No Stratification Without Representation

GERDUS BENADÈ, PAUL GÖLZ, and ARIEL D. PROCACCIA, Carnegie Mellon University

Sortition is an alternative approach to democracy, in which representatives are not elected but randomly selected from the population. Most electoral democracies fail to accurately represent even a handful of protected groups. By contrast, sortition guarantees that *every* subset of the population will in expectation fill their fair share of the available positions. This fairness property remains satisfied when the sample is stratified based on known features. Moreover, stratification can greatly reduce the variance in the number of positions filled by any unknown group, as long as this group correlates with the strata. Our main result is that stratification cannot increase this variance by more than a negligible factor, even in the presence of indivisibilities and rounding. When the unknown group is unevenly spread across strata, we give a guarantee on the reduction in variance with respect to uniform sampling. We also contextualize stratification and uniform sampling in the space of fair sampling algorithms. Finally, we apply our insights to an empirical case study.

1 INTRODUCTION

Does democracy need elections? The very question might seem absurd — so much has democracy become synonymous with the electoral process. We do not dispute the value of representation, i.e., of distilling from a large populace a smaller body of representatives. Elections are one tool for achieving this. However, they are only useful insofar as the chosen representatives have the ability and legitimacy to make decisions for the community.

Alas, both effectiveness and legitimacy have increasingly come under siege: For one, the constant need for self-promotion undermines the credibility of politicians and with it their legitimacy. It also hampers their capacity to make unpopular decisions, immobilizing Congress for ever longer periods of time before every election. There are many more examples: Around the world, recent elections have rewarded populist candidates who skillfully stoke public emotion, a skill that need not reflect a capacity to govern. Finally, the underrepresentation of women, racial minorities, and social milieus risks making legislatures less receptive to the needs of large parts of the population — and lays bare the remoteness of the representatives from the people.

Pointing out the shortcomings of electoral democracy is easy but futile in the absence of alternatives. Are there other ways of selecting representatives in a democracy? In the last decades, a growing coalition of academics and activists has been advancing *sortition* as one such option [3, 7, 11, 12, 17, 21, 22], i.e., picking the representatives at random from the population. A panel selected by sortition is composed of people of all walks of life and gives every citizen a chance at participation. Obviously, these representatives are not selected for political skill or experience. For this reason, sortition often accompanies a heavy focus on providing representatives with informational material, access to experts, and opportunities to deliberate with other representatives — which is why some also speak of *deliberative democracy* [12]. In contrast to opinion polls and plebiscites, sortition reveals an *informed* opinion, which can act as a proxy for the will of a

(utopian) perpetually-informed public. Nonetheless, this proposal is radical in that it is predicated on the ability of average citizens to govern themselves and their peers.

It may, therefore, come as a surprise that there is extensive historical precedent for governance by sortition — and that present-day experiments give grounds for optimism. In 4th-century BC Athens, the Council of 500 and the People's Court, described by contemporary sources as equal to the better-known Assembly, were comprised of randomly selected citizens [14]. Sortition also played a role in the Italian city-states of the Renaissance, most notably in Venice, where the doge was selected through a complicated sequence of rounds alternating election and sortition [16]. After sortition was abandoned in favor of elections in the aftermath of the American and French revolutions, it was largely forgotten and persisted chiefly as a method for selecting jurors.

Over the last two decades, sortition saw a revival in the formation of citizens' panels around the world, with encouraging outcomes. In his excellent survey on sortition, Van Reybrouck [22] gives a detailed account of several of these experiments, in which such randomly selected panels deliberated on topics as varied as energy policy, electoral reform, and gay marriage.

Given our limited space, we will focus on the Citizens' Assembly in Ireland,[1] which took place over twelve weekends between 2016 and 2018. The Irish parliament established this panel of 99 citizens and charged it with formulating policy recommendations on multiple topics of societal importance. The most prominent issue on the agenda was the legal status of abortion. At the time, the Irish constitution granted an equal right to life to an embryo and its mother, and abortion was illegal even in extreme cases. Over five weekends, the panel met with experts to receive legal, medical, and ethical information, heard advocacy groups, and deliberated. The final recommendation — to legalize abortions up to the twelfth week without indication of specific reasons — was more liberal than observers expected.[2] Nevertheless, the constitutional amendment gained 67% approval in a 2018 referendum, and passed into law. In anonymous comments gathered after the vote [9, Appendix B], participants expressed varied and nuanced positions on the issue, and showed an exceptional level of ownership of the process. The Irish government is yet to act on the panel's other recommendations, but this experiment has already shown that a cross-section of society can make competent decisions on complex issues.

A unique advantage of sortition is that it affords fair representation *to all possible groups in the population*, in expectation. As long as every agent has the same probability of being selected, any subpopulation M is expected to send $\frac{|M|}{n} k$ representatives to a council of size k, where n is the total number of agents. This is a drastic improvement over the status quo, which widely fails to provide similar representation even to high-profile groups such as women or racial minorities.

Still, these guarantees only hold *ex ante*, i.e., in expectation over the random sampling. Panels are typically selected to be large enough that any given group is unlikely to be grossly over- or underrepresented. Still, practitioners seem unhappy with the plausible deviations in demographic representation. To mitigate these deviations, they often *stratify* their samples — for example, by choosing half of the representatives among women and half among men. Assuming that this reflects the composition of the population, stratification does not change an individual's probability of being selected, nor the expected representation of groups. However, it guarantees *ex-post* fair representation to the strata, i.e., to women and men.

But what if the exact group is unknown to the stratifier? We would argue that the largest need for equal representation concerns political convictions. If the panels are constituted repeatedly, fluctuations in demographic representation might impact optics; the guarantee on expected representation ensures that all demographic groups wield their fair share of power over time. This argument does

(a) M contains 90% of women, 10% of men.

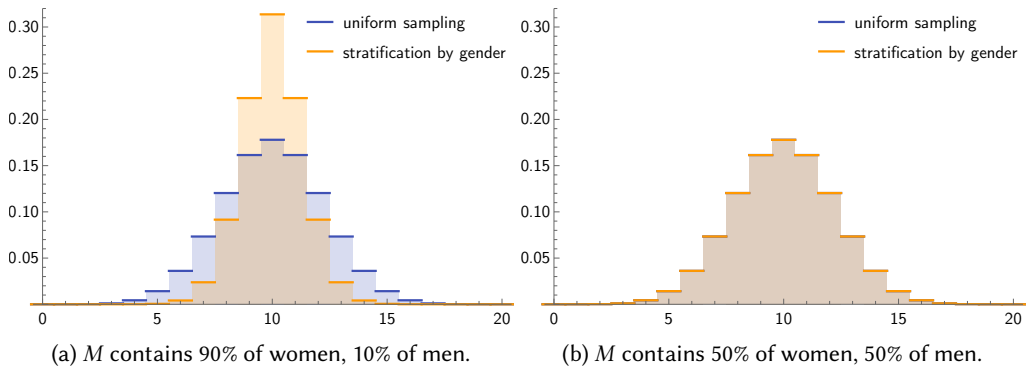(b) M contains 50% of women, 50% of men.

Fig. 1. Probabilities of different numbers of agents from M in the panel. The population consists of 500 women and 500 men; the panel has size 20; 500 agents are in M, distributed among genders as indicated.

not apply to political opinions: Certainly, fair expected representation is a big improvement over the status quo, and we know of no other system that could give similar guarantees. But a negative fluctuation in representation can prevent a suggestion from gaining traction or could afford an undeserved majority to the other side. If the opinion was relevant to the topic of this specific panel, higher representation in future panels might not compensate for a bad decision.

We cannot directly stratify for these opinions either since they are hidden features — informed opinions will only be revealed through information and deliberation. Furthermore, we might not even know which opinions will be relevant, and the opinions could be too numerous to stratify for all of them.

Nonetheless, we show that stratification is an effective tool for promoting fair representation of such groups, by reducing the *variance* of their representation. Taking the perspective of the organizer of a citizens' panel, our aim is to

> ... characterize the effect of stratification on the variance of the number of representatives from *unknown* groups, and to demonstrate that this knowledge can help more accurately represent opinion holders.

1.1 Our Approach and Results

The key insight underlying our work is that the benefits of stratification extend, beyond the strata themselves, to all groups which correlate with the strata. Informally, this observation has been made as early as 1972 [17]. Consider our earlier example of stratification by gender. Let M capture a certain political opinion held by, say, half of the population, and let a random variable A denote the number of agents from M in the panel. If the opinion is highly prevalent among women and rare among men, the distribution of A is much more concentrated with stratification than without, as can be seen in Fig. 1a. Quite amazingly, there is virtually nothing to be lost in stratifying: If, as in Fig. 1b, M is equally split between the strata — which is the worst case — stratification only increases the variance by a minuscule amount.

Transcending our toy example, these observations suggest a strategy of more elaborate stratification than is common today: When planning a citizens' panel, an organizer can use public data to partition the population into many small strata, each corresponding to a small number of seats in the panel. The goal is to group together citizens who are as similar as possible. Since demographic data is highly predictive of political views [18, p. 92], most M of interest should "polarize" the strata, i.e., most strata should either have a very high or very low prevalence of M, in which case we profit

from the reduction in variance observed above. To do so, a panel organizer needs to reason about the variance of A, which can be difficult.

Our main contribution is a tight bound showing that — up to a factor very close to one — stratification cannot increase variance. Crucially, we propose a way of rounding seat allocations for the strata, which maintains equal probability of participation for all agents, and still satisfies the above guarantee. To gauge the benefits of stratification, we give a second upper bound which characterizes the reduction in variance due to stratification in terms of the concentration of M in every stratum.

Next, we explore the space of sampling algorithms that uphold expected representation for all groups, but are not necessarily based on stratification. We show that no such algorithm dominates any other algorithm; thus, we need to assume that relevant M will correlate with visible features. Finally, we show that uniform sampling is optimal from a worst-case perspective. Since stratified sampling can *never* increase variance by more than a minuscule amount, all our sampling algorithms provide close-to-optimal worst-case guarantees.

Finally, we investigate the effect of stratification on variance using a large dataset containing information about demographic features and political attitudes. We find that random stratifications are helpful overall, but that only a few of them lead to significant reductions. In a case study simulating the situation of a panel organizer, we find that, using insights from our theoretical analysis, a human stratifier can simultaneously reduce the variance with respect to multiple unknown, attitude-based groups. Compared to uniform sampling, these decreases in variances correspond to an increase in panel size by multiple positions. Manual stratification also clearly outperforms a simple stratification by gender and race. Stratifications automatically generated via k-means clustering fall short of the manual stratification, but show promise.

1.2 Related Work

While randomization is a common tool in computational social choice [5], few papers explicitly consider sortition. One exception is an article by Walsh and Xia [23], in which they prefix standard voting rules with random subsampling of the voters. They find that sortition can add hardness of manipulation to otherwise manipulable voting rules. Thus, they consider sortition as an additional step in choosing a societal outcome, whereas we are interested in the composition of the panel itself.

Saran and Tumennasan [19] study which social choice rules can be implemented in Nash equilibrium under sortition. Agents can vote strategically, but must commit to a vote before knowing who is chosen into the panel. Then, under mild assumptions, every rule that can be implemented by consulting the entire population can also be implemented under (uniform) sortition, as long as panels have at least size four.

We share our goal of fairly representing subgroups with multiple areas in social choice, namely apportionment and multi-winner elections. In apportionment, the task is to distribute positions in a legislature between groups such that the allocation is approximately proportional to the group sizes (see Balinski and Young [2, Appendix A] for an overview). While all standard apportionment mechanisms are deterministic, Grimmett [13] proposes a randomized apportionment mechanism. Whereas, in our setting, agents are part of many groups, apportionment generally considers every agent as belonging to a single state or party.

This restriction is lifted in the *party-approval elections* recently introduced by Wilker [24]. In this setting, agents can belong to multiple known groups, and each group receives a nonnegative number of positions in the committee. All representation axioms that can be ensured in this setting provide an agent with guarantees on the number of positions summed up over all of the agent's

groups. By contrast, a sortition panel is composed of agents who themselves belong to multiple groups, and our representation guarantees hold for every group simultaneously.

Party-approval elections, in turn, are a special case of multi-winner elections [1]. Since they allow agents to directly approve arbitrary subsets of candidates, these preferences cannot directly be interpreted as group memberships. By constraining the set of possible preferences, one recovers apportionment [6] or party-approval elections.

On a technical level, our work is connected to questions of (stratified) sampling in statistics [10, 15]. Whereas we want a proportional representation of a certain feature in our sample, pollsters use the sample to obtain an accurate estimate of the prevalence of the feature. In this framework, our requirement of fair expected representation translates into requiring an unbiased estimator. In both cases, mechanisms leading to lower variance are preferred, and it is known that sampling from a continuous pool cannot increase variance, which we will show again in Section 3. Despite these similarities, our setting is more restrictive, since we cannot weight agents differently, which is an important technique in polling. While, in final votes, weighting representatives might be defensible, a representative's influence on a debate cannot be weighted.

2 PRELIMINARIES

Denote with $N := [n]$ the population, by which we mean the set $\{1, \dots, n\}$ for some $n \in \mathbb{N}$. Out of these n agents, we will randomly sample a panel of size k. An unknown set of agents $M \subseteq N$ with $|M| = m$ share a hidden feature, which we hope to represent as fairly as possible.

