Random-Sample Voting

More democratic, better quality, and far lower cost

David Chaum

ABSTRACT: Random-sample voting can be used locally, nationally, regionally, non-geographically and even globally. It is more trustworthy than current elections, yet has one thousand times lower cost.

As a new member of the democracy tool kit, it makes practical for the first time: petitions of government that prove majority support; binding consultations of constituents by officials or parties; tournaments of votes that elect the most important and clearly stated proposals to be put to full vote; juries for public policy issues with unprecedented resistance to manipulation; and even Athenian-style direct democracy that is well suited to the scale and complexity of society today.

Voters may be better motivated and informed since each vote carries more weight and each voter can more meaningfully research and consider the single issue that voter is asked to help decide. Increased voter confidence, resulting from the increased quality of the election process, may also enhance voter participation. Anyone can verify online that nobody, not even those running the election, could have manipulated randomness of voter selection or integrity of outcome. Ballots are sent voters by paper mail and they vote them online, but a novel technique presented makes vote buying ineffective.

After introducing the concept from several perspectives, ten attributes of election quality are proposed. Next the operational aspects and security properties are sketched, historical context discussed, and a variety of adoption scenarios laid out. Appendix 1 argues that the system compares favorably to voting today for each of the ten election-quality attributes. Appendix 2 gives implementation details, already tested in governmental and small-scale test elections, proving technical and practical feasibility.

INTRODUCTION

"Random-sample voting" can be introduced as the polling of a random subset of voters that is at least as secure against abuse as current elections. This white paper aims to show that random-sample voting can already be deployed in a variety of significant ways, with or without participation of government—there is no catch or major impediment. It also aims to explore the advantages and potential of this new approach. (The aim, however, is neither values neutrality nor a manifesto for a particular system.)

Only a small number of voters need be sampled, depending on how close the contest, yet achieved is overwhelming confidence. For instance, if the margin is at least ten percent, then merely 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 of the whole population is in favor—even with an electorate of millions or billions. This dramatic reduction in the number of voters needed in each election, compared to elections today, means a similarly dramatic reduction in election cost.

A major benefit of such low cost, and also taking advantage of the small number of voters involved in an election, is that many more elections can be conducted, both in parallel and more often. This in turn means that each voter need 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 that issue, since voters can use the web (including to access experts who may have additional input discussed later) to investigate and deliberate even complex issues with unprecedented ease and depth.

Another benefit of the dramatic cost reduction is an opening up of who can conduct elections. Any interested group can create their own election initiatives, if they can bear the modest expense, and without requiring assistance from government. A new type of election infrastructure could, as another example, allow all interested parties to submit questions and would then conduct mini-elections between proposed questions in a tournament process until one question is voted most wanted and is then 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, highlights reasoned sincere policy preferences and—even if national leaders are never selected this way—is potentially enormously beneficial to society.

Representative democracy, by no means the only type of democracy, is based on a right to vote for candidate representatives and that every such vote should be counted. This hard-won mechanism, from which political legitimacy of most governments today is derived, sadly appears to be failing.

Trends over the last several decades bear out this precipitous decline. Trust in elected governments has plummeted, at least for major democracies. The number of and, according to the various published ratings, almost all metrics of democracies are also dropping, so much so this has been called the "democracy deficit." So-called "responsiveness" of public policy to the values of society is perhaps the key figure of merit for a democracy. Yet today most government policy is non-responsive on almost all major issues. Examples include income distribution, human rights, war/peace, environmental and other catastrophic and systemic risks. All this is surprising in the face of huge advances of information technology, bringing unprecedented ability to provide transparency, tackle complexity and to scale.

Common explanations for the apparent failing of representative democracy today are the inherent low utility to the individual of actually casting a ballot especially in view of the choices presented and the corrupting influence of economic power in such elections. An alternate explanation is that transparency through information technology is at last exposing the failure of representative democracy to scale.

The solution introduced here, in contrast, uses proven techniques to obviate corrupting influences on governance and make each vote more meaningful, so that policy can converge with widely held values —while actually benefiting from transparency and the scaleability of information technology—in a way that can address the complexity of society today.

The fundamental challenge of democracy, as long understood, is providing full enfranchisement and preventing corruption while incorporating specialized knowledge. Enfranchisement as a value, one citizen one vote, has remained unchanged since Classical Athens (though then citizenship was defined comparably to the US prior to 1868). Random selection in Athens provided full enfranchisement while accurately reflecting the will of the citizenry, without involving every citizen in every decision.

Corruption then, defined as abuse of public office, was believed prevented also through random selection. There were a few officials whose positions required specialized knowledge, but addressing

their potential corruption proved more difficult. Specialized knowledge is now needed much more widely to arrive at beneficial policies, owing to the far greater complexity and scale of society. This might seem to suggest that preventing corruption is fundamentally incompatible with democracy today.

Fortunately, however, we have reached a kind of critical mass in public access to specialized knowledge—widespread use of the web. This powerful access, together with the new technology introduced here for secure and robust publicly-verifiable random polling, now once again allows a full and practical solution to the democracy challenge.

More generally, the new type of election introduced allows continuous, effective and indisputable translation of the will of the people into policy, whether through government or alongside government.

To whatever extent binding, random-sample voting can give powerful voice to issues while gaining acceptance and raising the bar on election-system quality. Gradually, as the techniques are proven out in practice, potentially refined, and acceptance broadens, various types of adoption by governments is at least plausible. Random-sample voting could then be used to qualify and even conduct various types of referenda and initiative, such as are used in the United States by many states and in many countries. Use for voting within legislative bodies would help them scale and provide crucial ballot secrecy.

A variety of other adoption scenarios hold promise for increasing participation, improving perception of fairness and efficacy, and more generally for strengthening the institutions and foundations of democracy. For instance, experience with more direct democracy, with its focus on issues over personalities, reportedly increases participation, public debate, and responsiveness of government.

Outside government, the techniques also open new possibilities. Issues cutting across national boundaries, including even election-related issues such as access to information, for instance in terms of privacy-protected/uncensored online access and insuring media integrity by funding based on voted ratings, could be realized even at the unprecedented scale of global votes.

Voter protection, ensured in part by public evidence that votes are correctly included in election results, is provided by technology that has been used to fill political offices in a United States city. (A further novel extension described ensures that vote buying is ineffective.) This strongest known technology for integrity and transparency of tallying, as explained later and detailed in an appendix, is incorporated here in random-sample elections. It is extended to ensure, also verifiable by anyone online, that which voters are included is truly random and cannot be manipulated. This prevents discrimination in selection or targeted influence of voters and provides a new level of enfranchisement.

