

source code review; (2) documentation review; (3) system level testing; (4) security/penetration testing; (5) privacy analysis; and (6) usability analysis. The functional examination was open to the public and was videotaped by Department staff.

18. As a result of the examination, several enhancements were made to facilitate proper handling of Pennsylvania's straight party voting method on the ExpressVote and ExpressVote XL. In addition, performance enhancements were made to the Electionware Reporting module which reads in results media from the voting machines and generates all jurisdiction-required election reports. EVS 6.0.2.1 voting system incorporated those enhancements into an updated release. The system components remained the same; the only change in the new release were the aforementioned software enhancements.

19. EVS 6.0.2.1 voting system underwent independent testing in September 2018 to obtain EAC certification and certification by the Pennsylvania Department of State. The Department's examiner determined that the EVS 6.0.2.1 voting system release complied with Article XI-A of the Pennsylvania Election Code and certified it on November 30, 2018. Likewise, on November 12, 2018, the EAC certified that EVS 6.0.2.1 complied with VVSG.

20. As a result of a petition by a group of individuals on July 17, 2019—over eight months after the certification by the state—the Pennsylvania Department of State undertook a re-examination of the ExpressVote XL.

21. The Department’s examiner, SLI Compliance, again performed a security review, functional review and documentation review of the ExpressVote XL on August 7 and 8, 2019. The re-examination focused on the petitioners’ allegations that the ExpressVote XL violated Sections 1107-A(1) and (12) of the Pennsylvania Election Code, 25 P.S. § 3031.7(1) & (12).

22. The Department and its examiner again determined that the ExpressVote XL satisfies the requirements of the Election Code and maintained its certification.

c. Additional Testing & Security Procedures

23. In addition to the above testing and certification steps, extensive testing of the voting system is performed before each election. For all types of ballots, jurisdictions perform pre-election logic and accuracy tests and post-election audits to ensure the accuracy of the tabulation process. Logic and accuracy tests verify the readiness of the system for the specific election to be conducted.

24. A post-election audit also verifies that the equipment and procedures used to count votes during an election worked properly, and that the election yielded the correct outcome. Logic and accuracy testing and post-election auditing

provide a testable and auditable method to verify that ballots are programmed correctly and counted as the voter intended.

III. The ExpressVote XL is Based on a Secure Design

25. Every election reinforces the importance of voting as the foundation of America's democracy. Nothing is more important to ES&S than maintaining the integrity of the voting process.

26. Plaintiffs' declarant states that it is "feasible for malware to cause the machines to print bar codes that corresponded to candidates the voter did not select." Declaration of Alex Halderman dated November 21, 2019, ("Halderman Decl.") ¶ 8. The declarant also claims that "it would be feasible for malware to compromise [the tabulator] and cause paper records that have been rejected by voters to be tabulated as well as those that have been accepted by voters." Halderman Decl. ¶ 7. Such a compromise of the system is not feasible as a practical matter.

27. There is no record that any of the hypothetical malicious cybersecurity breaches Plaintiffs describe have ever taken place, either in actual elections or in testing of the ExpressVote XL. The design of the ExpressVote XL, coupled with all the other layers of security that protect ES&S's voting systems, means that any of these breaches are a practical impossibility.

28. Election systems provided by ES&S employ a concept called “security in depth” which utilizes layers of overlapping security measures to ensure that the compromise of a single protection feature does not result in the compromise of the system’s integrity.

29. System-level protections include hardened software for programming the election, encryption and digital signing of data to and from the machines, use of authorized data transfer media, and locks and seals for physical protection.

30. Voting machines that are a part of this system utilize many of these same measures to prevent “hacking” or cybersecurity compromises. In addition to those mentioned, voting machines certified to the VVSG1.0 standard are never connected to the internet.

31. ES&S voting machines also have multiple layers of security on the data, including unique encryption keys for every election. The machines will only accept the certified type of USB media and will ignore all other types. The machines also verify that the media has been programmed for the expected purpose, was programmed by the trusted software, and has not been modified in any way.

32. The ExpressVote XL provides the VVSG-required ability to validate the operating system and application software installed on the machine, allowing for audits before and after each election.

33. All election programming data is stored and encrypted until it is loaded into random access memory for use during machine operation. At such time, its digital signature is validated to confirm that the data is from the trusted source and has not been altered. All vote and results data generated is encrypted and digitally signed.

34. The ExpressVote XL generates its own public/private key pair for each election and digitally signs each artifact (log files, vote records, results files, etc.) generated by the device during the election. The digital signatures are validated by the system before any artifact is processed, ensuring that should unauthorized access of a unit occur, it will be detected and reported to the user so no other units can be affected through data transfer.

35. The ExpressVote XL generates a detailed audit log of all actions and events that have occurred on the unit, which can be printed at any time. Every action and event, including access attempts, access of system functions and errors is logged and timestamped. The log is digitally signed each time it is updated to allow detection of any potential tampering with the entries. The vote data is encrypted and digitally signed.

36. ES&S also conducts thorough security reviews of its entire supply chain to ensure that every component is trusted, tested and free of malware. Every single item and manufacturer is approved and under engineering revision control to

ensure that only authorized components and firmware are used in the production of voting machines.

37. Moreover, even in the extremely unlikely circumstance that these security measures were breached, such breach would be identified in a post-election recount or audit. Because the ExpressVote XL provides a voter-verified paper record of every vote, election officials would be able to accurately recount the votes using the verified paper record.

IV. The ExpressVote XL Matches the Barcode on the Paper Record to the Candidate Names

38. ES&S considers paper records to be critical for auditing. As such, in 2018 ES&S decided to no longer sell paperless voting machines as the primary voting device for any jurisdiction given that it is difficult to perform a meaningful audit without a paper record of each voter's selections. Using a physical paper record sets the stage for all jurisdictions to perform statistically valid post-election audits should they choose.

39. The paper record of votes created by the ExpressVote XL includes both printed candidate names and barcodes that are read by a tabulator. Barcodes are a trusted, tested, universal technology used in a variety of ways across many different industries to improve safety, accuracy, speed and efficiency. DMVs, pharmacies, hospitals, banks and food manufacturers all use barcodes. They are

reliable and have been used commercially for more than 50 years. They are used extensively in the medical field where errors could be catastrophic.

40. A barcode is a pattern code—a group of lines or blocks and spaces that represent specific characters—that computers can read automatically to identify information in a database. Because barcodes offer a reliable way to accurately read information, the technology eliminates the possibility of many kinds of human error (e.g., poorly marked ballots) and provide a layer of tamper resistance, as they are virtually unmodifiable, especially when employing a check digit or signature. Displayed along with human-readable text, summary cards with barcodes are fully auditable.

41. Whether votes are cast on a hand-marked paper ballot or a machine-marked paper ballot, when paper ballots are tabulated by machines, barcodes are used to count votes.

42. Both hand-marked paper ballots and machine-marked paper ballots are secure methods of casting a vote. Vote counting machines (called tabulators) are used in both instances and read barcodes in the same way they read the oval positions on a paper ballot; thus, a summary card with barcodes contains the same essential data as on a hand-marked ballot. In the case of both hand-marked and machine-marked paper ballots, these lines translate to numbers that are grid

coordinates and those grid coordinates correspond to a candidate name in a database.