Let $\mathcal{U}(N, k)$ be the *uniform sampling algorithm*, which returns a uniformly random subset of k agents. Notice that $\mathbb{P}[x \in \mathcal{U}(N, k)] = k/n$ for all $x \in N$. The number of agents with the hidden feature selected by $\mathcal{U}$ is $U_M^{n,k} := |\mathcal{U}(N, k) \cap M|$. Since $\mathcal{U}_M^{n,k}$ follows a hypergeometric distribution,

$$\mathrm{Var}(U_M^{n,k}) = k \, \frac{m}{n} \, \left(1 - \frac{m}{n}\right) \frac{n-k}{n-1}.$$

In general, $\mathcal{A}(N, k)$ will denote a random process which generates subsets of N of size k. As above, we set $A_M^{n,k}$ for the number $|\mathcal{A}(N, k) \cap M|$ of agents from M in the panel. To ensure fair representation to every possible subset $M \subseteq N$ of agents, we constrain such processes to satisfy $\mathbb{E}[A_M^{n,k}] = \frac{m}{n} k$ for all $M \subseteq N$. By linearity of expectation, this is equivalent to every agent $x \in N$ being selected with equal probability

$$\mathbb{P}[i \in \mathcal{A}(N, k)] = \frac{k}{n}. \tag{1}$$

Our goal is to find sampling algorithms $\mathcal{A}$ which, for unknown and arbitrary M, reduce the variance of $A_M^{n,k}$, compared to $U_M^{n,k}$.

We will be particularly interested in stratified sampling algorithms. Such an algorithm defines a partition of N into ℓ strata, i.e., $N = N_1 \,\dot\cup\, N_2 \,\dot\cup \cdots \dot\cup\, N_\ell$. Let $n_i := |N_i|$ and $m_i := |N_i \cap M|$ for all $i \in [\ell]$. If the agents with the hidden feature were proportionally distributed across strata, we would expect stratum i to contain $\frac{n_i}{n} m$ of them. Define ϵ_i as the difference between the actual and expected number of agents with the hidden feature in stratum i, i.e., $m_i = \frac{n_i}{n} m + \epsilon_i$. Note that $\sum_{i \in [\ell]} \epsilon_i = 0$.

After stratification, we assume that $\mathcal{A}$ will select $\tilde{k}_i$ agents out of each stratum i uniformly at random. Conceptually, $\tilde{k}_i$ should be $k_i := \frac{n_i}{n} k$, which would guarantee that every agent is selected with the appropriate probability. However, k_i may not be integer, in which case it is necessary to randomly "round"[3] k_i to get $\tilde{k}_i$. Because the sampling happens independently in each stratum, the

---

[3] In the rounding scheme proposed in Section 4, $\tilde{k}_i$ can take on integer values between $\lfloor k_i \rfloor - 1$ and $\lceil k_i \rceil + 1$, so our use of the term is slightly non-standard.

variance of $A_M^{n,k}$ is

$$\sum_{i=1}^{\ell} \tilde{k}_i \frac{m_i}{n_i} \left(1 - \frac{m_i}{n_i}\right) \frac{n_i - \tilde{k}_i}{n_i - 1},$$

conditioned on the choice of the $\tilde{k}_i$.

3 WARMING UP IN A CONTINUOUS WORLD

We hope to employ stratification to decrease the variance of the number of selected agents possessing a hidden feature. There are two ways in which the discrete nature of our setting complicates studying this variance, however.

First, the $\tilde{k}_i$ draws from a single stratum (or the k draws for uniform sampling) are not independent, since the same agent cannot be chosen multiple times. This leads to variance terms of the form $\tilde{k}_i \frac{m_i}{n_i} (1 - \frac{m_i}{n_i}) \frac{n_i - \tilde{k}_i}{n_i - 1}$ rather than $\tilde{k}_i \frac{m_i}{n_i} (1 - \frac{m_i}{n_i})$. Since the $\tilde{k}_i$ are typically much smaller than the n_i, these correction factors do not matter, but they suffice to make the variances less well-behaved.

Second, the indivisibility of agents forces us to round the k_i. Again, this rounding should not change the big picture, but showing this requires careful analysis.

To understand the high-level impact of rounding on variance, we will for now ignore these complications. We study a setting in which the agents form a continuum. To make the connections between the models more suggestive, we reuse the notation from our standard model, but would like to point out where the continuous setting differs: Algorithms can now return any multiset with cardinality k and support N. Accordingly, the uniform mechanism now returns the collection of k independent uniform draws[4] from N and has a variance of $k \frac{m}{n} (1 - \frac{m}{n})$. Each agent x is expected to appear $\frac{k}{n}$ times in the panel, and this is what we require of all algorithms in this setting. All stratified-sampling algorithms will simply have $\tilde{k}_i = k_i$ and sample from every stratum as the uniform algorithm samples from the whole population. Thus, the variance of stratified sampling is $\sum_i k_i \frac{m_i}{n_i} (1 - \frac{m_i}{n_i})$.

Surprisingly, in this setting, the benefits of stratification come entirely for free! More precisely, stratification never increases the variance over uniform sampling — but can bring it all the way down to zero if every stratum contains either only members of M, or only members of its complement. The same argument can be applied to any preexisting stratum to argue for further subdivision, down to the level of individual agents. Notably, this observation does not extend to the discrete case, where we will see that such extreme stratification leads to a potentially large increase in variance due to rounding.

PROPOSITION 3.1. *In the continuous setting, for any stratified-sampling algorithm $\mathcal{A}$ and any M, it holds that $Var(A_M^{n,k}) \leq Var(U_M^{n,k})$.*

PROOF.

$$\mathrm{Var}(A_M^{n,k}) = \sum_{i=1}^{\ell} k_i \frac{m_i}{n_i} \left(1 - \frac{m_i}{n_i}\right)$$

$$= \sum_{i=1}^{\ell} \frac{n_i}{n} k \frac{m_i}{n_i} \left(1 - \frac{m_i}{n_i}\right)$$

$$= \frac{k}{n} \sum_{i=1}^{\ell} m_i \left(1 - \frac{m_i}{n_i}\right)$$

---

[4]This is not equivalent to returning a uniformly chosen multiset. For example, the multiset $\{1, \dots, k\}$ is $k!$ times more likely than the multiset containing the first agent k times.

$$\leq \frac{k}{n} \mu,$$

where μ is the optimal value of

$$\max_{m_i} \left\{ \sum_{i=1}^{\ell} \left(m_i - \frac{m_i^2}{n_i} \right) : \sum_{i=1}^{\ell} m_i = m \right\}.$$

In our setting, the m_i must be integers between 0 and n_i, but relaxing this requirement to $m_i \in \mathbb{R}$ does not invalidate the inequality.

Since the maximized function is concave (and smooth), and since the equality constraint is affine, the KKT conditions are sufficient for a global maximum. For this, we need a real constant λ which equals $-1 + 2 \frac{m_i}{n_i}$ for all i, which means that $\frac{m_i}{n_i} = \frac{m_j}{n_j}$ for all $i, j \in [\ell]$. This can easily be reconciled with the constraint $\sum_i m_i = m$ by choosing the common value of the $\frac{m_i}{n_i}$ as $\frac{m}{n}$. In other words, the worst case variance occurs when the hidden feature appears in every stratum with the same density as in the overall population. It follows that

$$\mathrm{Var}(A_M^{n,k}) \leq \frac{k}{n} \sum_{i=1}^{\ell} m_i \left(1 - \frac{m}{n} \right) = \frac{k}{n} \left(1 - \frac{m}{n} \right) \sum_{i=1}^{\ell} m_i = \frac{k}{n} \left(1 - \frac{m}{n} \right) m = \mathrm{Var}(U_M^{n,k}). \qquad \square$$

For the rest of the paper, we return to our standard discrete setting.

4 MAIN RESULT: THE VARIANCE OF STRATIFIED SAMPLING

In this section, we transport the result of Proposition 3.1 from the continuous utopia to our discrete reality. We first discuss rounding, then turn to establishing an upper bound on the variance of stratified sampling.

4.1 Block Rounding

Arguably the most significant obstacle to studying the variance of stratified sampling is the construction of a rounding mechanism. How should we round k_i to get $\tilde{k}_i$? To satisfy Eq. (1), every stratum i must satisfy $\mathbb{E}[\tilde{k}_i] = k_i$. Simultaneously, the rounding must be dependent to ensure that $\sum_i \tilde{k}_i = k$.

How we do this can seriously affect the variance. Consider the case where half of the strata are full of agents in M and the other half contain none of them. Then, after rounding, $A_M^{n,k}$ is deterministic and has zero variance. All variance comes from the rounding process, which directly assigns between 0 and $O(\ell)$ additional seats to M. Since the rounding decisions are not independent, the introduced variance may be in $O(\ell^2)$, which can drastically exceed the variance of uniform sampling. Ideally, our rounding mechanism would never add more variance than what lies between the real variance and the variance of uniform sampling on the same M. Since we were not able to make such an argument for off-the-shelf dependent randomized rounding mechanisms as defined by Chekuri et al. [8], we propose our own rounding mechanism, which we refer to as *block rounding*.

Imagine lining up the agents, stratum by stratum, in the order of strata. As shown in Fig. 2, draw a line every $\frac{n}{k}$ agents to obtain a block for each of the k open seats.

Some of these blocks only contain agents from a single stratum (for example, the first, third, fourth, and seventh blocks in Fig. 2). Such a seat will certainly be filled with an agent from that stratum. For each stratum i, denote the number of seats guaranteed to be filled by one of its members by g_i.

We make the innocuous assumption (as explicitly done in Theorem 4.1) that $n_i \geq \frac{n}{k}$, i.e., each stratum is large enough that it is entitled to at least one seat in expectation. It follows that blocks of size $\frac{n}{k}$ can intersect at most two strata.
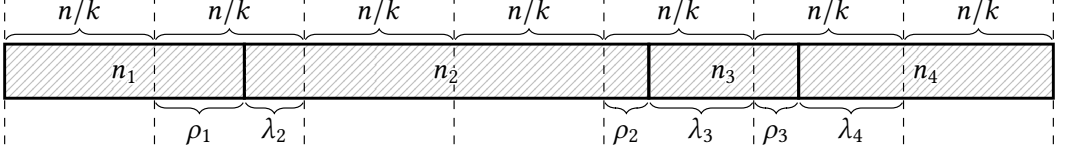
Fig. 2. Example illustrating block rounding, with $\ell = 4$ strata (boxes) and $k = 7$ many blocks (areas between dashed lines). The indicated λ_i and ρ_i should be divided by $\frac{n}{k}$ such that, for example, $\rho_1 + \lambda_2 = 1$. Both λ_1 and ρ_4 are zero.

Consider a block that intersects two strata i and $i + 1$. Let the fraction of the block intersecting stratum i be ρ_i and the fraction intersecting stratum $i+1$, λ_{i+1}. The seat corresponding to this block's seat will be drawn from stratum i with probability ρ_i and else (with probability λ_{i+1}) from stratum $i + 1$, and this decision is made independently for each block. We will record these options through binary random variables R_i and L_{i+1}, specifically, $R_i = 1$ with probability ρ_i and $L_{i+1} = 1 - R_i$. When the end of a stratum lines up with the border of a block, the corresponding variables are zero.

Thus, we choose $\tilde{k}_i := g_i + L_i + R_i$ many agents from stratum i.[5] It is easy to see that $\mathbb{E}[\tilde{k}_i] = k_i$ and that $\sum_i \tilde{k}_i = k$. Furthermore, all L_i and R_j belonging to different blocks are independent, whereas for all i,

$$CoVar(R_i, L_{i+1}) = \mathbb{E}[R_i\, L_{i+1}] - \mathbb{E}[R_i]\,\mathbb{E}[L_{i+1}] = \mathbb{E}[0] - \rho_i\,\lambda_{i+1} = -\rho_i\,(1 - \rho_i).$$

4.2 Variance Upper Bound

Using the foregoing rounding scheme construction, we are able to formulate and prove our main result.

THEOREM 4.1. *For some n and k, let $\mathcal{A}(N, k)$ be a stratifying algorithm based on block rounding. Suppose that every stratum i has size $n_i \geq \frac{n}{k}$, i.e., that the expected number k_i of selected representatives from i is at least 1. Then, for any $M \subseteq N$,*

$$Var\left(A_M^{n,k}\right) \leq \frac{n-1}{n-k}\, Var\left(U_M^{n,k}\right).$$

If we assume that k grows much slower than n, $\frac{n-1}{n-k}$ is essentially 1 for large n. For example, for the Irish Citizens' Assembly, this number is around 1.00003. Thus, we nearly recover what we found in the continuous case, i.e., that stratification can only decrease variance.

Turning to the theorem's proof, we advise the interested reader to pay attention to the ϵ_i terms, as defined in Section 2. We find their appearance, and subsequent disappearance, particularly satisfying.

PROOF OF THEOREM 4.1. Since we fix the algorithm and M, we will simplify the notation by setting $A := A_M^{n,k}$ and $U := U_M^{n,k}$.

By the law of total variance,

$$Var(A) = \mathbb{E}[Var(A \mid L_1, \ldots, L_\ell, R_1, \ldots, R_\ell)] + Var(\mathbb{E}[A \mid L_1, \ldots, L_\ell, R_1, \ldots, R_\ell]), \tag{2}$$

where the outer expectation and variance range over the choices of the L_i and R_i.

