Random-Sample Elections

Far lower cost, better quality and more democratic

David Chaum

ABSTRACT: A Random-Sample Election can be conducted locally, nationally, regionally, or even globally, with results that are more irrefutable than those of current elections but at less than one-thousandth of the cost.

As a new member of the democracy toolbox, such elections hold great promise, for instance making practical today: Petitions of government that prove majority support. Binding consultations of constituencies by officials or parties. Ladder competitions that elect the most important and clearly stated issues to be put to vote. Juries for public policy issues with unprecedented resistance to manipulation. Even, in the extreme, full Athenian-style direct democracy, practical at the scale and complexity of society today.

Voters may be better motivated and informed since each vote carries more weight and each voter can meaningfully investigate and study the single issue that voter is asked to help decide. Voter confidence in the enhanced integrity of the election process may also enhance participation. Anyone can verify online that neither randomness of voter selection nor integrity of outcome can have been manipulated, even by governments. The ballots are sent voters by paper mail, but vote buying is made ineffective by a novel technique.

The approach is compared with elections today in terms of nine attributes of election quality, the operational concept along with its properties sketched, historical context discussed, and various adoption scenarios laid out. An appendix gives implementation details, based on systems already proven in governmental elections, sufficient to show technical and practical feasibility.

It is accepted as democratic when public offices are allocated by lot;
and as oligarchic when they are filled by election.

—Aristotle (Politics 1301a28-35)

INTRODUCTION

A new way is introduced here for translating the “will of the people” into governance, the essential characteristic of democracy. This white paper aims to show that the new approach can already be deployed in a variety of ways and without need to involve governments—there is no catch. It also aims to explore the advantages and potential of the new approach to whatever extent adopted by government. The aim, however, is neither values neutrality nor a manifesto for a particular ultimate system.

The number of voters sampled can be small, depending on how close the contest, yet give overwhelming confidence. For instance, if the margin is at least ten percent, then a thousand votes will

likely yield a result that itself, without any assumption about the margin and with only a one-in-a-million chance of error, establishes that a majority are in favor—even with an electorate of millions or billions. This dramatic reduction in the number of voters participating in each election compared to a conventional election today yields a substantially proportionate reduction in cost.

A general benefit of such low cost, and also taking advantage of the small number of voters needed for an election, is that many more elections can be conducted, both in parallel and more often. This in turn means that each voter would typically only be involved in an election relatively infrequently. It also means that each election can be on a single issue, allowing voters ample opportunity to research and consider the issue. The mechanics of supplying ballots would be through paper mail; nevertheless, with today's widely adopted information technology, such as the web, search, forums, email access to experts, and so on, voters will be able with the increased time and focus to investigate and deliberate even complex issues with unprecedented ease and depth.

Another benefit of the extreme cost reduction is an opening up of who can conduct elections. Any interest group can create their own election initiatives, if they can bear the modest expense, without requiring permission or assistance from government. A new type of election infrastructure could, as another example, allow all interested parties to submit questions, and then it would conduct mini elections between proposed questions until one prevails and is put to a larger vote. Political parties or candidates for office could also conduct their own non-governmental elections and even agree in advance to be bound by the results. Every use helps establish efficacy, increases acceptance, and—even if national leaders are never selected this way—is potentially hugely beneficial to society.

Representative democracy, by no means the only type of democracy, is based on a right to vote for representatives and that every vote should be counted. This hard-won mechanism of universal suffrage, from which political legitimacy of most governments is extracted, may unfortunately be failing.

Trends over the last few decades bear this out. Trust in elected governments has plummeted, at least for major democracies. The number and, according to the various published ratings, almost all metrics of democracies are also declining, the so-called "democracy deficit." Divergence of public policy and fundamental values of society is accelerating, for instance on issues such as income distribution, human rights, war/peace, and environment. All this in spite of huge technology advances, especially for information, with its unprecedented ability to provide transparency, tackle complexity and scale.

Common explanations for the failure of representative democracy are the low utility to the individual of actually casting a ballot with the choices available and the corrupting influence of economic power in such elections. The approach introduced here, in contrast, promises to obviate corrupting influences on governance and to make each vote more meaningful so that policy can converge with widely held values—while actually benefiting from information technology and transparency and scaling to meet complexity.

Furthermore, using the technique of democracy outside governments opens new possibilities. Issues cutting across national boundaries, including even election-related issues such as access to information, for instance in terms of anonymous/uncensored surfing and media supported by voted quality, become something even a modestly-well-funded non-profit organization could put to global vote.

To whatever extent binding, random-sample elections can give powerful voice to issues while gaining acceptance and raising the bar on election system quality and performance. Gradually, as the techniques are proven out in practice, refined, and acceptance broadens, adoption in governmental elections is at

least plausible. The techniques could be used for such things as filling more positions by election. They are also a good fit for the various types of referenda and initiative, such as used in many countries and in the ten western most US states, allowing the making of major policy decisions. They then hold promise for increasing participation, improving perception of fairness and efficacy of the process, and strengthening the foundations of democracy and arguably society more generally.

This new type of election, more fundamentally, allows continuous, effective and indisputable monitoring of the will of the people on a very wide range of issues. It thus offers an alternative to representative democracy—at least, to whatever extent it may be desirable, solving a challenging "open problem"—Athenian democracy at the scale and complexity of governments today.

This new approach can also improve voter protection. Elections in which voters are able to verify convincing online evidence that their votes were correctly included in election outcomes have already been used to fill political offices in a United States city. This strongest known techniques for integrity and transparency of elections, as explained later and detailed in the appendix, is used to implement the new approach. An important property achieved, and verifiable by anyone online, is that which voters are included is truly random and cannot be known in advance, let alone intimidated or excluded by anyone. An extension ensures that vote buying is ineffective. An optional extension can even hide the identity of voters in perpetuity.

