Motivations, Challenges, and Future Potential of Responsible Fair Clustering Work

Brian Brubach

Wellesley College

What we've seen so far…

- Several classic clustering problems
- A current taxonomy of fairness definitions in clustering
- Algorithmic approaches to solving fair clustering problems

Where we're heading…

- **Outcomes** → Summarize risks and responsibilities that motivate careful and thoughtful approaches to fair clustering
- **Methods** → Examine the challenges and pitfalls we can encounter doing meaningful and impactful fairness work
- **Next steps** → Explore the vast frontier of future work that arises from interdisciplinary engagement and specific applications

- **Goal** → Inspire and prepare you to dive into the next iteration of fair clustering work

Running example topics we'll use

- Criminal justice applications and algorithmic risk assessment
  - Clear interaction with vulnerable and marginalized populations
  - Popular public datasets used in fairness research (e.g., COMPAS)
  - Overlooked work in other disciplines (criminology, psychology, sociology, etc.)
- Educational and political districting
  - Applications with clustering as a hard decision point
  - Less commonly associated with clustering
  - Engaging problems with unique needs
- Supervised learning
  - Longer history of algorithmic fairness than clustering
  - Inspired a lot of fair clustering ideas

Where we're heading…

- **Outcomes** → Summarize risks and responsibilities that motivate careful and thoughtful approaches to fair clustering
- **Methods** → Examine the challenges and pitfalls we can encounter doing meaningful and impactful fairness work
- **Next steps** → Explore the vast frontier of future work that arises from interdisciplinary engagement and specific applications

Ethical questions for working on sensitive subjects

- By design, algorithmic fairness interacts with vulnerable and marginalized communities
- How can we ensure that we serve and give back to these communities?
  - Avoid exploiting for research topics and publications
- How can we avoid harming these communities?

- AAAI code of ethics:
  - https://www.aaai.org/Conferences/code-of-ethics-and-conduct.php
  - “1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.”
  - “1.2 Avoid harm.”

Harmful fairness example: “ban the box”

- Problem → Employers discriminate against applicants’ criminal history
  - Racial disparities in criminal justice propagated into employment
- Fair solution → Ban employers from asking about criminal history
- New problem → Evidence that employers start using race as a proxy for the unknown variable (criminal history), increasing racial discrimination
  - Agan and Starr (2018) “Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment”
  - Kleinberg and Mullainathan (EC 2019) “Simplicity creates inequity: implications for fairness, stereotypes, and interpretability”

More harmful fairness examples

- Incorporating temporal analysis shows that demographic parity and equal opportunity in fair
  - Liu, Dean, Rolf, Simchowitz, and Hardt (ICML 2018) “Delayed Impact of Fair Machine Learning”
- *“Imposing a fairness constraint can make the disadvantaged group worse off if the fairness constraint and the utilities of the population mismatch.”*
  - Ben-Porat, Sandomirskiy, and Tennenholtz (AAAI 2021) “Protecting the Protected Group: Circumventing Harmful Fairness”

Contrast with work that had a clear positive impact

- Identified discrimination in a healthcare algorithm due to biased proxy variable
  - Obermeyer, Powers, Vogeli, and Sendhil Mullainathan (Science 2019) “Dissecting racial bias in an algorithm used to manage the health of populations”
  - Authors worked with manufacturer improve variable choice and reduce bias

Where we're heading…

- **Outcomes** → Summarize risks and responsibilities that motivate careful and thoughtful approaches to fair clustering
- **Methods** → Examine the challenges and pitfalls we can encounter doing meaningful and impactful fairness work
- **Next steps** → Explore the vast frontier of future work that arises from interdisciplinary engagement and specific applications

Considerations for modeling and problem formulation

- Specificity → Applications and context matter
  - Beneficial to define and model fairness for a specific social problem
  - Application details can suggest fascinating new challenges to solve
  - General purpose abstractions can be useful, but are often over-sold
  - Use caution mapping ideas from fair classification to fair clustering

Considerations for modeling and problem formulation

- Non-modularity → Fairness interventions do not act in a vacuum
  - Broader context and upstream/downstream effects are important
  - Can't simply swap an "unfair" algorithm for a "fair" one
  - Different bad inputs require different fair algorithms
  - How the algorithm's output is used must also be considered

- Example considering downstream effects
  - Kannan, Roth, and Ziani (FAccT 2019) "Downstream effects of affirmative action"

Considerations for modeling and problem formulation

- Interdisciplinarity → Read and collaborate outside CS
  - Important to know your limits as a researcher (and reviewer)
  - Respect prior work in other fields and avoid reinventing the wheel
  - Understand what compromises are most acceptable when ideal can’t be achieved
  - Establish what is and isn’t allowed in practice

Considerations for modeling and problem formulation

- Stakeholders → Real people are involved
  - Who is this for?
  - Who are we being fair to?
  - What do they want?
  - How do they want fairness defined?

