

Kim and Zauberman (2025)

BAFM Discussion

Charlie Rafkin
Stanford → UBC

September 2025

This Presentation

- 1 Quick overview of paper
- 2 High-level: Relationship to behavioral public economics
- 3 Stylized welfare calibration → remaining missing pieces → extensions

Paper in One Slide

Treatments vary piece rate, linear tax rate, and redistribution share (w, τ, T)

- **Five arms:** (5, 0, 0), (3, 0, 0), (5, 2, 0), (5, 2, 0.5), (5, 2, 1)
- **Outcomes:** captchas solved, perceived fairness

Core results contrasting (5, 2, 0), (5, 2, 0.5), (5, 2, 1)

- **Targeted redistribution** vs. **withheld**: little effect on work, fairness
- **Targeted redistribution** vs. **equal redistribution**:
reduces captchas 9% ($p = 0.13$), fairness 0.2 s.d. ($p < 0.01$)
- Also: **withheld** reduces effort and sentiment even when $w - \tau$ the same

Why this paper is important: in classical models, only w and τ enter utility

- Gives new evidence on social/non-welfarist utility from tax-and-transfer policy
[e.g., Cappellen et al., 2007, 2013, 2020; Weinzierl, 2014, 2017; Stantcheva, 2021; Ambuehl et al., 2024]

This Presentation

- ① Quick overview of paper
- ② High-level: Relationship to behavioral public economics
- ③ Stylized welfare calibration → remaining missing pieces → extensions

How This Paper Relates to Public Finance Research

Optimal tax literature writes down expressions like $\mathcal{T}^*(u, e; \theta)$

- For utilities u , behavioral responses e , and policy parameters θ

How This Paper Relates to Public Finance Research

Optimal tax literature writes down expressions like $\mathcal{T}^*(u, e; \theta)$

- For utilities u , behavioral responses e , and policy parameters θ

This paper makes two creative intellectual moves

- Put **taste** or **distaste** for direct taxation into the utility function: $u(\cdot) \rightarrow u(\cdot, \theta)$
- Utility, effort depend on tax rates τ **and** redistribution (to others): $\theta = (\tau, T)$

How This Paper Relates to Public Finance Research

Optimal tax literature writes down expressions like $\mathcal{T}^*(u, e; \theta)$

- For utilities u , behavioral responses e , and policy parameters θ

This paper makes two creative intellectual moves

- Put **taste** or **distaste** for direct taxation into the utility function: $u(\cdot) \rightarrow u(\cdot, \theta)$
- Utility, effort depend on tax rates τ **and** redistribution (to others): $\theta = (\tau, T)$

... and then **measures** them with real choices

What Can Lab Experiments Contribute? Strengths

Objects: $u(\cdot, \tau, T)$, $e(\cdot, \tau, T)$ for tax rates τ and redistribution function T

Outcomes: $du/d\tau$, $de/d\tau$, du/dT , de/dT

What can we learn from observing these parameters in mTurk workers?

- Set-up has two advantages relative to observational settings
- **First:** hard to observe direct effect of government policy on fairness utility $du/d\theta$
 - Not usually revealed by choices, absent structural assumptions
 - ✓ Surveys can help: sentiment outcomes
- **Second:** clean experimentation with controlled (τ, T)
 - Many tax reforms and analyses munge both

What Can Lab Experiments Contribute? Limitations

Naturally, the lab has some limitations

- Relative to observational analysis, **less suitable** for measuring $de/d\tau$
 - The incentive impacts of labor-income taxation is a classic topic in public economics (Theory review: Piketty and Saez, 2013; Empirics review: McClelland and Mok, 2012)
 - Observational settings give plenty of naturalistic variation and real-effort outcomes

What Can Lab Experiments Contribute? Limitations

Naturally, the lab has some limitations

- Relative to observational analysis, **less suitable** for measuring $de/d\tau$
 - The incentive impacts of labor-income taxation is a classic topic in public economics (Theory review: Piketty and Saez, 2013; Empirics review: McClelland and Mok, 2012)
 - Observational settings give plenty of naturalistic variation and real-effort outcomes
- Even for lab outcomes, this is a **stylized “economy”**: hard to extrapolate to real world
 - Effort in short mTurk jobs: less relevant for choices about many occupations
 - Fairness measures are coarse, not incentivized, subject to demand effects
 - Taxes and redistribution made salient → probably amplifies behavioral responses
 - Redistribution function more coarse than multidimensional, nonlinear real-world taxation