Future weather data or stock market prices on an upcoming date certain provide the randomness. Like other secure systems, data is "committed to" by being publishing in unchangeable encrypted form in advance. It's like the encrypted data is the result of potentially suspect coin flips still hidden by hands on wrists, with weather or stock data the random "call" made just before the coins are revealed to ensure unmanipulatability of the outcome. Because the decryptions can be checked by anyone online, the process is fully transparent. Such outcomes determine what to decrypt in the transparent audits that don't compromise ballot secrecy but that do ensure outcome integrity extremely effectively. Also, voters can verify that they are picked just as fairly by such outcomes as if they had flipped coins themselves with predefined patterns of heads and tails resulting in being selected to vote.

This work is organized in five sections, each intended to mainly stand alone. In the first section, nine desirable attributes for quality of public-sector elections are introduced and used to compare current elections with the system introduced here. The attributes are then recast slightly so that the system introduced achieves the full desiderata.

The second section summarizes the technical aspects of the system for the non-specialist reader. Emphasis is on the ideas behind some of the more innovative aspects of its security. The various categories of participants in the system are laid out, the new properties achieved are stated informally, and sample-sizes are derived from first principles. The system's technical feasibility is shown in an appendix detailing it in terms of component parts similar to those of systems already used in governmental elections.

The third section considers context in terms of the historical and current mechanics of democracy. A variety of situations in which the approach can be applied outside government as well as scenarios for gradual integration into government are sketched in the fourth section. Finally, the fifth section touches on some broader perspectives.

ELECTION QUALITY

Desirable characteristics of public-sector elections, defining what is here called their "quality," include: (a) high voter turnout, (b) well informed voters, (c) effectiveness of results in shaping governance, (d) resistance to manipulation through advertising and electioneering, (e) indisputability of tally, (f) protection against voter corruption or coercion, (g) resistance to voter fraud, (h) decisiveness, and (i) low cost.

Current elections perform poorly against every one of these nine positive attributes.

Election reform, however, has been notoriously difficult and slow. One explanation is that improving integrity, and by extension other aspects of elections, implicitly criticizes the soundness of the very system that selected those to positions that can block change. Political parties are also sensitive to changes in voter demographics and potential for manipulation, which can result from even small changes in election mechanisms.

Other reasons for the sluggishness of reform may include economic interests. Relatively large amounts of money have been spent on election equipment in some countries, notably the United States, Brazil and India, while recurring expenditures on administration are even higher in aggregate. Elections today entail massive campaign spending, which in large measure benefit media and opinion experts. (Such advertising and electioneering outlays, however, may be considered undesirable because of the conflict of interest they raise between officeholders/parties and the public with respect to various sponsors and lobbies.)

It has been posited that individual voters expending effort to learn about the options presented them in conventional elections is not economically rational, at least not as far as the probability of having an impact. This analysis may account in part for low voter-turnout. It also predicts compounding of the problem of meaningful voter participation, as technology continues to increase the complexity of governance and policy options.

Current elections will here be called "mass" elections to contrast with "random-sample" elections, which may broadly be characterized generally as a poll of a sample of the population that is at least as secure against abuse as traditional secret-ballot elections.

The significance of each vote is elevated in a random-sample election, compared to a mass election. This allows those voters who are selected to rationally put substantial effort into informing themselves, deliberating and voting, rising to the occasion much like members of juries are known to. Additionally, participants in elections optionally could be paid, but in a way that's verifiably independent of how they vote, detailed later, such as is common with juries and as was practiced in ancient Greece. Even without compensation, a significant boost can accordingly be expected in effective "turnout" and extent to which all contests/questions on ballots are voted. (High voter turnout [a].)

Other current impediments include unclear statement of issues or bundles of issues that are too complex. The obfuscation of beneficiaries of government policy through voluminous regulation, comprised of convoluted and archaic language, is well known. Random-sample elections can allow voters to peer more deeply into complex ballot questions. To the extent this new type of election is used to frame ballot questions, such as through question ladders mentioned above and discussed further below, it also holds the potential to directly return us to the days when ballot language and legislation were more comprehensible.

Thus, random-sample elections both raise the level of issues that can meaningfully be put before voters and hold potential to bring the statement of issues closer to what voters can understand. (Well informed voters [b].)

The combination of increased participation and meaningful deliberation just described allows random-sample election results to be more fine-grained and useful than those of mass elections in informing and steering governance. Moreover, the episodic election of parties and representatives creates lag, discontinuities and short-term focus of public policy; whereas continual polling facilitated by random-sample elections may foster gradual evolution of longer-term policy while providing lower lag in response to changing circumstances. (Effectiveness of results in shaping governance [c].)

Advertisements and other types of media and campaigning are able to influence mass election outcomes perhaps surprisingly well in this age of unprecedentedly widespread and selective access to information. (In the US, for instance, expenditure for related advertising trumps the mechanics of elections by a significant factor thereby creating decisive leverage for sponsors and lobbies.) This may be due to the shallow diligence of rational voters mentioned above, but it also derives from the episodic nature of election cycles. (Such cycles also in a related aspect focus perhaps too much of elected officials' time and attention on campaigning instead of governing.) Influencing an ongoing as opposed to an episodic polling process through media and campaigning is far more costly and difficult, especially when voters are motivated and able to investigate in depth, and random-sample elections may thus reduce such influence. (Resistance to manipulation through advertising and electioneering [d].)

The security features detailed below bring the integrity of a random-sample election up to at least par with those of the best polling-place mass elections, such as so-called cryptographic end-to-end systems. In some such elections that have been conducted, the correctness of the tally is proved mathematically in a way that can be verified easily by voters and also significantly by anyone interested enough to download or even write a modest amount of software. (Indisputable/trustworthy tally [e].)