School districting: the problem

- The U.S. public education system is divided into districts
  - “Clusters” of households whose children attend the same school system
- Funding, etc. is not distributed equitably
- Districts are segregated by demographics such as race, income level, etc.
  - Contributes to inequity via factors such as local property taxes funding schools

School districting: the fair clustering solution

- Hammer → Fair clustering
- Nail → Unfair school districts
- Solution → Redraw school districts using fair clustering techniques
  - E.g., draw “group fair” districts with demographic parity for race or income
  - Maybe also bound geographic and population size of each cluster

- What are the limitations and concerns of this approach?

- Are we serving all wants/needs of stakeholders?
  - Maybe some communities don’t want to be split or combined

School districting: the problem with the fair clustering solution

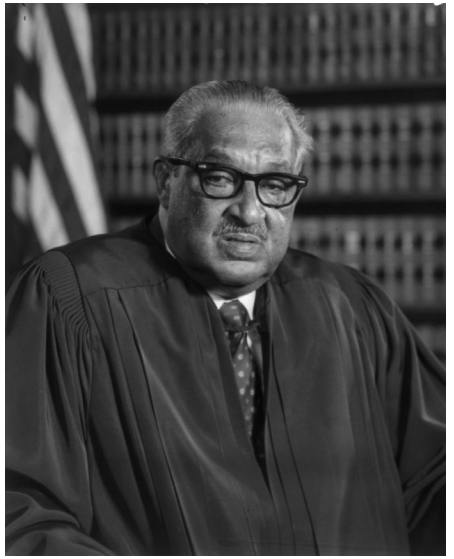
- Need to factor in costs and logistics
  - Some works do this
- People move for schools
  - Families will leave a district if schools are perceived to be “bad” (or they don’t like the demographics of the student body)
  - Families will move to be near “good” schools
- Solution may not be desired by stakeholders
- Solution doesn’t engage with broader context of the problem or downstream effects
- Solution doesn’t engage with legal history
  - Echoes historical solution of bussing

School districting: very brief legal history of bussing

- Brown v. Board of Education (1954)
  - Made de jure school segregation illegal in the U.S.
  - De facto segregation exists to this day
- Bussing
  - Reassign students to better balance racial demographics
  - Similar to group fair clustering
- Milliken v. Bradley (1974)
  - Asked whether students could be bussed across district lines between urban Detroit and its suburbs
  - Ruling prevented states from bussing across district lines
  - Allowed families to avoid integration by moving districts
  - May have contributed to “white flight” and increased housing segregation

School districting: Thurgood Marshall's dissent in Milliken

"Moreover, the result of a Detroit-only decree to this Court would be to increase the flight of whites from the city to the outlying suburbs, compounding the effects of the present rate of increase in the proportion Negro students in the Detroit system. Thus, even if a plan were adopted which at its outset provided in every school a 65 Negro, 35 white racial balance mix in keeping with the Negro-white population of the total school population, such a system would in short order devolve into an all-Negro system. For these reasons, the Detroit-only plan simply has no hope of achieving actual desegregation."



- From Thurgood Marshall's dissent in Milliken v. Bradley (1974)
- Recording from: https://www.oyez.org/cases/1973/73-434

School districting: potential future directions

- Ask experts in education policy how fair clustering might help them
- Learn about and factor in the needs of all stakeholders from parents to school administrators
- Incorporate game theoretic aspects into the model
- Draw “fairest” possible hypothetical districts for the purpose of comparing to and evaluating current districts
  - Can provide data/analysis to argue for systemic changes
  - Can reveal evidence of discrimination
- Explore how clustering can interface with potential systemic changes
  - E.g., housing regulations
- Consider a related problem like school assignment within a district

Evaluation, experiments, and datasets

- Lots of room for improvement in empirical evaluation
  - Methods-based work positions datasets as decontextualized benchmarks
  - Can lead to unconvincing empirical evaluation

