\Gamma(x-r)$	Yes	everywhere decreasing + single-peaked
Simple Loss Aversion with Diminishing Sensitivity	$-\alpha^{-1}(1\{x < r\}\Lambda)(r-x)^{\alpha}$	2.2 Fails	N/A
Loss Aversion with Gain Utility & Diminishing Sensitivity	$\alpha^{-1}(\eta)(x-r)^{\alpha}$, if $x \geq r$ $-\alpha^{-1}(\eta+\Lambda)(r-x)^{\alpha}$, if $x < r$	2.2 Fails	N/A
Two-Dimensional Loss Aversion, (r_x, r_y) on budget constraint	$1\{x < r_x\}\Lambda_x(x-r_x)$ $+1\{y < r_y\}\Lambda_y(y-r_y)$	Yes	single-peaked
Two-Dimensional Loss Aversion with Gain Utility, (r_x, r_y) on budget constraint	$(\eta_x + 1\{x < r_x\}\Lambda_x)(x-r_x)+$ $(\eta_y + 1\{y < r_y\}\Lambda_y)(y-r_y)$	Yes	depends on parameters
Two-Dimensional Loss Aversion, any (r_x, r_y)	$1\{x < r_x\}\Lambda_x(x-r_x)$ $+1\{y < r_y\}\Lambda_y(y-r_y)$	1.3 Fails	N/A

Notes: The table summarizes the formulations of reference-dependent payoffs considered in the Appendix. Column (1) shows the functional form of reference-dependent payoffs for each formulation. Columns (2) and (3) describe the features of each formulation that pin down the sign of key welfare effects: whether the formulation satisfies Assumptions 1 and 2, and the which of the three possibilities for v_x obtains.

Flexible Reduced Form: Details ▸ Back

- We focus henceforth on $\beta \in [0, 1] \implies v$ is single-peaked.
  - $\beta < 0$ would generate extreme policy recommendations, and
  - *Multi-dimensional* KT91 payoff tends to be single-peaked
- Our formulation as a linear approximation of any formulation satisfying A1 & A2.
  - The approximation is quantitatively exact in the reference domain R.
  - Non-linearities become more important, quantitatively, the larger is $|x(p, r) - r|$, due e.g. to
    - Whether units of gains and losses $\mu(z)$ are units of the good or utils (see Kőszegi-Rabin 2006, Proposition 2)
    - Potentially also diminishing sensitivity, if we relax A2.2.
- A restriction Kőszegi & Rabin (2006) impose on differences in payoffs across dimensions would essentially imply $\beta = 0.5$.

Flexible Reduced Form: Details ▸ Back

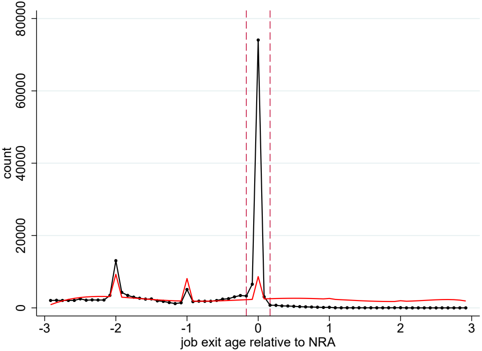
- We focus henceforth on $\beta \in [0,1] \implies v$ is single-peaked.
  - $\beta < 0$ would generate extreme policy recommendations, and
  - *Multi-dimensional* KT91 payoff tends to be single-peaked
- Our formulation as a linear approximation of any formulation satisfying A1 & A2.
  - The approximation is quantitatively exact in the reference domain R.
  - Non-linearities become more important, quantitatively, the larger is $|x(p,r) - r|$, due e.g. to
    - Whether units of gains and losses $\mu(z)$ are units of the good or utils (see Kőszegi-Rabin 2006, Proposition 2)
    - Potentially also diminishing sensitivity, if we relax A2.2.
- A restriction Kőszegi & Rabin (2006) impose on differences in payoffs across dimensions would essentially imply $\beta = 0.5$.