So-called "improper influence" of voters refers generally to vote buying and coercion of voters. Variants of both types of influence are known in practice in mass elections at polling-places as well as in current vote-by-mail. Vote buying is essentially obviated under random-sample elections by flooding the market with willing sellers of decoy votes that will not be counted but that are enduringly indistinguishable from genuine votes, as detailed later; also, potential coercers are unable to find victims since they are unable to effectively learn who has received a ballot. Influence of groups of voters is also known, where for instance one locality is favored over others because of certain voting patterns; however, integrity of results in random-sample elections obviates mass elections' need to break results down by locality. (Protection against voter corruption or coercion [f].)

Abuses sometimes called "voter fraud" involve voting by those who are unregistered or vote more than once. Doubt is sometimes raised as to whether such "retail" threats significantly affect election outcomes; however, there has in some cases been apparent concern sufficient to cause restrictions on participation and even disruption of civil governance. The system introduced here substantially reduces such concerns, as it does not allow voters to select themselves, but instead itself selects voters from the rolls verifiably at random. Of course any voting system's resistance to voter fraud is only as good as its roster of registered voters. (Resistance to voter fraud [g].)

Decisiveness, the ability of an election to come to a conclusive result, is limited by what may be called

its “precision”: how close a contest it can meaningfully adjudicate. Precision has both a technical and a policy dimension. Best practices with conventional elections include thresholds, though they are often set on unverifiable totals with poor accuracy and thus are of questionable value. In case of ties, unverifiable and often flawed random procedures are required in some cases under current law. In theory mass elections could properly adjudicate extraordinarily close contests, even up to tie, yet their precision in practice does not justify this and actually makes it undesirable since this can amplify the efficacy of even small manipulations.

With random-sample elections, it would be preferable to discard and re-run with different parameters when a result falls below a preset threshold of sample size or confidence level. The need for a decisive outcome, however, may also be reduced by the decoupling with election cycles already mentioned. A random-sample election could also be structured so that if the tally is too close, the result is in effect verifiably-randomly determined by reverting to randomness when the underlying precision is exceeded, a better way to structure a vote that must be decisive by a date certain. (Decisiveness [h].)
Cost, compared to a mass election can, as mentioned, be extremely low because the number of voters is far less and cost is primarily per voter. (Low cost [i].)

Thus, random-sample elections come significantly closer to the desiderata above, as more generally stated: (a) higher participation ratios, (b) better informed voters rationally motivated to delve into issues, (c) increased effectiveness of results in shaping governance, (d) improved resistance to manipulation through advertising/campaigning, (e) increased indisputability and trustworthiness of results, (f) anonymity of voters with unsaleable votes, (g) voter fraud only through improper voter rolls, (h) equivalent but safer decisiveness, and all with (i) significantly reduced direct and overall cost.

INFORMAL SUMMARY OF TECHNICAL CONCEPT

Indisputability of random-sample elections derives from public verifiability that:
(i) Only votes from the un-manipulatable random selection of voters are counted.
(ii) The tally correctly and un-manipulably reflects the votes cast.
(iii) Vote buying (and optionally other “improper influence” of voters) is impractical.
(iv) Votes can remain unlinkable to voters, at least to all but the election authority.

How each of these four properties is achieved may be summarized as follows:

(i) A pre-agreed public random event, such as stock-market or weather data on a set future date, determines which voters are to receive ballots that will be counted. How the event data will be interpreted is further randomized in a way that is publicly commited before the event. Thus, voters can be selected essentially independently and uniformly from voter rolls (such lists of registered voters are typically available in the US for political purposes, such as getting out the vote).

(ii) Although the choice of voters is publicly verifiable as having been selected by the results of the random process, cryptography used in the commitments to how the data is interpreted hides the voter identities. The ballots mailed to the randomly-selected voters, however, will be known to those voters as being extremely likely to be counted, as a consequence of published data that can be audited online by anyone, such as used in elections with Scantegrity in Takoma Park, Maryland.

(iii) To keep vote-buying in check, any registered voter can request a “decoy” ballot. These will not be counted but will be distinguishable only by the requesting voter, even after the election. Requestors of

decoys will presumably try to sell them to buyers, especially buyers trying to manipulate an election opposite the requestor's preference. Those running the election would ideally monitor the casting of decoy ballots and the market for them and endeavor to keep decoy supply such that vote-buyer offers are below a price likely to entice randomly-selected voters.

(iv) To allow verification that the selected voters did in fact at least receive the opportunity to vote, the identity of selected voters (indistinguishably mixed with those requesting decoy ballots) can be made public once voting is over. In an optional variant, a number of "verifiers" are selected at random, each provided after close of polls with the identity of a different one of the voters, and later only the identities of these verifiers, not the voters, are made public. Each verifier is instructed to contact that verifier's voter to check whether the voter has in fact cast a ballot—and to raise an alarm if the ballot was not at least received by that voter. This allows verifiers to be checked on by the public while essentially keeping the identity of voters from becoming public.

Compensation of voters or verifiers was mentioned above as an option. Verifiers, as well as optionally voters who answer verifier queries, may collect rewards conditioned on meeting certain public criteria that obviously must avoid biasing the outcome. Such criteria could, for instance, include consistency of various responses or extra effort by the voter. To allow authentication between verifier and voter and then provide evidence of the successful verification, verifiers can be issued parts of a numeric confirmation code only known fully to the voter. Additionally, if ballots are mailed "signature required," then the authority has some recourse against a voter falsely crying foul.

The participants in a random-sample election may be divided into five categories: a single election authority, three very likely disjoint sets of members of the public, and an open-ended collection of auditors. The five categories (and their roles) are:

the election authority (commit to encrypted data initially, receive requests for decoy ballots, monitor public random events, mail ballots, decrypt part of data per audit driven by public random data);
randomly-selected voters whose votes will be counted (receive a ballot in the mail, send vote codes in, optionally check codes online and/or respond to inquiry);
self-selected requesting voters whose votes will not be counted (apply online, receive a decoy ballot, and try to sell the ballot);
optionally, randomly selected verifiers who do not vote but rather check that a corresponding voter did in fact participate (contact the voter whose information they are provided, provide part of code to voter, check part of code obtained from voter, return full code online); and
members of the public who audit the election process (run open-source or self-written software that uses published keys to decrypt published data and check that it is in accordance with protocol, published random values, and election outcome).

