Mechansim Design for Social Good

Provision and Targeting for Vulnerable Populations

EC 2020 Tutorial, June 25 and 26 Part 2B

Sera Linardi

University of Pittsburgh

Sam Taggart

Oberlin College

Part IIb: Theoretical issues in information acquisition.

Goal of this session

So far.

- **Day 1:** Targeting toolbox.
- **Previous session:** Behavioral considerations.

This session: Strategic and computational issues in PMT and CBT.

- **Proxy means testing:** Lessons from strategic classfication.
- **Community-based targeting:** Learning from local data.

Case Study: SSDI

Income support, targeted at people with disabilities.

Application Process:

- Interview with evaluator, extensive paperwork.
- 5-month waiting period w/ no gainful employment.
- Screening based on medical history.

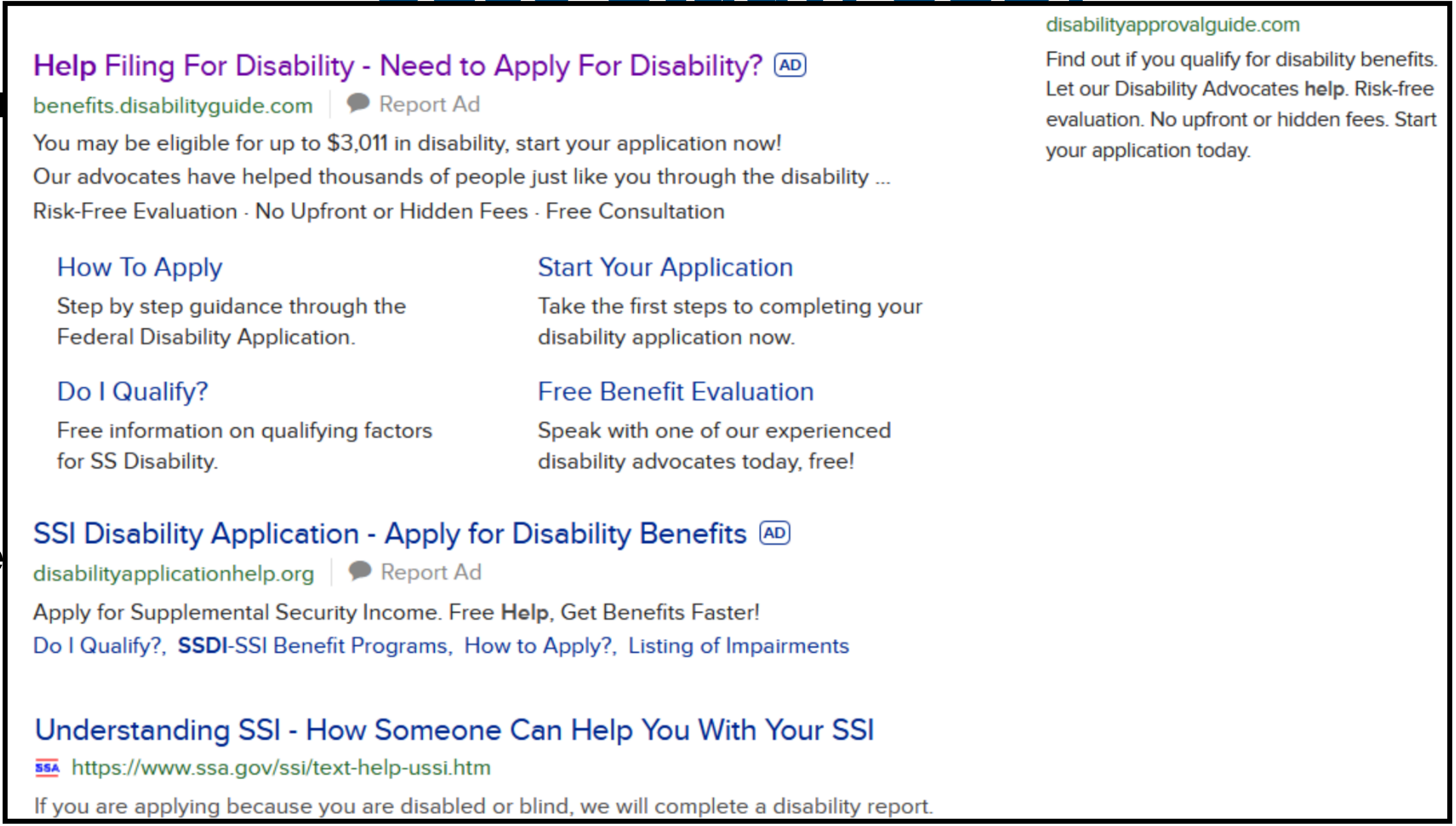
Observations: applicants manipulate

- labor supply [Maestas et al., *AER* 2000]
- application quality

Case Study: SSDI

Desc

Obse



Eligibility Manipulation

Labor Distortion:

- US Social Security [Friedberg, *R. Econ. and Stat.* 2000]
- UK Working Families Tax Credit [Blundell and Hoynes 2004]

PMT Standard Practice: Choose features that are harder to manipulate.

Challenge: How to design your targeting if you expect manipulation.

Tradeoffs.

- explanatory power
- manipulation cost

Strategic Classification

[Hardt et al., ITCS 2016]

Idea: Treat targeting as a learning problem.

- training is from honest data
- testing is on manipulated data

Data points = Individuals in population.

Strategic Classification

[Hardt et al., ITCS 2016]

Idea: Treat targeting as a learning problem.

Learning environment:

Each individual has:

- features = points in R^n
- eligibility in {0,1} ("low income")

Underlying joint distribution D

Strategic Classification

[Hardt et al., ITCS 2016]

Idea: Treat targeting as a learning problem.

Learning environment:

Each individual has:

- features = points in R^n
- eligibility in {0,1}

Underlying joint distribution D

Training stage:

- learner receives m samples (x_i,y_i) (initial survey)
- learner selects linear classifier h

Strategic Classification

[Hardt et al., ITCS 2016]

Idea: Treat targeting as a learning problem.