Blockchain values, on a future date-certain, provide the randomness. In advance of that future date, data is “committed” to by being publishing in unchangeable encrypted form. It’s as if the encrypted data is the result of a potentially suspect coin flip still hidden by a hand covering it on a wrist, with blockchain data the random “call” made after the flipped coin can no longer be changed but before its heads or tails face is revealed—thereby ensuring unmanipulatability of the combined outcome. Because the mathematically-unique decryptions can be checked by anyone online, the process is fully transparent. Such randomness also determines what to decrypt in audit, ensuring election integrity with ballot secrecy. It also lets voters verify that they are picked just as fairly as if they had flipped coins themselves, with predefined patterns of heads and tails resulting in being selected to vote.

This white paper is in five sections, each intended mainly to stand alone. In the first section below, ten desirable attributes for quality of public-sector elections are introduced. (These attributes underly a

detailed comparison of the approach introduced here with current elections in Appendix 1, which concludes that random-sample elections are at least as good across all attributes once generalized.)

The second section introduces the novel security concepts of the system for the non-specialist reader. The various categories of participants in the system are laid out, properties achieved stated informally, and sample-size derived from first principles. (The system's technical feasibility is shown in Appendix 2, which details its various aspects and parts and provides a complete recipe for the current software and operational deployment.)

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

1. ELECTION QUALITY

Desirable characteristics of public-sector election mechanics, 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) ensured voter access to vote casting, (h) resistance to voter fraud, (i) decisiveness, and (j) low cost.

Current elections perform egregiously poorly against every single one of these ten 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 the positions of power that enable them to block change. The power of election administration is considerable and perhaps explanatory. Also, political parties worry about changes in voter demographics, 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 significant campaign spending, which in large measure benefit media and various 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 sponsors and lobbies.)

See Appendix 1 for the detailed comparison in which it is argued that random-sample elections compare favorably to current elections for each of the ten above (a-j) attributes of election quality.

2. INFORMAL SUMMARY OF TECHNICAL CONCEPT

One way to look at random-sample elections is in terms of who is supposed to do what related to the election. Another is in terms of what properties are achieved by the election protocol these parties should follow, even if those running the election try to cheat. Related is how many votes are needed for a positive result to be convincing. Each of these considerations is taken up in turn below.

Voters who receive ballots by paper mail and who then cast their votes online are the main parties in random-sample voting. Voters are afforded the option to check that their ballots have codes printed correctly and that the codes for the votes they cast are properly recorded online. Another novel and unique security feature of these elections is that anyone can request what will be called a "decoy" ballot that can be voted, such as according to the want of a vote buyer; the requestor of the decoy ballot, however, knows that the vote will remain uncounted even though a vote buyer should never be able to learn that what they have purchased is actually a decoy. Also, since ballots are sent to randomly selected people, false identification is less an issue.

The "election authority" is the party responsible for conducting the election according to protocol. Anyone is allowed to be one of the "auditors" of an election and check data published by the election authority. Auditors aim to ensure that the election authority can neither influence the selection of voters nor change the outcome of votes cast by voters. Any such malfeasance requires deviation from protocol and, with extremely high probability, would result in auditors obtaining irrefutable evidence of it.

Indisputability of random-sample election integrity derives from the protocol conducted by the election authority with voters, requestors of decoy ballots, and auditors, that is laid out in detail in Appendix 2. As a result of this protocol, audit can ensure the following four main properties:

(i) *The selection of voters is at random and cannot be chosen in advance or manipulated:* Voters should be selected independently and uniformly from the list of potential voters in a manner that even the election authority cannot manipulate and these voters should remain anonymous at least during the voting. To ensure this, the election authority publishes first in effect a commitment to a random re-ordering of the public list of voters. This commitment is encrypted so that it remains hidden until decrypted, but the way it is encrypted ensures that it decrypts uniquely. After this, the blockchain determines which positions in the still secret re-ordered voter list are to receive ballots. Only after the election, to allow checking with voters, is the committed secret ordering decrypted and revealed.

(ii) *Voters are unlinkable to their votes while ensuring votes are recorded as cast:* Voters cast votes using "vote codes," unique to ballot serial number and to "voter choice," such as between "yes" and "no." These vote codes are printed on each ballot next to the corresponding voter choice. Vote codes that are voted are posted online, but which actual voter choice corresponds with which vote code cannot be learned from data online. To allow verification that the correct selection of voters did in fact at least receive the opportunity to vote, the re-ordered voter list is made public once voting is over. (In an optional variant, "verifiers" are selected at random and after close of polls each is provided with the identity of a different one of the voters; later verifier identities but not voter identities are made public.)

(iii) *The tally correctly reflects the votes cast by the selected voters:* A voter receiving a ballot can check that the vote codes printed on it are correctly paired with the adjacent printed voter choices. The voter can also check that the vote code that voter has chosen is correctly posted online. Correctness of the tally is ensured by a series of encrypted commitments linking vote codes to the vote outcomes tallied. Enough of these linked commitments are decrypted responsive to the blockchain after the election to give high assurance that the tally of these outcomes is correct, but not enough are decrypted to reveal which vote code corresponds with which vote. (Such verifiability is similar to that provided voters in elections conducted using Scantegrity in the city of Takoma Park, Maryland.)

(iv) *Vote buying is impractical:* To keep vote-buying in check, any eligible voter can request a decoy ballot. These will not be counted. Requestors of decoys will presumably try to sell the decoys to vote buyers, knowing that these ballots will remain indistinguishable from actual ballots in principle forever.

Those running the election should monitor the market for decoy ballots and endeavor to keep supply such that vote-buyer offers are below a price likely to entice the actual randomly-selected voters to sell. (Since a voter typically would sell to a buyer on the opposite side of an issue, a "proof-of-decoy" is sent separately to the original requestor so only that requestor can be sure the ballot is in fact a decoy.)

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 tosses 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 needed. 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.

As with any election, turnout determines how compelling the vote is with respect to the entire population. With random-sample voting, high turnout can be expected at least because of the weight carried by each vote, the origin of questions put to vote, and the high election quality generally.

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 tournament or ongoing polling, are all corroborated by subsequent votes. Generalization to super-majorities, multi-way contests, and none-of-the-above winners are all easily implemented though perhaps unnecessary as they can be constructed from binary contests.