One way to determine how many votes are needed to establish a majority is familiar from coin tossing. Few would dispute that the odds are overwhelming that 20 tosses of a fair coin would include at least one tail, since the chance of all heads is less than one in a million. The same odds hold, for example, for: 22 flips having at least two tails, 25 flips three tails, and 100 flips 28 tails. (In the language of statistics, the relative frequency in the sample space of 100 independent fair coin flips of the event defined by less than 28 tails, calculated using the binomial cumulative distribution with success probability one half, is less than one over one million.)

Suppose organizers of an election believe that about three quarters of voters are in favor. If their belief were reasonable, when they collect 100 randomly-sampled votes it is likely that less than 28 voted

against—and such an outcome would itself establish, independent of their belief, and with overwhelming odds, that at least a simple majority are in favor. If they believed support is less than 75%, then larger samples are required. For instance, since 191 out of 300 gives the same overwhelming odds, 2/3 support would likely return a suitable number of affirmations in 300 or so votes. For 60% support, 1000 voters would mean 576 affirmations establish a one-in-a-million chance of error.

Questions that are believed too close may be recast to provide more economical margins. Less compelling odds may also be acceptable or can even result in overwhelming confidence when votes earlier in a series, such as with a ladder or ongoing polling, are all corroborated by subsequent votes. Generalization to super-majorities, multi-way contests, and even correcting for the distribution of voters using different rosters, are all possible though perhaps unnecessary.

CONTEXT

Random-Sample Elections may be appreciated in context of the historical and current technics of democracy.

Much of what is regarded as Western culture and civilization, including philosophy, constitutional law, science, mathematics, medicine, sculpture, theater, music, literature, sports, and arguably even the phonetic alphabet—akin to the printing press or Internet of its time—can be traced to the 170 years of democracy during the "classical period" of Ancient Greece. All governance, not only adjudication of criminal and civil disputes, was decided by large, randomly-selected juries without judges and using only yes or no votes. (Such juries could, for instance, be invoked by any citizen to decide constitutionality of legislation and criminal penalties for legislators for even only proposing unconstitutional bills.) A then appreciated attribute of their practice of random selection to fill most government posts, called "sortition," was its resistance to lobbying and corruption—something present in their democracy but now lost in our "representative democracy."

Juror selection in common law countries, albeit unverifiable and heavily post-culled, is all that remains of Ancient Greek democratic mechanics. Random selection of citizens, however, is widely used by government much more heavily today: conscription for military service, for instance of over fifteen thousand US conscripts who lost their lives in Vietnam; random selection of tax audit subjects; selection of citizens in policing generally, for instance at airport checkpoints; and random selection of poll workers, such as in Brazil. Safety testing and regulation by government in medical, food, transportation, and other sectors is often based on random selection or sampling (as is research underlying much scientific advance). More particularly relevant, random selection is required by law in various jurisdictions around the world in ways that can directly affect election outcome, such as random listing order of candidates, random selection in transferable vote counting, and random choice in case of apparent tie.

Random selection is also used heavily by political parties, candidates for office, and interest groups—in electioneering—far less so, however, beyond trying to influence election outcomes. Respondents in so-called "public opinion polling" are in principle selected at random and asked to quickly answer a series of questions related to candidates and issues of the day. As currently practiced, however, such surveys are neither transparent, trusted, nor generally believed worthy of public trust.

Public opinion polls are traditionally conducted by banks of questioners calling more or less random phone numbers using a script aimed at identifying a random voter among those present in the

household. Mobile phones have rendered this even more difficult and online pools of persons and even persons selected in shopping malls are also used as respondents. Demographic data obtained from respondents, often of dubious quality, is often used presumably in an attempt to correct for the bias in the sampling method, but based on unpublished models that may themselves bias outcome. Turnouts are very low. Respondents have little if any time to explore the meaning or consequences of questions, let alone think about, research or deliberate on, issues in such "opinion" surveys. (The related field called "deliberative democracy" is premised on the utility of group settings and has even been used in binding public-sector elections, but suffers from high cost and is subject to well-studied techniques developed for manipulating juries.)

Surveys have other well-known problems as well. How questions are worded and sequenced is a notorious form of bias. Questioners also bias the results, whether they intend to or not, by how they communicate. The published results of multiple such polls often deviate significantly on the same question, depending on the orientation of the entity conducting or sponsoring the survey, contributing to erosion of public confidence in such polls generally. Opinion polling predicting election results, a significant revenue source for polling organizations, is at best as flawed as the underlying elections. It has a self-fulfilling effect, and consequently some countries ban publication of its results close to elections. It has even been suggested that Asimov's 1955 short story "Franchise," about an artificial intelligence that votes for a whole society based on a few quirky questions asked of an apparently randomly-selected citizen, is a kind of half-hearted straw-man argument by scenario against public opinion surveys.

The techniques presented here could be taken to be merely an incremental improvement on public opinion polling, but are better understood instead as potentially providing so large a qualitative improvement, even surpassing the quality of mass elections themselves, that they offer really new options for democracy.

Some countries, such as Switzerland, routinely have binding referenda on questions at a national level; many countries have them only rarely; and most countries, including the United States, have no provision for national referenda, though the ten westernmost US states do provide for initiatives and many even on constitutional changes. The techniques introduced here go a long way toward mooting traditional first order reservations about referenda: the lack of deliberation, inability of the public to deal with complex issues, and ease with which the public can be manipulated temporarily by media campaigns. Ironically, these same concerns, obviated by random-sample elections, are present in representative democracies, where policies often seem driven by short-term waves of public opinion.