- Let’s look at some common dataset issues related to
  - Poor quality of popular public datasets
  - Common misunderstanding/mistakes
  - Mismatched fairness definitions
  - Lack of real world meaning in experiments
  - Benchmarks and community norms

Dataset issue 1: noisy demographic labels

- Important to know where sensitive features in dataset come from
- Criminal justice datasets often have noisy race labels
- Popular “German Credit” dataset has coding errors in sex label
  - Grömping (2019) “South German Credit Data: Correcting a Widely Used Data Set”
- Limits the ability to evaluate demographic fairness with these datasets

Dataset issue 2: weird features & the Adult dataset

- Adult dataset → Publicly available census income dataset
  - Intended for predicting whether income exceeds $50K/year
  - Commonly used as a fairness benchmark
- Fair clustering works → Typically use only numerical features
  - age, fnlwgt, education-num, capital-gain, and hours-per-week
  - Most seem like reasonable features, but what is fnlwgt?
- fnlwgt feature → Lots of issues
  - Orders of magnitude larger than other features and must be normalized to avoid dominating other features
  - Not clear that it is appropriate to use as a feature regardless
  - Meant to be a weight associated with an entry, not a feature
  - Should probably be used when sampling small subset

Dataset issue 2: weird features & the Adult dataset

"The UCI Adult feature "fnlwgt". This column is actually not a demographic feature of an individual but a weight value computed by the Census Bureau to make the sample representative for the US population. We compared the "fnlwgt" data to all weight variables available in IPUMS CPS but did not find an exact match. The closest match is the variable "UH_WGTS_A1", which has a similar distribution. Since we did not identify an exact match for "fnlwgt" and the variable is not a property of an individual, we do not utilize it further in our experiments."

- Ding, Hardt, Miller, and Schmidt (NeurIPS 2021) "Retiring adult: New datasets for fair machine learning"

Dataset issue 3: meaning & equality of opportunity

- Equality of opportunity → Fairness definition from supervised learning
  - People whose true label is “positive” should have an equal chance of being classified as positive regardless of group membership
  - I.e., avoid discrimination in false negatives, but don’t worry about inequality in false positives
  - E.g., qualified job candidates get equal chance at interviewing
  - Hardt, Price, and Srebro (NeurIPS 2016) “Equality of opportunity in supervised learning”
- Wording of definition assumes “positive” label is a good outcome
- Challenge → “Positive” label is a bad outcome in many datasets
  - Leads to backwards experiments if not corrected
  - E.g., don’t give people “equal opportunity” to be labeled high risk and denied bail

Dataset issue 4: fairness mismatch & Diabetes dataset

- Diabetes dataset → Publicly available diabetes patient dataset
  - Commonly used as a fairness benchmark
  - Used to evaluate group fairness in clustering with sex as the sensitive
- Problem→ Health conditions affect men and women differently
  - Balancing male/female membership in clusters could be inappropriate and harmful at worst
- Important to consult an expert before using health/medical data to ensure the fairness intervention makes sense

Dataset issue 5: benchmarks break all the rules

- Many previous issues and similar mistakes are enshrined as benchmarks
  - Reviewers expect to see these experiments replicated in new work
- Ideal to supplement classic benchmarks with better experiments
  - Benefit of pursuing a specific application with relevant data