3. CONTEXT

Random-sample voting has context in the historical and current technics of democracy.

Much of what is regarded as Western culture and civilization, including philosophy, constitutional law, mathematics, science, 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 verifiably-randomly-selected juries without judges or lawyers and using only "yes" "no" secret-ballot votes. (Such juries, sometimes made up of 1001 jurors, could for instance be invoked by any citizen to decide punishment of a legislator for merely proposing a bill deemed unconstitutional and, in case of a guilty verdict, a second vote decided between penalties proposed by each party; other large, randomly-selected juries, called "nomothetai," had the exclusive and final say in revising laws.) A then appreciated attribute of their practice of random selection to fill most government posts, called "sortition," was its resistance to corruption of office holders. (Posts requiring specialized knowledge, such as related to drinking water infrastructure or naval architecture, were filled by election, though with short terms and audit by randomly selected auditors.)

Juror selection in common law countries, albeit heavily post-culled and with unverifiable randomness, and with communication allowed between jurors, 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 order for listing of candidates on ballots, 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 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 deliberate influence and the well-studied techniques developed for manipulating assembled 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. 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 survey 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 current elections themselves and with potential for endogenously framing issues, 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 many US states do provide for initiatives even for 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 voting, are present in representative democracies, where legislation often at least seems to be 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 form of questions put to the electorate. Thus, these second-order concerns may also be obviated.

A related perhaps tertiary problem is what are called election paradoxes (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 voting system 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. Random-sample voting makes such multiple contests practical.

Random-sample voting, according to the Condorcet Jury Theorem, is robust against errors by voters—and resistant to targeted manipulation of opinion—while it taps the wisdom of the crowd.

4. ADOPTION SCENARIOS

One example early-adoption scenario for random-sample voting, flowing from the above discussion, is interest-group initiated surveys of the public on significant issues. Such an election offers the group a way to make an irrefutable statement about the will of a majority of the electorate—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.

Other example uses include public opinion surveys related to current elections. For one thing, these would be more resistant to manipulation and thus potentially of more interest and less likely to be banned. For another, many current elections around the world are disputed, such as by allegations of mixtures of technical fraud as well as polling place discrimination. A parallel random-sample election with improved turnout might neither validate nor invalidate a current election, but it could provide irrefutable evidence of the will of the electorate and thereby prevent exploitation of disputed elections.

A different example use, one that is generative of ideas or suggestions and competitively filters them, is a "tournament" of crowd-sourced issues and wordings that compete to be put to general vote. Such a competition endogenously and transparently arrives at specific language for the questions that prevail. The sample size could be small with larger margins for the initial contests and increase as questions move up. Electoral rules suitable for allocation, such as the variously rediscovered techniques like "approval," "range" or "bee" voting, can be more efficient than a tournament when the number of input questions is large. Questions can be allowed nearly free entry to the selection process by so-called "CAPTCHAS" or letting voters each only pay a limited amount towards the cost of a question.

Specialized knowledge can be incorporated in a variety of ways. Delegated or liquid proxy voting, Delphi panels of experts, and other panel consensus mechanisms, can provide questions and even propose answers—enriching the process but without danger of being able to manipulate or control or otherwise corrupt outcomes. Then random-sample voters would be free to follow the resulting recommendations. Rewarded expertise, such as so-called "prediction markets," could also be applied to prospective public votes, with named persons making predictions and earning economic rewards when accurate. Any impending vote can be expected to stimulate public debate and motivate participation, especially by experts and those with a particular interest in the outcome—thereby tapping hidden knowledge—and such debate can then inform voters when making the deciding random-sample vote.

Elections can be binding on non-governmental organizations. Use within smaller populations, such as volunteer organizations, universities, or corporations, for instance, may allow a wide range of less contentious issues to be efficiently decided by what are in effect randomly-constituted committees, where near unanimity can keep sample size quite small. Other similar examples include: a philanthropic organization allowing a target population to determine use of funds or a political movement deploying the technique internally first and then later advocating wider use.

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 random-sample voting. 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, constituents could be given a veto on any vote for war (interestingly a right granted leading women in the Iroquois nation).

Use by existing governments to strengthen democracy is another type of transitional scenario. When, for example, a popular vote is required by law, yet a full election would be too costly or time consuming, a random-sample vote could be used instead. Culling candidates—greatly improving political party mechanics and reducing media control—allows more voter focus on candidate qualifications and policy positions, such as by using tournament or approval voting rules. Initiative or referenda today require submission of signatures to qualify; however, a random-sample election can be a less easily corrupted and less costly way to qualify a question for a full vote. A further improvement on today's initiatives would, for each voter, randomly assign a single ballot question to that voter.

The main way minorities are protected in a democracy is constitutionally guaranteed rights. Changes in such rights could be safeguarded, to protect against easy manipulation during short-lived periods such as after events, by requiring a continuous period of support established by random-sample voting.

Groups not well aligned with the largely geographical hierarchy of governments today often seek a means to express their will in a vote. Some regions or groups of cities, for instance, do not fit withing governmental jurisdictions; they could conduct elections in order to appoint persons to represent them externally or set internal policy. A legislative body could also use random-sample voting to create irrefutable yet unbiased votes on major policy issues.

There has never been a global referendum. Major issues are increasingly global, but binding democracy does not currently exist between or above nation states. Moreover, democracies, whose sovereignty derives from 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 is no shortage of issues that would likely engender substantial global interest and support, one type of issue specially relevant here relates to information policy and rights: voters being able to obtain information without risk and arrangements favoring transparency and quality of information available, mentioned in

the introduction, are particularly relevant to effective elections and thus to democracy itself.

The cost of 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.) Election workers, perhaps a small team including a translator if needed, would visit the residence to collect and document the vote.

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 it could start really using random-sample voting. What kinds of elections might be considered? Simply deciding the same questions typical of current 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. 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. This can yield more responsive and fine-grained representative democracy.

5. DISCUSSION

Conventional secret-ballot elections today represent at best a paradigm extracted from a somewhat arbitrary choice of era. The technology is about 150 years old, introduced in the U.S. about when male citizens were enfranchised, though women's suffrage took another 50 years. Currently information technology is used to manipulate political processes: for example in redistricting, interest group monitoring of legislator votes, ferreting out and influencing public opinion using random sampling combined with big data and social media. Meanwhile, many voters are only allowed a high-tech simulation, seen through the screens of voting machines, of this paradigm, that was promulgated well after the US constitution but before the first telephone call.