It is known that those who control what candidates or questions are put to vote, how contests are grouped and ordered, and when polls are held, can significantly influence outcomes, even with more direct systems like referenda. Dramatically lower cost might allow referenda to be conducted without the need to aggregate or sequence contests, fit a sparse schedule, or even let government control the questions put to the electorate. Thus, these second-order concerns may also be obviated.

A related perhaps tertiary problem is what are called election (or more generally aggregation) paradoxes, ways to combine policy choices so that preference for various combinations seem surprisingly at odds with at least some intuition about the original individual policy choices. Arrow's impossibility result from social choice theory is a well known example, but the more indirect the bigger the problem. Such issues are avoided when independent or "separable" yes/no policy decisions are voted individually but in parallel, and non-separable issues voted sequentially. Using random sample elections, these separate contests can readily be conducted in practice and at low cost.

ADOPTION SCENARIOS

Example early adoption scenarios for random-sample elections flowing from the above discussion include interest-group initiated surveys of the public on a particular issue. Such an election offers its sponsors a way to make a statement about the will of the electorate that is irrefutable—far more compelling than petitions and perhaps more effective and less damaging than protests. It could also legitimize a spokesperson or organization, such as to serve as a representative in a consultative process.

Another example use is public opinion surveys related to mass elections. For one thing, this would be more resistant to manipulation and thus potentially of more interest and less likely to be banned. For another, many mass elections around the world are disputed, such as by a mixture of allegations of technical fraud and registration and polling place discrimination. A parallel random-sample election with improved turnout might neither validate nor invalidate a mass election, but it could perhaps more importantly provide indisputable information on the will of the electorate.

An example use that is generative of ideas or suggestions and competitively filters them, is a ladder of contests to establish issues to be subjected to a general vote. Such a ladder simultaneously and transparently arrives at specific language for questions that emerge from it. The initial rung of candidate questions could be submitted—with a low barrier, such as "CAPTCHAS," a minimum of supporters, or a small fee—by anyone. The sample size could be small and margins required large at the lower rungs and increase as successful wordings move up the ladder. One variation allows each person to pay only a limited amount towards a question, such that when a question receives sufficient funding it is put to vote; another variation uses delegated or liquid proxy voting to fill lower rungs. Providing such avenues for self-selected supporters to push questions to vote advantageously allows motivated or expert groups a way to develop questions without allowing such groups to control outcomes.

The scale of electorates need not be huge, however. Use with smaller populations, such as large volunteer organizations, universities, or corporations, may allow a wide range of less contentious issues to be efficiently decided by what are in effect small, randomly-constituted committees, where near unanimity can keep sample size quite small.

One example kind of transitional scenario towards full use in governance lets candidates for political office or political parties commit in advance or later simply opt to use a random-sample election. They could use it to determine one or more aspects of how they exert their elected power, such as by voting a certain way in a legislative body or implementing a policy. For instance, guaranteeing constituents veto on any vote for war, a right interestingly granted women in the representative democracy of the Iroquois nation.

Limited initial use by governments is another type of transitional scenario. When, for example, a particular issue must under law be subject to vote, and a mass election is too costly compared to the significance of the issue or the delay in conducting a mass election is problematic, a random-sample election may be used instead. Another example limited use is an initial culling of candidates, perhaps a way to greatly improve political party mechanics, allowing more focused debate and voter focus on a smaller set of candidates chosen without bias. A conventional initiative or referenda usually requires submission of a certain number of signatures, presumably to enhance the chance that the measure is worthy of public attention and winnable. Allowing a random-sample election as a trigger (to say nothing of its ultimate superiority in conducting the final vote) would do a better job and at lower cost.

Yet another kind of transitional scenario involves groups not well aligned with the largely geographical hierarchy of governments today. Some regional groups, for instance, are split between governmental jurisdictions; their elections could be allowed to define their own boundaries. A related use is appointing external representatives of such groups.

There has never been a global referendum. Major issues are increasingly global, but democracy does not currently exist between or above nation states. Moreover, democracies, whose sovereignty derives from that of their citizens, may be hard pressed to justify opposition to the studied and proven will of a majority of the citizens of each major country, at least with respect to global issues. While there are a variety of huge issues that would no doubt engender substantial global interest and support, one kind of issue specially relevant here relates to information policy and rights. Voters being able to obtain information without risk and arrangements that promote transparency and quality of information available, illustrated above, are particularly relevant to effective elections and thus democracy itself.

The cost of even a global election might be as low as ten million dollars for a sample size of a thousand or so, since the lion's share of voters would be in countries with adequate infrastructure. In countries without such infrastructure, voters can still be identified as the person living in a dwelling selected randomly from those in a randomly selected region. (Rules for automatically selecting regions of similar population from satellite imagery and for ordering dwellings within regions would ideally be fixed in advance.) Then an election worker, perhaps with a translator, would go and collect the vote. Example topics include Internet privacy and access, which are additional examples of issues relevant to elections themselves, as well as income distribution, environmental, human rights, drug and food policy as well border and sovereignty issues.

What if someday a government wished to, perhaps because of expression of the will of the electorate after positive experience with non-governmental elections, change the legal framework so that they could start really using random-sample elections. What kinds of elections might be considered? Simply deciding the same questions typical of mass elections would probably rank rather low, as there would be little advantage and less voter involvement. Major policy decisions are an obvious choice for referenda. Switzerland's what is sometimes called "half-direct" democracy lets voters approve laws and this could also take advantage of the increased number of elections possible with random-sample elections. Another example is appointment of the key persons actually running governments, such as cabinet member heads of major departments, key legislative committee leadership roles, and even supreme court judges. The increased number of elections could allow such more fine-grained democracy.