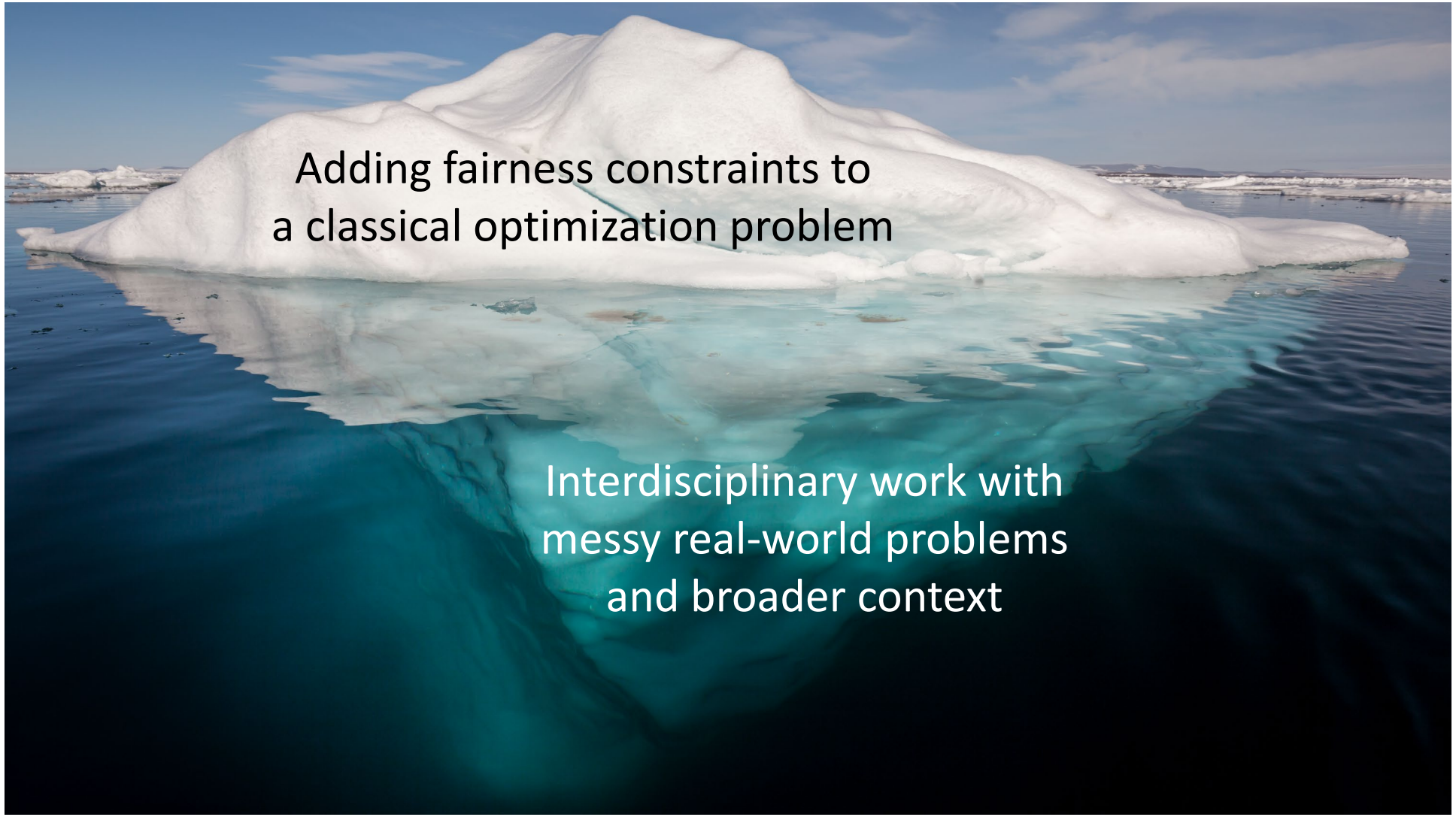
Concluding thoughts on datasets and evaluation

- Recuring questions to consider:
  - How was data prepared? Who reported sensitive features?
  - What does it “mean” to cluster this data this way?
  - What do domain experts think of this experiment?
- More reading on datasets:
  - Fabris, Messina, Silvello, and Susto (arxiv 2022) “Algorithmic Fairness Datasets: the Story so Far”
  - Bao, Zhou, Zottola, Brubach, Desmarais, Horowitz, Lum, Venkatasubramanian (NeurIPS 2021) “It's COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks”
  - Gebru, Morgenstern, Vecchione, Wortman Vaughan, Wallach, Daumé Iii, and Crawford (Comm. ACM 2021) “Datasheets for datasets”

Where we're heading…

- **Outcomes** → Summarize risks and responsibilities that motivate careful and thoughtful approaches to fair clustering
- **Methods** → Examine the challenges and pitfalls we can encounter doing meaningful and impactful fairness work
- **Next steps** → Explore the vast frontier of future work that arises from interdisciplinary engagement and specific applications