43. On a hand-marked paper ballot, often referred to as an oval-filled ballot, voters make their choices by filling in by hand the ovals next to their desired selection. When tabulating a hand-marked paper ballot, the system works as follows:

- a. There is a barcode, typically called a code channel, along the left edge of the ballot and the top and/or bottom of a hand-marked paper ballot that indicates the ballot that is being counted.
- b. When a voter hand marks the oval next to a particular candidate and inserts the hand-marked paper ballot into a tabulation machine, that tabulation machine does not read the candidate's name.
- c. In fact, the tabulation machine does not recognize the text of the candidate's name at all, nor does it even determine that the oval position for a particular candidate is next to the candidate's name. Instead, the tabulation machine reads the code channel to determine the ballot being read and then retrieves the programmed location of all of the ovals. It then reads the ovals and sends a list of the oval positions that are marked.
- d. Even though the candidate's name appears on the ballot, it is actually the name in the database corresponding to the programmed oval position for which a vote is recorded.

44. Machine-marked paper ballots work the same way, except that the voter makes their choices by touching a screen by hand, or by using an assistive device, instead of holding an ink pen. For example:

- a. When the voter selects a particular candidate's name on the touch screen, the marking device prints out a paper record with

the text of the candidate's name along with a barcode corresponding to the coordinates of that candidate's location.

- b. When the paper record is inserted into the tabulator, the tabulator performs the same actions as it does with the hand-marked paper ballot. It reads the master barcode, which performs the same purpose of identifying the exact ballot style being counted. It then reads the selection barcodes and returns a list of marked candidate locations.
- c. The candidate names in the database corresponding to each of the positions returned has a vote recorded for those candidates.

45. Just as in the case of hand-marked paper ballots, the tabulation machine in the ExpressVote XL is only looking for grid coordinates. The casted vote records from both examples are identical.

46. A mismatch between the barcode on the paper record and the words on the paper record has never occurred on an ExpressVote or an ExpressVote XL, although it has happened many times on oval-filled ballot systems.

47. For oval-filled ballots, such an aberration could occur in two ways. First, the election coding could be changed after the ballots are printed, such that the machine is programmed with coding that does not match the printed ballots. If testing is not performed correctly, the discrepancy will not be caught. Second, if oval-filled election ballots are created by hand and the coding (i.e., the locations of the ovals) is manually programmed to match this layout, human error can cause a discrepancy.

48. These aberrations are not possible for the ExpressVote or the ExpressVote XL, because the ballot is not pre-printed, and the ballots and coding are sent as part of the same data set to the machine.

49. In the case of both a hand-marked ballot and a ballot marked by a ballot-marking device (“BMD”), the voter can only validate the human-readable content. The voter cannot determine that the tabulator portion of the process correctly captured their intent and properly assigned the votes to the proper candidates. The ExpressVote XL does, however, allow a voter to read the selections encoded in the barcodes back to them using the display or audio capabilities which provide a verification that typical ballot scanners do not provide for oval-filled ballots.

V. The ExpressVote XL Printer Does Not Touch the Paper a Second Time After the Barcode is Affixed.

50. As Plaintiffs’ declarant states, after the voter verifies his or her selections on the paper record, the paper record passes back through the machine on its way to a collection bin. Halderman Decl. ¶ 6. The declarant continues that “[i]t would be feasible for malware to tamper with this function and cause the printhead to add additional races or selections to the paper after the voter has reviewed it.” *Id.* This is not plausible, for several reasons.

51. The ExpressVote XL creates a paper ballot based on a voter’s selections, which is tabulated when the voter affirms that he/she is ready to cast. It

allows voters to review their selections before printing for tabulation on scanners. The print head is located such that, after the initial print, the vote summary record passes to the collection bin without making contact with the print head again during the vote summary record deposit process.

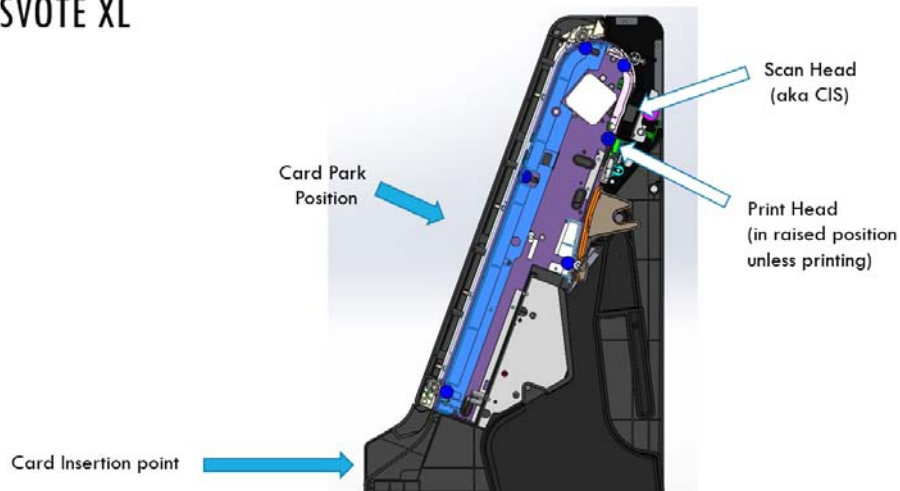
52. To ensure that the printer does not touch the paper again after the voter certifies, the print head has sensors to indicate that it has been raised off the path, and it would generate an error if it did not successfully raise. In addition, the paper moves more quickly when it moves under the printer after verification. If the printer was in the down position when the card was passed through while being cast into the bin, the paper transport motor would stall, and the result would be a card jam. That would raise an error. The print head is always raised from the paper path when not printing. Images of this aspect of the ExpressVote XL are below:



53.

Paper Path Module (PPM)

EXPRESSVOTE XL



54. Further, the printing process is audible and therefore able to be detected.

55. Even if the print head touched the paper as it passed through a second time, it is extremely unlikely that this could alter a vote. First, the printer prints the

card from bottom to top. Any over-printing as the card is ejected would require the printing to occur from top to bottom. Second, the barcodes that are used include a check digit, such that the barcode is nearly impossible to modify. Finally, to do so would require that the card pass through the paper path with no shifting or skewing, which is also virtually impossible.

56. In addition, the system can be programmed to leave no spaces between human-readable contest or candidate text, such that there is no available space to add a human-readable selection.

57. Finally, the master barcode indicates the number of selection barcodes that are present. If a barcode were added, subsequent scans would fail validation, and the added barcode would be detected.

58. The ExpressVote XL design renders it is extremely unlikely that an erroneous vote (i.e., one that the voter rejected) could nonetheless be tabulated. Such a scenario would require the voting machine's security to be compromised, which is equally possible on oval-filled ballot scanners, and, indeed, on any system that relies on technology for tabulation of votes.

59. Finally, in the nearly impossible scenario that a malicious agent could somehow evade all of these security features, the discrepancy between the human-readable record and the altered bar codes would be easily detectable in an audit.

VI. The Northampton Election is Not a Basis to Decertify the ExpressVote XL Statewide

60. The issues affecting the 2019 election in Northampton County are unrelated to the complaints that Plaintiffs now assert in their motion. Notably, the ExpressVote XL was used successfully in Philadelphia's 2019 elections.