Flexible Reduced Form: Details ▸ Back

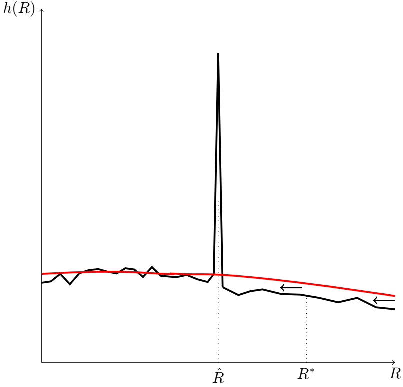
- We focus henceforth on $\beta \in [0,1] \implies v$ is single-peaked.
  - $\beta < 0$ would generate extreme policy recommendations, and
  - *Multi-dimensional* KT91 payoff tends to be single-peaked
- Our formulation as a linear approximation of any formulation satisfying A1 & A2.
  - The approximation is quantitatively exact in the reference domain R.
  - Non-linearities become more important, quantitatively, the larger is $|x(p,r) - r|$, due e.g. to
    - Whether units of gains and losses $\mu(z)$ are units of the good or utils (see Kőszegi-Rabin 2006, Proposition 2)
    - Potentially also diminishing sensitivity, if we relax A2.2.
- A restriction Kőszegi & Rabin (2006) impose on differences in payoffs across dimensions would essentially imply $\beta = 0.5$.