Wide open frontier of future work in fair clustering

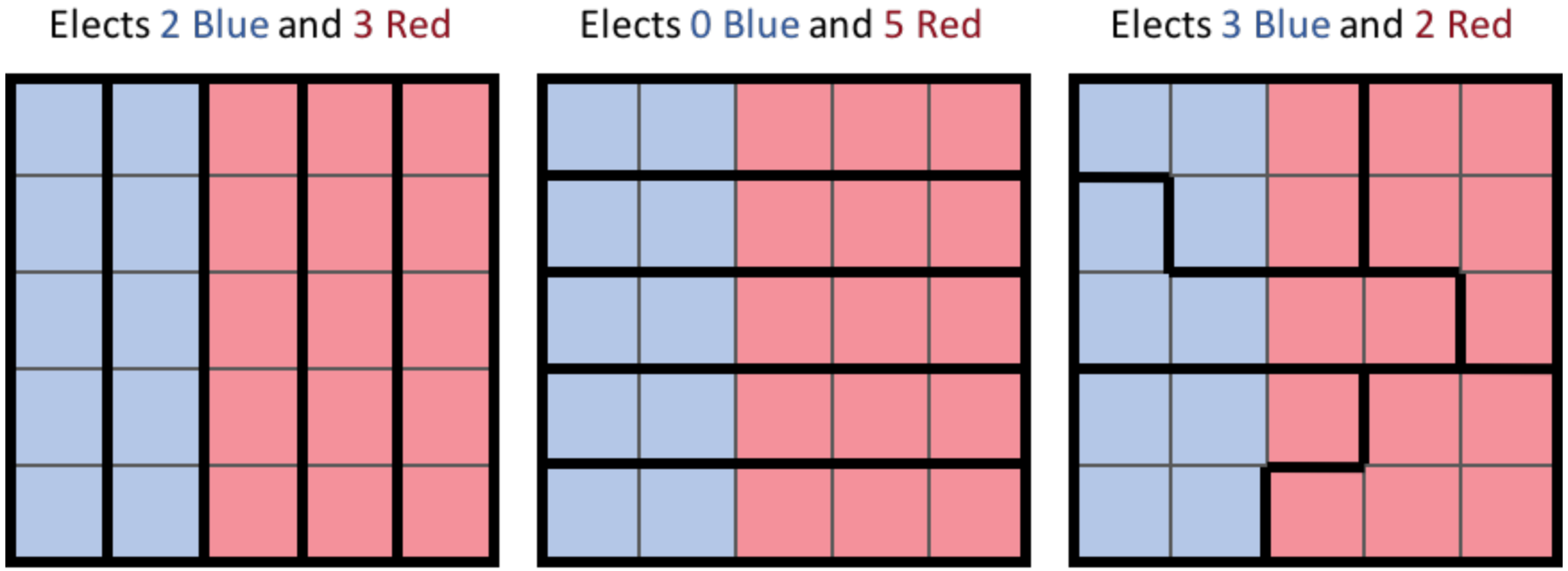


Recommendations for future work

- Find a specific social problem/application/domain that interests you and learn about it from sources outside CS academia
- Study every facet of a problem and find a new perspective to explore
- Seek interdisciplinary collaborations and expert input
- Consider broader context and upstream/downstream impact
- Build models and fairness definitions grounded in evidence and need
- Start with a problem or dataset in mind
  - Nail first as opposed to hammer first

Political districting: the problem

- U.S. states are divided into single-member, winner-take-all districts
- Each district elects a representative to congress
- Voters/parties/groups are disenfranchised through gerrymandering
- Open problem for over 200 years



Political districting: what is fair?

- Balanced population
- Preserving communities (e.g., towns) in same district
- Compact districts or k-median objective
- Competitive districts
- Majority minority districts
- Proportional representation

Political districting: what is unfair?

- Wasted votes/large “efficiency gap”
- Weird looking districts
- One party/group getting disproportionate share of representation
- One party/group disenfranchised
- “Safe” districts for incumbents
- “Kidnapping” incumbents
- Redistricting maps that are outliers with respect to a random sample of legal maps

Political districting: considering different perspectives

- Algorithmic fairness → Inspire new fairness definitions and civil rights
  - Brubach, Chakrabarti, Dickerson, Khuller, Srinivasan, and Tsepenekas. (ICML 2020) "A pairwise fair and community-preserving approach to k-center clustering"
- Impossibility results → Show which fairness goals are impossible
  - E.g., can't guarantee that all pairs of voters who share same district in a majority of possible maps will share district
- Downstream effects → Ask how voter incentives are affected
  - Brubach, Srinivasan, and Zhao (EC 2020) "Meddling metrics: the effects of measuring and constraining partisan gerrymandering on voter incentives"
- Systemic changes → Show benefits of an alternative system
  - Garg, Gurnee, Rothschild, and Shmoys (arxiv 2021) "Combatting Gerrymandering with Social Choice: the Design of Multi-member Districts"

Thanks!

Questions? Discussion?

Thanks!

Questions? Discussion?

Thanks!

Questions? Discussion?