DISCUSSION

Conventional secret-ballot mass elections are a paper-based technology that is 150 years old (concurrent in the U.S. with extension of the franchise to male citizens, roughly, although women's suffrage took another 50 years). Today's sophisticated information technology, however, is used to manipulate political processes by those with more direct access (for example in redistricting, ferreting out and influencing public opinion using random sampling combined with big data, and interest group monitoring of legislator votes). Meanwhile, voters have so far been left with a technology for voting—or a high-tech simulation of it, which makes the point that it is actually only a paradigm for voting—that was developed and promulgated well after the US constitution but before the first telephone call.

Random-sample elections can thus be interpreted more broadly as providing a way forward, from our

current paradigm-induced disparity in access to the power of information technology, towards allowing effective voter steering of governance.

In future there may be some who long to vote in mass elections, perhaps romanticizing about the act of casting a secret ballot in person among one's neighbors or at least the chance to submit a vote no matter how futile. But there will likely be few who would oppose the deeper and wider and more continuous monitoring of the will of the electorate provided by random-sample elections—once efficacy is established—at least informing if not being binding on government.

CONCLUSION

Random-sample elections offer practical low-cost yet unprecedented quality for almost any election, with a range of immediate applications, and each election potentially a step towards more effective and finer-grained democracy at scale.

APPENDIX

DETAILED ELECTION PROCESS

A random-sample election can be conducted as detailed in the diagram on the last page. It shows, using Scantegrity I in an Eperio format, an example of the more general concept. This example, described next, conveys the underlying ideas concretely and is well suited to practical implementation.

Each voter receives by mail a paper form like that on the left side of the diagram. The voter first chooses freely one of the two ballots printed on the double-ballot form received, in the example either serial number #100a or #100b. The voter then enters this ballot serial number online. To cast his or her vote, the voter next enters online the unique "vote code" printed on the chosen ballot form adjacent the desired "YES" or "NO" vote. At this point, for the ballot serial number not voted, both vote codes and their corresponding votes are displayed, so voters can check that they have a real ballot and help audit that ballots do in fact pair votes with codes correctly. (The red ovals highlight the example choices made by a voter, with "A" being the choice of the upper ballot on the double-ballot form and "B" the "NO" vote and its vote code.)

The election authority posts data online in a format allowing audit of election integrity while protecting ballot secrecy and not revealing which ballots are decoys. This posted data consists of multiple copies of a table of data, which includes ballot and vote information. Before posting, the election authority separately transforms each identical copy of the original table with three transformations, based on secrets it has randomly chosen. The first transformation is "row-shuffling," resulting in an unpredictable re-arrangement of the rows. A particular row thus almost always appears in different positions in each table. The second transformation randomizes pairs of numbers in adjacent columns, called "summands," but in a way that keeps their sum unchanged, examples of which are described below. The third transformation, yielding the posted data, encrypts each column as a whole using a secret key unique to that column and table. The class of encryption used ensures that the tables are unalterably committed to, as it offers only a single possible decryption per encryption.

In the specific example illustrated, 250 transformed copies of the original table are posted, each made up of 8 columns, with two rows for each ballot. This data could even be handled by a spreadsheet that supports suitably encrypted columns. The example sends double-ballots to 1,000 voters, divided between real and decoy in some proportion. The double-ballot form results in 2,000 ballot serial numbers and accordingly 4,000 rows per table. The 1,000 ballots in the example are distributed among the 10,000 potential voters, whose addresses are selected from the posted list, here called the "roster," sometimes elsewhere called a voter or registration or electoral roll or file.

(The red ovals follow the example vote along its row in table number one, the table shown at the top of the stack: the ballot serial number and vote code in oval "C"; and in "D" a "NO" vote marked "VOTED." The second summand pair "E" is 0000 and 5555, with the same sum as the other pairs for this double-

ballot, 5555. The entry in the list of third summands, corresponding to serial #100a, “F,” is 2222. Summing all three summands, 0000+5555+2222=7777, yields the position number “G” in the “voter roster” address list to which this ballot should be mailed.)

Making sure all the encrypted columns are published before the final random draw, and which columns to decrypt then being controlled by that random draw, prevents the election authority from posting anything but correct tables, otherwise the authority would very likely be detected in audit. This is because the final draw determines unpredictably which one of the five “batches” each table ends up in, and each batch has a different combination of columns decrypted and these combinations cover all columns in an overlapping and interlocked way. An election authority posting table content that deviates enough to likely alter the outcome or manipulate the choice of voters would be detected with overwhelming odds; extremely few of the vast number of ways the tables can be divided allow any particular deviation to go undetected.

The combinations of columns decrypted per batch, however, reveal neither how any single ballot serial number was voted nor which serial numbers are decoys (even when considering what can be linked by unique values that appear in multiple batches).

The first two batches are processed so voters can check that their raw vote data are published correctly on the electronic bulletin-board and also that the ballots were printed without interchanging “YES” and “NO.” This is shown as “audit casting & printing.” Then “audit voter selection & tally” opens the third and fourth batch combinations to reveal the real votes comprising the tally. It also separately reveals the decoy votes. It further establishes that the two summands for real votes are correctly copied from those committed before the first draw. Finally, “reveal all voters” discloses the voter identities.

This final step of the election is similar to what in mass elections is sometimes called certification. It reveals all roster entries to which ballots were to be mailed and the respective serial numbers—but of course nothing about how the ballots were voted or whether the voters/ballots were real or decoy. This is done by decrypting the serial number column and a pair of summand columns of the fifth batch. The alternative to such certification mentioned earlier, where verifiers provide anonymity of voters, can be added using an elaboration of the same techniques.* That option can also be included in more distributed variants where the election authority is replaced by a multiparty computation that protects privacy from all but collusion or compromise of a majority of potentially multiple election authorities including the option of a separate election authority for each portion of the electorate.