Random-sample voting can thus be interpreted more broadly as providing a way forward, from our current paradigm-based disparity in access to power through information technology, towards governance that is responsive to the will of an engaged electorate.

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 some of one's neighbors or at least the chance to submit a vote on everything no matter how futile. But hopefully there will only be few who oppose the deeper and wider and more continuous monitoring of the will of a more engaged electorate provided by random-sample voting—once efficacy is established—at least informing if not being binding on governance.

CONCLUSION

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

APPENDIX 1

COMPARING RANDOM-SAMPLE VOTING WITH CURRENT ELECTIONS

The ten desired attributes of mass elections listed in the "Election quality" section above are in turn compared to random-sample voting below. These desirable attributes are restated at the end of this Appendix in language that is more general and inclusive of the benefits of random-sample voting.

Current elections can be called "mass" elections to contrast with "random-sample" 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, such as is common with juries and as was practiced in ancient Greece. In some countries voting is required by law; however, random voting may be a larger problem for coerced voting than with paid voting. Even without compensation, and independent of whether voting is mandatory, 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 to election efficacy include unclear statement of issues or even complex bundling of issues. Obfuscation for example of beneficiaries of government policy through voluminous regulation, comprised of convoluted and archaic language, is well known. Random-sample voting lets voters 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 tournaments mentioned in the introduction and discussed in "Adoption scenarios" section, it also holds the potential to directly return us to the day when ballot language and legislation were more comprehensible. Thus, random-sample voting both raises the level of issues that can meaningfully be put before voters and holds significant promise for bringing 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, as far as 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 voting's low cost 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 greatly exceeds that on mechanics of elections, thereby creating decisive leverage for sponsors and lobbies.) This may be due to the shallow diligence of rational voters mentioned above, but it also results from the episodic nature of election cycles. (Such cycles also in a related aspect disadvantageously focus perhaps 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; random-sample voting may thus reduce such influence. (*Resistance to manipulation through advertising and electioneering* [d].)

The security features of random-sample elections is the same as that of the best polling-place mass elections, such as so-called "cryptographic end-to-end systems." Some such mass elections have been conducted. The correctness of the tally has been 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. Unfortunately, the security of all but these few mass elections has remained far lower and relies on unproven and unverifiable trust in such things as polling-place procedures, chain-of-custody for ballots, correctness of proprietary software run on unverifiable platforms, and so forth. (*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 with random-sample voting by flooding the market with willing sellers of decoy votes that will not be counted but that are enduringly indistinguishable from genuine votes; also, potential coercers are unable to find victims since they are unable to effectively learn who has received a ballot. (This also prevents leaking of preliminary election outcome while polls are still open.) 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 reveal how results break down by locality. (*Protection against voter corruption or coercion* [f].)

Selective denial of voter access to ballot casting is common in mass elections. Physical intimidation of certain demographics along access routes to polling places, outside of polling places, or even within polling places, are known even in countries with developed civil society infrastructure. Non-opening or under-equipping of polling places with particular demographic biases also impacts election outcomes. Calls or letters misdirecting voters to wrong polling times or place are often discovered after elections. The system presented here obviates attacks related to misdirection, by including direct addresses on ballot forms; it also obviates the above denial of access attacks since there are no poling places. Longer voting periods, and even rolling elections, can be expected with random-sample voting, making blocking of online access also considerably more difficult than with online mass elections. (*Ensuring access to vote casting* [g].)

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, in some cases there has apparently been concern sufficient to cause restrictions on participation and even disruption of civil governance. The system introduced here substantially reduces such problems, 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, something shared by both types of elections. (*Resistance to voter fraud* [h].)

Decisiveness is the ability of an election to come to a conclusive result. In theory mass elections can adjudicate extraordinarily close contests, even up to tie, yet in practice this may be undesirable since it can amplify the efficacy of even small manipulations. With random-sample voting, when a result falls below a preset threshold of sample size or confidence level, it can be much more easily discarded and re-run with different parameters. (*Decisiveness* [i].)

Cost, compared to a mass election can, as mentioned in the Introduction, be extremely low because the number of voters is far less and cost is primarily per voter. (*Low cost* [j].)

Thus, random-sample elections come significantly closer to the above desiderata when somewhat more generally stated: (a) high effective voter turnout, (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) reduced opportunity for selective denial of voter access, (h) voter fraud only through improper voter rolls, (i) equivalent but safer decisiveness, and all with (j) significantly reduced direct and overall cost.

APPENDIX 2

DETAILED ELECTION PROCESS

A random-sample election can be conducted as illustrated in the diagram on the last page. It shows an example protocol including all the data and an example ballot for an election, using what might be called an "Eperio-style variant of Scantegrity I". This concrete example, introduced in the "Informal overview of technical concept" section above, is well suited to practical implementation and has already been used in elections. Its basic ingredients and techniques are first described from a number of perspectives highlighted below, followed by the complete step-by-step protocol recipe.

In terms of the voting experience, each voter receives by mail a paper form like that shown on the upper left side of the diagram. The voter first chooses freely, and ideally randomly, one of the two ballots printed attached on the form received, in the example either serial number #100a or #100b. The voter then enters serial number of the chosen ballot online. To cast his or her vote, the voter next enters online the unique "vote code" printed on the chosen ballot adjacent the desired "YES" or "NO" vote. For the ballot serial number not voted, at least some digits of both vote codes and their corresponding votes are displayed to voters at some point online, so voters can help check that ballot printing does in fact pair votes with codes correctly. If inconsistency were to be detected, voters would have the printed form half as convincing evidence of improper printing that reveals nothing about the vote cast. All other audits can be conducted independently online by all interested parties on behalf of all voters.

(The red ovals in the diagram follow the example choices made by an example voter through the process, with "A" the choice of the upper ballot on the form and "B" the "NO" vote and its vote code.)

In terms of who does what to during an election, the parties were already introduced in the "Informal summary of technical concept" section and what each does can be further sketched as follows:

The election authority (called "EA" for short)—commit to data in advance by posting it in encrypted form, post information in advance defining blockchain data as well as times, receive requests for decoy ballots from voters, monitor decoy market, print real and decoy ballots intermixed, mail ballots to voters, and reveal during audit keys that decrypt exactly those values previously committed to once selected by pre-defined blockchain data values;

Randomly-selected voters whose votes will be counted—receive a ballot in the mail from the EA, vote by providing vote-codes online, and optionally check codes/ballots online and/or respond to audit inquiry;

Self-selected requestors of decoy ballots—request decoy ballot from the EA, receive decoy ballot in mail as if an actual ballot, check separately-received proof of decoy from EA, and try to sell decoy ballot/vote; and

Self-selected persons who audit the election process—run open-source or self-written software that decrypts (using those keys selected for release by pre-defined blockchain data) the encrypted data published by the EA and checks its consistency with: results of public random data, posted vote codes, published encrypted data, and keys to published encrypted data once released.