Learning environment:	**Training stage:**	**Test stage:**
Each individual has: - features = points in R^n - eligibility in {0,1} Underlying joint distribution D	- learner receives m samples (x_i,y_i) - learner selects linear classifier h	- learner draws fresh data point (x,y) - goal: maximize $Pr[h(x)=y]$

Strategic Classification

[Hardt et al., ITCS 2016]

Idea: Treat targeting as a learning problem.

Learning environment:

Each individual has:

- features = points in R^n
- eligibility in {0,1}

Underlying joint distribution D

Training stage:

- learner receives m samples (x_i, y_i)
- learner selects linear classifier h

Test stage:

- learner draws fresh data point (x,y)
- x moves to new set of features z(x)
- learner outputs h(z(x))

Strategic Classification

[Hardt et al., ITCS 2016]

Idea: Treat targeting as a learning problem.

Learning environment:

Each individual has:

- features = points in R^n
- eligibility in {0,1}

Underlying joint distribution D

Training stage:

- learner receives m samples (x_i,y_i)
- learner selects linear classifier h

Test stage:

- learner draws fresh data point (x,y)
- x moves to new set of features z(x)
- learner outputs h(z(x))

benefits

manipulation cost

Objectives

- objective of x: maximize $u(x) = \underbrace{\mathbb{I}(h(z(x))=1)}_{\text{benefits}}-\underbrace{c(z(x),x)}_{\text{manipulation cost}}$ (knows h)
- objective of learner: maximize $\Pr_{x\sim D}[h(z(x))=y]$ (knows c but not D)

Solution: “Move the Goalposts”

[Hardt et al., ITCS 2016]

Def. c is linearly separable if it is of the form $c(x,y) = \max(0,\langle\alpha,y-x\rangle)$ for some α.

Ex. α_1 = cost to “borrow kids,” α_2 = worsen home exterior

Theorem (informal). For separable cost functions and linear hypotheses, a near-optimal hypothesis can be learned efficiently in the strategic environment.

benchmark manipulated, knows D

Solution: “Move the Goalposts”

[Hardt et al., ITCS 2016]

Def. c is linearly separable if it is of the form $c(x,y) = \max(0, \langle \alpha, y-x \rangle)$ for some α.

Ex. α_1 = cost to “borrow kids,” α_2 = worsen home exterior

Theorem (informal). For separable cost functions and linear hypotheses, a near-optimal hypothesis can be learned efficiently in the strategic environment.

Algorithm (informal).

- Select hypothesis $\langle \alpha, y \rangle \geq t$ that does best on training data.
- “Move the goalpost”: $\langle \alpha, y \rangle \geq t^* + 1$

Solution: "Move the Goalposts"

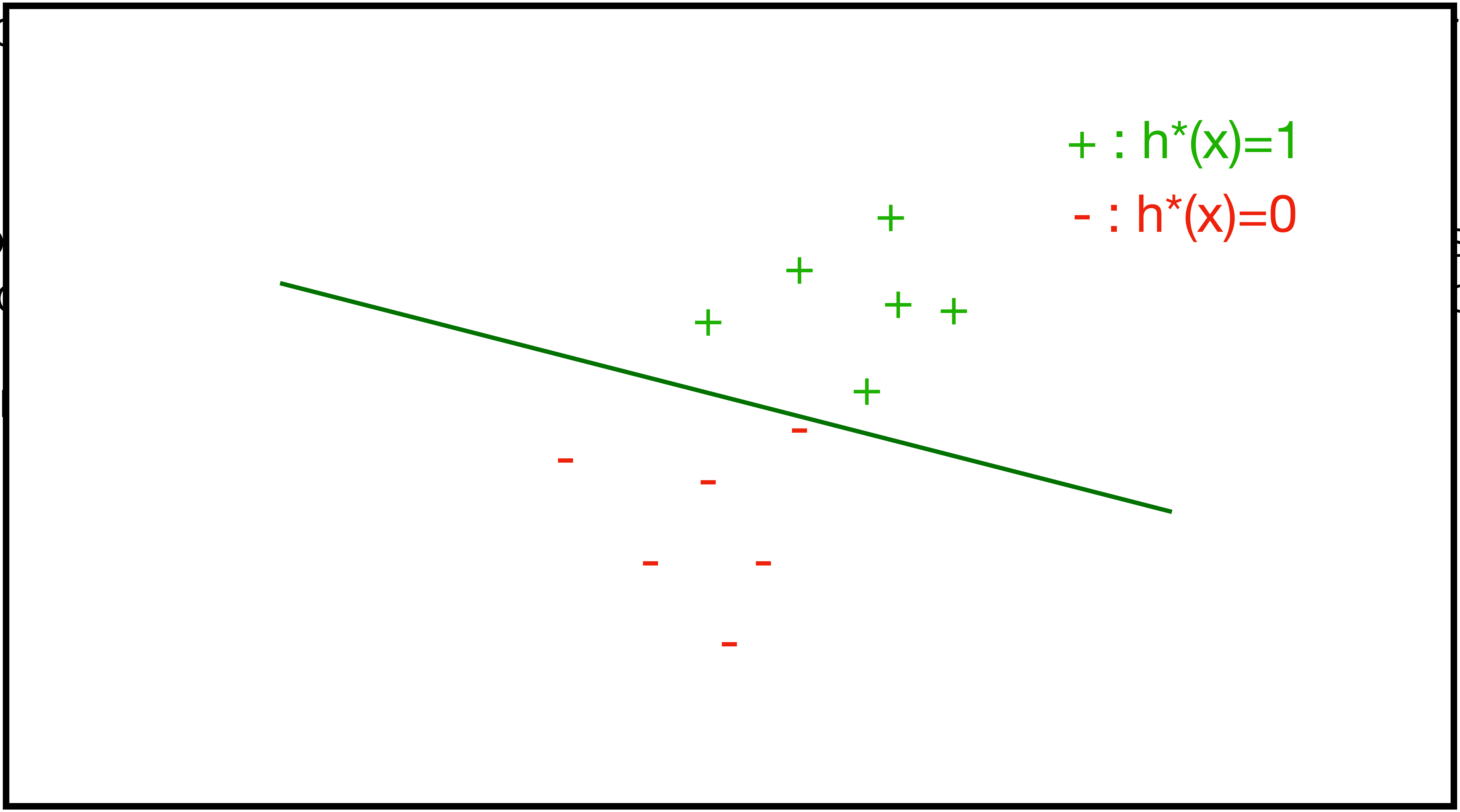
[Hardt et al., ITCS 2016]

Def. α some α.