---

[5]Going forward, we assume that this is indeed feasible, i.e., that $\tilde{k}_i$ never exceeds n_i. This can be enforced through extremely mild assumptions, e.g., by requiring $k \leq n/3$ for all i.

Recall that $\tilde{k}_i = g_i + L_i + R_i \in \mathbb{N}_0$ is the (rounded) number of agents sampled from stratum i. We bound the first summand:

$$\mathbb{E}[Var(A \mid L_1, \ldots, L_\ell, R_1, \ldots, R_\ell)]$$

$$= \mathbb{E}\left[\sum_i \tilde{k}_i \frac{m_i}{n_i}\left(1 - \frac{m_i}{n_i}\right)\frac{n_i - \tilde{k}_i}{n_i - 1}\right]$$

$$\leq \mathbb{E}\left[\sum_i \tilde{k}_i \frac{m_i}{n_i}\left(1 - \frac{m_i}{n_i}\right)\right] \qquad (\tilde{k}_i \in \mathbb{N}_0, \text{ thus } \tilde{k}_i = 0 \text{ or } \tilde{k}_i \geq 1)$$

$$= \mathbb{E}\left[\sum_i (g_i + L_i + R_i)\frac{m_i}{n_i}\left(1 - \frac{m_i}{n_i}\right)\right]$$

$$= \sum_i k_i \frac{m_i}{n_i}\left(1 - \frac{m_i}{n_i}\right) \qquad \text{(linearity of expectation)}$$

$$= \sum_i \frac{n_i}{n} k \frac{m_i}{n_i}\left(1 - \frac{m_i}{n_i}\right)$$

$$= \frac{k}{n}\sum_i m_i\left(1 - \frac{m_i}{n_i}\right)$$

$$= \frac{k}{n}\sum_i \left(\frac{n_i}{n}m + \epsilon_i\right)\left(1 - \frac{\frac{n_i}{n}m + \epsilon_i}{n_i}\right)$$

$$= \frac{k}{n}\sum_i \frac{m}{n}n_i + \epsilon_i - \frac{m^2}{n^2}n_i - 2\frac{m}{n}\epsilon_i - \frac{\epsilon_i^2}{n_i}$$

$$= \frac{k}{n}\left(\left(\frac{m}{n}\sum_i n_i\right) + \left(\sum_i \epsilon_i\right) - \left(\frac{m^2}{n^2}\sum_i n_i\right) - \left(2\frac{m}{n}\sum_i \epsilon_i\right) - \left(\sum_i \frac{\epsilon_i^2}{n_i}\right)\right)$$

$$= \frac{k}{n}\left(m + 0 - \frac{m^2}{n} - 0 - \sum_i \frac{\epsilon_i^2}{n_i}\right)$$

$$= k\frac{m}{n}\left(1 - \frac{m}{n}\right) - \frac{k}{n}\sum_i \frac{\epsilon_i^2}{n_i}.$$

Recall that the uniform algorithm had a variance of

$$Var(U) = k\frac{m}{n}\left(1 - \frac{m}{n}\right)\frac{n - k}{n - 1}.$$

Thus, the first summand of Eq. (2) can be bounded as

$$\mathbb{E}[Var(A \mid L_1, \ldots, L_\ell, R_1, \ldots, R_\ell)] \leq \frac{n-1}{n-k}Var(U) - \frac{k}{n}\sum_i \frac{\epsilon_i^2}{n_i}. \qquad (3)$$

Now, consider the second summand of Eq. (2), which was $Var(\mathbb{E}[A \mid L_1, \ldots, L_\ell, R_1, \ldots, R_\ell])$.

$$Var(\mathbb{E}[A \mid L_1, \ldots, L_\ell, R_1, \ldots, R_\ell])$$

$$= Var\left(\sum_i \frac{m_i}{n_i}\tilde{k}_i\right)$$

$$= Var\left(\sum_i \left(\frac{m}{n} + \frac{\epsilon_i}{n_i}\right)(g_i + L_i + R_i)\right)$$

$$= Var\left(\underbrace{\sum_i \left(\frac{m}{n} + \frac{\epsilon_i}{n_i}\right)g_i + \frac{m}{n}\sum_i(L_i + R_i)}_{\text{constant}} + \sum_i \frac{\epsilon_i}{n_i}L_i + \sum_i \frac{\epsilon_i}{n_i}R_i\right)$$

$$= Var\left(\sum_i \frac{\epsilon_i}{n_i}L_i + \sum_i \frac{\epsilon_i}{n_i}R_i\right).$$

We can rewrite the latter variance of a linear combination as

$$= \sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 Var(L_i) + \sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 Var(R_i) + 2\sum_{i,j} \frac{\epsilon_i}{n_i}\frac{\epsilon_j}{n_j} CoVar(L_i, R_j)$$

$$+ 2\sum_{i<j} \frac{\epsilon_i}{n_i}\frac{\epsilon_j}{n_j} CoVar(L_i, L_j) + 2\sum_{i<j} \frac{\epsilon_i}{n_i}\frac{\epsilon_j}{n_j} CoVar(R_i, R_j)$$

$$\leq \frac{1}{2}\sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 + 2\sum_{i,j} \frac{\epsilon_i}{n_i}\frac{\epsilon_j}{n_j} CoVar(L_i, R_j) + 2\sum_{i<j} \frac{\epsilon_i}{n_i}\frac{\epsilon_j}{n_j} CoVar(L_i, L_j)$$

$$+ 2\sum_{i<j} \frac{\epsilon_i}{n_i}\frac{\epsilon_j}{n_j} CoVar(R_i, R_j),$$

where the last step used the fact that binary variables have variance of at most $\frac{1}{4}$. As noted, nearly all of these L_i and R_i are independent, and thus, their covariance is zero. This allows us to drop most covariance terms, with only $\ell - 1$ many remaining. All other rounding methods we considered correlate more pairs of variables, leading to bounds on the second term that might far exceed the first term.[6] Block rounding allows us to continue

$$= \frac{1}{2}\sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 + 2\sum_{i=1}^{\ell-1} \frac{\epsilon_i}{n_i}\frac{\epsilon_{i+1}}{n_{i+1}} CoVar(R_i, L_{i+1})$$

$$= \frac{1}{2}\sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 - 2\sum_{i=1}^{\ell-1} \frac{\epsilon_i}{n_i}\frac{\epsilon_{i+1}}{n_{i+1}} \rho_i(1 - \rho_i)$$

$$\leq \frac{1}{2}\sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 + 2\sum_{i=1}^{\ell-1} \frac{|\epsilon_i|}{n_i}\frac{|\epsilon_{i+1}|}{n_{i+1}} \rho_i(1 - \rho_i)$$

$$\leq \frac{1}{2}\sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 + \frac{1}{2}\sum_{i=1}^{\ell-1} \frac{|\epsilon_i|}{n_i}\frac{|\epsilon_{i+1}|}{n_{i+1}}.$$

For all real numbers a and b, $a^2 + b^2 - 2ab = (a-b)^2 \geq 0$, and thus, $\frac{1}{2}(a^2 + b^2) \geq ab$. Accordingly,

$$\leq \frac{1}{2}\sum_i \left(\frac{\epsilon_i}{n_i}\right)^2 + \frac{1}{4}\sum_{i=1}^{\ell-1} \left(\left(\frac{|\epsilon_i|}{n_i}\right)^2 + \left(\frac{|\epsilon_{i+1}|}{n_{i+1}}\right)^2\right)$$

---

[6]While dependent rounding [8] guarantees negative correlation, this is not helpful when two strata have different $sgn(\epsilon_i)$.

$$= \frac{1}{2} \sum_i \left(\frac{\epsilon_i}{n_i} \right)^2 + \frac{1}{4} \sum_{i=1}^{\ell-1} \left(\left(\frac{\epsilon_i}{n_i} \right)^2 + \left(\frac{\epsilon_{i+1}}{n_{i+1}} \right)^2 \right)$$

$$= \sum_i \left(\frac{\epsilon_i}{n_i} \right)^2 - \frac{1}{4} \left(\left(\frac{\epsilon_1}{n_1} \right)^2 + \left(\frac{\epsilon_\ell}{n_\ell} \right)^2 \right)$$

$$\leq \sum_i \left(\frac{\epsilon_i}{n_i} \right)^2$$

$$\leq \frac{k}{n} \sum_i \frac{\epsilon_i^2}{n_i}, \tag{4}$$

where the final step used the assumption $n_i \geq \frac{n}{k}$. Inserting this bound and Eq. (3) into Eq. (2), we obtain the desired bound of

$$Var(A) \leq \frac{n-1}{n-k} Var(U). \qquad \square$$

As we mentioned above, the theorem shows only a small multiplicative loss. Can we make it even smaller? No: The following example shows that the bound given by Theorem 4.1 is tight.

Example 4.2. Let k divide n, and let all $\ell := k$ strata have equal size $\frac{n}{k}$. Note that this setting will not lead to rounding. Furthermore, let all strata have equal $m_i =: m_0$. Under uniform sampling, the variance is $k \frac{m_0 k}{n} (1 - \frac{m_0 k}{n}) \frac{n-k}{n-1}$. Under stratification, the variance is

$$\ell \left(1 \frac{m_0}{n/k} \left(1 - \frac{m_0}{n/k} \right) \frac{n/k - 1}{n/k - 1} \right) = k \frac{m_0 k}{n} \left(1 - \frac{m_0 k}{n} \right).$$

We conclude that there are instances where Theorem 4.1 is tight.

If we ignore the $\frac{n-1}{n-k}$ factor, stratification does not increase the variance. This is reassuring, but we consider stratification in the hope that it will *reduce* variance. If we stratified well and managed to concentrate the hidden feature in some strata and not in others, we would like a guaranteed improvement in variance over uniform sampling. As it turns out, the proof of Theorem 4.1 immediately gives us such a bound, just by assuming slightly larger strata.

COROLLARY 4.3. *In the setting of Theorem 4.1, assume furthermore that every stratum has size $n_i \geq \alpha \frac{n}{k}$ for some constant $\alpha \geq 1$. Then, for any $M \subseteq N$,*

$$Var \left(A_M^{n,k} \right) \leq \frac{n-1}{n-k} Var \left(U_M^{n,k} \right) - \left(1 - \frac{1}{\alpha} \right) \frac{k}{n} \sum_i \frac{\epsilon_i^2}{n_i}.$$

PROOF. In Eq. (4) in the proof of Theorem 4.1, use $n_i \geq \alpha \frac{n}{k}$ instead of $n_i \geq \frac{n}{k}$. $\square$

This bound generalizes the previous one and, as all ϵ_i are 0 in Example 4.2, it is also tight. But it is more interesting to look at this bound in settings with rounding. In the proof of Theorem 4.1, we saw that the expectation of the variance was essentially[7]

$$\frac{n-1}{n-k} Var \left(U_M^{n,k} \right) - \frac{k}{n} \sum_i \frac{\epsilon_i^2}{n_i}.$$

The additional $\frac{1}{\alpha} \frac{k}{n} \sum_i \frac{\epsilon_i^2}{n_i}$ in the bound of Corollary 4.3 accounts for the variance of expectation, i.e., for the increase in variance caused by rounding. The following example demonstrates that this treatment is nearly tight.

---

[7]Up to factors of $\frac{n_i - \tilde{k}_i}{n_i - 1}$, which are essentially 1 for large n/k.
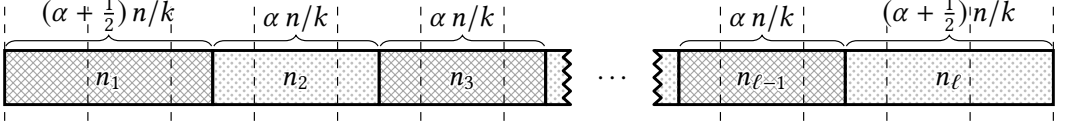
Fig. 3. Example 4.4 for $\alpha = 2$. All agents in hatched strata are in M; none of those in dotted strata are.

Example 4.4. For a given integer $\alpha \geq 1$, let $\ell := \frac{k-1}{\alpha}$ be an even integer. Set $n_1 = n_\ell = (\alpha + \frac{1}{2})\frac{n}{k}$ and all other n_i to $\alpha \frac{n}{k}$, and assume that these numbers are integers as well. Let M be exactly all strata with odd indices. This implies that $m = \frac{n}{2}$, thus $\frac{n-1}{n-k} Var(U_M^{n,k}) = \frac{k}{4}$. The strata are laid out as in Fig. 3: $\ell - 1$ of the blocks of Section 4.1 are split half-half between an all-in-M and a none-in-M stratum while all other blocks are fully in M or not in M. The variance is the same as that of a sum of $\ell - 1$ independent Bernoulli trials with probability $1/2$, which is $\frac{\ell-1}{4} = \frac{k}{4\alpha} - \frac{1+\alpha}{4\alpha}$.