In terms of overall election timeline, the centerpiece time interval is when voters are allowed to cast their votes online. There are two other prescribed time intervals, one preceding voting and one following it, during which random values are harvested. The two gaps between these three intervals as

well as both the period immediately preceding and following them are when the EA publishes data.

The first of these four publications by the EA defines the election and locks in a hidden mapping from random values to voters. The second publication allows decoy voters to be fixed, which is followed by printing and mailing of ballots. The third publication, following the voting interval, translates the vote-codes cast by voters into conformance with the data previously published by the EA.

After the final interval, as its fourth publication, the EA is required to reveal some but not all keys used to encrypt previously published commitments. Exactly which keys will be determined by random draw. This release of keys allows auditors to "spot check" with extreme effectiveness that all the encrypted data published by the EA was according to protocol. Combined with cross-checking what voters reported online and report when they are queried by auditors, this confirms correctness both of the random selection of voters and the published tally—but does not compromise ballot secrecy.

In terms of the protocol to be followed by the election authority, an election takes place in nine steps. These are enumerated in the frieze across the top of the diagram and are conducted separately, one after another. Before these steps can start, particulars defining the election must be fixed. These comprise the roster of eligible voters, the maximum number of ballots, the exact language of the question and response options, how the EA will authenticate its postings, when polls will be open, the random draw sources and dates, where voters can vote online, and where all election data will be posted.

The first step consists of the EA using its own private source of random numbers to create tables of data and then independently encrypting and publishing each column of each table (shown yellow). The secrecy of the random numbers used by the EA to create this data protects secrecy of who voted (at least until after close of polls) and which way they voted and indistinguishability of decoys. The second step is a public random draw, detailed below, the result of which is preferably locked in by the EA reformatting and publishing it. In the third step, responsive to requests for decoy ballots, the EA publishes additional encrypted columns (shown green), assigning the decoy voters to the decoy ballot slots that were dispersed indistinguishably among the real (i.e. non-decoy) ballots in the first step. The third step also includes the EA physically printing and mailing ballots.

The fourth step, the actual voting from opening of polls to closing of polls, accepts online submission of votes on a so-called "electronic bulletin-board," which makes all the serial numbers and vote codes submitted available to any auditor. This is followed by the fifth step, which translates the resulting posted vote codes and serial numbers into two additional columns (shown blue) that are encrypted and posted to complete the encrypted tables.

The sixth step, like the second, is an independent public random event. It determines exactly which keys the EA must release for each of the three remaining steps. The keys the EA releases in the seventh step let anyone crosscheck consistency between the voter bulletin-board's public data and the published encrypted data; the public yet unpredictable nature of this choice of which parts the EA must make decryptable allows anyone to confirm that the bulletin board is consistent with all the encrypted data. The eighth step is the release of further keys that, in addition to allowing checking internal consistency of the published tables, reveal the tally. The ninth and final step is the release of just enough additional keys to reveal the voters, so that anyone can check with voters to ensure that they did in fact receive the correct ballots, without revealing which ballots were decoys or who voted which way.

In terms of data posted by the election authority, apart from definition of the election and reformatted public random data, all that the EA posts for an election is a number of identically-formatted encrypted

tables and eventually keys decrypting some of their encryptions. More specifically, each table consists of eight columns and each column of each table is encrypted with a different secret column key chosen randomly by the election authority; the authority only reveals some of these column keys, corresponding to some subset of tables, and only during audit, with the choice of which keys to reveal determined by the final draw.

To create each table, the EA first builds a "canonical" table that simply lists in a standard order all the serial numbers, vote codes, possible votes, and a pair of numbers. Then, the EA transforms the table, using its own secret random parameters, in order to to keep secret the linking between voters and votes, which ballots are decoys, and (at least until after close of polls) who voted. One transformation is "row-shuffling," resulting in an unpredictable re-arrangement of the rows, so a particular row almost always appears in a different position in the table. Another transformation further randomizes the pairs of random numbers, called the first and second "summands," but keeps their sum unchanged. (The random values of the summands are the same per serial number at this point.) The third and final transformation encrypts each column as a whole, using a secret key that is unique to this table and column. To protect election integrity, the class of encryption used ensures that column content is unalterably committed to, as mathematically there is only a single decryption possible per encryption.

The specific example illustrated has 250 tables, the number recommended for an election of any size. This particular example sends a "double-ballot" (two "single-ballots," printed attached, whose serial number differs only in the "a" or "b" suffix) form to each of 1,000 voters. (The double-ballots are divided somehow between real and decoy.) The double-ballots thus result in 2,000 distinct ballot serial numbers and, since there are two choices per ballot, 4,000 rows per table. The 1,000 double-ballots are each assigned a voter among the example 10,000 potential voters on the posted list (sometimes called a voter or registration or electoral roll or file), which will be called the "roster." Instead of posting all columns of the tables at once, four columns are posted first (yellow, 1, 3, 5, and 6), then two more to select the decoys (green, 7 and 8) before voting but after the initial draw, and finally the remaining two (blue, 2 and 4) after voting to lock-in the votes.

The sum of all three summands that correspond to a serial number yields the position in the roster of the voter who should get that ballot with that serial number. (This sum is the same for both ballots of the double-ballot sent the voter.) If the sum is less than the number of voters, it is simply the row number of the voter in the roster, numbered from zero to the number of voters minus one; if, however, the sum of the three numbers is larger, it is reduced to fit using so-called "modular arithmetic." This means that each summand can influence the result to be any voter (just as allowing a choice from zero to eleven can shift the hour on the face of a clock to any hour). Ensuring that the EA choice of summands does not result in any voter receiving more than one ballot is accomplished by audit.

(The red ovals in the diagram follow the example vote already mentioned along its row in table number one: in oval "C" the ballot serial number and vote code; 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. Simply 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. These example summands were chosen for convenience with repeated digits and not requiring carry; however, as summands are chosen from zero to the number of entries in the roster minus one, 0 to 9,999,999 in the example, carry and modular arithmetic would be needed for typical values.)