Bunching and the Dimensions of Loss Aversion

▸ Back to Theory ▸ Back to Empirical



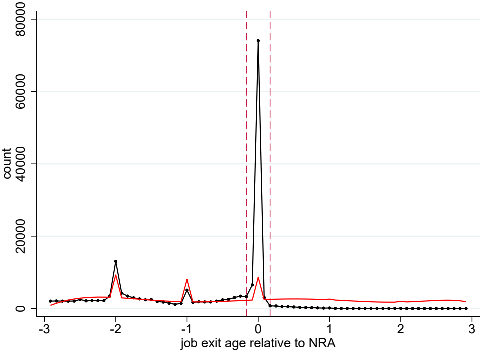
(a) Empirical Density



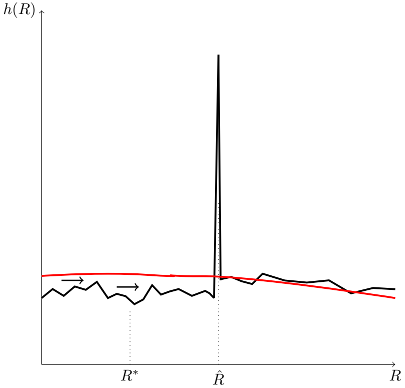
(b) Loss Aversion in Leisure

Bunching and the Dimensions of Loss Aversion

▸ Back to Theory ▸ Back to Empirical



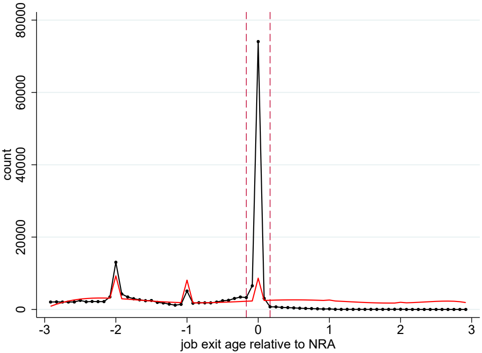
(a) Empirical Density



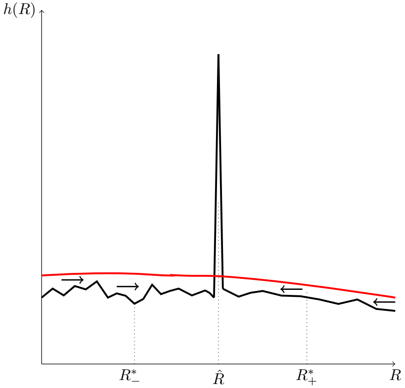
(b) Loss Aversion in Consumption

Bunching and the Dimensions of Loss Aversion

▸ Back to Theory ▸ Back to Empirical



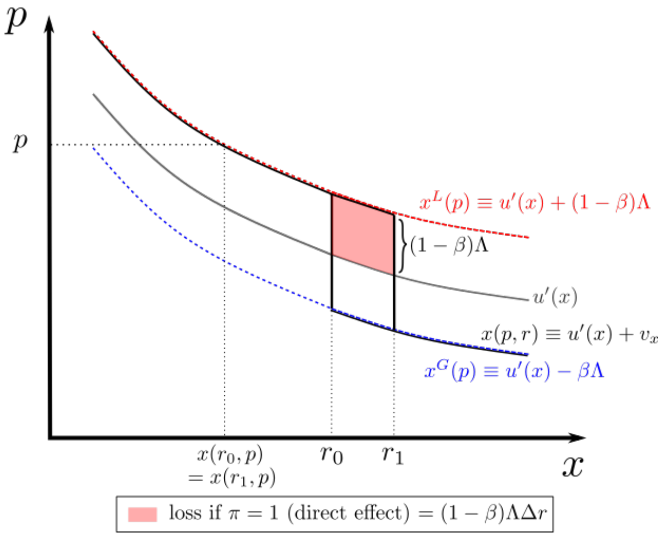
(a) Empirical Density



(b) Loss Aversion in Both Dimensions

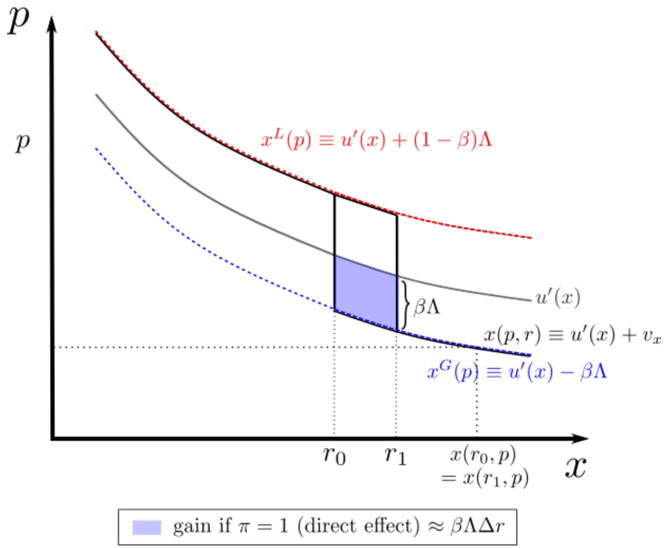
Welfare Effect of Increasing r: Loss Domain

Back



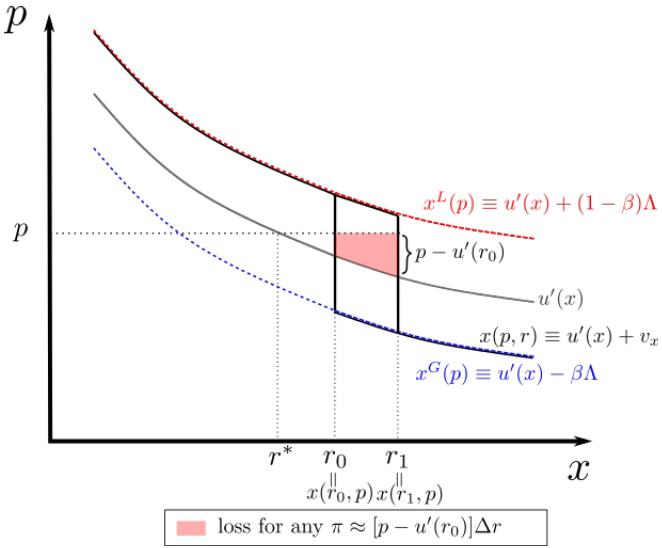
Welfare Effect of Increasing r: Gain Domain