The example election is detailed according to the nine sequential steps, indicated in the figure by numbers in blue discs and matching numbers in parenthesis below, as follows:

(1) Four encrypted columns (shown yellow) are posted by the election authority in this first step. The election authority chooses the row shuffling for each of the 250 tables, ideally independently and uniformly at random. The two pre-draw summand columns, 5 and 6, are filled randomly but with the same sum for all instances of the same ballot. This can, for example, be done by choosing all the summands randomly and then adjusting each second summand so that the sum for every row matches that for the corresponding ballot in the first table. The authority then posts, for each table, the four separately encrypted columns, 1, 3, 5 and 6. The roster of voters is also locked at this time.

(2) Public random values are drawn after step (1) is completed, in this initial draw, so that they were unpredictable during step (1). The list consists of third summands (shown purple) each labeled by its respective ballot serial number. The audit of step (8) ensures that the election authority could not have

manipulated the selection of voters receiving real ballots, because when the first two summands for real ballots are posted, the list of third summands is unpredictable. This list may simply be defined as successive binary string outputs of the initial draw, skipping strings that are repetitions or too large a binary value (though step (1) could lock a more efficient algorithm). These draws can be shared across multiple elections, though then offer practical but less than optimal probabilities. To ensure no voter gets more than one ballot, in the unlikely event that there are repeats in the sums, the election authority marks as cancelled every such row of every table and these rows are ignored in steps (3) through (8).

(3) Two additional encrypted columns (shown green) are posted by the election authority, who also then prints and mails the double-ballot forms to voters. The authority first copies the pre-draw summands, which are for real ballots, from columns 5 and 6 into columns 7 and 8, respectively. Instead of copying the "DECOY BALLOT" indication, the authority makes up and inserts the corresponding post-draw summands. A post-draw summand for a decoy ballot in, say, column 7, is first chosen by the authority at random. The summand for the corresponding column 8 is then computed by the authority so that, along with the summand labeled by the ballot serial-number in the list of third summands now known from step (2), the sum of all three summands is the position of a desired decoy voter in the voter roster. These positions in the roster, along with the similarly computed positions for the real ballots, yield the addresses for mailing the respective ballots. Decoy ballot roster positions, whether requested or chosen at random, that would result in more than one ballot per address are denied or at least excluded from the tables. (Those chosen at random could be supplied under separate cover with all its row numbers as a kind of proof that their ballot is in fact a decoy and they should try to sell it.)

(4) Voters cast votes by posting, on the electronic bulletin-board, the ballot serial-numbers they wish to vote along with corresponding vote codes. The cast vote itself, however, is neither entered online nor included on the bulletin-board, preventing voters from using prior-knowledge of vote codes to prove how they voted to vote buyers. (An alternative, that prevents vote buyers from being able to require a choice of ballot half after learning what's printed on both halves, informs the voter online of the choice of half made by a public draw after the ballots are mailed.) Each voter is requested to check that the un-voted ballot on the double-ballot form is correctly posted with vote codes matching votes, as mentioned earlier. The bulletin-board publicly posts all this data for use in audit step (7).

(5) Once the polls close, two columns are posted by the election authority(shown blue). Column 2 shows serial numbers and vote codes for all un-voted ballots, allowing cross-checking with the audit step (7). Column 4 marks the rows whose vote codes were voted, so that after the decryptions in step (8), the real votes can be tallied.

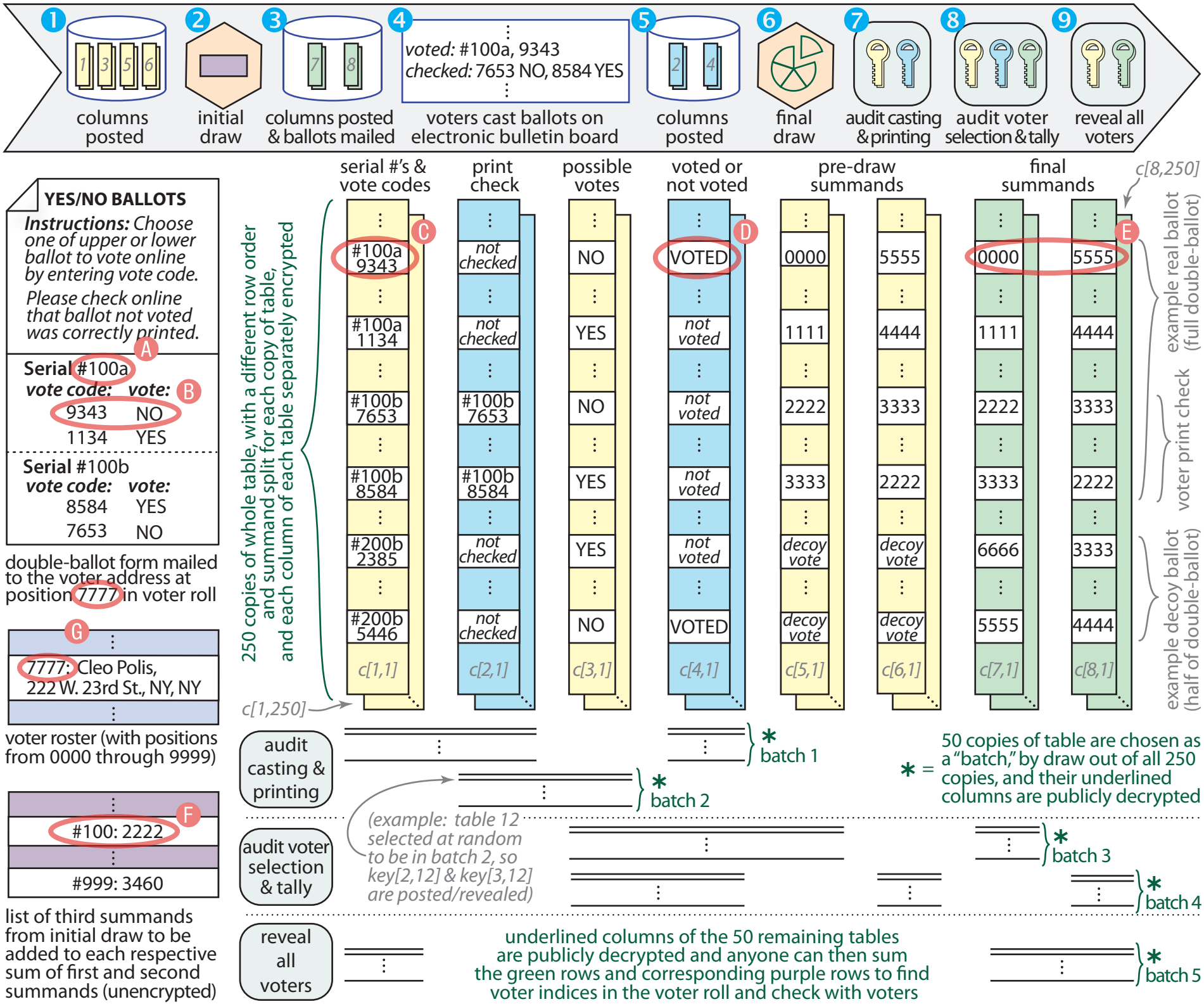
(6) The final public random draw is performed after the outputs of step (5) are posted. This draw is used to unpredictably divide the 250 tables into five batches: 50 tables for each half of the first two audits, with the remaining 50 for step (9). The rule defining this division places the first 50 table numbers that appear in the successive draw bytes into the first batch, the second into the second, and so on. This draw is ideally completed quickly, so that audit can begin shortly after polls close, but need yield only a relatively small amount of random output that can be shared across multiple concurrent elections.