Since the $\frac{\epsilon_i}{n_i}$ alternate between $\frac{1}{2}$ and $-\frac{1}{2}$, the bound of Corollary 4.3 reduces to

$$\frac{k}{4} - \left(1 - \frac{1}{\alpha}\right) \frac{k}{n} \sum_i n_i \left(\frac{\epsilon_i}{n_i}\right)^2 = \frac{k}{4} - \left(1 - \frac{1}{\alpha}\right) \frac{k}{4n} \sum_i n_i = \frac{k}{4\alpha},$$

which only leaves an additive gap of $\frac{1+\alpha}{4\alpha} \in (1/4, 1/2]$ to the actual variance of $A_M^{n,k}$. This gap is independent of the instance size, so if we fix α and let k go to infinity, the ratio between our bound and the actual variance converges to 1.

5 GENERAL SAMPLING ALGORITHMS

So far, we have compared stratified sampling to the baseline of uniform sampling. Since these are the options widely used in practice, this was a natural choice. Nevertheless, we should reassure ourselves that we do not overlook novel sampling algorithms that might surpass both in terms of variance.

In this section, we consider general algorithms $\mathcal{A}(N, k)$, not necessarily based on sampling, for some $k \geq 2$.

Every such algorithm is uniquely defined by probabilities p_C for all $C \in \binom{N}{k}$, where each p_C gives the probability of choosing the panel C.[8] In the following, C always ranges over $\binom{N}{k}$. We continue to require that the algorithm preserve expectations, which is equivalent to selecting each agent $x \in N$ with probability $\sum_{C \ni x} p_C = \frac{k}{n}$.

After fixing $\mathcal{A}$, fix a subset M of N. Since N and k are clear, we drop the superscript from $A_M^{n,k}$ and extend the notation $A_{M'} := |\mathcal{A} \cap M'|$ to other subsets M' of N. Furthermore, we set $A_x := A_{\{x\}}$ for all $x \in N$. Then, we can write the variance of A_M as

$$Var(A_M) = Var\left(\sum_{x \in M} A_x\right) = \sum_{x \in M} \underbrace{Var(A_x)}_{= \frac{k}{n}\left(1 - \frac{k}{n}\right)} + 2 \sum_{x < y \in M} \underbrace{CoVar(A_x, A_y)}_{= \mathbb{P}[A_x \wedge A_y] - \left(\frac{k}{n}\right)^2}$$

$$= |M| \frac{k}{n} - |M|^2 \left(\frac{k}{n}\right)^2 + 2 \sum_{x < y \in M} \mathbb{P}[A_x \wedge A_y]. \tag{5}$$

For specific priors on M, some complicated $\mathcal{A}$ might reduce the expected variance better than any algorithm based on stratification. In general, however, we show that no such algorithm is a

---
[8]For a set N and an integer k, $\binom{N}{k}$ denotes the set of all subsets of N with cardinality k.

clearly superior choice. For this, we show that no algorithm $\mathcal{A}$ dominates another algorithm $\mathcal{A}'$, i.e., strictly decreases the variance on some M without increasing the variance on any M; the proof is relegated to Appendix A.1.

PROPOSITION 5.1. *Let there be two algorithms $\mathcal{A}(N, k)$ and $\mathcal{A}'(N, k)$, and let their corresponding random variables be called A_M and A'_M, respectively. If $Var(A_M) < Var(A'_M)$ for some M, there is an M' such that $Var(A_{M'}) > Var(A'_{M'})$.*

Thus, no matter how we choose to stratify, we have the peace of mind that no other sampling strategy is universally preferable.

Another reason for selecting an algorithm outside of stratification might be if it had a much better worst-case guarantee. In other words, such a hypothetical algorithm would guarantee a low variance to all groups M in the population, which might be attractive in the absence of much information about M. As we show in Proposition 5.2 (whose proof is relegated to Appendix A.2), uniform sampling turns out to be optimal from this worst-case perspective. By extension, since Theorem 4.1 guarantees that stratified sampling can only be marginally worse on any M, stratified sampling must also be competitive.

PROPOSITION 5.2. *Uniform sampling minimizes $\max_{M \subseteq N} Var(A_M)$ among all algorithms.*

Interestingly, uniform sampling is not the only algorithm with this property: As shown by the following example, there are non-uniform sampling algorithms that have the same variance as uniform sampling on every $M \subseteq N$, at least for some values of n and k.

Example 5.3. Let $n := 6$ and $k := 3$. Let $\mathcal{A}(N, k)$ be the algorithm which picks one of the following ten triangles uniformly at random and returns its vertices:
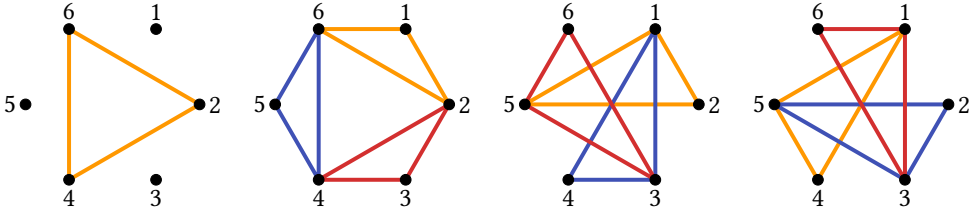
In this representation, we can see that it is obtained from the uniform mechanism by dropping every second rotation of the sets, for instance, the set $\{1, 2, 3\}$.

One can verify that every $1 \le x \le 6$ appears in the result with probability $\frac{k}{n} = \frac{1}{2}$. Furthermore, every set x, y is jointly selected with probability $\binom{k}{2} / \binom{n}{2} = \frac{1}{5}$, just as by the uniform algorithm. Thus, by Eq. (5), the variance equals the variance of the uniform algorithm for every M.

6 EXPERIMENTS

As we showed in the previous sections, there is practically no harm in stratifying to a very fine granularity. If the strata correlate well with all relevant M, the variance can be reduced considerably.

In this section, we investigate how such stratifications can be found without knowing M. Any such argument need rely on the correlation between visible features and M in a population. We explore these relationships using data from the *General Social Survey* (GSS) [20] of the years 2014 and 2016.[9] The subpopulations that we use of each year contain $n = 1714$ and $n = 1956$ agents,

---

[9]The GSS is representative of the adult population living in US households (subject to being able to complete the survey in English or Spanish). Because we do not weight participants, we cannot claim representativeness. Our sample likely underrepresents individuals living in large households and individuals who initially refuse to participate in the survey.
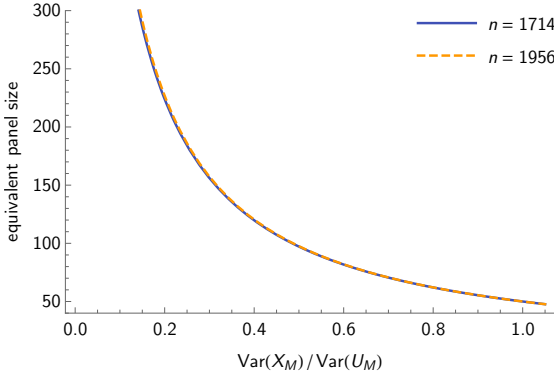
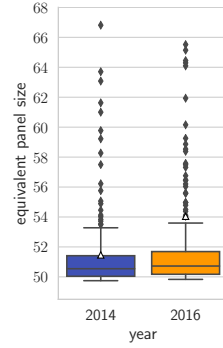Fig. 4. Conversion between ratio of variances and equivalent panel size for $k = 50$, independent of m.

Fig. 5. Equivalent panel size for 250 random stratifications per year. Omitting three outliers per year above 68. Triangles mark means.

respectively. In all of our experiments, we set $k = 50$. This is a compromise between having a reasonably large panel without the panel being too large of a fraction of the population.

Since ratios of variances are hard to interpret, we benchmark stratifications by their *equivalent panel size*. For a given stratification and group $\emptyset \subsetneq M \subsetneq N$, we can compute its normalized variance $Var(A_M^{n,k}/k)$. Its equivalent panel size k' is the panel size such that the uniform mechanism has the same normalized variance $Var(U_M^{n,k'}/k')$.[10]

Fig. 4 shows the conversion between ratios of variances and equivalent panel sizes for our choice of n and k. As shown in Appendix A.3, the equivalent panel size can be directly computed as

$$\frac{n\,k}{(n-k)\,Var(A_M^{n,k})/Var(U_M^{n,k}) + k}.$$

If a stratification has an equivalent panel size of $50 + x$ on a relevant M, the stratification saved x seats without reducing the accuracy of M's representation. Based upon the Irish Citizens' Assembly's cost of roughly €10,000 per participant,[11] high equivalent panel sizes imply a significant reduction in the cost of a citizens' panel due to stratification.

Our simulation code, along with the exact experimental setup as an IPython notebook, are available at https://github.com/pgoelz/sortition.

6.1 Random Stratification

We begin with a quantitative overview over the effect of fine-grained stratification on variance. For this, we benchmark random stratifications based on demographic features with respect to groups M induced by random other features of the GSS dataset.

We identify 14 "demographic" features, capturing age, gender, race, social class, education, region of living, general happiness, religion, party affiliation, number of children, marital state, veteran status, urban-rural divide, and whether the individual was born abroad (see Appendix D.1). Our stratifications are induced by a random permutation of these features. We traverse the features in order, in every step subdividing every stratum greedily by the current feature, subject to the

---

[10]We allow equivalent panel sizes to be fractional, interpolating the uniform variance using the formula $k\,m/n\,(1-m/n)\,(n-k)/(n-1)$.

[11]According to https://www.citizensassembly.ie/en/About-the-Citizens-Assembly/Background/Procurement-and-Costs/, the Assembly spent around €1 million in the categories "conference/catering and accommodation", "reimbursement of travel and other expenses", and "recruitment of members/facilitation and notetaking services".

constraint that no stratum be smaller than n/k. The groups M are determined by a random column other than the demographic features. We pick a random individual, and define M to either be all agents with lower or larger value in this column.[12] Since similar features are usually encoded with sequential numbers, we hope that this will lead to relatively coherent groups.

As guaranteed by Theorem 4.1, no stratification increases variance by more than a factor of $\frac{n-1}{n-k}$, which is approximately 1.03 for both years. Expressed in equivalent panel sizes, this guarantees sizes of at least 48.6 in 2014 and of 48.8 in 2016. As shown in Fig. 5, we did not observe any panel sizes close to this theoretical lower bound, with all values staying above 49.7. A large fraction of equivalent panel sizes are above 50, implying that even random stratification is more likely to be beneficial than harmful. The mean equivalent panel size is about 51.5 (2014) and 54.0 (2016), but the latter is skewed by a few large outliers. Still, a majority of stratifications give modest improvements — around one seat-equivalent — over uniform distributions.

Clearly, it is possible to increase the equivalent panel size further. It remains to be seen, however, whether targeted stratification can achieve this simultaneously for a broad set of unknown features.

6.2 Case Study: Comparison of Stratification Methods

To show this, we recreate the situation of a panel organizer. Hopefully, using insights from our theoretical analysis, a human stratifier can decrease the variance for unknown groups M, and go beyond uninformed stratification, say by the intersections of gender and race. We also experiment with automatic stratification based on k-means clustering (where the number of clusters is ℓ, not k).

One of the authors took the role of the stratifier. Until the stratification was complete, we isolated him from information about the dataset (including the analyses of the previous section). Our pre-committed experimental setup can be found in Appendix C.

We want to benchmark the different approaches on groups M that might be relevant in a political context. Thus, we identified 10 "attitude" features for each of the years and projected them into binary features (see Appendix D.2). These attitude features reflect opinions touching a wide range of issues, including social liberties, economic policy, penal law, and trust in institutions. No pair of attitude features has a higher (anti-)correlation than 37%, with most pairs being well below that. For each year, a random subset of five attitude features was made available to the stratifier and the clustering algorithm. Their stratifications should have low variance for the remaining attitude features, of which the stratifier does not even know the category name. The revealed attitudes allow the stratifier to build an intuition for the political topology and to get an impression of the kinds of groups he will have to accommodate. In practice, similar information (and likely more) is available from experience, polling, and knowledge about upcoming issues. The rationale for the manual stratification is documented in Appendix E. Important elements of the high-level approach — such as the granularity of stratification, the goal of polarizing strata, and the sequential ordering of the strata — were directly motivated by our technical analysis.