Back

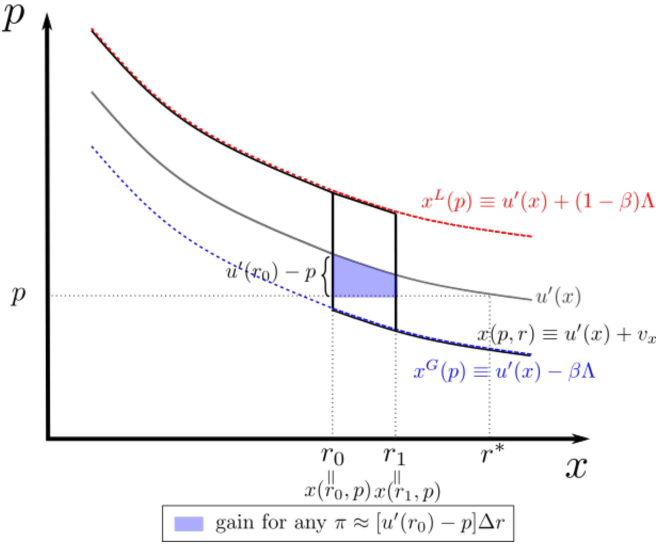


Welfare Effect of Increasing r: Reference Domain, $r > r^*$

▸ Back

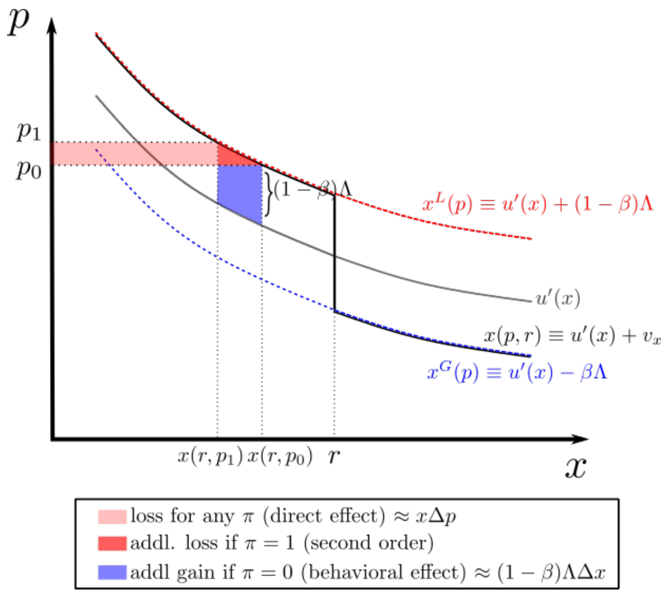


Welfare Effect of Increasing r: Reference Domain, $r < r^*$ ▸ Back



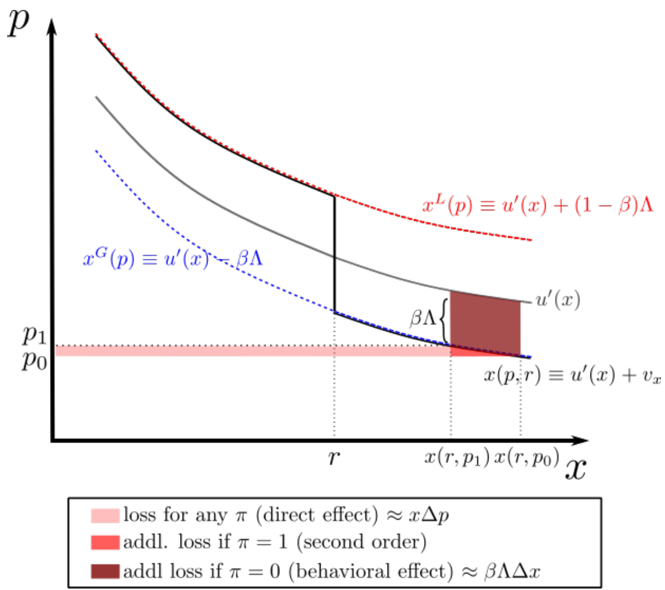
Welfare Effect of Increasing p: Loss Domain

Back



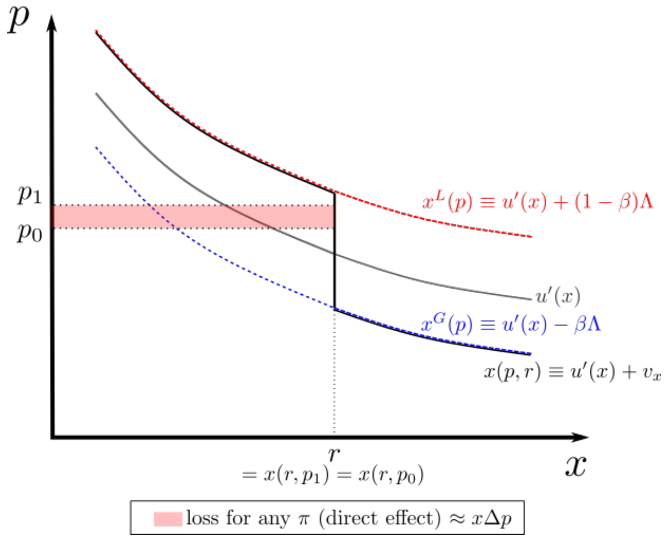
Welfare Effect of Increasing p: Gain Domain