Ex. α

Theo ses, a near- onment.

Algo



Solution: “Move the Goalposts”

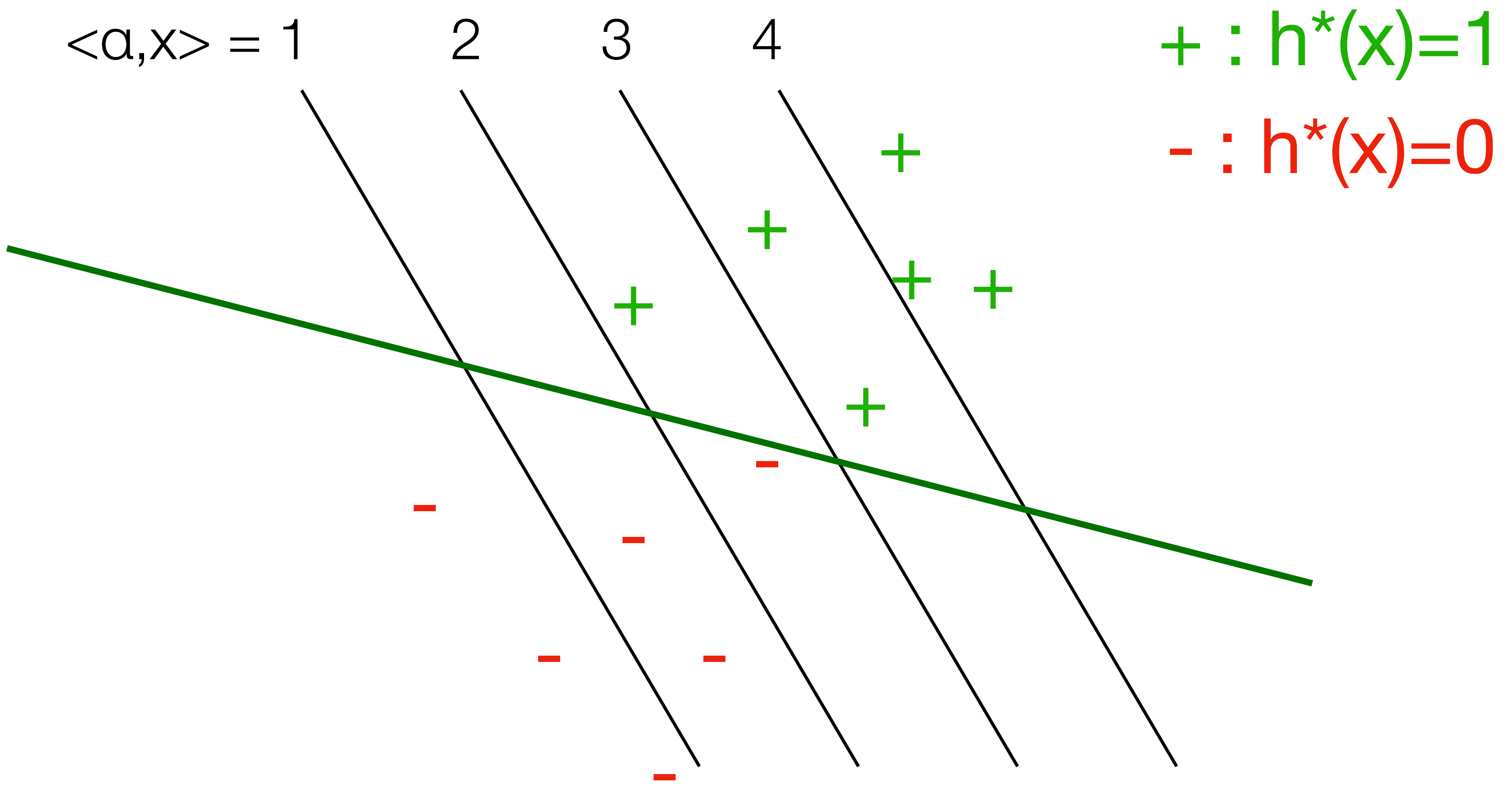
[Hardt et al., ITCS 2016]

Def. α … some α.

Ex. α

Theo… ses, a
near-… onment.

Algo



Note: Without loss
to choose $\langle \alpha, x\rangle \geq t$

Solution: “Move the Goalposts”