We also consider a stratification based on k-means clustering. Since the demographic and revealed features form a high-dimensional space, a clustering algorithm might be better at identifying coherent subgroups than a human. We request $\ell := 48$ clusters, the maximum number for which each stratum can still have a size of at least n/k. As documented in the experimental setup, we translate the features into real vectors in a relatively naïve way. We do not attempt to scale features differently and simply optimize the squared Euclidean distance. A more principled approach to clustering might lead to better results, but is outside of the scope of this paper. We slightly deviated

---

[12]We do not consider columns with more than 10% missing entries, agents who do not have a value in the given column, and groups M containing no or all agents.
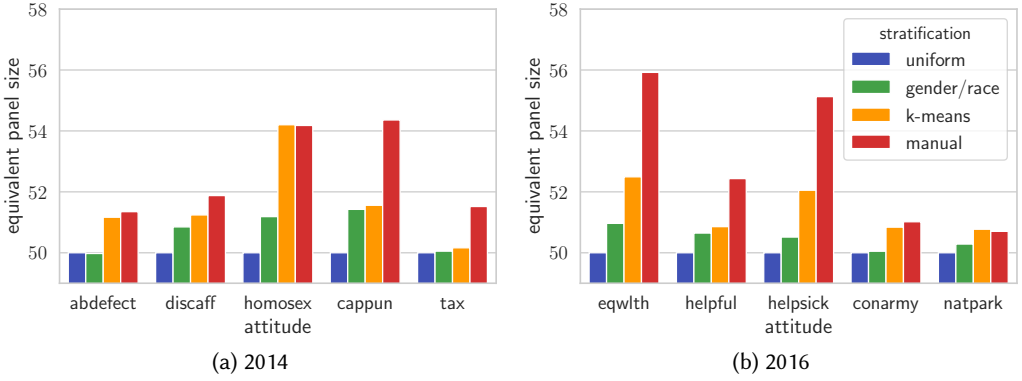
Fig. 6. Equivalent panel size for different attitude groups and different stratification mechanisms.
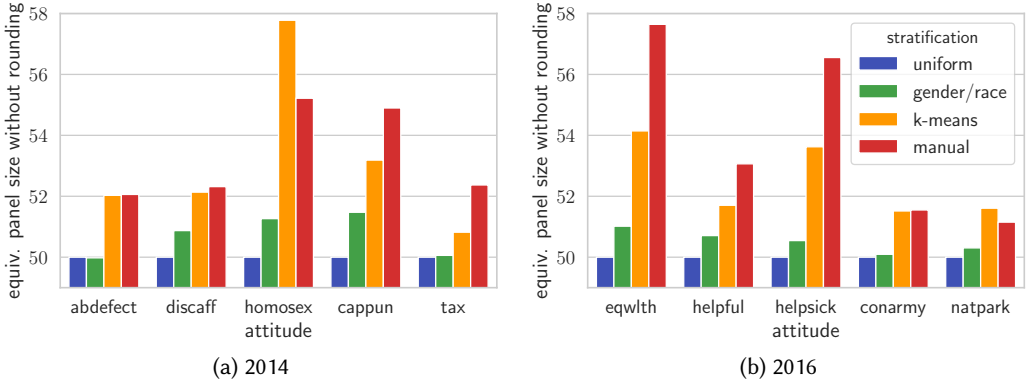

Fig. 7. Equivalent panel size for different attitude groups and different stratification mechanisms, based only on the expectation of variance, ignoring the variance of expectation.

from our experimental setup by using *constrained k*-means clustering [4] instead of *balanced k*-means clustering. While balanced clustering constrains all clusters to have size between $\lfloor n/\ell \rfloor$ and $\lceil n/\ell \rceil$, we only require the lower bound. We decided to do so after observing strongly non-contiguous clusters formed by balanced clustering even on two-dimensional toy examples.

Finally, the gender-race stratification gives one stratum to each combination of white/black/other and male/female. Stratifying by these categories can be easily implemented, and might be a stratification scheme chosen for current citizens' panels.

As displayed in Fig. 6, we find that manual stratification clearly outperforms all other stratification approaches. For all tested groups, variance of distribution decreases over the baseline of uniform sampling. In four out of the six cases, this increase corresponds to an increase in panel size of more than four seats, i.e., by 8%. Clustering performs second best and marginally beats manual stratification for two attitude groups. For several groups M however, clustering performs comparably to simple gender-race stratification, even though the manual stratification was able to make larger gains. Stratifying by race and gender leads to a slight increase in variance for one group, but still looks worthwhile given the improvements on other groups. Nevertheless, both manual or clustering-based stratification are clearly preferable when possible.
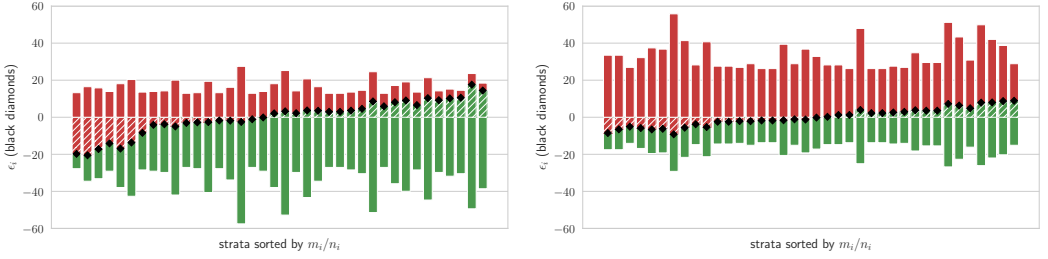
(a) M: "Government should reduce income inequality: response between in favor and indifferent" (*eqwlth*).

(b) M: "US spending too little on parks and recreation" (*natpark*).

Fig. 8. Distribution of different M over manual 2016 stratification. Overall height of bars is strata size n_i, height of green part (lower end to diamond) is m_i. Position of black diamonds corresponds to ϵ_i.

How would these results improve in absence of rounding? Depending on how much the equivalent panel size increases without rounding, it might be more or less worthwhile to optimize future stratifications in this direction. To examine this, we decompose the variance as in Eq. (2) and completely ignore the variance of expectation. As can be seen in Fig. 7, this leads to significant gains for both manual stratification and clustering. While the overall ordering of approaches remains the same, the clustering approach clearly profits more, with an average gain of about 1.3 seat-equivalents as compared to the 0.8 seat-equivalents gained by manual stratification. At first glance, this might be surprising because k-means leads to near-equal strata sizes. However, this does not imply that the strata boundaries line up nicely with the blocks since (due to divisibility issues) there can be at most 48 strata with size at least n/k. The variance of expectations is driven by terms of the shape $\rho_i (1 - \rho_i)$ and of the shape $(m_i/n_i - m_{i+1}/n_{i+1})^2$. The distribution of these terms for both approaches is displayed in Appendix B.2. The lining up of strata with blocks is reflected in the first kind of terms; neither kind of stratification has a clear edge there. Instead, the large rounding losses of k-means match its significantly higher terms of the second kind, which implies that it could profit from optimizing the order of strata. Adjacent strata should be similar because rounding between strata with similar concentrations of M adds less to the variance of expectation. Our manual stratifier explicitly tried to do this, and seems to have been successful. In Appendix B.3, we illustrate this difference for the feature *homosex*. In general, we see that rounding should be considered when stratifying, both with respect to block alignment and ordering.

We predicted that the key to low variance is that relevant M polarize as many strata as possible. We inspect the manual stratification of the 2016 dataset to see whether different levels of polarization explain the difference in success between attitudes like *eqwlth* and *natpark*. Looking at Fig. 8, we indeed see a pronounced difference. For the feature *eqwlth* (Fig. 8a), a substantial number of strata are polarized in each direction. By contrast, the feature *natpark* (Fig. 8b) appears in many strata in a concentration m_i/n_i similar to the global concentration m/n. In the continuous setting, the relative reduction in variance is given by the ratio between $\frac{k}{n} \sum_i \frac{\epsilon_i^2}{n_i}$ and $k \frac{m}{n} (1 - \frac{m}{n})$. This indicator for the polarization in both stratifications captures the superior performance on *eqwlth* (14.9%) over that on *natpark* (3.7%). Corresponding values for all stratifications and features are available in Appendix B.1.

Finally, we wanted to see whether our human stratifier can stratify better when given the question defining a group M, without having information about which individuals belong to this group. The previous experiment models the situation of a general-purpose panel whose topics of discussion are not fixed at sampling time. Revealing the question defining M could then reflect the situation of a panel convening to debate a fixed issue without specialized polling information on this issue.
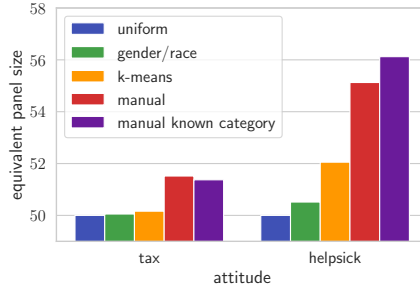
Fig. 9. Equivalent panel sizes for different stratifications, including manual stratification when knowing category description.

For each year, we revealed one random category description: *tax* in 2014 and *helpsick* in 2016. In both cases, our stratifier adapted the general-purpose stratification created earlier, and attempted to stratify on available features that seemed most relevant to the revealed category. While the specialized stratification added around one seat-equivalent for the *helpsick* group, variance slightly increased on the *tax* group. Further experimentation will be needed to see whether knowing the category consistently helps in manual stratification.

7 DISCUSSION

In this paper, we examined the effect of stratification in a sortition setting. We formulated the goal of minimizing the variance for opinion groups, proposed a low-variance rounding scheme, and characterized the variance under stratification. We have applied these contributions in a case study, and expect them to be of value to practitioners. It seems to us that there is much more potential for the fields of artificial intelligence and computational social choice to contribute to sortition, starting with the following issues:

Stratification by machine learning. While our clustering approach fell short of manual stratifications, there is ample room for improvement. A principled distance measure in the space of demographics and attitudes and a better treatment of rounding may allow automated clustering to surpass manual approaches. Beyond that, a better solution to the machine learning problem in Section 6.2 should distinguish between the demographic and revealed attitude features and leverage the different types of information they offer.

Optimal stratification and sampling. Say that our belief about M is given as an independent probability of membership for every agent. Which stratification has the lowest expected variance over M? To decrease variance, it may be useful not only to negatively correlate similar agents — as we do by placing them in the same stratum — but also to positively correlate agents at opposite ends of the opinion spectrum. Even if this is optimal, will the fairness of such an approach still be convincing? Otherwise stratification, which is already in wide use, might be the better choice.

Refusal to participate. We took it as an axiom that every agent should be selected with equal probability. But how should we react when sampled agents decline to participate? Measures like financial compensation are important for reducing the frequency of refusals, but there will always be legitimate reasons preventing an agent from participating. If panels are repeated regularly, one could allow agents to defer participation. Otherwise, one might at least replace the unavailable agent by someone maximally similar, rather than than resampling from the population at large. Both

approches lead to issues of fairness and manipulation. Can we formulate mathematical guarantees for such settings?

ACKNOWLEDGMENTS

REFERENCES

[1] H. Aziz, M. Brill, V. Conitzer, E. Elkind, R. Freeman, and T. Walsh. 2017. Justified Representation in Approval-Based Committee Voting. *Social Choice and Welfare* 48, 2 (2017), 461–485.

[2] M. L. Balinski and H. P. Young. 2010. *Fair Representation: Meeting the Ideal of One Man, One Vote.* Brookings Institution Press.

[3] T. G. Bouricius. 2013. Democracy through Multi-Body Sortition: Athenian Lessons for the Modern Day. *Journal of Public Deliberation* 9, 1 (2013), article 11.

[4] P. S. Bradley, K. P. Bennett, and A. Demiriz. 2000. Constrained K-Means Clustering. Microsoft Research Technical Report 2000-65.

[5] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice.* Cambridge University Press.

[6] M. Brill, J.-F. Laslier, and P. Skowron. 2018. Multiwinner Approval Rules as Apportionment Methods. *Journal of Theoretical Politics* 30, 3 (July 2018), 358–382.

[7] J. Burnheim. 1985. *Is Democracy Possible? The Alternative to Electoral Politics.* University of California Press.

[8] C. Chekuri, J. Vondrak, and R. Zenklusen. 2009. Dependent Randomized Rounding for Matroid Polytopes and Applications. arXiv:0909.4348.

[9] Citizens' Assembly. 2017. Final Report on the Eighth Amendment of the Constitution.

[10] T. Dalenius. 1950. The Problem of Optimum Stratification. *Scandinavian Actuarial Journal* 1950, 3-4 (1950), 203–213.

[11] G. Delannoi and O. Dowlen (Eds.). 2010. *Sortition: Theory and Practice.* Imprint Academic.

[12] J. S. Fishkin. 2009. *When the People Speak: Deliberative Democracy and Public Consultation.* Oxford University Press.

[13] G. Grimmett. 2004. Stochastic Apportionment. *The American Mathematical Monthly* 111, 4 (2004), 299–307.

[14] M. H. Hansen. 1991. *The Athenian Democracy in the Age of Demosthenes: Structure, Principles, and Ideology.* Blackwell.

[15] D. G. Horvitz and D. J. Thompson. 1952. A Generalization of Sampling Without Replacement From a Finite Universe. *J. Amer. Statist. Assoc.* 47, 260 (1952).

[16] M. Mowbray and D. Gollmann. 2007. Electing the Doge of Venice: Analysis of a 13th Century Protocol. In *Proceedings of the 20th IEEE Computer Security Foundations Symposium.* 295–310.

[17] D. C. Mueller, R. D. Tollison, and T. D. Willett. 1972. Representative Democracy via Random Selection. *Public Choice* 12, 1 (1972), 57–68.

[18] Pew Research Center. 2014. Beyond Red vs. Blue: The Political Typology.

[19] R. Saran and N. Tumennasan. 2013. Whose Opinion Counts? Implementation by Sortition. *Games and Economic Behavior* 78 (2013), 72–84.

[20] T. W. Smith, M. Davern, J. Freese, and M. Hout. 2018. General Social Surveys, 2014 and 2016.

[21] P. Stone. 2011. *The Luck of the Draw: The Role of Lotteries in Decision Making.* Oxford University Press.

[22] D. Van Reybrouck. 2016. *Against Elections: The Case for Democracy.* Random House.

[23] T. Walsh and L. Xia. 2012. Lot-Based Voting Rules. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems.* 603–610.

[24] K.-H. Wilker. 2019. *Axioms for Approval-Based Committee Voting Rules.* Master's thesis.