Back



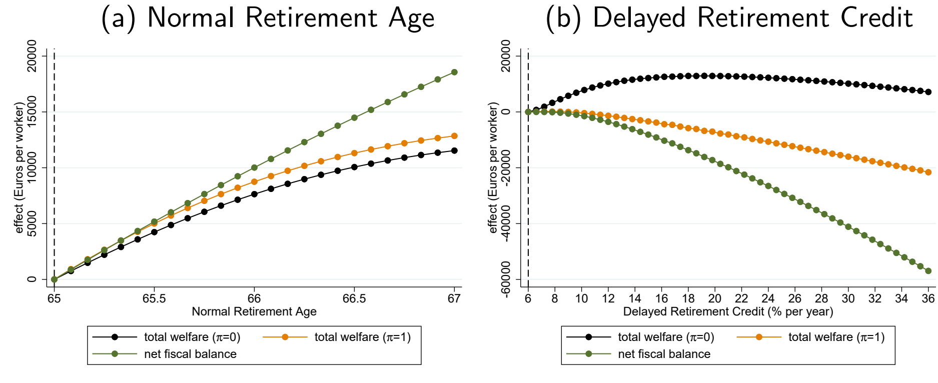
Welfare Effect of Increasing p: Reference Domain

Back



Extended Simulations

Back



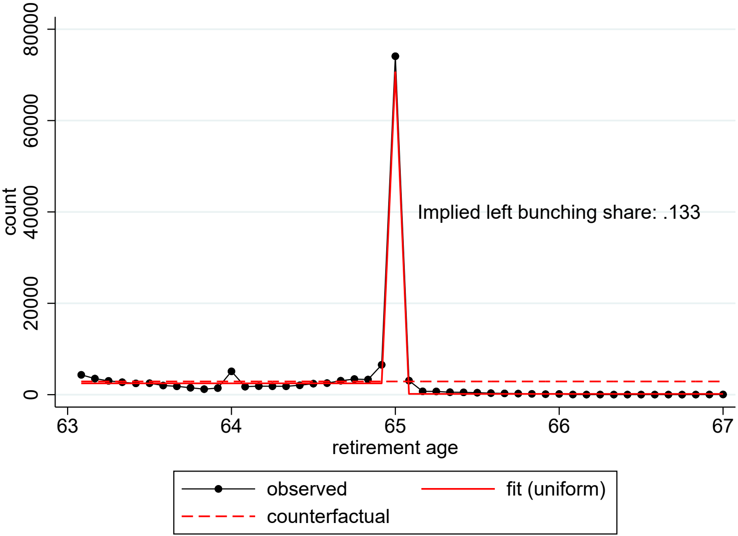
Institutional Linkage Between NRA and Benefits

	Policy 1: Normal Retirement Age to 66	
	Stylized scenario: without benefit cut	Realistic scenario: with benefit cut
Contributions collected	+2,359	+2,359
Benefits paid	+3,999	+7,658
Net fiscal effect	+6,358	+10,017
Worker consumption	+4,230	+571
Disutility from work	–2,950	–2,950
Worker welfare ($\pi = 0$)	+1,280	–2,379
Ref. dep. disutility from work	–6,835	–6,835
Ref. dep. utility from ref. point	+7,946	+7,946
Worker welfare ($\pi = 1$)	+2,391	–1,268
Total welfare ($\pi = 0$)	+7,638	+7,638
Total welfare ($\pi = 1$)	+8,749	+8,749

Back

Two-Dimensional Loss Aversion: Estimating the Left Bunching Share

Back



Further Questions ▸ Back

- For reference dependence in general
  - Reference point formation: when can policy establish and shift ref points
  - Use other tools from behavioral public economics to analyze payoff formulation and/or welfare
    (e.g. Chetty Looney Kroft 2009; Allcott Lockwood Taubinsky 2019; Allcott & Kessler 2019; Goldin & Reck 2020)
  - Welfare economics of reference dependence *under uncertainty*
- For optimal statutory retirement ages
  - Left vs right bunching in other contexts
    - Why do we see so much right bunching for German NRA?
    - Framing of incentives vs location relative to intrinsic optima
  - With multiple potential reference points (e.g. Early & Normal Retirement Age), what do people use?
  - Dynamics/inertia and reforms (e.g. Gelber, Jones, Sacks 2020)