(7) The first audit is ideally completed shortly after close of polls. Columns 1, 2 and 4 are publicly decrypted for the first batch of 50 tables and columns 2 and 3 for the second 50. Cross checking with the bulletin-board ensures consistency of which serial numbers were voted. It also allows the printed pairing of vote codes to cleartext votes that voters saw on the bulletin board to be cross-checked with columns posted before the ballots were mailed.

(8) The second audit can be conducted when the tally is to be revealed. Columns 3, 4, 5 and 7 are publicly decrypted for the third batch of 50 tables and 3, 4, 6 and 8 for the fourth batch. Rows marked "VOTED" in column 4, but not marked "DECOY VOTE" in column 5 or 6, are valid votes and their summation is the tally. The last two decrypted columns, in either case, 5, 7 or 6, 8, should be identical, unless a "DECOY VOTE" appears in the pre-draw column of the pair. This verifies that, as mentioned, at least those rows with votes in the tally were in fact committed to before the initial draw and consequently that the choice of real voters determined by the tables is un-manipulable, even though the election authority can freely choose the decoy voters. This audit could even be repeated with additional tables for extra certainty later.

(9) The positions in the voter roster containing addresses to which ballots were sent are revealed by the election authority in this final certification audit step, but without revealing whether the ballots were real or decoy or how they were voted, also as mentioned earlier. The remaining batch of 50 tables is publicly decrypted in columns 1, 7 and 8. By using a serial number entry in the first column to look up a third summand in the published list and summing that with the summands in the other columns, anyone can check uniqueness of unmarked voters and locate the addresses in the voter roster to which ballots should have been sent. They can then verify, such as by alerting voters, checking mail receipt signatures or online receipt records, or even by directly asking voters, that these voters did in fact receive ballots reflected on the bulletin-board.

* Each row entry of column 5, 6, 7 and 8 is appended with a "verifier summand," selected at random just as with the non-decoy summands. The original summands for columns 7 and 8 are encrypted with a special unique key per row. When column 7 or 8 is opened in step 8, the "master key" for all its original summands is opened and they are revealed. When column 7 and 8 are opened in step 9, however, the master key is not published but used instead to compute the individual sub keys for the original summands and these keys are encrypted and posted using the corresponding unique key that was mailed at ballot mailing time to the respective verifier.


1 columns posted
2 initial draw
3 columns posted & ballots mailed
4 voted: #100a, 9343
checked: 7653 NO, 8584 YES
voters cast ballots on electronic bulletin board
5 columns posted
6 final draw
7 audit casting & printing
8 audit voter selection & tally
9 reveal all voters
YES/NO BALLOTS
Instructions: Choose one of upper or lower ballot to vote online by entering vote code.
Please check online that ballot not voted was correctly printed.
A
Serial #100a
vote code: vote:
B
9343 NO
1134 YES
Serial #100b
vote code: vote:
8584 YES
7653 NO
double-ballot form mailed to the voter address at position 7777 in voter roll
G
7777: Cleo Polis, 222 W. 23rd St., NY, NY
voter roster (with positions from 0000 through 9999)
F
#100: 2222
#999: 3460
list of third summands from initial draw to be added to each respective sum of first and second summands (unencrypted)
250 copies of whole table, with a different row order and summand split for each copy of table, and each column of each table separately encrypted
serial #'s & vote codes
print check
possible votes
voted or not voted
pre-draw summands
final summands
c[8,250]
C
D
E
#100a 9343
#100a 1134
#100b 7653
#100b 8584
#200b 2385
#200b 5446
not checked
NO
YES
VOTED
not voted
0000
1111
2222
3333
5555
4444
6666
decoy vote
c[1,1]
c[2,1]
c[3,1]
c[4,1]
c[5,1]
c[6,1]
c[7,1]
c[8,1]
c[1,250]
example real ballot (full double-ballot)
voter print check
example decoy ballot (half of double-ballot)
audit casting & printing
batch 1
batch 2
audit voter selection & tally
(example: table 12 selected at random to be in batch 2, so key[2,12] & key[3,12] are posted/revealed)
batch 3
batch 4
reveal all voters
batch 5
* = 50 copies of table are chosen as a "batch," by draw out of all 250 copies, and their underlined columns are publicly decrypted
underlined columns of the 50 remaining tables are publicly decrypted and anyone can then sum the green rows and corresponding purple rows to find voter indices in the voter roll and check with voters