A OMITTED PROOFS

A.1 Proof of Proposition 5.1

It holds that

$$\sum_{x<y} \mathbb{P}[A_x \wedge A_y] = \mathbb{E}\left[\sum_{x<y} A_x A_y\right] = \sum_{x<y} A_x A_y = \binom{k}{2} = \sum_{x<y} \mathbb{P}[A'_x \wedge A'_y]. \tag{6}$$

By assumption, $Var(A_M) < Var(A'_M)$, and thus, by Eq. (5),

$$\sum_{x<y\in M} \mathbb{P}[A_x \wedge A_y] < \sum_{x<y\in M} \mathbb{P}[A'_x \wedge A'_y].$$

Combined with Eq. (6), this shows that

$$\sum_{\substack{x<y \\ \{x,y\}\not\subseteq M}} \mathbb{P}[A_x \wedge A_y] > \sum_{\substack{x<y \\ \{x,y\}\not\subseteq M}} \mathbb{P}[A'_x \wedge A'_y].$$

In particular, there must be some $M' = \{x, y\}$ such that $\mathbb{P}[A_x \wedge A_y] > \mathbb{P}[A'_x \wedge A'_y]$. By Eq. (5), $Var(A_{M'}) > Var(A'_{M'})$. □

A.2 Proof of Proposition 5.2

Let $M^\star$ be a set maximizing $Var(U_M)$, and denote $m^\star := |M^\star|$. We claim that there is $M' \subseteq N$ such that $|M'| = m^\star$ and $Var(A_{M'}) \geq Var(U_{M'})$. Indeed, it holds that

$$\sum_{M\in\binom{N}{m^\star}} \sum_{x<y\in M} \mathbb{P}[A_x \wedge A_y] = \sum_{x<y\in N} \underbrace{\left|\left\{M \in \binom{N}{m^\star}\,\middle|\, \{x, y\} \subseteq M\right\}\right|}_{=\binom{n-2}{m^\star-2}} \cdot \mathbb{P}[A_x \wedge A_y]$$

$$= \binom{n-2}{m^\star - 2} \sum_{x<y\in N} \mathbb{P}[A_x \wedge A_y]$$

$$= \binom{n-2}{m^\star - 2}\binom{k}{2}.$$

Thus, there is an M' of size $m^\star$ such that

$$\sum_{x<y\in M'} \mathbb{P}[A_x \wedge A_y] \geq \frac{\binom{n-2}{m^\star-2}\binom{k}{2}}{\binom{n}{m^\star}} = \binom{m^\star}{2}\frac{\binom{k}{2}}{\binom{n}{2}} = \sum_{x<y\in M'} \mathbb{P}[U_x \wedge U_y],$$

where U_x is the equivalent of A_x for uniform sampling. Thus, by Eq. (5), we have that $Var(A_{M'}) \geq Var(U_{M'})$.

To complete the proof, note that, by symmetry, $Var(U_{M^\star}) = Var(U_{M'})$, as $M^\star$ and M' have equal cardinality. It follows that $Var(A_{M'}) \geq Var(U_{M^\star})$. □

A.3 Derivation of Formula for Equivalent Panel Size

For convenience, let $\gamma := \frac{Var(A_M^{n,k})}{Var(U_M^{n,k})}$. Then,

$$Var(U_M^{n,k'}/k') = Var(A_M^{n,k}/k)$$

$$\text{iff} \qquad \frac{Var(U_M^{n,k'})}{k'^2} = \frac{Var(A_M^{n,k})}{k^2}$$

$$\text{iff} \qquad \frac{Var(U_M^{n,k'})}{k'^2} = \frac{\gamma \, Var(U_M^{n,k})}{k^2}$$

$$\text{iff} \qquad \frac{k' \, \frac{m}{n} \left(1 - \frac{m}{n}\right) \frac{n-k'}{n-1}}{k'^2} = \frac{\gamma \, k \, \frac{m}{n} \left(1 - \frac{m}{n}\right) \frac{n-k}{n-1}}{k^2}$$

$$\text{iff} \qquad \frac{n - k'}{k'} = \frac{\gamma \, (n - k)}{k}$$

$$\text{iff} \qquad \frac{n}{k'} = \frac{\gamma \, (n - k) + k}{k}$$

$$\text{iff} \qquad k' = \frac{n \, k}{\gamma \, (n - k) + k} \qquad \square$$

B ADDITIONAL FIGURES

B.1 Strata Polarization

(a) 2014

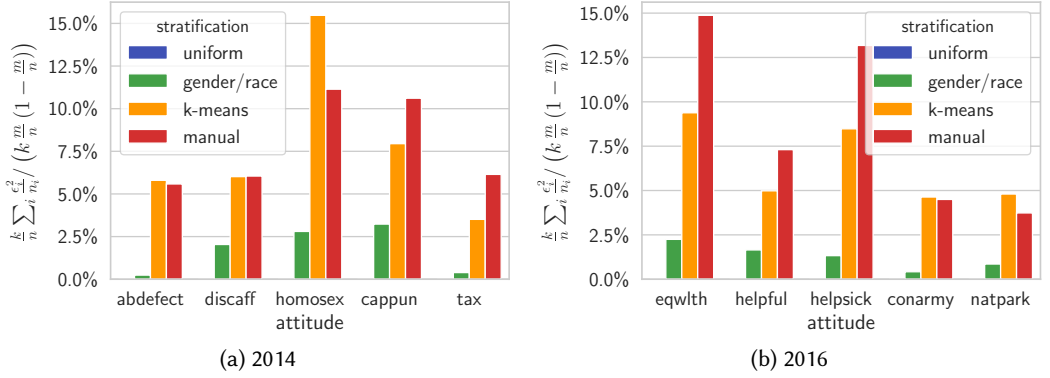(b) 2016

Fig. 10. Relative reduction in variance that each stratification would have led to in a continuous setting. The observed reductions closely match the increases in panel size when ignoring rounding variance (see Fig. 7).
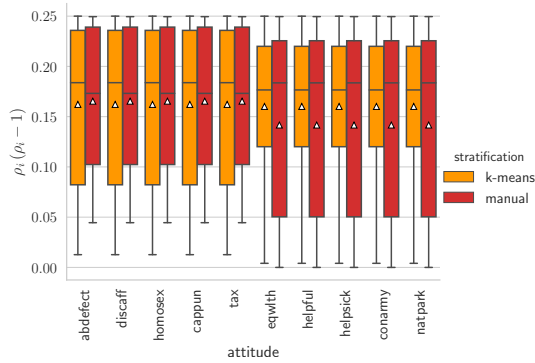
B.2 Sources of Rounding Losses

Fig. 11. Distribution of alignment variance $\rho_i \, (1 - \rho_i)$ between pairs of adjacent strata. Means marked by triangles.
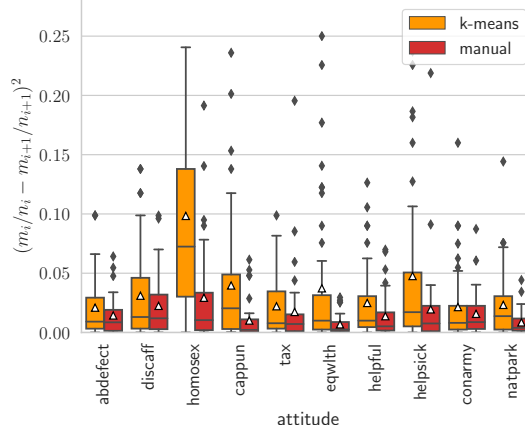
Fig. 12. Distribution of squared difference of concentration between pairs of adjacent strata. Means marked by triangles. Three positive outliers omitted.

B.3 Stratification in Order for Attitude *homosex*

(a) manual stratification
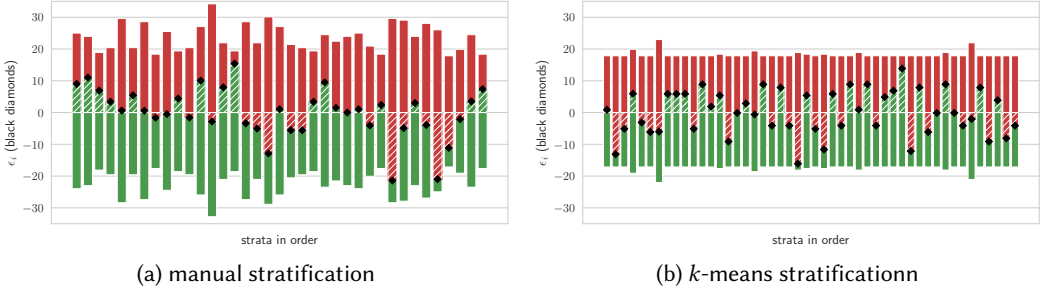
(b) *k*-means stratificationn

Fig. 13. Illustration of the manual and *k*-means stratification for the feature *homosex*. Overall height of bars is strata size n_i, height of green part (lower end to diamond) is m_i. Position of black diamonds corresponds to ϵ_i. In contrast to Fig. 8, strata are in order of the rounding layout.

The clustering stratification has more strata and leads to significantly higher polarization on *homosex* (see Fig. 10). However, adjacent concentrations fluctuate wildly (see also Fig. 12), leading to excessive rounding losses. As a result, the actual reduction in variance is virtually identical to that of the manual stratification (see Fig. 6).

C EXPERIMENTAL SETUP

We use data from the General Social Survey (GSS). Two of the authors take the role of experimenter A and B.

Experimenter A chooses

- for each of the two years of 2014 and 2016, two of the three *ballots*,[13]
- a set of "demographic" features, which may also include non-attitude variables such as happiness or health status, and

---

[13]GSS partitions their data into three random subsets of approximately equal size, the ballots. As many questions are only presented to two ballots, we need to restrict the dataset.

- a set of 10 "attitude" variables for each of the two years, which reflect attitudes and opinions. The attitude variables chosen for both years are disjoint. Not all of these variables must directly be political, but should have relevance to political questions. Variables should capture attitudes towards different areas of public life and should not be obviously related. All variables are projected into binary variables by selecting a subset of values that are mapped to 1. By convention, answers of the shape "don't know", "not applicable", or "no answer" belong to value 0.

All identified variables should be available for a large fraction of agents in the respective ballots of the year. Agents lacking demographic information might be excluded from the dataset. Agents are not weighted, which means that the sample is not representative of the US population.

Experimenter B has no detailed prior knowledge of the GSS dataset (including potential categories for attitude variables). B receives both year's datasets (without the third ballot that was not chosen), with columns for agent identity, for all demographic features, and for a subset of the attitude features to be determined as follows: For both years, a randomly selected subset of 5 features is revealed. For both datasets, B designs a stratification, i.e., a partition of the agents, with the goal of reducing the variance on the non-revealed attitude features (of which B neither knows the identity nor membership). The number of seats is fixed at $k = 50$; the number and size of strata can be set at B's discretion. Note that the stratification may explicitly depend on the revealed attitude features. For both years, after committing to a strafication, B receives the description of one random non-revealed attitude category (but no information on the responses themselves). B proposes a specialized stratification, which is only supposed to minimize the variance for the newly-revealed category. Only after committing to all stratifications for both years may B learn about the remaining attitude features and their variances, or get more information about the GSS dataset.

We compare the variance for the non-revealed variables of his stratification with the baseline of uniform sampling and with stratification by all combinations of gender and race (all combinations with less than n/k agents are combined into a single stratum).

As a final point of comparison, we cluster using the balanced K-means algorithm. Clustering receives demographic and revealed attitude features. Binary features are encoded as 0-1, one-dimensional features are evenly spaced between 0 and 1, with "don't know", "not applicable", and "no answer" entries at 0.5, and other features are encoded as a sequence of binary variables with the previously mentioned answers represented as 0 everywhere.

D DESCRIPTION OF FEATURES FOR CASE STUDY

This section lists the demographic and attitude features chosen for the experiment in Section 6.2. The descriptions of the features are taken, and in a few cases slightly adapted, from the GSS codebook and `https://gssdataexplorer.norc.org/`.

B initially received the description in Appendix D.1 and the descriptions in Appendix D.2 marked as "(revealed)". The dataset included the complete GSS columns of the demographic features and the binarized columns for the revealed attitude features. In the last stage, B received the two descriptions marked "(hidden, revealed description in second stage)" and their corresponding age, but did not receive further column data.