What Can Lab Experiments Contribute? Limitations

Naturally, the lab has some limitations

- Relative to observational analysis, **less suitable** for measuring $de/d\tau$
 - The incentive impacts of labor-income taxation is a classic topic in public economics (Theory review: Piketty and Saez, 2013; Empirics review: McClelland and Mok, 2012)
 - Observational settings give plenty of naturalistic variation and real-effort outcomes
- Even for lab outcomes, this is a **stylized “economy”**: hard to extrapolate to real world
 - Effort in short mTurk jobs: less relevant for choices about many occupations
 - Fairness measures are coarse, not incentivized, subject to demand effects
 - Taxes and redistribution made salient → probably amplifies behavioral responses
 - Redistribution function more coarse than multidimensional, nonlinear real-world taxation
- **Bull case**: Shed light on signs and perhaps rough relative magnitudes of these forces

This Presentation

- ① Quick overview of paper
- ② High-level: Relationship to behavioral public economics
- ③ Stylized welfare calibration → remaining missing pieces → extensions

Welfare Impact of Redistribution

Where we are going:

- **Calibration exercise:** proof-of-concept $\rightarrow$ welfare impact of redistribution dW/dT
- **Challenges:** show how improved measurement would connect theory & data

Welfare Impact of Redistribution

Where we are going:

- **Calibration exercise:** proof-of-concept $\rightarrow$ welfare impact of redistribution dW/dT
- **Challenges:** show how improved measurement would connect theory & data

Utility: $u_i(\cdot; \tau, T) = \underbrace{(w - \tau)c_i^*}_{\text{wage utility}} - \underbrace{e_i(c_i^*)}_{\text{effort cost}} + \underbrace{b(T)p_{i \leq T}(c_i^*)}_{\text{redistrib}} + \underbrace{f_i(\tau, T)}_{\text{fairness utility}}$

- T is the share of people who get redistribution $\in \{0, b(T)\}$
- $R(\tau, T)$ is total taxes collected, $b(T) = R(\tau, T)/T$ is benefit given
- Uncertainty about rank i via p can give effort response $\frac{dc_i^*}{dT} > 0$ (theoretically unsigned)

$\hookrightarrow$ No behavioral forces necessary, standard incentive effect can reduce effort

Social welfare: $W = \int \lambda_i u_i(\cdot; \tau, T) di$

Welfare Impact Equation

Welfare impact of changing redistribution threshold T :

$$\frac{dW}{dT} = \underbrace{b(T) (\lambda_T - \bar{\lambda})}_{\text{Targeting}} + \underbrace{R_T \bar{\lambda}}_{\text{Effort}} + \underbrace{\int_i \lambda_i \frac{\partial f_i}{\partial T} di}_{\text{Fairness}}$$

- ① **Targeting effect:** raising T gives to the “marginal” enrollee, who is less needy than inframarginals ($\searrow W$)
- ② **Effort effect:** Raising T affects labor supply (effort) and revenues R_T (here, $\nearrow W$)
- ③ **Fairness effect:** Raising T affects direct “fairness utility” ($\nearrow W$)
- ✓ **Envelope Theorem:** changing T only affects private utility via direct effects and revenues, not via re-optimization

Calibration exercise: toward dW/dT

Consider **welfare impact** of moving from 50% to 51% redistribution ($\approx dW/dT \times 0.01$)

- Tradeoffs: ↗ reported fairness, ↗ work, ↘ targeting

Toy calibration: Linear welfare weight schedule

- Calibrate marginals' welfare weight as 2/3 of avg inframarginals': $\lambda_{0.5} = 1 = 2\bar{\lambda}/3$
- Use moments from the experiment to calibrate effort, fairness response
- Welfare **gain** of new revenue: $R_T \bar{\lambda} dT \approx 0.048c$ (society WTP 0.048 cents per capita)
- Welfare **cost** of reduced targeting is $b(T)(\lambda_T - \bar{\lambda})dT \approx -0.167$ cents
- What about society's willingness to pay for the fairness **gain** f_T ??