61. There were two issues with the 2019 Northampton County election. First, some contests reported incorrect results for some candidates. The paper ballots printed by the ExpressVote XL were used as a backup, and those ballots were used to correctly tabulate the votes.

62. Thorough analysis and extensive testing following the election established that the issue was caused by a ballot layout technique that had not been properly tested and that caused the system to not attribute votes to certain candidates.

63. In particular, the ballot included instructional text on the ballot. Although the ballot was displayed correctly on the screen and on paper ballots, the instructional text error created a misalignment in the database that resulted in votes not being attributed to candidates in a particular position relative to the instructional text. This was confirmed through testing that replicated the exact ballot used in Northampton County.

64. In this replication, although the voter's selections displayed correctly on the screen and on the printed ballot, the system assigned the votes to the

instructional text, thereby preventing the votes from registering for the intended candidate.

65. When the instructional text was removed in a recoded version of the election made for analysis purposes, the results were captured accurately and showed all of the corresponding vote totals for all candidates. This instructional text issue was ultimately human error in using a layout technique that was not fully tested. Tabulators like the high-speed scanners used by the customer do not display the ballot and do not consider the instructional text; they therefore counted the votes without issue by simply rescanning the paper ballots printed by the ExpressVote XL.

66. Second, some voters reported difficulty in making selections on the machines' touchscreens. In particular, some voters reported that their touch was not registering the proper location, or at all, and that the touch screen was too sensitive. They were nonetheless able to verify their votes with the paper ballots printed by the ExpressVote XL.

67. Extensive testing indicated that the primary cause of difficulties with the touch screen displays was that some machines (as many as 30%) were configured improperly at the factory prior to delivery to Northampton County. In addition, the layout of the ballot unnecessarily pushed voting selection areas to the very edge of the touch screen where the screens can be less responsive.

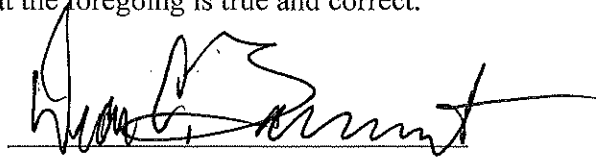
68. The ExpressVote XL uses infrared touch screen technology. Because the sensors cover the entire border, it is possible for selections within the border area to not register properly. Small boxes within the candidate selection box may have also caused an inadvertent selection, by prompting voters to make selections that were on the “border” of another selection option. These issues will be resolved with an improved ballot layout and corrected machine configuration.

69. Despite the issues with that specific election, the results, as obtained through the ExpressVote XL machines, were still verifiably accurate. Each voter had the opportunity to read and verify their ballot before they cast it, and the paper ballots that the voters read were used to tabulate the total.

70. In addition, the paper ballots were scanned, tabulated by election officials, and audited for accuracy. The audit confirmed the accuracy of the election results.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on DECEMBER 12, 2019.

A handwritten signature in black ink, appearing to read "Dean C. Baumert", is written over a horizontal line. The signature is stylized and cursive.

Dean C. Baumert

EXHIBIT 3



Home › Resources › Voting Equipment › Election Systems and Software (ES&S) › ES&S ExpressVote XL

ES&S ExpressVote XL



With the EVS 6.0.0.0 system certified by the EAC on July 2, 2018, ES&S introduced the ExpressVote XL, a hybrid paper-based polling place voting device that provides a full-face 32-inch interactive touchscreen and incorporates the printing of the voter's selections on a vote summary card and tabulation scanning into a single unit. The ExpressVote XL is set up for each election using a proprietary USB drive that remains in the machine and collects the tabulated cast vote records. The compartment enclosing the USB drive, the memory module, two other ports and the administration switch is equipped with a barrel lock on the top front of the unit inside the privacy curtain with the voter.

Voting sessions are initiated by the insertion of an un-voted summary card. The ExpressVote XL guides voters through the ballot selection process with screen prompts, symbols and optional audio. It can be programmed to display ballots and instructions in up to 16 languages and allows voters to choose large text and/or high-contrast text if desired. When large text is selected the ballot will run to as many screens as are needed. Write-in votes are entered using a virtual keyboard displayed on the touchscreen. The content of write-in votes is not captured in memory and cannot be read back from the vote summary card for audio verification.

When the voter has finished making their selections, the ExpressVote XL prints a vote summary card, which is internally processed for tabulation and then presented to the voter for review behind a plastic window. The summary card is not physically accessible to the voter unless the voter chooses to spoil the summary card, which requires a poll worker to enter the booth to eject it. If the voter chooses to cast the vote, the vote summary card is deposited in a removable card container attached to the ExpressVote XL cart and the tabulation is recorded as a cast vote record. Vote summary cards are collected in the cartridge in the order that they were cast. The cartridge can be sealed before mounting in the machine so that it remains sealed when removed after polling.



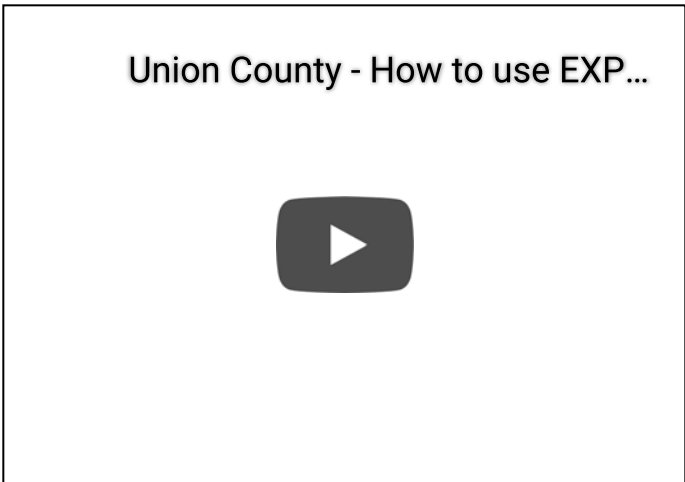
The vote summary card does not replicate the format of the ballot as displayed on the touchscreen nor the ballots used in the jurisdiction for absentee or provisional voting. The vote summary cards list the voter’s selections and the names of contests for which the voter made selections. There is no indication of contests for which the voter did not make a selection. The selections are encoded into barcodes, which are also printed on the vote summary card above the human readable summary of the voter’s selections. The ExpressVote XL tabulates the information encoded in the barcode and not the human readable summary.

Each ExpressVote XL unit weighs 296 pounds and is equipped with four large wheels. Each unit self-contains its own supplies, including privacy curtain. It includes a secondary printer to print a zero tape at opening of the polls and a summary tape at closing. The ExpressVote XL is intended to run on AC power but includes a battery rated for 2 hours; extra batteries are available as an option.

The ExpressVote XL generates a detailed audit log of all actions and events that have occurred on the unit similar to the Real Time Audit Log (RTAL) used with the iVotronic DRE, which can be printed at any time. Every action and event, including access attempts, access of system functions and errors, is logged and timestamped.