D.1 Demographic Features

```
age:
18, ..., 88: exact age
89: includes all older ages
```

sex:
1: male
2: female

race: self-categorization
1: white
2: black
3: other

class: self-categorization
1: lower class
2: working class
3: middle class
4: upper class

degree: highest degree
0: less than high school
1: high school
2: junior college
3: bachelor's
4: graduate

region: region of interview
1: New England (Maine, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode
   Island)
2: Middle Atlantic (New York, New Jersey, Pennsylvania)
3: East North Central (Wisconsin, Illinois, Indiana, Michigan, Ohio)
4: West North Central (Minnesota, Iowa, Missouri, North Dakota, South Dakota,
   Nebraska, Kansas)
5: South Atlantic (Delaware, Maryland, West Virginia, Virginia, North Carolina,
   South Carolina, Georgia, Florida, District of Columbia)
6: East South Central (Kentucky, Tennessee, Alabama, Mississippi)
7: West South Central (Arkansas, Oklahoma, Louisiana, Texas)
8: Mountain (Montana, Idaho, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico)
9: Pacific (Washington, Oregon, California, Alaska, Hawaii)

happy: Taken all together, how would you say things are these days--would you say
that you are very happy, pretty happy, or not too happy?
1: very happy
2: pretty happy
3: not too happy

relig: What is your religious preference? Is it Protestant, Catholic, Jewish, some
other religion, or no religion?
1: protestant
2: catholic
3: jewish
4: none

5: other
6: buddhism
7: hinduism
8: other eastern religions
9: islam
10: orthodox-christian
11: christian
12: native American
13: inter-nondenomenational

partyid: Generally speaking, do you usually think of yourself as a Republican,
Democrat, Independent, or what?
0: strong democrat
1: not very strong democrat
2: independent, close to democrat
3: independent (neither, no response)
4: independent, close to republican
5: not very strong republican
6: strong republican
7: other party, refused to say

childs: How many children have you ever had? Please count all that were born alive
at any time (including any you had from a previous marriage).
0, ..., 7: exact number
8: eight or more

marital: Are you currently--married, widowed, divorced, separated, or have you
never been married?
1: married
2: widowed
3: divorced
4: separated
5: never married

vetyears: Have you ever been on active duty for military training or service for two
consecutive months or more? IF YES: What was your total time on active duty?
0: no active duty
1: less than 2 years
2: 2-4 years
3: more than 4 years
4: some, don't know how long

srcbelt:
1: Central city of 12 largest SMSAs (standard metropolitan statistical area)
2: Central city of the remainder of the 100 largest SMSAs
3: Suburbs of 12 largest SMSAs
4: Suburbs of the remaining 100 largest SMSAs
5: Other Urban (counties having towns of 10,000 or more)

6: Other rural (counties having no towns of 10,000 or more)

born: Were you born in this country?
1: yes
2: no

D.2   Attitude Features

D.2.1   2014.

gunlaw_bin (revealed)
Would you favor or oppose a law which would require a person to obtain a police permit before he or she could buy a gun?
1: favor

racopen_bin (revealed)
Suppose there is a community-wide vote on the general housing issue. There are two possible laws to vote on (READ CATEGORIES A & B). Which law would you vote for?
A. One law says that a homeowner can decide for himself whom to sell his house to, even if he prefers not to sell to (Blacks/African-Americans)(Whites (if Respondent is Black))
B. The second law says that a homeowner cannot refuse to sell to someone because of their race or color
1: can't discriminate

getahead_bin (revealed)
Some people say that people get ahead by their own hard work; others say that lucky breaks or help from other people are more important. Which do you think is most important?
1: luck or help / both equally

colcom_bin (revealed)
Now, I should like to ask you some questions about a man who admits he is a Communist. Suppose he is teaching in a college. Should he be fired, or not?
1: not fired

libmslm_bin (revealed)
Now consider a Muslim clergyman who preaches hatred of the United States. If some people in your community suggested that a book he wrote which preaches hatred of the United States should be taken out of your public library, would you favor removing this book, or not?
1: not remove

abdefect (hidden)
Please tell me whether or not you think it should be possible for a pregnant woman to obtain a legal abortion if...
A. If there is a strong chance of serious defect in the baby?
1: no

discaff (hidden)
What do you think the chances are these days that a white person won't get a job or promotion while an equally or less qualified black person gets one instead? Is this very likely, somewhat likely, or not very likely to happen these days?
1: very / somewhat likely.

homosex (hidden)
What about sexual relations between two adults of the same sex--do you think it is always wrong, almost always wrong, wrong only sometimes, or not wrong at all?
1: (almost) always wrong, sometimes wrong

cappun (hidden)
Do you favor or oppose the death penalty for persons convicted of murder?
1: oppose

tax (hidden, revealed description in second stage)
Do you consider the amount of federal income tax which you have to pay as too high, about right, or too low?
1: too high

D.2.2   2016.

polhitok_bin (revealed)
Are there any situations you can imagine in which you would approve of a policeman striking an adult male citizen?
1: no

contv_bin (revealed)
I am going to name some institutions in this country. As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? TV.
1: great deal/only some

conmedic_bin (revealed)
I am going to name some institutions in this country. As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? Medicine.
1: only some (neither "a great deal" nor "hardly any")

helpblk_bin (revealed)
Some people think that (Blacks/African-Americans) have been discriminated against for so long that the government has a special obligation to help improve their living standards. Others believe that the government should not be giving special treatment to (Blacks/African-Americans).
A. Where would you place yourself on this scale, or haven't you made up your mind on this?
1: gov helps blacks, more convinced than "agree with both"

pornlaw_bin (revealed)

Which of these statements comes closest to your feelings about pornography laws?
A. There should be laws against the distribution of pornography whatever the age.
B. There should be laws against the distribution of pornography to persons under 18.
C. There should be no laws forbidding the distribution of pornography.
1: A

eqwlth (hidden)
Some people think that the government in Washington ought to reduce the income
differences between the rich and the poor, perhaps by raising the taxes of wealthy
families or by giving income assistance to the poor. Others think that the
government should not concern itself with reducing this income difference between
the rich and the poor. Here is a card with a scale from 1 to 7. Think of a score of
1 as meaning that the government ought to reduce the income differences between rich
and poor, and a score of 7 meaning that the government should not concern itself
with reducing income differences. What score between 1 and 7 comes closest to the
way you feel?:
1: between "favor reducing income inequality" (1) and "split between both" (4)

helpful (hidden)
Would you say that most of the time people try to be helpful, or that they are
mostly just looking out for themselves?
1: helpful

helpsick (hidden, revealed description in second stage)
In general, some people think that it is the responsibility of the government in
Washington to see to it that people have help in paying for doctors and hospital
bills. Others think that these matters are not the responsibility of the federal
government and that people should take care of these things themselves. A. Where
would you place yourself on this scale, or haven't you made up your mind on this?
1: between "people help selves" and "agree with both"

conarmy (hidden)
I am going to name some institutions in this country. As far as the people running
these institutions are concerned, would you say you have a great deal of confidence,
only some confidence, or hardly any confidence at all in them?
M. Military
1: only some / hardly any

natpark (hidden)
We are faced with many problems in this country, none of which can be solved easily
or inexpensively. I'm going to name some of these problems, and for each one I'd
like you to name some of these problems, and for each one I'd like you to tell me
whether you think we're spending too much money on it, too little money, or about
the right amount. [...] Are we spending too much, too little, or about the right
amount on (ITEM)?
O. Parks and recreation
1: too little

E PROTOCOL FOR MANUAL STRATIFICATION

The stratification by hand was guided by the observation that the ideal outcome (in terms of variance) would be k equally sized strata each consisting either entirely of people with the feature M, or containing no such people. We will attempt to create small strata with either very high or very low concentrations of people with the hidden feature to benefit from the reduction in variance caused by large absolute values for the ϵ_i's promised by Eq. (4) and Corollary 4.3.

There is some tension between having small strata with large $\epsilon_i's$ to minimize the variance due to Eq. (4) and minimizing the part of the variance due to rounding, which can be minimized by having fewer, larger strata and imposing $n_i \geq \alpha n/k$. In the absence of rounding, we see from Eq. (3) that the variance of a large stratum i has the same for as for uniform sampling (up to a multiplicative factor), and that further substratification can lead to gains of the form $\frac{k_i}{n_i} \sum_j \frac{\epsilon_j^2}{n_j}$, where j ranges over the substrata of stratum i. For this reason, we generally favor small strata and attempt to mitigate the variance introduced by rounding by labeling the strata in such a way that adjacent strata are as similar as possible.

The strategy will generally be to group based on demographics, followed by the revealed preferences. Since it is unknown which of the revealed preferences most distinguishes M, responses to a variety of these questions will be used throughout the stratification.

E.1 2014

This data set contains 1 714 responses, so every stratum should have at least 35 people in order to satisfy the $n_i \geq n/k$ assumption underlying our theory.

The first step is to find which demographic features are predictive of the revealed preferences. The features that correlate most with each of the revealed preferences are given below, along with the single feature used to create a decision stump classifier predicting the value of the revealed preference (indicated with an asterisk).[14]

- libmslm : colcom* (0.37), degree (0.29), race (-0.13), class (0.12)
- colcom: libmslm* (0.37) , degree (0.27), age (-0.13). Notably uncorrelated with gunlaw, racopen
- getahead: libmslm* (0.11), sex (-0.09), degree (0.08), colcom (0.07)
- gunlaw: partyid* (-0.25), srcbelt (-0.15) , race (0.13), sex (0.11), born (.11)
- racopen: partyid* (-0.17), srcbelt (-0.1) , gunlaw (0.09), race (0.09)

We observe that partyid, race, srcbelt, degree, class, age and sex are the demographic features that most predict your revealed preferences. Interestingly, we also detect a trend among the issues asked about: For the issues that are often politicized, like gun laws (gunlaw) and race (racopen), the political party you identify with (partyid) and where you live (srcbelt) play the most important role in determining preferences. The other three issues are broader, less polarized and not as often discussed in the media, and we find that people's opinions are determined by education level, age and sex. If M were known, the extent to which it is a political issue will determine which of these demographic features to focus on. In the case of an unknown M, we will stratify based on both types of features whenever possible.

To determine which demographic features to use in the first two levels of the stratification, we randomly select a revealed preference (gunlaw), and train an entropy-minimizing tree classifier with maximum depth 2 to predict the reported preferences. Based on this, we first stratify based on race and partyid when possible.

---

[14]For this exploration, as for the others in the manual stratification section, WEKA 3.8, available from the University of Waikato, was used.

The remainder of the stratification was done by hand, while trying to balance the types of demographic features and revealed preferences used. To satisfy the constraint on stratum sizes, 4 people were moved from stratum (15) to (14), and 3 from (31) to (30). Finally, as mentioned before, in an attempt to minimize the variance that will come from using block rounding, we label the strata in such a way that adjacent strata are adjacent leaves of the decision tree and therefore (hopefully) as similar as possible. The resulting stratification rules may be seen below.

(1) Race: white, Born: non-US, Degree: no more than high school (0,1), Colcom: 1

(2) Race: white, Born: non-US, Degree: no more than high school (0,1), Colcom: 0

(3) Race: white, Born: non-US, Degree: other

(4) Race: white, Born: US, Partyid: dem (0,1), Degree: high school, Class: working (1,2,3), colcom: 1

(5) Race: white, Born: US, Partyid: dem (0,1), Degree: high school, Class: working (1,2,3), colcom: 0

(6) Race: white, Born: US, Partyid: dem (0,1), Degree: high school, Class: upper (4)

(7) Race: white, Born: US, Partyid: dem (0,1), Degree: post-hs, Gender: male

(8) Race: white, Born: US, Partyid: dem (0,1), Degree: post-hs, Gender: female

(9) Race: white, Born: US, Partyid: dem (0,1), Degree: graduate

(10) Race: white, Born: US, Partyid: rep (5,6), Degree: max hs (0,1), Srcbelt: city (1,2,3)

(11) Race: white, Born: US, Partyid: rep (5,6), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Racopen: 1, Colcom: 1

(12) Race: white, Born: US, Partyid: rep (5,6), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Racopen: 1, Colcom: 0

(13) Race: white, Born: US, Partyid: rep (5,6), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Racopen: 0

(14) Race: white, Born: US, Partyid: rep (5,6), Degree: post-hs (2,3,4), Racopen: 1, Gunlaw: 1

(15) Race: white, Born: US, Partyid: rep (5,6), Degree: post-hs (2,3,4), Racopen: 1, Gunlaw: 0

(16) Race: white, Born: US, Partyid: rep (5,6), Degree: post-hs (2,3,4), Racopen: 0

(17) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: city (1,2,3), Gunlaw: 1

(18) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: city (1,2,3), Gunlaw: 0

(19) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Gunlaw: 1, Colcom: 1

(20) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Gunlaw: 1, Colcom: 0, Racopen: 1

(21) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Gunlaw: 1, Colcom: 0, Racopen: 0

(22) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Gunlaw: 0, Libmslm: 1

(23) Race: white, Born: US, Partyid: ind (2-4,7), Degree: max hs (0,1), Srcbelt: rural (4,5,6), Gunlaw: 0, Libmslm: 0

(24) Race: white, Born: US, Partyid: ind (2-4,7), Degree: post-hs (2,3,4), Srcbelt: city (1,2,3)

(25) Race: white, Born: US, Partyid: ind (2-4,7), Degree: post-hs (2,3,4), Srcbelt: rural (4,5,6), Gunlaw: 1, Age: 0-49

(26) Race: white, Born: US, Partyid: ind (2-4,7), Degree: post-hs (2,3,4), Srcbelt: rural (4,5,6), Gunlaw: 1, Age: 50+

(27) Race: white, Born: US, Partyid: ind (2-4,7), Degree: post-hs (2,3,4), Srcbelt: rural (4,5,6), Gunlaw: 0

(28) Race: black, Partyid: dem (0,1), Degree: hs (1), Gunlaw: 1
(29) Race: black, Partyid: dem (0,1), Degree: hs (1), Gunlaw: 0
(30) Race: black, Partyid: dem (0,1), Degree: other (1, 2-4), Srcbelt: city (1,2,3)
(31) Race: black, Partyid: dem (0,1), Degree: other (1, 2-4), Srcbelt: rural (4,5,6)
(32) Race: black, Partyid: other (2-7), Srcbelt: city (1,2,3)
(33) Race: black, Partyid: other (2-7), Srcbelt: rural (4,5,6)
(34) Race: other, Partyid: dem (0,1)
(35) Race: other, Partyid: other (2-7), Getahead: 1
(36) Race: other, Partyid: other (2-7), Getahead: 0, Colcom: 1
(37) Race: other, Partyid: other (2-7), Getahead: 0, Colcom: 0

E.1.1 Revealed M. The hidden feature for 2014 was revealed to be responses to the question 'Do you consider the amount of federal income tax which you have to pay as too high, about right, or too low?'