In terms of implementation system security, all secrets of the EA can be stored in encrypted form

between steps. An election committee then meets for each step and supplies so-called “passphrases” in person to a fresh installation of a computer that they witness. This computer then decrypts the stored EA state, conducts the step, outputs the new encrypted state for storage, and wipes its own memory.

The online bulletin board need only be provided by the EA with checksums for vote codes and respective serial numbers. This lets the bulletin board confirm to voters that they have entered the vote code correctly even though there are many vote codes per checksum. The bulletin-board would only be able to learn codes voted and thus unable to create or change votes. Voters are instructed to write their voted vote-code on the unvoted single-ballot, destroy the voted single-ballot to protect the secrecy of how they voted, but then keep the unvoted single-ballot until they can check it once the choices on the unvoted single-ballots are publicly linked to vote codes during audit.

Random-sample elections lack the time urgency of mass elections and accordingly are less subject to so-called “denial of service” attacks online, which are typically short lived. Discreditation by voters would involve claims that they did not receive ballots or that the forms were printed incorrectly. These are effectively addressed by certified mailing of ballots and standard document security techniques applied to ballot forms. The half-ballot each voter is instructed to keep can serve as evidence of malfeasance without betraying how the voter voted.

If someone were to wish to bias the selection of voters or change the election outcome, at least with a significant probability of success, compromising all EA committee members would not help. One potential type of attack, however, would be to somehow change the random draws. Other potential attacks would involve somehow tricking enough voters into not following protocol or into falsely believing that they are in secured communication with the voter bulletin board. Auditors would use so-called “public-key digital signature” authentication specified in the election definition in order to authenticate all information that they receive from the EA, bulletin board, and public random sources.

If someone were to wish to learn how particular voters have voted or to distinguish decoy ballots, one attack would be to seek passphrases from sufficient committee members, though this should become impossible once enough members erase their passphrases after the election. Another attack would be to somehow tap or subvert the computer used by the committee at a meeting.

If someone were to wish to merely learn the recipients of ballots early, they could do so by observing the envelope printing or mailing. Using this information to try to influence voters, however, might be risky since it could be revealed by voters or even by decoy voters aiming to entrap.

In terms of the public random draws, called out as Step (2) and Step (6) in the diagram and recipe below, these are based on publicly-verifiable yet unpredictable data. The blockchain at a time certain can be an ideal source for this. (Example sources that have also been proposed are closing prices of a public market or temperatures from a number of cities.) What will be called here the “beacon” uses a pre-defined method that completely determines the public outcome from the definition of such an unpredictable value, once it becomes available after the predefined time of the corresponding “draw” of random bits. The initial draw determines the public list of third summands (in effect a permutation of the voter list), used to make the selection of voters unpredictable and unmanipulatable by the election authority; the final draw determines which of the column keys are revealed for audit.

In terms of audit, this process is divided among the last three steps. Step (7) makes sure that the vote-codes cast by voters recorded on the bulletin board were correctly copied into the encrypted tables and that the printed ballots correctly associate votes with vote codes. Step (8) reveals the tally. It also

checks that the real ballot rows were not changed when the decoy ballots were interspersed. Step (9) lets anyone check with voters to make sure that they did in fact receive the correct double-ballot. These steps correspond roughly to customary phases in mass elections, where preliminary checking is done before the results are announced and an extensive canvasing precedes final certification of the result.

The surprising public auditability of election integrity while maintaining ballot-secrecy, decoy-indistinguishability, and, at least temporarily, voter-anonymity, results from use of a so-called "you cut and I choose" protocol: the EA posts the encrypted tables but the random beacon chooses which columns to open. This prevents the EA from posting anything but correct tables—otherwise the discrepancy would, at least with extremely high probability, be detected. The reason is because the final draw determines unpredictably which one of the five "batches" each table ends up in and each batch decrypts a different combination of columns and those combinations cover all columns in a sufficiently 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; very 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 ballot serial number was voted nor which serial numbers are decoys.

Specifically, Step (6) divides the 250 tables into five "batches," each of which contains 50 tables, but in a way that was completely unpredictable to the EA when the tables were committed to by being published in Steps (1), (3), and (5). Each batch corresponds to a particular selection of table columns whose keys are to be revealed by the EA during a corresponding step, in effect "opening" the already committed but hidden content of the selected columns to public inspection. Two batches of tables are opened in Step (7), two in Step (8), and the remaining batch finally in Step (9). The diagram shows lines below the tables grouped in batches, indicating which columns are opened for which batches.

The first audit, Step (7), labeled "audit casting & printing" in the diagram, reveals keys for two batches to allow checking consistency between the tables, the voter bulletin-board, and ballot forms, without revealing the tally or voters. The first batch of columns opened lets anyone check that the vote codes, as posted on the bulletin-board and identified by serial number, are recorded correctly and marked as "voted" (and "not checked") in the decrypted data. The second batch of columns opened lets anyone cross-check, at least for those ballots not voted but checked by voters, that the serial numbers and corresponding vote codes were associated with the correct "YES" and "NO" ballot choice in the encrypted data, ensuring that the ballots were printed with the correct association of codes to votes.

The middle audit, Step (8), "audit tally & voter selection," decrypts the third and fourth batch columns to reveal the votes so that they can be tallied while revealing the decoys to ensure that they are not tallied. Anyone can then add up the revealed votes and check that each of the 100 tables yields the same election result. These two batches also let anyone check that the list of who should get a ballot was correctly copied from the real voters committed to before, and randomized by, the first draw—even though it lets pre-arranged interspersed positions be filled by whatever decoy voters the EA chooses.

The last audit, Step (9), "reveal all voters," opens the fifth and final batch and thereby lets anyone see who which ballot should have been sent to, so that voters can be contacted and asked whether they received the correct serial number on their double-ballot. It discloses all roster entries to whom ballots were to be mailed and the respective serial numbers, but nothing about how the ballots were voted or whether the voters/ballots were real or decoy. It decrypts the serial number column and a pair of summand columns. Also checked (assuming that any duplicate real ballots, those that would result in more than one ballot of any type per voter, should verifiably be deleted as detailed below) is that no

voter should get more than one ballot.

In terms of properties achieved, the three audit steps address the first three properties already mentioned in the "Informal summary of technical concept" section above:

Property (i), "The selection of voters is at random and cannot be chosen in advance or manipulated," is addressed by Step (7), since the third summands are unpredictable and unmanipulably posted once the first and second summands have, in advance of this step, been publicly committed to. (No voter receives more than one ballot, as Step (8) and Step (9) ensure that ballots were not sent to those voters chosen more than once.)