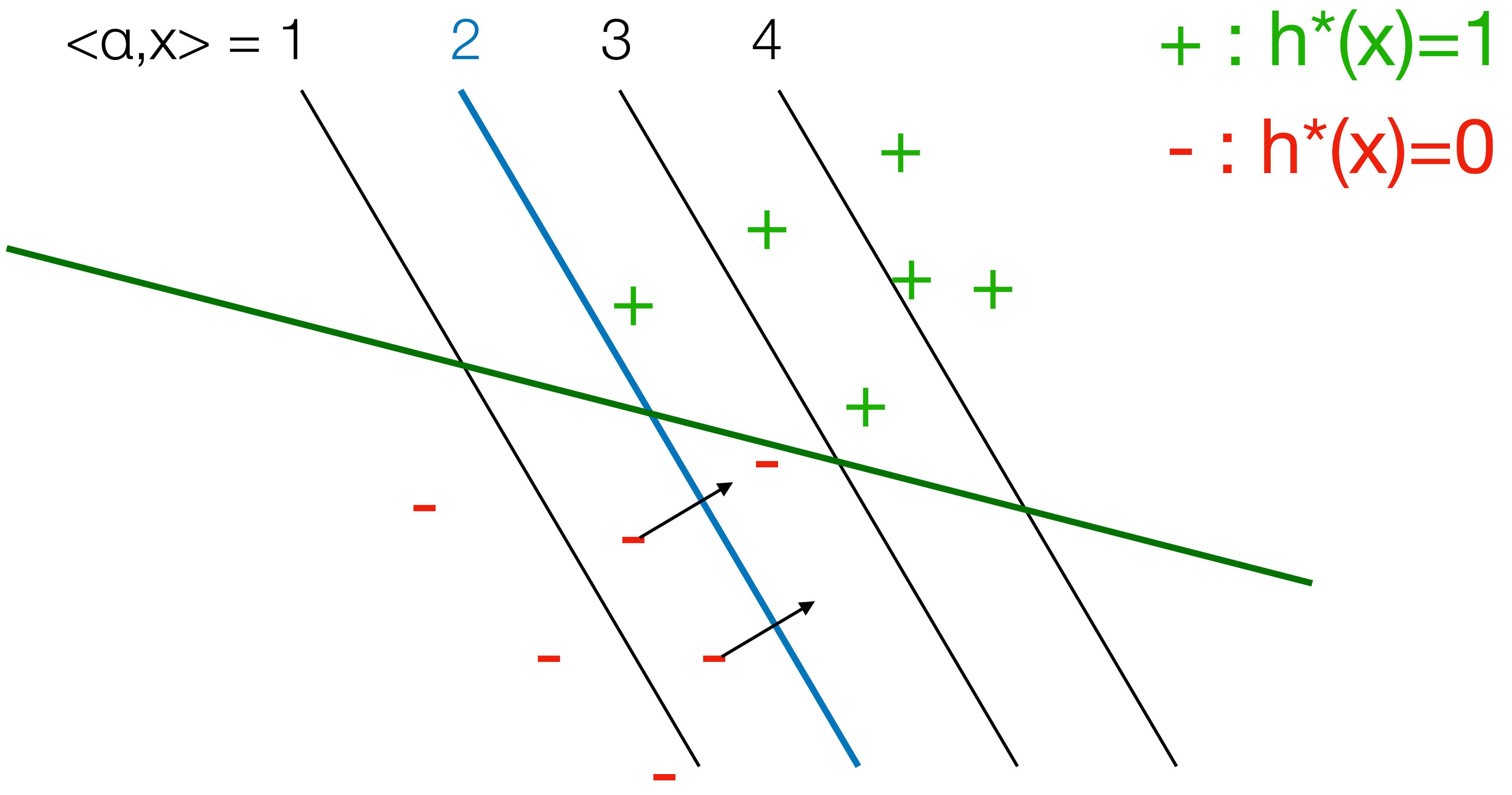
[Hardt et al., ITCS 2016]

Def. α … some α.

Ex. α

Theo … ses, a
near- … onment.

Algo



Note: Without loss
to choose <α,x> ≥ t

Solution: “Move the Goalposts”

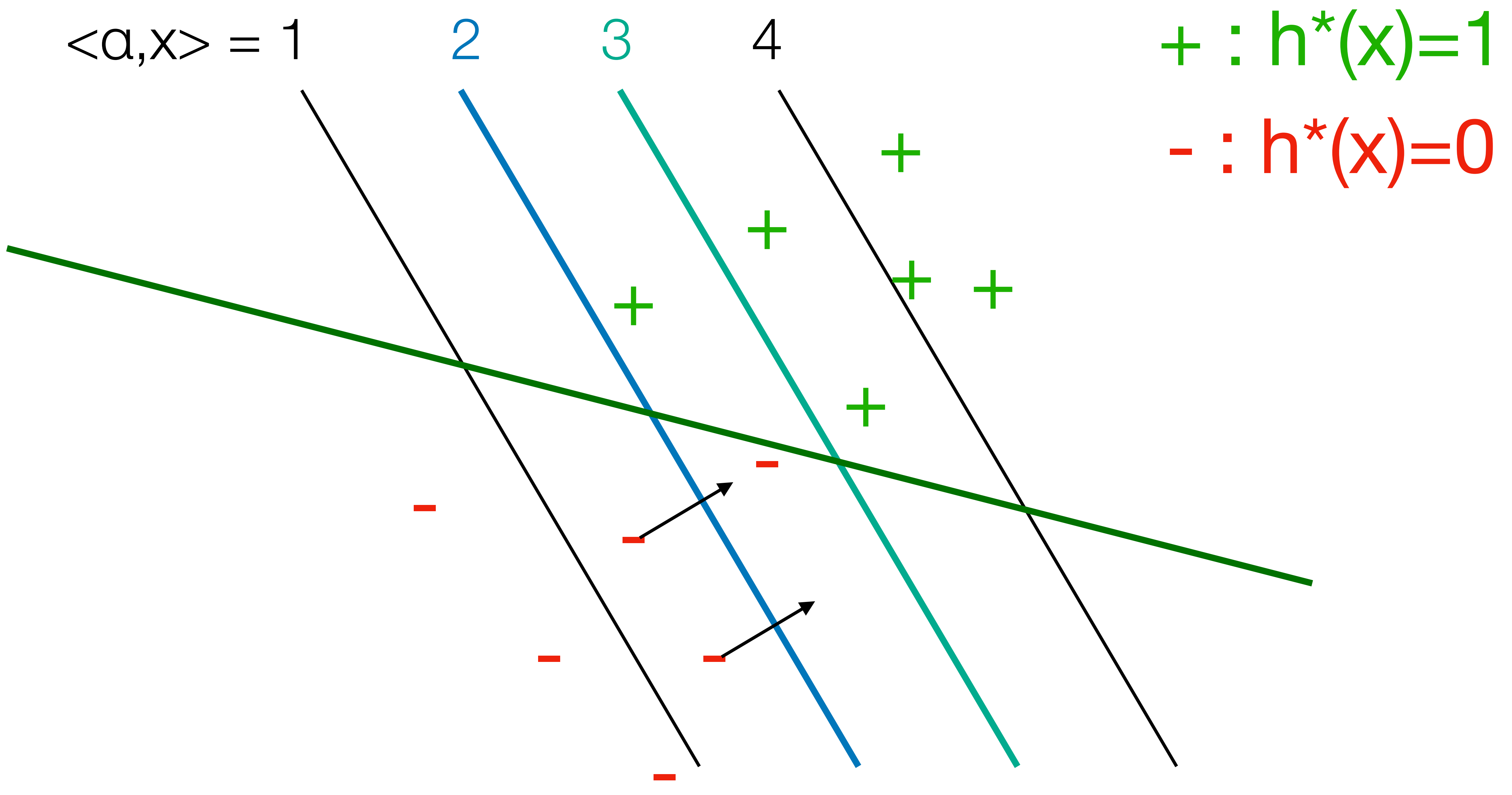
[Hardt et al., ITCS 2016]

Def. α … some α.