A brief video of the voting process for the ExpressVote XL from Union County NJ:

A full demo of the ExpressVote XL by the Philadelphia City Commission:





MENU

- [VerifiedVoting.org](#)
- [Verified Voting Foundation](#)
- [The Verifier](#)
- [Blog](#)
- [The Voting News](#)
- [Take Action](#)
- [Donate](#)
- [Contact Us](#)

© Copyright 2019, Verified Voting Foundation, Inc. All rights reserved, although reprint permission granted for nonprofit purposes with attribution to Verified Voting Foundation, Inc.

EXHIBIT 4

Can Voters Detect Malicious Manipulation of Ballot Marking Devices?

Matthew Bernhard, Allison McDonald, Henry Meng, Jensen Hwa, Nakul Bajaj*, Kevin Chang, J. Alex Halderman

University of Michigan *The Harker School

Abstract—Ballot marking devices (BMDs) allow voters to select candidates on a computer kiosk, which prints a paper ballot that the voter can review before inserting it into a scanner to be tabulated. Unlike paperless voting machines, BMDs provide voters an opportunity to verify an auditable physical record of their choices, and a growing number of U.S. jurisdictions are adopting them for all voters. However, the security of BMDs depends on how reliably voters notice and correct any adversarially induced errors on their printed ballots. In order to measure voters’ error detection abilities, we conducted a large study ($N = 241$) in a realistic polling place setting using real voting machines that we modified to introduce an error into each printout. Without intervention, only 40% of participants reviewed their printed ballots at all, and only 6.6% told a poll worker something was wrong. We also find that carefully designed interventions can improve verification performance. Verbally instructing voters to review the printouts and providing a written slate of candidates for whom to vote both significantly increased review and reporting rates—although the improvements may not be large enough to provide strong security in close elections, especially when BMDs are used by all voters. Based on these findings, we make several evidence-based recommendations to help better defend BMD-based elections.

I. INTRODUCTION

The threat of election hacking by hostile nations has prompted a major push to ensure that all voting systems in the United States have voter-verifiable paper trails, a defense recommended by the National Academies [36], the Senate Select Committee on Intelligence [53], and nearly all election security experts. Guided by past research [8], some states and localities are implementing paper trails by deploying ballot-marking devices (BMDs). In these systems, the voter makes selections on a computer kiosk, which prints a paper ballot that the voter can review before inserting it into a computer scanner to be counted [56]. BMDs have long been used as assistive devices for voters with disabilities, and a growing number of jurisdictions are purchasing them for use by all voters [24], [25], [37].

BMDs have the potential to provide better security than direct-recording electronic voting machines (DREs), which maintain the primary record of the voter’s selections in a computer database and often lack a voter-verifiable paper trail. Numerous studies have demonstrated vulnerabilities in DREs that could be exploited to change election results (e.g., [11], [23], [31], [35]). In contrast, BMDs produce a physical record of every vote that can, in principle, be verified by the voter and manually audited by officials to confirm or correct the initial electronic results.

However, BMDs do not eliminate the risk of vote-stealing attacks. Malware could infect the ballot scanners and change the electronic tallies—although this could be detected by rigorously auditing the paper ballots [50]—or it could infect the BMDs themselves and alter what gets printed on the ballots. This latter variety of cheating cannot be detected by a post-election audit, since the paper trail itself would be wrong, and it cannot be ruled out by pre-election or parallel testing [51]. Instead, BMD security relies on voters themselves detecting such an attack. This type of human-in-the-loop security is necessary in many systems where detection and prevention of security hazards cannot be automated [18]. However, as several commentators have recently pointed out [7], [20], [51], its effectiveness in the context of BMDs has not been established.

Whether such a misprinting attack would succeed without detection is highly sensitive to how well voters verify their printed ballots. Every voter who notices that their ballot is misprinted and asks to correct it *both* adds to the evidence that there is a problem *and* requires the attacker to change an additional ballot in order to overcome the margin of victory. Consider a contest with a 1% margin in which each polling place has 1000 voters. If voters correct 20% of misprinted ballots, minimal outcome-changing fraud will result in an average of 1.25 voter complaints per polling place—likely too few to raise alarms. If, instead, voters correct 80% of misprinted ballots, polling places will see an average of 20 complaints, potentially prompting an investigation. (We model these effects in Section V.) Despite this sensitivity, voters’ BMD verification performance has never before been experimentally measured.

In this paper, we study whether voters can play a role in BMD security. We first seek to establish, in a realistic polling place environment, the rates at which voters attempt to verify their printed ballots and successfully detect and report malicious changes. To measure these, we used real touch-screen voting machines that we modified to operate as malicious BMDs. We recruited 241 participants in Ann Arbor, Michigan, and had them vote in a realistic mock polling place using the ballot from the city’s recent midterm election. On every ballot that our BMDs printed, one race was changed so the printout did not reflect the selection made by the participant.

We found that, absent interventions, only 40% of participants reviewed their printed ballots at all, only 6.6% reported the error to a poll worker, and only 7.8% correctly identified it on an exit survey. These results accord with prior studies that found poor

voter performance in other election security contexts, such as DRE review screens [1], [15] and voter-verifiable paper audit trails (VVPATs) [48]. The low rate of error detection indicates that misprinting attacks on BMDs pose a serious risk.

The risks notwithstanding, BMDs do offer practical advantages compared to hand-marked paper ballots. They allow voters of all abilities to vote in the same manner, provide a more user-friendly interface for voting, and more easily support complex elections like those conducted in multiple languages or with methods such as ranked choice [44]. BMDs also simplify election administration in places that use vote centers [56], which have been shown to reduce election costs and lower provisional voting rates [28], [42], as well as in jurisdictions that employ early voting, which can improve access to the ballot [30].

Given these advantages and the fact that BMDs are already in use, the second goal of our study was to determine whether it might be possible to boost verification performance through procedural changes. We tested a wide range of interventions, such as poll worker direction, instructional signage, and usage of a written slate of choices by each voter.

The rate of error detection varied widely with the type of intervention we applied, ranging from 6.7% to 86% in different experiments. Several interventions boosted review rates and discrepancy reporting. Verbally encouraging participants to review their printed ballot after voting boosted the detection rate to 14% on average. Using post-voting verbal instructions while encouraging participants to vote a provided list of candidates raised the rate at which voters reported problems to 73% for voters who did not deviate from the provided slate.

These findings suggest that well designed procedures can have a sizable impact on the real-world effectiveness of voter verification. We make several recommendations that election officials who already oversee voting on BMDs can employ immediately, including asking voters if they have reviewed their ballots before submission, promoting the use of slates during the voting process, informing voters that if they find an error in the printout they can correct it, and tracking the rate of reported errors. Our recommendations echo similar findings about the most effective ways to alert users to other security hazards (i.e., in context [12] and with active alerts [21]) and redirect them to take action.

Although our findings may be encouraging, we strongly caution that much additional research is necessary before it can be concluded that any combination of procedures actually achieves high verification performance in real elections. Until BMDs are shown to be effectively verifiable during real-world use, the safest course for security is to prefer hand-marked paper ballots.

Road Map Section II provides more background about human factors and security and about previous work studying the role of voter verification in election security. Section III describes our experimental setup, voting equipment, and study design. Section IV presents our results and analyzes their significance. Section V provides a quantitative model for BMD verification security. Section VI discusses the results, avenues for future work, and recommendations for improving the verifiability of BMDs. We conclude in Section VII.