Property (ii), "Voters are unlinkable to their votes while ensuring votes are recorded as cast," still holds even after all the auditing is done. This is addressed by batches 2, 3, and 4 being the only batches including actual votes (column 3) and batch 2 only linking to details of non-voted ballots, and batches 3 and 4 only linking to single random summands and/or decoy indications—and by columns decrypted in these batches being linked neither to voted vote-codes nor voted summand pairs.

Property (iii), "The tally correctly reflects the votes cast by the selected voters," is addressed by three aspects of the audits. First, that columns 2 and 3 correctly link "YES" and "NO" to the unique vote codes printed. This is established with high probability by voters checking their unvoted single-ballot against data revealed in Step (7). Second, that these "YES" and "NO" postings are correctly linked to voted but non-decoy column 5 or 6 entries and that all the tables of the third and fourth batches have the same tally, both as checked by audit Step (8). And third, that all voters listed in the roster at the positions determined by the column 5 and 6 summand pairs (which, as verified in Step (8), highly-likely match column 7 and 8 pairs) were provided non-decoy ballots with corresponding serial numbers (which, as verified in audit Step (7), highly-likely match column 2 serial numbers) as verified in audit Step (9).

In terms of potential extensions, distributed variants are possible where the election authority is replaced by a multiparty computation that protects privacy from all but collusion or compromise of a majority of multiple election authorities, including the option of election authorities per portion of the electorate. The bulletin board can also be distributed, multiple ballots on a form obviated, and confirmation codes provided to voters. Scratch-off on vote codes gives tangible evidence if the EA were to cast a ballot instead of the voter. Proofs that a ballot will not be counted can be provided to decoy voters at time of ballot issue by separate channel; such a proof, not includd in the basic protocol here for clarity, can be as simple as a list of rows where the ballot appears and a pre-voting audit of columns 1 and 5 based on a draw after the proofs are received. The number of ballots can be hidden until close of polls by marking column 6 and 8 as "VOID" for some ballots. The optional variant to canvasing mentioned earlier, where verifiers provide anonymity of voters, can be substituted by using an elaboration of the techniques detailed below.*

The complete detailed step-by-step recipe, based on the ingredients and diagram already introduced for conducting high-quality random-sample elections, is now detailed in the nine already-mentioned steps:

(1) [EA] Four encrypted columns (yellow, 1, 3, 5 and 6) are posted by the EA in this first step. To generate these, the EA first forms a four-column canonical table. Column 1 of this table is filled by assigning serial numbers for all ballots sequentially down the rows, two for each "a" followed by two for each "b" suffix, and assigning a different random vote-code per row. Column 3 is then filled simply by alternating "YES" and "NO" per row. Next the desired number of decoy ballots determines how many of the adjacent four-row double-ballot "blocks," those with the same numerical serial number, are

randomly selected and marked “decoy ballot” in columns 5 and 6. The remaining blocks each get instead a random summand copied across all four of their column 5 entries and another independent random summand copied across all four of their column 6 entries, where summands are chosen from zero up to one less than the number of voters on the roster.

To form each of the actual 250 tables posted, the EA begins with a copy of the canonical table and transforms it, first with two randomizations and then with encryption. One randomization permutes the rows, with all four column entries making up a row remaining together in whatever row they are moved to. The other randomization changes the summands of columns 5 and 6, while preserving the sum per row, such as by adding a random value to one summand and subtracting the same value from the other (with summands always represented modulo the roster size). The authority finally then separately encrypts each column 1, 3, 5 and 6 with an independent secret key. Once all 1,000 columns are encrypted using their respective one of the 1,000 keys, the EA posts all 250 encrypted tables.

(2) [beacon; EA] Initial random numbers are drawn in a public and unpredictable manner once both Step (1) is completed and the roster of voters (blue rectangle, left side below ballot) is fixed. These numbers consist of third summands (purple rectangle, lower left corner) each labeled by its respective unique double-ballot serial number. The public rule defining how these third summands are formed from the draw, fixed in advance of this step, preferably ensures that they are independently and uniformly distributed from zero to one minus the number of roster entries. The rule may be as simple as taking successive fresh, suitably-sized strings of raw random-draw bits while skipping in their entirety those strings representing a value equal to or larger than the number of voters on the roster (though a more efficient algorithm could be specified in the election definition).

(3) [EA] Two additional encrypted columns (green, 7 and 8) are posted by the EA, who also then prints and mails out the double-ballot forms. To generate the two columns for a particular table, the EA first copies all the pre-draw summands from columns 5 and 6 to form new respective columns 7 and 8 entries. Where column 5 and 6 are marked “decoy ballot,” however, the authority places a random summand in column 7 and then computes the respective column 8 value so that the sum of all three summands is the position in the roster of the decoy voter assigned that block. (Since each voter is to receive at most one ballot, in the unlikely-at-scale event that column 7 and 8 entries would cause a voter to receive more than one real ballot, all rows corresponding to such ballots include both the summands and the mark “DUPLICATE” in columns 7 and 8; an agreed policy limiting the frequency of voter selection across multiple elections can also be verifiably realized by such marking; and decoys are not assigned, though statistically insignificant at scale, to voters who will receive real ballots.) Once all 250 columns 7 and 8 are formed in this way, the election authority encrypts each, using a respective one of the 500 independent secret keys, and then posts the encryptions. After printing the double-ballot forms, the authority finds the respective mailing address for each by looking up the third summand in the summand list by its serial number and then using the sum of the three summands, modulo the roster size, as the row number in the roster.

(4) [voters; bulletin-board] Voters cast votes by posting on the electronic bulletin-board the ballot serial number along with the corresponding vote code each wishes to vote. Voters are requested to check that the unvoted single-ballot of the double-ballot form they received is posted with its vote codes correctly matching votes in Step (7). The bulletin-board posts all voted serial numbers and corresponding vote codes.

(5) [EA] Upon close of polls, the EA forms the two remaining columns (blue, 2 and 4), encrypts them using the final 500 respective independent keys, and posts the result. Column 2 records pairs of serial

numbers and vote codes for the non-voted ballots of each voted double-ballot. Column 4 marks, without distinguishing real from decoy ballots, the rows whose vote codes were voted. The remaining entries are filled “not checked” for column 2 and “not voted” for column 4.