Challenge: Willingness to Pay for Fairness

In this model, fairness utility not revealed by choice: missing “utility scale”

- Fairness outcome is an advantage of the lab, but welfare interpretation is not clean

Other approaches:

- Directly elicit WTP or choice over different pay schemes → reveals utility scale
- Elicit WTP for different mental states, beliefs about impact of pay on fairness
(Bernheim, Kim, and Taubinsky, 2025)
- Postulate another model (e.g., let fairness utility depend on c)

Unsatisfying “solution”: e.g., Anders and Rafkin (2024) on welfare stigma

- Calibrate utility scale, benchmarking to more familiar quantity
- If 25% of utility from wage change comes from fairness ($\phi = 0.25$), then $f_T dT \approx 0.166$

These (heroic) assumptions yield a normative payoff

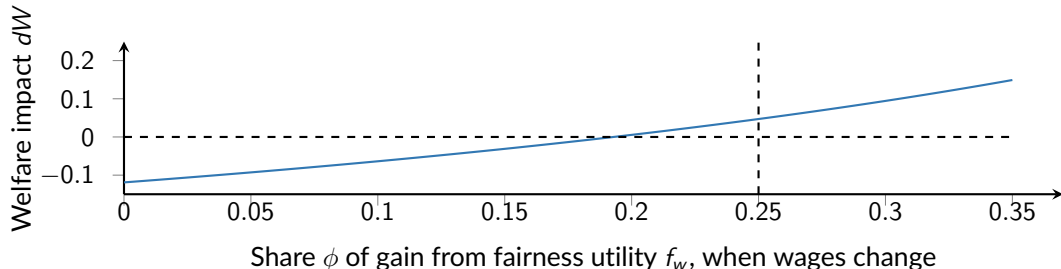
$$\text{Overall welfare impact } dW \approx \underbrace{\text{Targeting}}_{= -0.167} + \underbrace{\text{Effort}}_{= 0.048} + \underbrace{\text{Fairness}}_{= 0.166} > 0$$

- Benefits exceed cost — and, absent fairness, we would have reached wrong conclusion

These (heroic) assumptions yield a normative payoff

$$\text{Overall welfare impact } dW \approx \underbrace{\text{Targeting}}_{=-0.167} + \underbrace{\text{Effort}}_{=0.048} + \underbrace{\text{Fairness}}_{=0.166} > 0$$

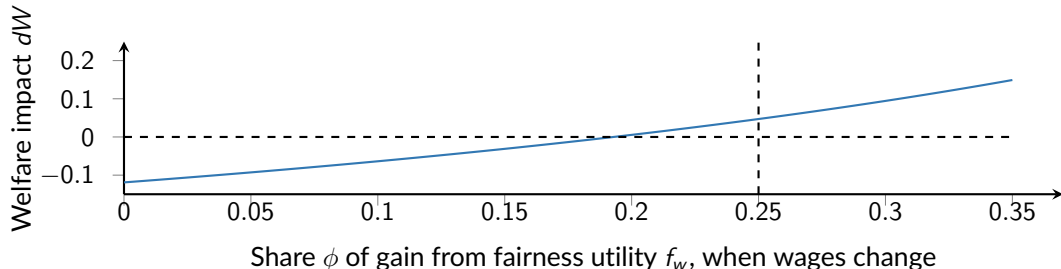
- Benefits exceed cost — and, absent fairness, we would have reached wrong conclusion



These (heroic) assumptions yield a normative payoff

$$\text{Overall welfare impact } dW \approx \underbrace{\text{Targeting}}_{=-0.167} + \underbrace{\text{Effort}}_{=0.048} + \underbrace{\text{Fairness}}_{=0.166} > 0$$

- Benefits exceed cost — and, absent fairness, we would have reached wrong conclusion



- $dW/dT > 0$, but fairness utility ϕ must be meaningful (possible it's much smaller)
- Fairness utility has large magnitudes because affects *all* in society

My normative exercise was unsatisfying: suggests new work

Measure missing policy-relevant parameters: especially WTP for change in fairness

My normative exercise was unsatisfying: suggests new work

Measure missing policy-relevant parameters: especially WTP for change in fairness

Related Q1: *Why* does effort fall with targeting?

- One interpretation: classical incentive effects (raising Q's about core mechanism)
- If effort costs depend on redistribution $T \rightarrow$ more terms in formulas

My normative exercise was unsatisfying: suggests new work

Measure missing policy-relevant parameters: especially WTP for change in fairness

Related Q1: *Why does effort fall with targeting?*

- One interpretation: classical incentive effects (raising Q's about core mechanism)
- If effort costs depend on redistribution $T \rightarrow$ more terms in formulas

Related Q2: How does redistribution affect welfare impact of linear tax reform $dW/d\tau$?

- Cross-derivative $\frac{d^2 f}{d\tau dT}$ not identified from experiment arms
- Requires jointly manipulating τ and $T \rightarrow$ one more arm, easy extension!

Conclusion

Zooming out: thought-provoking work

- Sheds light on **targeting**, **equity**, and **efficiency**
- Taking seriously the psychic costs and benefits of taxes/transfers: advantage of the lab setting, would have **major implications** for public economics
- **But huge measurement challenges** → more research
- Excited about the broader agenda!