II. BACKGROUND AND RELATED WORK

A. Human-Dependent Security

Elections fundamentally depend on having humans in the loop—as Stark [51] notes, the voter is the *only one* who knows whether the ballot represents their intended vote—and the success or failure of election security has the potential to have history-altering effects. The type of risk posited by Stark, wherein voters do not check their paper ballots to ensure the BMD has correctly represented their selections, is a post-completion error [14], in which a user makes a mistake (or fails to verify the correctness of something) *after* they have completed the main goal of their task. Voters who forget or do not know to verify the correctness of a paper ballot after they have entered their selections on a BMD miss a critical step in ensuring the accuracy of their vote. We therefore explore how to communicate this risk to voters.

Cranor [18] describes five ways that designers can communicate risk to a user who needs to make security decisions:

- 1) *Warnings*: indication the user should take immediate action
- 2) *Notices*: information to allow the user to make a decision
- 3) *Status indicators*: indication of the status of the system
- 4) *Training*: informing users about risks and mitigations before interaction
- 5) *Policies*: rules with which users are expected to comply

Implementing indicators that reveal meaningful information to voters about the security status of a BMD would be next to impossible, as security issues are often unknown or unforeseen to the operators. Although voter education about the importance of verification might be an effective form of training, significant coordination would be necessary to enact such a scheme at scale. Therefore, we focus in this study on the effectiveness of warnings issued through poll worker scripts and polling place signage.

A warning serves two purposes: to alert users to a hazard, and to change their behavior to account for the hazard [62]. There are many barriers to humans correctly and completely heeding security warnings. Wogalter proposes the Communication-Human Information Processing (C-HIP) Model [61] to systematically identify the process an individual must go through for a warning to be effective. The warning must capture and maintain attention, which may be difficult for voters who are attempting to navigate the voting process as quickly as possible. Warnings must also be comprehensible, communicate the risks and consequences, be consistent with the individual's beliefs and attitudes toward the risk, and motivate the individual to change—all of which are substantial impediments in an environment with little to no user training and such a broad user base as voting.

To maximize effectiveness, warnings should be contextual, containing as little information as necessary to convey the risk and direct individuals to correct behavior [12], [61]. Voters are essentially election security novices; Bravo-Lillo et al. [12] found that, in the context of computer security, advanced and novice users respond to warnings differently. Most significantly, novice users assessed the hazard *after* taking action, whereas

advanced users assessed the hazard *before* engaging in the activity.

There may be effective ways to improve voter verification performance. Many studies have applied lessons from Cranor, Wogalter, and Bravo-Lillo et al. to help humans make secure choices in different contexts, including phishing [21], [41], browser warnings [2], [46], [52], app permissions [3], [40], and operating system interfaces [13]. In the context of phishing warnings, for example, Egelman et al. [21] found that users were far more likely to heed an active warning, or a warning that disrupted their workflow, than a passive warning. This suggests that similar interventions applied in a polling place may have a significant effect on voters' ability to review and verify their BMD ballots.

Our study contributes to this literature by exploring the effects of several modalities of warnings (oral and visual) on human detection of malicious ballot modification.

B. Voter-Verifiable Paper and Ballot-Marking Devices

A guiding principle in election security is that voting systems should be *software independent* [47]: that is, any software errors or attacks that change the reported election outcome should be detectable. Bernhard et al. [9] note that elections backed by a voter-verifiable paper record are currently the only known way to provide robust software independence. Like BMDs, voter-verifiable paper audit trails (VVPATs) and hand-marked paper ballots are widely used in an attempt to achieve software independence. However, each poses a different set of usability and accessibility challenges.

Hand-marked paper ballots record the voter's selections without the risk of having a potentially compromised computer mediating the process. However, voters often make mistakes when filling out ballots by hand that can lead to them being counted incorrectly or ruled invalid [27]. Moreover, many voters have difficulty marking a paper ballot by hand due to a disability or a language barrier. Ballots in the U.S. are among the most complex in the world, further magnifying these difficulties [38].

VVPAT technology also suffers from noted usability, privacy, and auditability problems [26]. Most implementations consist of clunky printer attachments for DREs that are difficult for voters to read, record votes in the order in which they are cast, and use a fragile paper tape. In laboratory studies, Selker et al. [48] and de Jong et al. [19] found that voters frequently did not review the VVPAT, with Selker finding that only 17% of voters detected changes between the selections they made on the DRE and those printed on the VVPAT. While there has been some criticism of Selker's findings and methodology [45], [49], their results broadly comport with work by Campbell et al. [15] and Acemyan et al. [1] about voters' ability to detect errors introduced in DRE review screens. The latter found that only 12–40% of participants successfully detected such errors.

In part due to the concerns raised by these studies, BMDs have become a popular choice for new voting system deployments in the United States. South Carolina and Georgia, together comprising nearly 9 million voters, recently adopted

BMDs statewide [24], [25], as have several counties and cities, including Los Angeles County, the largest single election jurisdiction in the U.S. [58].

There has been vigorous debate among election security experts as to whether BMDs can provide software-independence (e.g., [7], [20], [51], [60]). However, the discussion has yet to be informed by rigorous experimental data. Our work seeks to fill that gap by contributing the first human-subjects study to directly measure the verification performance of voters using BMDs under realistic conditions and with a variety of potential procedural interventions.

III. MATERIALS AND METHODS

Our goals in this work were to empirically assess how well voters verify BMD ballots and whether there are steps election officials can take that will enhance verification performance. To these ends, we conducted a between-subjects study where we tested several hypotheses in a simulated polling place, following the best practices recommended by Olemba et al. [39] for election human-factors research. The study design was approved by our IRB.

We sought to answer several questions, all of which concern the rate at which voters are able to detect that a BMD-printed ballot shows different selections than those the voter picked:

- What is the base rate of error detection?
- Is error detection impacted by:
 - Ballot style?
 - Manipulation strategy?
 - The manipulated race's position on the ballot?
 - Signage instructing voters to review their ballots?
 - Poll worker instructions?
 - Providing a slate of candidates for whom to vote?

In order to answer these questions in an ecologically valid way, we attempted to create an environment that closely resembled a real polling place. Nevertheless, it is impossible for any experiment to fully recreate what is at stake for voters in a real election, and so study participants may have behaved differently than voters do in live election settings. We went to extensive lengths to mitigate this limitation, and we find some data to support that we did so successfully (see Section VI-A). We used real (though modified) voting machines, printers and paper stock from deployed BMD systems, a ballot from a real election, and ballot styles from two models of BMDs. We conducted the study at two city library locations, one of which is used as a polling place during real elections.

A. The Polling Place

To provide a realistic voting experience, we structured our simulated polling place like a typical BMD-based poll site. Three investigators served as poll workers, following the script in Appendix A. Library patrons who were interested in voting began at a check-in table, where they were greeted by Poll Worker A and asked to sign an IRB-approved consent form. Participants were told they would be taking part in “a study about the usability of a new type of voting machine” and instructed



Fig. 1: *Polling Place Setup*. We established mock polling places at two public libraries in Ann Arbor, Michigan, with three BMDs (*left*) and an optical scanner and ballot box (*right*). Library visitors were invited to participate in a study about a new kind of election technology. The BMDs were DRE voting machines that we modified to function as malicious ballot marking devices.

on how to use the equipment, but they were not alerted that the study concerned security or that the BMDs might malfunction.