Ex. α

Theo … ses, a
near- … onment.

Algo



Note: Without loss
to choose <α,x> ≥ t

Solution: “Move the Goalposts”

[Hardt et al., ITCS 2016]

Def. c is linearly separable if it is of the form $c(x,y) = \max(0,\langle\alpha,y-x\rangle)$ for some α.

Ex. α_1 = cost to “borrow kids,” α_2 = worsen home exterior

Theorem (informal). For separable cost functions and linear hypotheses, a near-optimal hypothesis can be learned efficiently in the strategic environment.

Algorithm (informal).

- Select hypothesis $\langle\alpha,y\rangle \geq t$ that does best on training data.
- “Move the goalposts”: $\langle\alpha,y\rangle \geq t+1$

Different papers, similar conclusions:

[Brückner and Scheffer, KDD 2011]

[Dalvi et al., KDD 2004]

Inequality

[Milli et al., FAT* 2019]

Q: Does strategic classification treat vulnerable populations fairly?

Two groups: A ("majority") and B ("vulnerable")

Welfare disparity: E[u(x) | + , A] - E[u(x) | + , B]

Inequality definitions:

Inequality in costs

$c_A(x,y) = \max(0, \langle \alpha, y-x \rangle)$

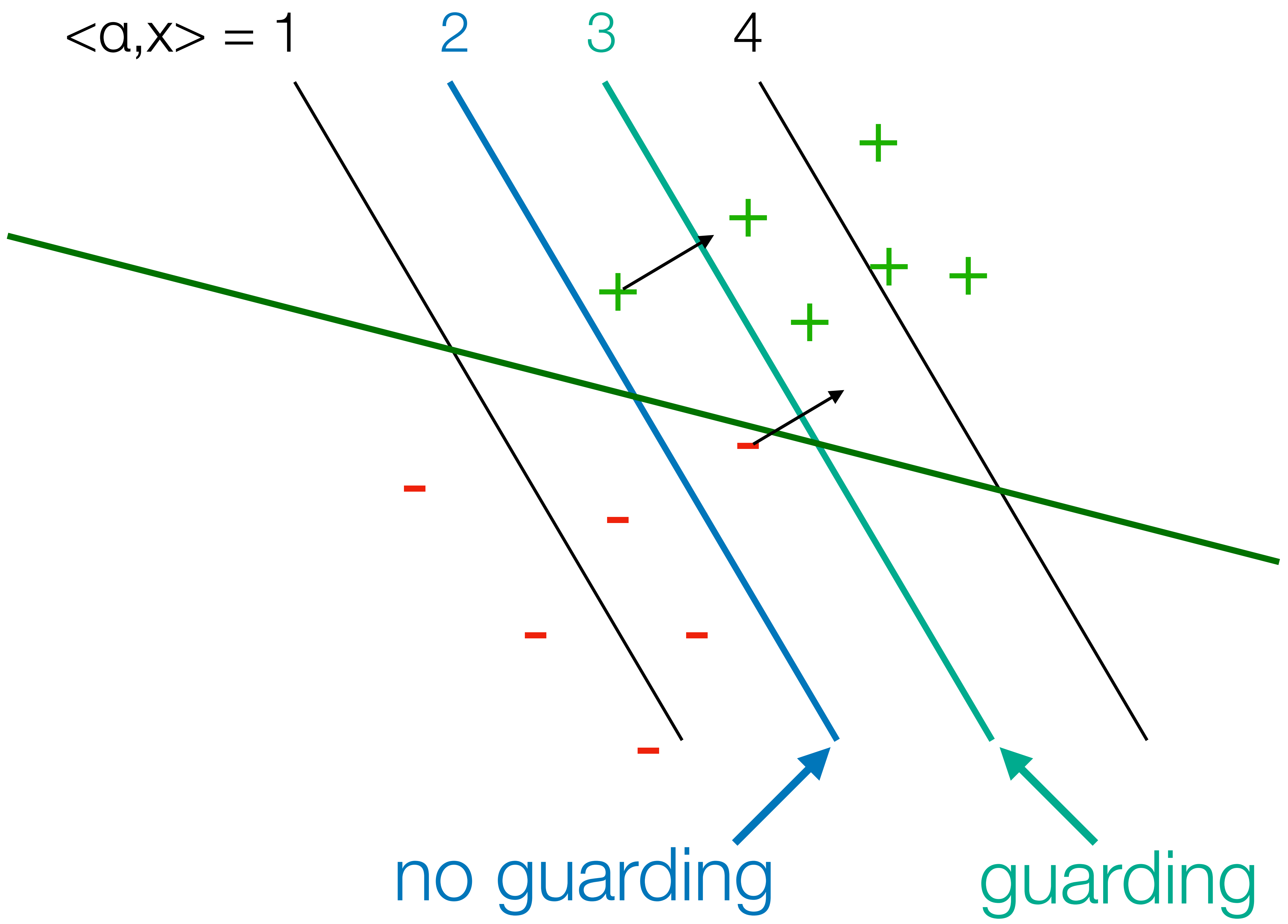
$c_B(x,y) = \max(0, \langle \rho\alpha, y-x \rangle)$ $\rho > 1$

Inequality in features: given "likelihood" $L(x) = \Pr[\, + \mid x \,]$

$\Pr[\, L(x) \leq q \mid + , A \,] \leq \Pr[\, L(x) \leq q \mid + , B \,]$ for all q

Inequality

Theorem: Between \ and \, under either notion of inequality (plus a regularity condition), welfare disparity $E[\, u(x) \mid +, A\,] - E[\, u(x) \mid +, B\,]$ increases.