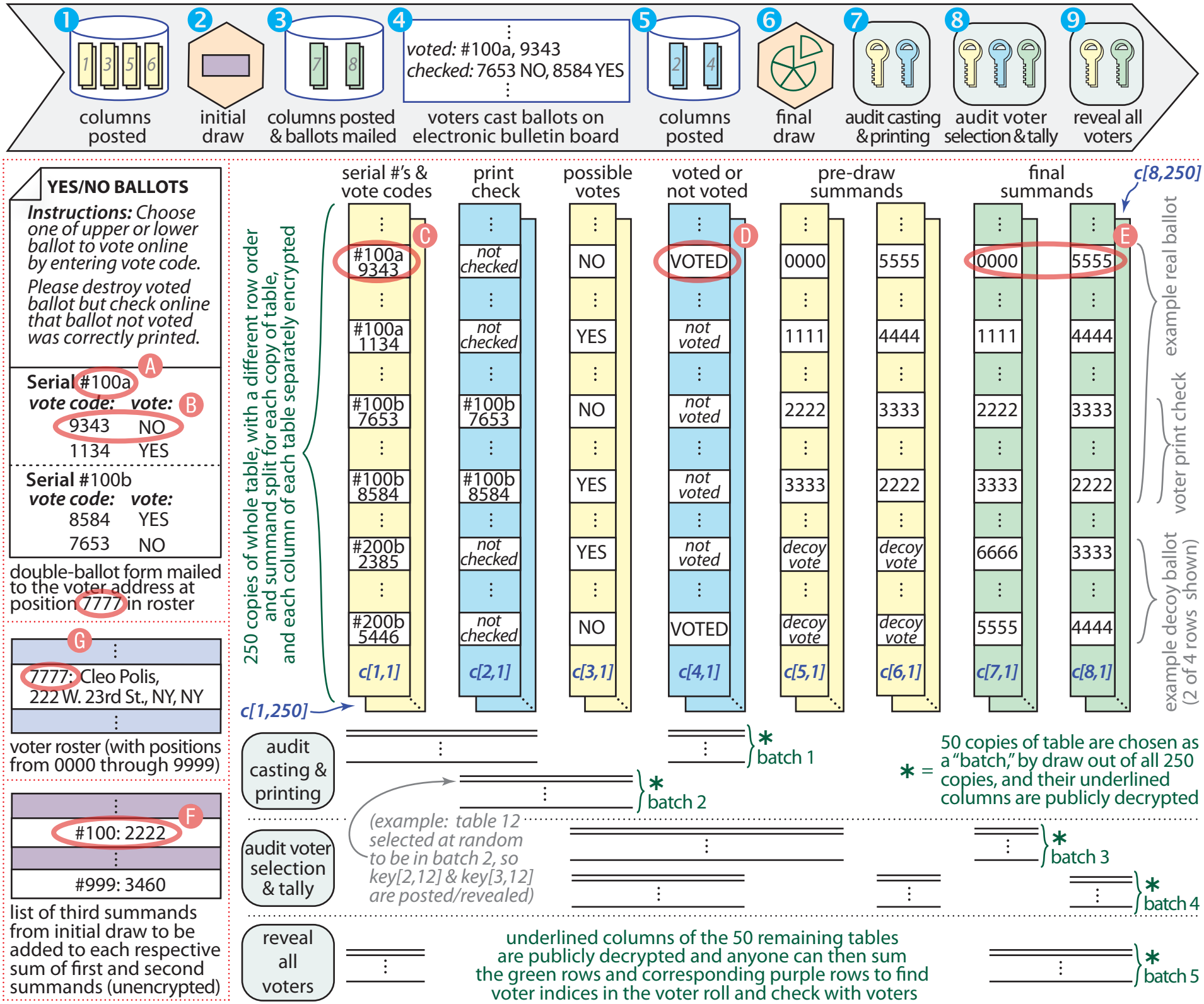
(6) [beacon; EA] The final public random draw is performed after the output of Step (5) is posted. This draw is used to unpredictably divide the 250 tables into five batches: 50 tables for each respective half of Step (7) and (8), with the remaining 50 for Step (9). The pre-defined rule for this division can be to place the first 50 table numbers that appear in the successive bytes of the random draw into the first batch, the second into the second, and so on, skipping any repeat or out-of-bound values. Each table appears in exactly one batch. So that audit can begin shortly after polls close, this draw is ideally completed quickly. (This draw can be shared by multiple concurrent elections and/or multiple such draws can be used for an election in a pre-defined way, such as allowing some tables to be assigned to batches early and others only later when different or more trustworthy random values are available.)

(7) [EA; voters; auditors] This first audit can be conducted as soon as Step (6) is completed. The EA releases keys allowing decryption of columns 1, 2 and 4 for the first batch of 50 tables and columns 2 and 3 for the second 50. The first batch allows voter and auditor crosschecking, bulletin-board against columns, of the voted and not voted serial numbers, with respective vote codes, as well as that serial numbers and vote codes of columns 1 and 2 match; the second batch allows crosschecking, bulletin board against columns, of the pairing of vote codes and votes printed on unvoted ballots.

(8) [EA; auditors] This second audit is 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. For the third batch, decrypted columns 5 and 7 should be identical, as should columns 6 and 8 for the fourth batch, unless “decoy ballot” appears in the first column of the pair (in which case “DUPLICATE” is not allowed). 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, which should be the same for all tables in the two batches.

(9) [EA; auditors] The positions in the voter roster containing addresses to which ballots were sent are revealed by the EA in this final/canvass audit step, but without revealing whether the ballots were real or decoy or how they were voted. The remaining batch of 50 tables is publicly decrypted in columns 1, 7 and 8. No voter should receive more than one ballot and each such ballot should have the same serial number for its block in each table of the fifth batch. 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, any auditor can locate the addresses in the voter roster to which ballots should have been sent. (Rows marked “DUPLICATE” should correspond to each voter who would have received more than one real ballot, and auditors check that all such rows should be so marked; but no such voter should be sent any ballot.) The auditor can then verify, such as by alerting voters, checking mail receipt signatures or online receipt records, or even by directly contacting voters, that these voters did in fact receive corresponding serial-numbered ballots and that none of these voters complained that serial numbers and vote codes were improperly 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 destroy voted ballot but 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 roster
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
decoy vote
5555
4444
6666
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
voter print check
example decoy ballot (2 of 4 rows shown)
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