Each participant received a voter access card with which to activate a BMD and was free to choose any unoccupied machine. There were three identical BMDs, as shown in Figure 1. On the last day of the study, one machine’s memory became corrupted, and it was removed from service; all votes that day were recorded on the other two machines.

The BMDs displayed contests in a fixed order, and voters made selections using a touch screen interface. After the last contest, the machines showed a review screen that accurately summarized the voter’s selections and highlighted any undervotes. The voter could return to any contest to change the selections. A “Print Ballot” button ended the voting session and caused a printer under the machine to output the paper ballot.

Participants carried their ballot across the polling place to the ballot scanner station, where they inserted them into an optical scanner that deposited them into a ballot box. Poll Worker B was stationed by the scanner and offered instructions if necessary. Next, the poll worker collected the voter access card and asked each participant to complete an exit survey using a laptop next to the scanning station. The survey was anonymous, but responses were keyed so that we could associate them with the voter’s on-screen selections, their printed ballot, and poll worker notes.

Poll Worker C, positioned separately from the other stations, acted as an observer. They verified that participants moved through the polling place stations sequentially, noted whether they spent time reviewing their printed ballots, and recorded whether they appeared to notice any abnormalities. The observer was also tasked with noting participant behavior, specifically how the participants completed each step in the voting process and any comments they made. The observer was available to answer participant questions and was frequently the poll worker participants approached upon noticing a discrepancy.

Like in a real polling place, multiple participants could progress through the voting process simultaneously. Occasion-

ally a one- or two-person line formed as participants waited to use the BMDs or the ballot scanner.

B. The Voting Machines

BMD voting systems are currently produced by several voting machine manufacturers, the largest of which is ES&S. Over a six month period, we repeatedly attempted to engage ES&S in discussions about acquiring samples of their equipment for this study. However, these attempts were ultimately not fruitful.

Instead, we utilized AccuVote TSX DRE voting machines, which we purchased on eBay and modified to function as BMDs. The TSX was first produced by Diebold in 2003 and is still widely deployed today. At least 15 states plan to use it in at least some jurisdictions in November 2020 [57].

The TSX runs Windows CE and is designed to function as a paperless DRE or a VVPAT system. We developed software modifications that allow it to print ballots in multiple styles using an external printer. This effectively converts the TSX into a BMD—and one we could easily cause to be dishonest—while preserving the original touch-screen interface used by voters.

In order to modify the machine, we built on techniques used by Feldman et al. [23]. We began by patching the firmware so that, when the machine boots, it attempts to execute a program provided on an external memory card. We used this functionality to launch a remote access tool we created, which allowed us to connect to the TSX over a network and perform file system operations, run applications, and invoke a debugger.

The TSXes in our polling place were connected to an Ethernet switch using PCMCIA network adapters. A Python program, running on a computer on the same network, used the remote access tool’s API to poll each machine for newly voted ballots. Whenever a ballot was cast, the program parsed the selections, generated a PDF file based on them, and sent it to a printer located underneath the appropriate voting machine. The program could be configured to apply different ballot styles and cheating strategies, depending on the experiment.

For every ballot, the program randomly selected one race to manipulate. In most experiments, selections could be changed in three ways: deselection in a voted-for race, selection in an unvoted-for race, or changing a selection to a different candidate. We ensured that some alteration would take place on every ballot. For example, in a vote-for-one race where the voter had made a selection, the algorithm would choose uniformly from the set of unselected choices plus no selection. One experiment used a different strategy, in which choices could only be deselected.

Both the voter’s original selections and the manipulated ballot were logged for later analysis. Each voting session was associated with a unique tracking number, which was printed on the ballot along with a timestamp and encoded as a barcode.

As the final step in the voting process, participants fed their printed ballots into an AccuVote OS optical scanner, a device used to tabulate votes in parts of 20 states [57]. The scanner was intended to add realism to the experiment, but AccuVote OSes are not capable of actually tabulating the ballot votes we used. Therefore, we modified the scanner so that it simply fed each ballot into the ballot box without counting it.

We mounted a barcode reader in a 3-D printed case above the scanner’s input tray and positioned it so that it would detect the ballot’s tracking barcode. (This setup can be seen in Figure 3.) When the barcode was read, a Raspberry Pi would activate the AccuVote OS’s feed motor to pull the ballot into the ballot box. The Raspberry Pi also displayed the ballot tracking number so that poll workers could associate the ballot with the participant’s exit survey response and the observer’s notes.

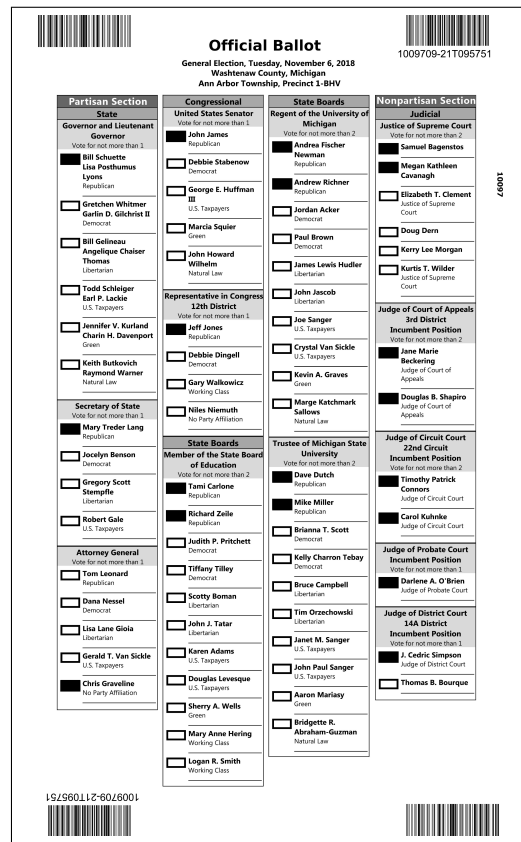
C. The Ballot

In order to ensure a realistic voting experience and increase participants’ psychological investment in the outcome of the mock election, we used races and candidates from the city’s actual ballot for the recent 2018 midterm election. For simplicity, we reduced the ballot to the first 13 races so that ballots would not require duplex printing or multiple pages.

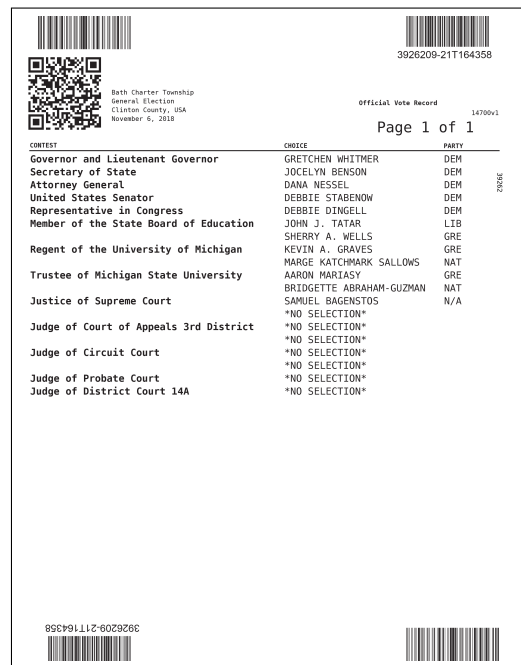
We tested two ballot styles, which are illustrated in Figure 2. One is a regular ballot that shows the entire set of candidates in every race. The other is a summary ballot, which shows only the voter’s selections or “NO SELECTION” if a choice is left blank. Most BMDs print ballots that resemble these styles.