Knowing M may allow us to come up with a better stratification. In this case, the hypothesis is that taxes is a rather political issue, and we expect preferences to behave similar to what they do for other political topics, as encoded in the features gunlaw and racopen. We therefore emphasize partyd, srcbelt and race in the stratification, additionally, it may be the case that class also plays a role in determining this preference. Education level (degree) was more important for the non-political issues, so it will be de-emphasized compared to the classification for an unknown M.

The resulting stratification can be seen below, 1 person was manually moved from stratum (16) to (17) to satisfy size constraints.

(1) Race: white, Born: non-US, Partyid: dem (0,1)
(2) Race: white, Born: non-US, Partyid: other (2-7), Class: low (1,2)
(3) Race: white, Born: non-US, Partyid: other (2-7), Class: upper (3,4)
(4) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: city (1,2,3), Class: lower (1,2)
(5) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: city (1,2,3), Class: upper (3,4), Getahead: 1
(6) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: city (1,2,3), Class: upper (3,4), Getahead: 0
(7) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Class: lower (1,2), Colcom: 1
(8) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Class: lower (1,2), Colcom: 0
(9) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Class: lower (1,2), Class: upper (3,4), Libmslm: 0
(10) Race: white, Born: US, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Class: lower (1,2), Class: upper (3,4), Libmslm: 1
(11) Race: white, Born: US, Partyid: rep (5,6), Srcbelt: city (1,2,3)
(12) Race: white, Born: US, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Class: lower (1,2), Gunlaw: 1
(13) Race: white, Born: US, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Class: lower (1,2), Gunlaw: 0
(14) Race: white, Born: US, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Class: upper (3,4), Gunlaw: 1
(15) Race: white, Born: US, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Class: upper (3,4), Gunlaw: 0
(16) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: city (1,2,3), Class: lower (1,2), Libmslsm: 1
(17) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: city (1,2,3), Class: lower (1,2), Libmslsm: 0
(18) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: city (1,2,3), Class: upper (3,4), Degree: max hs (0,1)
(19) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: city (1,2,3), Class: upper (3,4), Degree: post hs (2,3)
(20) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Class: lower (1,2), Gunlaw: 1, Colcom: 1

(21) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Class: lower (1,2), Gunlaw: 1, Colcom: 0

(22) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Class: lower (1,2), Gunlaw: 0

(23) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Class: upper (3,4), Gunlaw: 1, Colcom: 1

(24) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Class: upper (3,4), Gunlaw: 1, Colcom: 0

(25) Race: white, Born: US, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Class: upper (3,4), Gunlaw: 0

(26) Race: black, Partyid: dem (0,1), Colcom: 1, Srcbelt: city (1,2,3)

(27) Race: black, Partyid: dem (0,1), Colcom: 1, Srcbelt: rural (4,5,6)

(28) Race: black, Partyid: dem (0,1), Colcom: 0, Srcbelt: city (1,2,3)

(29) Race: black, Partyid: dem (0,1), Colcom: 0, Srcbelt: rural (4,5,6)

(30) Race: black, Partyid: other (2-7), Srcbelt: city (1,2,3)

(31) Race: black, Partyid: other (2-7), Srcbelt: rural (4,5,6)

(32) Race: other, Partyid: dem (0,1)

(33) Race: other, Partyid: other (2-7), Getahead: 1

(34) Race: other, Partyid: other (2-7), Getahead: 0, Gunlaw: 1

(35) Race: other, Partyid: other (2-7), Getahead: 0, Gunlaw: 0

E.2 2016

This dataset contains 1 956 responses, implying that every stratum should have at least 40 people in order to satisfy the $n_i \geq n/k$ assumption underlying our theory.

We follow the same methodology as for 2014, and find the following demographic features correlate most strongly with the revealed preferences (an asterisk again denotes the feature deemed most useful by a decision stump classifier).

- pornlaw: age (0.21), sex (0.17), relig* (-0.16), marital (-0.16)
- helpblk: partyid* (-0.32), marital (0.13), polhitok (0.12), degree (0.11), srcbelt (0.11)
- conmedic: pornlaw (0.09), marital* (-0.09)
- contv: conmedic (-0.08), polhitok (0.07) , partyid (-0.07). age*
- polhitok: race* (0.23), degree (-0.23), born (0.16), partyid (-0.15)

We observe that people's confidence in television media and the medical system is generally pretty weakly correlated with their demographic features. This suggests that these topics give us a new axis along which to stratify that is not captured by the demographic features, however, it also means that for M highly correlated with demographic features these features may not be useful. It is noteworthy that although these features are indicative of a general outlook on life, they are uncorrelated with happiness.

We will stratify according to sex, race, age, degree, relig, marital, partyid, srcbelt and born. Based on the output of a tree-classifier trained to predict responses to polhitok, we start by stratifying based on degree and race. Since these features, together with partyid, do not predict responses to conmedic and pornlaw, we focus on the responses to these two revealed preferences for the remainder of the stratification. The final stratification rules may be seen below. As before, it was necessary to manually balance strata sizes by moving 2 random people from stratum (3) to (4), 2 from (18) to (19) and 3 from (33) to (34).

(1) Degree: no high school, Race: white , Pornlaw: 1

(2) Degree: no high school, Race: white , Pornlaw: 0

(3) Degree: no high school, Race: other , Partyid: democrat (0,1)

(4) Degree: no high school, Race: other , Partyid: other

  (5) Degree: high school, Race: white, Partyid: dem (0,1), Conmedic: 1, Age: 0-47
  (6) Degree: high school, Race: white, Partyid: dem (0,1), Conmedic: 1, Age: 48+
  (7) Degree: high school, Race: white, Partyid: dem (0,1), Conmedic: 0, Age: 0-49
  (8) Degree: high school, Race: white, Partyid: dem (0,1), Conmedic: 0, Age: 49+
  (9) Degree: high school, Race: white, Partyid: rep (5,6), Conmedic: 1, Age: 0-49
(10) Degree: high school, Race: white, Partyid: rep (5,6), Conmedic: 1, Age: 50+
(11) Degree: high school, Race: white, Partyid: rep (5,6), Conmedic: 0, Age: 0-52
(12) Degree: high school, Race: white , Partyid: rep (5,6), Conmedic: 0, Age: 52+
(13) Degree: high school, Race: white, Partyid: ind (2-4,7), Conmedic: 1, Age: 0-49, Gender: male
(14) Degree: high school, Race: white, Partyid: ind (2-4,7), Conmedic: 1, Age: 0-49, Gender: female
(15) Degree: high school, Race: white, Partyid: ind (2-4,7), Conmedic: 1, Age: 50+, Gender: male
(16) Degree: high school, Race: white, Partyid: ind (2-4,7), Conmedic: 1, Age: 50+, Gender: female
(17) Degree: high school, Race: white, Partyid: ind (2-4,7), Conmedic: 0, Age: 0-49
(18) Degree: high school, Race: white , Partyid: ind (2-4,7), Conmedic: 0, Age: 50+, Gender: male
(19) Degree: high school, Race: white , Partyid: ind (2-4,7), Conmedic: 0, Age: 50+, Gender: female
(20) Degree: high school, Race: black, Partyid: dem (0,1), Conmedic: 1
(21) Degree: high school, Race: black, Partyid: dem (0,1), Conmedic: 0
(22) Degree: high school, Race: black, Partyid: 0ther (2-7), Conmedic: 1
(23) Degree: high school, Race: black, Partyid: other (2-7), Conmedic: 0
(24) Degree: high school, Race: other
(25) Degree: post high school, Race: white, Partyid: dem (0,1), Conmedic: 1
(26) Degree: post high school, Race: white, Partyid: dem (0,1), Conmedic: 0
(27) Degree: post high school, Race: white, Partyid: rep (5,6), Conmedic: 1
(28) Degree: post high school, Race: white, Partyid: rep (5,6), Conmedic: 0
(29) Degree: post high school, Race: white, Partyid: ind (2-4,7), Age:0-30
(30) Degree: post high school, Race: white, Partyid: ind (2-4,7), Age: 31-61, Srcbelt: city (1,2,3)
(31) Degree: post high school, Race: white, Partyid: ind (2-4,7), Age: 31-61, Srcbelt: rural (4,5,6)
(32) Degree: post high school, Race: white, Partyid: ind (2-4,7), Age: 62+
(33) Degree: post high school, Race: black
(34) Degree: post high school, Race: other
(35) Degree: graduate, Race: white, Partyid: dem (0,1)
(36) Degree: graduate, Race: white, Partyid: rep (5,6)
(37) Degree: graduate, Race: white, Partyid: ind (2-4,7)
(38) Degree: graduate, Race: other

E.2.1 Revealed M. The revealed issue for 2016 is the following question: 'In general, some people think that it is the responsibility of the government in Washington to see to it that people have help in paying for doctors and hospital bills. Others think that these matters are not the responsibility of the federal government and that people should take care of these things themselves. Where would you place yourself on this scale, or haven't you made up your mind on this?'

    Of the available revealed preferences, conmedic also deals with the health care system, and may provide insight into this topic. The previous stratification placed a pretty large emphasis on conmedic, we will retain this in the new stratification. This topic also has to do with people's opinions on welfare systems, so responses to blkhelp may be relevant. The most significant demographic indicators for blkhelp and conmedic are partyid, marital, degree and srcbelt. Beyond that, the creation of a global health care system is a pretty political issue, so we suspect that divisions along typical political lines (race, srcbelt, partyid) would be helpful.

Overall this stratification turned out very similar to the previous one and may be seen below. To satisfy size constraints, 3 people were manually moved from stratum (22) to (23) and 5 from stratum (28) to (29).

(1) Race: white, Partyid: dem (0,1), Srcbelt: city (1, 2,3), Conmedic: 1

(2) Race: white, Partyid: dem (0,1), Srcbelt: city (1, 2,3), Conmedic: 0

(3) Race: white, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Conmedic: 1, Helpblk: 1

(4) Race: white, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Conmedic: 1, Helpblk: 0

(5) Race: white, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Conmedic: 0, Helpblk: 1

(6) Race: white, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Conmedic: 0, Helpblk: 0

(7) Race: white, Partyid: rep (5,6), Srcbelt: city (1, 2,3), Conmedic: 1

(8) Race: white, Partyid: rep (5,6), Srcbelt: city (1, 2,3), Conmedic: 0

(9) Race: white, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Conmedic: 1, Class: lower (1,2)

(10) Race: white, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Conmedic: 1, Class: upper (3,4)

(11) Race: white, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Conmedic: 0, Class: lower (1,2)

(12) Race: white, Partyid: rep (5,6), Srcbelt: rural (4,5,6), Conmedic: 0, Class: upper (3,4)

(13) Race: white, Partyid: ind (2-4,7), Srcbelt: city (1, 2,3), Conmedic: 1, Class: lower (1,2)

(14) Race: white, Partyid: ind (2-4,7), Srcbelt: city (1, 2,3), Conmedic: 1, Class: upper (3,4)

(15) Race: white, Partyid: ind (2-4,7), Srcbelt: city (1, 2,3), Conmedic: 0, Class: lower (1,2)

(16) Race: white, Partyid: ind (2-4,7), Srcbelt: city (1, 2,3), Conmedic: 0, Class: upper (3,4)

(17) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 1, Class: lower (1,2), Age: 0-54, Degree: hs (1)

(18) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 1, Class: lower (1,2), Age: 0-54, Degree: other

(19) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 1, Class: lower (1,2), Age: 55+

(20) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 1, Class: upper (3,4), Age: 0-54

(21) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 1, Class: upper (3,4), Age: 55+

(22) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 0, Class: lower (1,2), Age: 0-57, Degree: hs (1)

(23) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 0, Class: lower (1,2), Age: 0-57, Degree: other

(24) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 0, Class: lower (1,2), Age: 58+,

(25) Race: white, Partyid: ind (2-4,7), Srcbelt: rural (4,5,6), Conmedic: 0, Class: upper (3,4)

(26) Race: black, Partyid: dem (0,1), Srcbelt: city (1,2,3), Conmedic: 1

(27) Race: black, Partyid: dem (0,1), Srcbelt: city (1,2,3), Conmedic: 0

(28) Race: black, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Conmedic: 1

(29) Race: black, Partyid: dem (0,1), Srcbelt: rural (4,5,6), Conmedic: 0

(30) Race: black, Partyid: other (2-7), Conmedic: 1

(31) Race: black, Partyid: other (2-7), Conmedic: 0, Age: 0-38

(32) Race: black, Partyid: other (2-7), Conmedic: 0, Age: 39+

(33) Race: other, Partyid: dem (0,1)

(34) Race: other, Partyid: other (2-7), Conmedic: 1

(35) Race: other, Partyid: other (2-7), Conmedic: 0