Inequality definitions:

Inequality in costs

$c_A(x,y) = \max(0, \langle \alpha, y-x \rangle)$

$c_B(x,y) = \max(0, \langle \rho\alpha, y-x \rangle) \quad \rho > 1$

Inequality in features: given "likelihood" $L(x) = \Pr[\, + \mid x\,]$

$\Pr[\, L(x) \le q \mid +, A\,] \le \Pr[\, L(x) \le q \mid +, B\,] \qquad \text{for all } q$

Interventions

[Hu et al., FAT* 2019]

Inequality in costs

$c_A(x,y) = \max(0, \langle \alpha, y-x \rangle)$

$c_B(x,y) = \max(0, \langle \rho\alpha, y-x \rangle)$ $\quad \rho > 1$

Theorem: There exists instances where the the learner improves their objective with subsidies, but both populations' utilities degrade.

Intervention: Subsidies

Subsidized costs for B:

$c_B(x,y) = \max(0, \langle \beta\rho\alpha, y-x \rangle)$ $\quad \rho > 1, \beta < 1$

New objective for learner:

$\Pr_{x\sim D}[h(z(x))=y] - \beta \boxed{\mathrm{cost}_B} \Pr[B]$

expected manipulation cost from B

Other Directions

Interventions: Beyond subsidies?

Targeting for interventions:

- Current approach: categorical.
- Are there better ways to target subsidies within B?

This model. Manipulation

- makes targeting harder
- otherwise irrelevant to learner

Payoff-relevant manipulations:

Manipulation gains in learner utility.

- [Kleinberg and Raghavan, EC 2019]
- [Haghtalab et al, IJCAI 2020]

Learning from Community Data

[Alatas et al., *AER* 2012]

Goal: Compare community-based targeting to a PMT.

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

What follows: Three observations from their data.

Learning from Community Data

[Alatas et al., *AER* 2012]

Goal: Compare community-based targeting to a PMT.

Data:

- Baseline: surveyed community members
    - consumption
    - social habits
    - impressions of others' wealth
- Community meeting: ranked village members by wealth
- PMT data

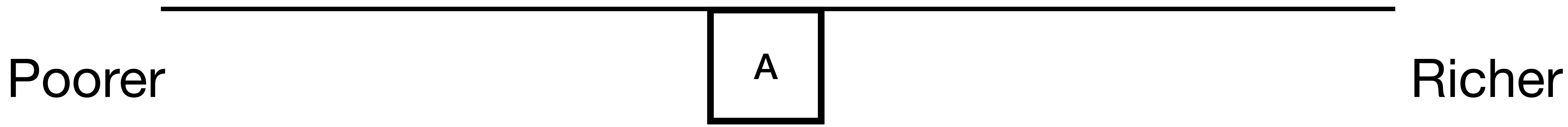
What follows: Three observations from their data.

Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

Ranking protocol:

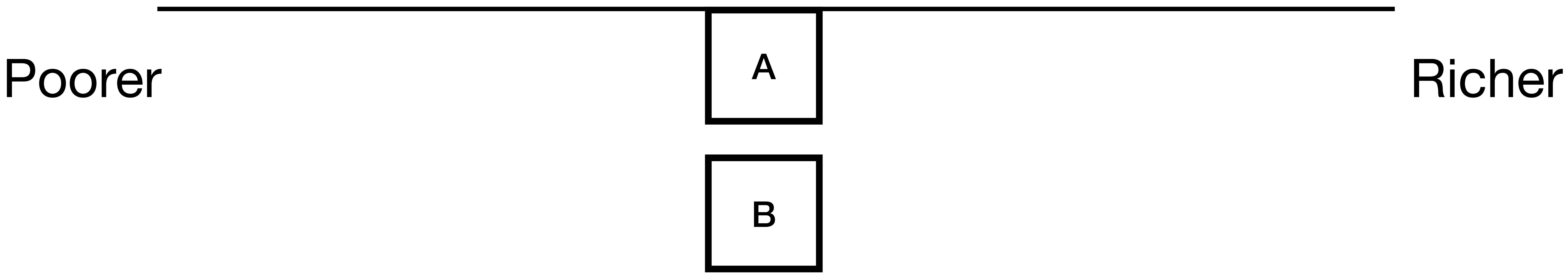


Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

Ranking protocol:

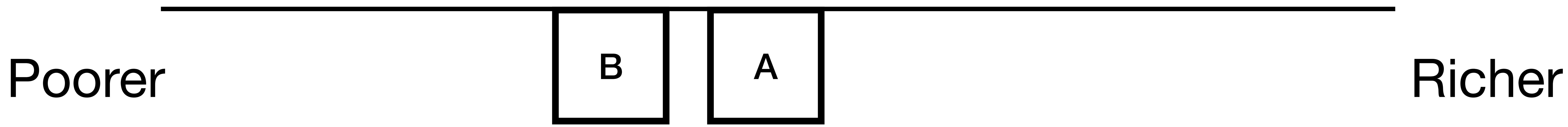


Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

Ranking protocol:

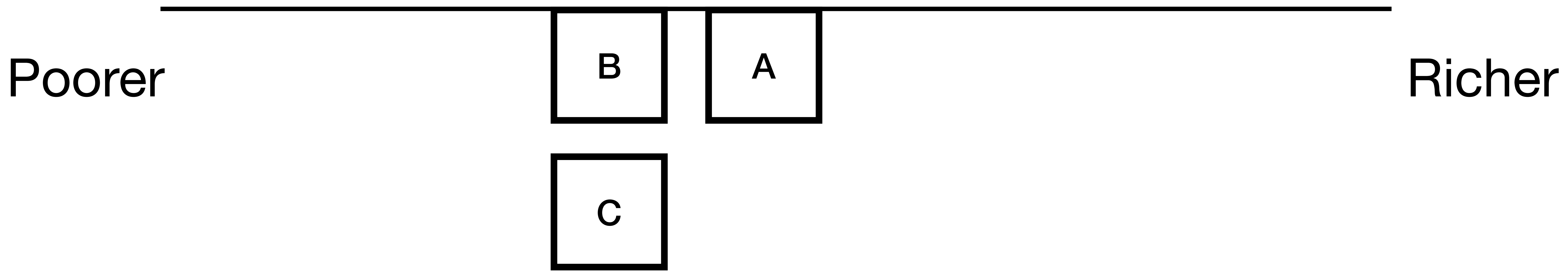


Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

Ranking protocol:

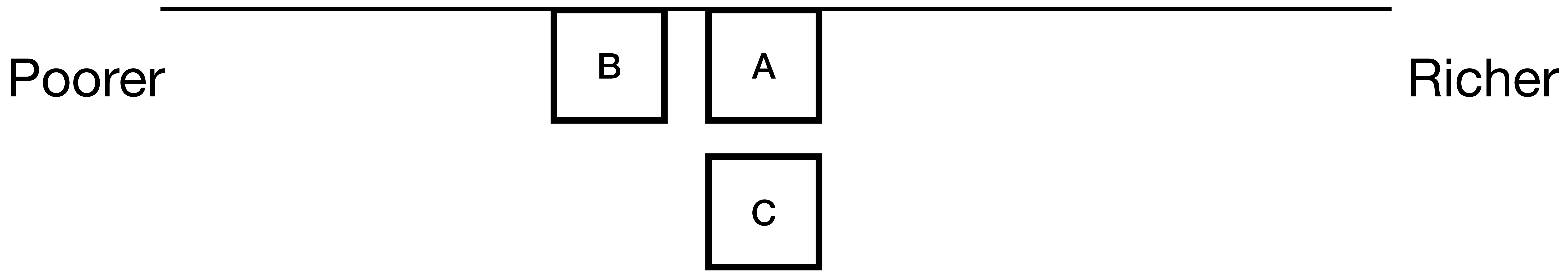


Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

Ranking protocol:

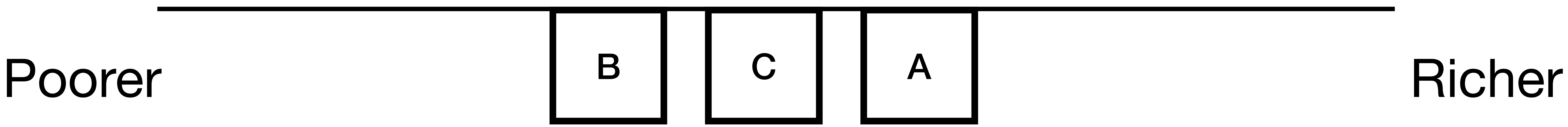


Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

Ranking protocol:



Learning from Community Data

Participatory Wealth Ranking:

- open-invitation community meeting
- group agrees on poverty definition
- group ranks members in community by wealth
- benefits given to bottom k

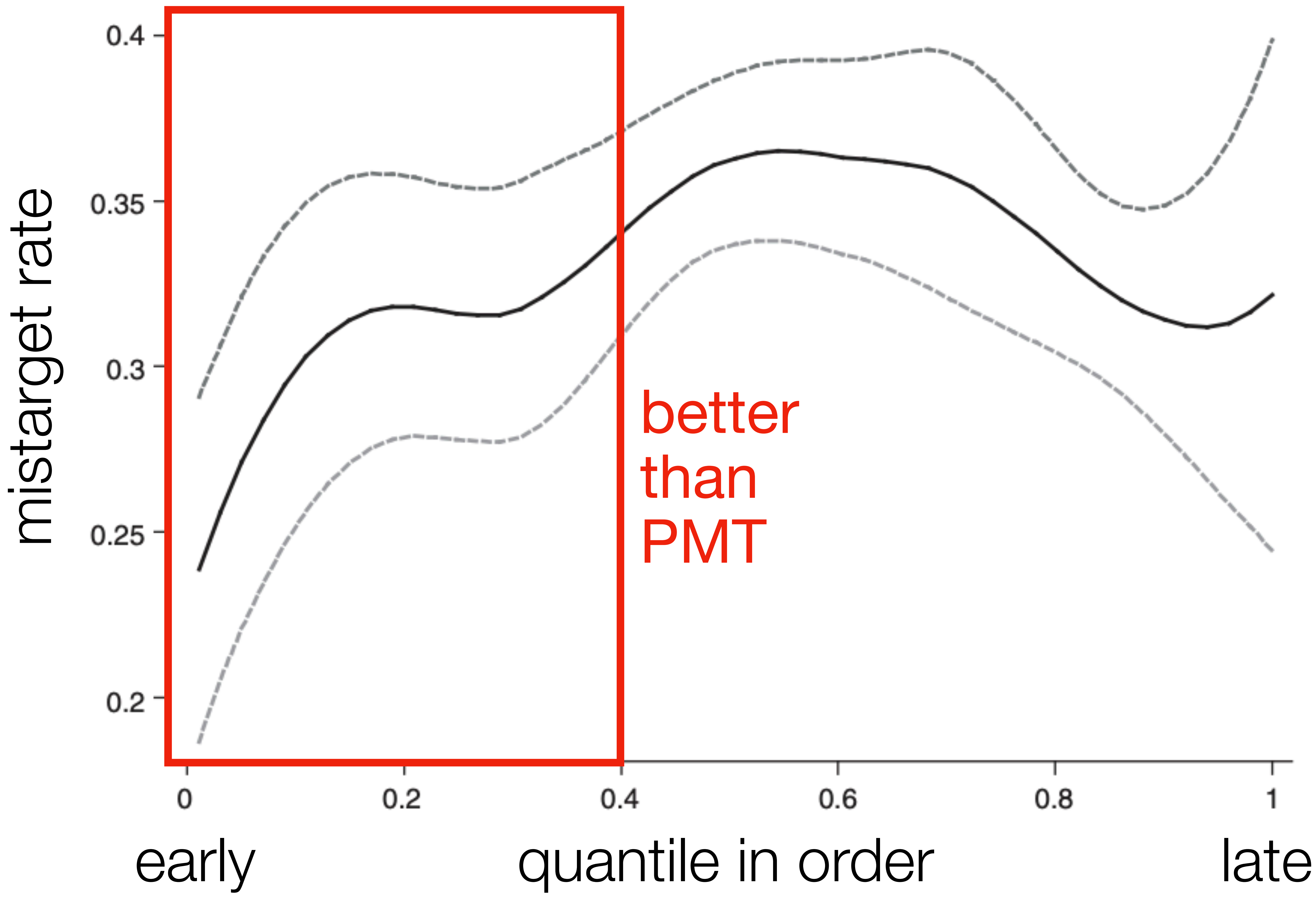
Ranking protocol:

- sequential search w/ short list
- binary search w/ long list

Thorough, but time-consuming.