The specific visual designs we used mimic ballots produced by two models of BMDs manufactured by Hart InterCivic, which also makes the voting equipment used in Ann Arbor. The regular style is also the same design as the hand-marked paper ballots most Ann Arbor voters use, ensuring that many participants found it familiar. These designs are used in jurisdictions that collectively have over 10 million registered voters [57].

The model of laser printer we used, Brother HL-2340, is certified for use with Clear Ballot’s ClearAccess BMD system [43], so we chose paper stock that meets the specifications for ClearAccess [16]. Summary ballots were printed on regular weight 8.5×11 inch letter paper, while regular ballots were printed on Vellum Bristol stock 67 pound 8.5×14 inch paper.



(a) Regular Ballot



(b) Summary Ballot

Fig. 2: *Ballot Styles*. We tested two ballot styles: (a) a regular style, resembling a hand-marked ballot; and (b) a summary style, listing only the selected candidates. Both had 13 races from the city’s recent midterm election. In one race, determined randomly, the printed selection differed from the voter’s choice.

D. Participants and Recruitment

To gather subjects for our study, we approached staff at the Ann Arbor District Library (AADL), who offered space for us to set up our mock precinct. We conducted a total of three days of data collection in July and September 2019 at two library locations: the Downtown and Westgate branches. The Downtown branch, where our study was held for two of the three days, is an official polling location during real elections.

The AADL advertised our study through its social media feeds and offered incentives to patrons for their participation, such as points for a scavenger hunt competition [5] and souvenir flashlights [6]. We also set up a fourth voting machine outside of the mock precinct where kids could vote in an election for mayor of the library’s fish tank.¹ Results from that machine were not used as part of this study, but it served as a recruitment tool for parents visiting the library with their children. In addition, we verbally recruited patrons who happened to be at the libraries during our study, using the script in Appendix B.

Participants were required to be at least 18 years of age and to sign an IRB-approved consent form. All data collected, including survey responses and behavioral observations, was completely anonymous. We informed participants that they were not required to vote their political preferences.

E. Experiments

To explore what factors affect voter verification performance, we devised nine experiments to run between subjects. In all experiments, for every participant, one selection that the participant made on the BMD was not accurately reflected on the printed ballot. Every participant within an experiment received the same instructions from the poll workers, following the script and variants in Appendix A.

The first three experiments were designed to measure verification in the absence of protective interventions. They varied the ballot style and manipulation strategy:

E1: Regular ballots We used the regular ballot style and the default manipulation strategy, in which a selection could be switched, deselected, or selected if left blank by the voter.

E2: Summary ballots We used the summary ballot style and the default manipulation strategy. As discussed in Section IV, we found no significant difference in error detection between regular ballots and summary ballots, so all subsequent experiments used summary ballots.

E3: Deselection only To assess the sensitivity of voters to the way their ballots were changed, we limited the manipulation to deselecting one of the voter’s choices at random.

Four further experiments tested interventions to determine if they improved error detection. We tried posting a sign and having poll workers give different instructions at various times:

E4: Signage A sign was placed above the scanner that instructed voters to check their printed ballots, as shown in



Fig. 3: *Warning Signage*. One of the interventions we tested was placing a sign above the scanner that instructed voters to verify their ballots. Signage was not an effective intervention.

Figure 3. We designed the sign following guidelines from the U.S. Election Assistance Commission [55].

E5: Script variant 1 During voter check in, the poll worker added this instruction: “Please remember to check your ballot carefully before depositing it into the scanner.”

E6: Script variant 2 When the voter approached the scanner, the poll worker said: “Please keep in mind that the paper ballot is the official record of your vote.”

E7: Script variant 3 When the voter approached the scanner, the poll worker said: “Have you carefully reviewed each selection on your printed ballot?”

The final two experiments assessed whether reminding participants of their selections during verification improved their performance. We gave voters a slate of candidates for whom to vote that they could carry with them throughout the voting experience. While we refer to this as a slate, a sample ballot that the voter filled in before voting could serve the same purpose. Every voter received the same slate (Appendix C), which was randomly generated and contained an even mix of parties.

E8: Slate with script variant 2 Voters were given the slate. Poll workers encouraged verification with script variant 2.

E9: Slate with script variant 3 Voters were given the slate. Poll workers encouraged verification with script variant 3.

¹Mighty Trisha unexpectedly beat Creepy Bob, leading some Bob supporters to complain that the results were fishy [4].

Experiment	N	Were observed examining ballot	Reported error on exit survey	Reported error to poll worker
<i>Without interventions:</i>				
E1: Regular ballots	31	41.9%	6.5%	6.5%
E2: Summary ballots	31	32.3%	6.5%	6.5%
E3: Deselection only	29	44.8%	10.3%	6.9%
Subtotal/Mean	91	39.7%	7.8%	6.6%
<i>With interventions:</i>				
E4: Signage	30	13.3%	3.3%	6.7%
E5: Script variant 1	30	46.7%	13.3%	6.7%
E6: Script variant 2	25	92.0%	16.0%	16.0%
E7: Script variant 3	31	38.7%	19.4%	12.9%
E8: Slate with script variant 2	13	100.0%	38.5%	38.5%
E9: Slate with script variant 3	21	95.2%	71.4%	85.7%
Subtotal/Mean	150	64.3%	24.0%	27.8%

TABLE I: *Verification Performance for Each Experiment.* Without interventions, participants’ verification performance was remarkably poor: only 7.8% noted on an exit survey that their ballots had been altered, and only 6.6% informed a poll worker (averaged across experiments). The various interventions we tested had widely different effects, ranging from no significant improvement (**E4**, **E5**) to a large increase in verification success (**E8**, **E9**).

IV. RESULTS

A. Participant Demographics

We recruited 241 participants. The vast majority (220, 91%) indicated that they were native English speakers; 19 reported speaking twelve other native languages, including Hungarian, Korean, and Arabic; and two subjects gave no response. Participants who disclosed their age ranged from 18 to 84 years old, with a mean of 43.7 and a median of 42; 15 subjects did not answer the question. The percentages that follow are out of the total number of responses to each question: Respondents identified as male (84, 35%), female (152, 64%), or other (3, 1%); two did not respond. Subjects reported their ethnicity as Caucasian (187, 80%), Asian (17, 7%), African American (6, 3%), Mexican American/Chicano (5, 2%), and Other Hispanic/Latino (9, 4%); others reported not having any of these ethnic backgrounds (2, 1%) or were multiracial (9, 4%). Participants reported their level of educational attainment as some high school (1, 0.4%), a high school diploma (4, 2%), some college (20, 8%), a two-year degree (10, 4%), a four-year degree (80, 33%), a master’s or professional degree (92, 38%), or a doctorate (34, 14%).