Degrading Accuracy

Question: How does targeting accuracy change during the meeting?



Observation: Protocol matters.

What is “Poor?”

Question: Did community incorporate information differently than PMT?

	Rank according to welfare metric			Targeting rank list in		
	Community survey ranks (r_c) (1)	Subvillage head survey ranks(r_e) (2)	Self-assessment (r_s) (3)	PMT villages (4)	Community villages (5)	Hybrid villages (6)
Log per capita consumption	0.176*** (0.008)	0.145*** (0.008)	0.087*** (0.004)	0.132*** (0.013)	0.197*** (0.014)	0.162*** (0.014)
Panel A. Household demographics						
Log HH size	0.164*** (0.011)	0.134*** (0.010)	0.073*** (0.006)	−0.028 (0.019)	0.154*** (0.019)	0.078*** (0.021)
Share kids	−0.125*** (0.021)	−0.094*** (0.021)	−0.037*** (0.012)	−0.296*** (0.035)	−0.068* (0.041)	−0.141*** (0.039)
Panel B. Ability to smooth shocks						
Elite connected	0.092*** (0.008)	0.044*** (0.009)	0.025*** (0.005)	0.062*** (0.016)	0.051*** (0.015)	0.043*** (0.015)
Total connectedness	−0.039*** (0.010)	−0.021** (0.009)	−0.015*** (0.005)	−0.016 (0.017)	−0.019 (0.017)	−0.054*** (0.019)
Number of family members outside subvillage	0.012*** (0.004)	0.010*** (0.003)	0.006*** (0.002)	0.020*** (0.006)	0.001 (0.006)	0.001 (0.006)
Participation through work to community projects	0.002 (0.011)	0.021** (0.010)	0.005 (0.006)	0.000 (0.018)	0.010 (0.019)	0.003 (0.019)
Participation through money to community projects	0.061*** (0.009)	0.041*** (0.009)	0.024*** (0.005)	0.056*** (0.016)	0.058*** (0.016)	0.034* (0.018)
Participation in religious groups	0.027*** (0.010)	0.033*** (0.010)	0.014** (0.006)	0.033** (0.016)	0.012 (0.017)	0.029 (0.017)

What is "Poor?"

Question: Did community incorporate information differently than PMT?

Panel C. Discrimination against minorities?						
Ethnic minority	−0.024*	−0.019	−0.003	0.012	−0.051**	−0.011
	(0.014)	(0.014)	(0.008)	(0.026)	(0.025)	(0.024)
Religious minority	0.012	−0.007	−0.014*	−0.018	0.025	0.012
	(0.018)	(0.017)	(0.008)	(0.030)	(0.032)	(0.033)
Panel D. Correcting for earnings ability						
HH head with primary education or less	−0.028***	−0.025***	−0.037***	−0.108***	−0.011	−0.066***
	(0.009)	(0.009)	(0.005)	(0.017)	(0.018)	(0.017)
Widow	−0.104***	−0.083***	−0.012	0.009	−0.108***	−0.026
	(0.014)	(0.014)	(0.008)	(0.027)	(0.024)	(0.028)
Disability	−0.045***	−0.037***	−0.026***	−0.079***	0.009	0.012
	(0.016)	(0.014)	(0.008)	(0.027)	(0.026)	(0.027)
Death	−0.041*	−0.031	−0.010	−0.111***	−0.013	−0.059
	(0.025)	(0.025)	(0.015)	(0.042)	(0.048)	(0.043)
Sick	−0.038***	−0.041***	−0.028***	0.007	−0.018	−0.044**
	(0.011)	(0.011)	(0.006)	(0.018)	(0.019)	(0.019)
Recent shock to income	−0.001	−0.005	−0.013**	−0.019	0.009	−0.012
	(0.009)	(0.009)	(0.005)	(0.016)	(0.016)	(0.017)
Tobacco and alcohol consumption	−0.0002***	−0.0002***	−0.0001***	−0.0002***	−0.0002***	−0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	5,337	4,680	5,724	1,814	1,876	1,889

Observation: Community maximized a different welfare function.

Who does the community learn from?

[Alatas et al., *AER* 2016]

Five observations about wealth impressions:

1. social proximity ⟶ more accurate
2. socially central ⟶ more accurate
3. individuals sometimes said “don’t know”
4. those who “did know” were sometimes wrong
5. less proximate ⟶ less certain

Reasonable conclusions:

- information is passed along social network
- transmission is noisy

Who does the community learn from?

[Alatas et al., *AER* 2016]

Question: Can network structure predict targeting accuracy?

Complex Approach:

- Estimate a structural model of learning on networks.
- Test if simulated diffusion predicts targeting accuracy.

Simple Approach:

- Identify coarse-grained properties of networks
  (avg. degree, clustering coefficient, …)
- Regress targeting accuracy on these properties.

Observation: Network structure matters a lot.

Open Problems for CBT

Protocol design: Can we better trade off thoroughness against fatigue?

Targeting for the community: How can we better learn and target to maximize a community's welfare function?

Predicting diffusion: Given a network structure, can we predict if CBT will work?

Predicting diffusion, simply: Are there easy-to-measure network properties that are predictive of CBT's success?

Acknowledgements

EC Tutorial Chairs: Sigal Oren, Brendan Lucier

MD4SG Leadership: Rediet Abebe, Irene Lo, Ana-Andreea Stoica

MD4SG Inequality Group: Especially Zoë Hitzig, Angela Zhou

Q+A