Most subjects indicated that they were registered to vote in the U.S. (220, 92%), had voted in a previous election (216, 91%), and had voted in the November 2018 midterm election (209, 87%). However, we note that, historically, 38–45% of non-voters have been found to falsely report having voted [10].

Compared to the population of Ann Arbor at the time of the 2010 census, our participant pool overrepresented Caucasians ($\Delta = 7.6\%$) and underrepresented African Americans ($\Delta = -4.4\%$) and Asians ($\Delta = -8.7\%$) [54]. The study population also overrepresented females ($\Delta = 13\%$) and underrepresented males ($\Delta = -16\%$) [59]. In other reported aspects, participants’

demographics resembled the population of Ann Arbor voters (the city is among the most highly educated in the U.S.) [33].

B. Verification Performance

To quantify verification performance, we collected three data points for each participant, which are summarized in Table I. First, an observer noted whether the subject appeared to examine the printed ballot for at least two seconds. Second, the exit survey asked, “Did you notice anything odd about your ballot?”, and we recorded whether the subject’s response corroborated the discrepancy (i.e., correctly articulated which race was changed). Third, we recorded whether subjects reported the ballot modification to a poll worker. Most experiments saw more participants identify discrepancies in the survey than were reported to poll workers, but these differences were not statistically significant. Where applicable, we refer to participants who by some means reported detecting the discrepancies as “noticers” and those who did not as “non-noticers”.

1) *Performance without interventions (E1–E3):* With no interventions, we found verification performance to be consistently poor. The three experiments involved 91 participants, and, averaged across the experiments, only 40% of participants examined their ballots, only 7.8% noted the error on the exit survey, and only 6.6% reported it to a poll worker. We did not find significant differences in performance between regular and summary ballots or between the tested attack strategies.

2) *Effectiveness of interventions (E4–E9):* The tested interventions resulted in a wide range of effect sizes. Neither signage (**E4**) nor poll worker instructions issued before the participant began voting (**E5**) yielded a statistically significant improvement to any aspect of verification performance. In

contrast, poll worker instructions issued *after* the ballot was printed (**E6** and **E7**) did have a positive effect, boosting reporting rates to 20% on the exit survey and 14% to poll workers (averaged across the experiments).

The largest performance gains occurred when participants were directed to vote using a slate of candidates (**E8** and **E9**). However, only **E9** produced a statistically significant difference in reporting rates (Fisher’s exact $p < 0.001$).² Averaged across both experiments, reporting rates increased to 55% on the exit survey and 62% to poll workers. **E8**, in which participants were directed how to vote using a slate of candidates, saw detection and reporting rates of 39%, which is similar to results for DRE review screen performance found by Campbell et al. [15] and Acemyan et al. [1], in studies that similarly directed participants how to vote. With script variant 3, the use of a slate produced a significant difference (comparing **E7** and **E9**, Fisher’s exact $p < 0.02$) for both review and report, but it did not produce a significant difference using script variant 2 (comparing **E6** and **E8**). This indicates that voters may be sensitive to the specific instructions they receive about reviewing their ballots.

C. Correlates

1) *Reviewing the ballot*: Reviewing the ballot at all was significantly correlated with error reporting (two-sample permutation test $p < 0.001$ with 10k repetitions). Some interventions do seem to promote reviewing: **E6**, **E8**, and **E9** saw significant increases (Fisher’s exact $p < 0.004$), although **E7** did not.

2) *Time to ballot submission*: Careful verification takes time, so one might expect that participants who noticed discrepancies took more time to cast their ballots. As an upper bound on how long subjects spent verifying, we calculated the time from ballot printing to ballot submission. (Due to clock drift on one of our machines, data from the third day of experiments was unusable, and consequently **E4** and **E7** are excluded from our timing analysis.) As expected, we find that noticers took an average of 121 s between printing and ballot submission (median 114 s), compared to only 43 s for non-noticers (median 32 s). This difference is statistically significant (two-sample permutation test $p < 0.004$, 10k iterations).

We compared the submission times for two sets of experiments: ones with extra instructions to the voter (**E5**, **E6**, **E8**, and **E9**; $N = 84$) and ones without (**E1**, **E2**, and **E3**; $N = 91$). The experiments that asked participants to review their ballots saw significantly more time spent between ballot printing and submission (two-sample permutation test $p < 0.004$, 10k iterations), an average of 83 s (median 72 s) compared to 50 s without (median 33 s).

Notably, participants who were given a slate of candidates to vote for had much higher submission times (two-sample permutation test $p < 0.004$, 10k iterations). Noticers in the slate experiments took an average of 119 s (median 111 s) and non-noticers averaged 55 s (median 52 s). This might be partly attributed to voters having to select unfamiliar candidates and wanting to check their work.

²All p -values were computed with a Bonferroni correction at a family-wise error rate of 0.05.

3) *Demographics*: Comparisons of detection rates across demographic groups revealed that a strong indicator for verification performance was voting experience. Subjects who reported being registered to vote ($N = 220$) detected errors with their ballots 19% of the time, while their those who did not ($N = 21$) detected errors 4.8% of the time. Those who reported voting previously ($N = 216$) caught ballot discrepancies in 19% of cases, again performing better than those who reported not voting before ($N = 25$), who detected an error in 4.0% of cases. If someone reported voting in the 2018 midterm election ($N = 209$), they detected problems with their ballot 20% of the time, whereas if they did not ($N = 32$), they detected problems 3.1% of the time. This may indicate that familiarity with the midterm ballot we used caused participants to feel more invested in the accuracy of their votes; however, we did not establish this to statistical significance.

Other demographic factors, such as age, education, ethnicity, and gender, had no correlation with detecting manipulation.

4) *Ballot position*: Noticing was correlated with ballot position (Pearson’s of -0.64), indicating that discrepancies in more prominent races are more likely to be noticed. (Race 0 was the first race on the ballot, so the number of noticers decreases as the race position increases, hence the negative correlation coefficient.) On our ballot, the first five races (Governor, Secretary of State, Attorney General, U.S. Senator, and Representative in Congress) were prominent partisan contests with a high likelihood of name recognition. In the experiments with no intervention (**E1–E3**), 37 participants had one of these races manipulated, and five reported the error on the exit survey, a rate of 14%. Additional experiments are necessary to establish the strength of this effect when combined with interventions.

5) *Undervotes*: A metric that may inform voters’ ability and willingness to verify their ballot is how much care they take in filling out the ballot. There are two metrics we use to examine this: whether a participant voted in every contest on the ballot, and whether the participant voted in every available position on the ballot (e.g., in a vote-for-two contests, the participant selected two choices). Table II shows the rates of voting in every race and every position on the ballot, with **E8** and **E9** removed as they directed participants to vote in every position. Voters who noticed discrepancies voted in every race or every position at a higher rate than those who did not, but not significantly so (likely due to our small sample size). Since these undervotes are visible to malware running on a BMD, this correlation could be exploited by an attacker to focus cheating on voters who are less likely to carefully verify, provided future work more firmly establishes this link.

	Overall	Noticers	Non-noticers
Every race	64.3%	73.9%	63.0%
Every position	43.0%	47.8%	42.4%

TABLE II: *Participant Attentiveness*. Voters who noticed the discrepancy tended to vote in every race and ballot position more often than those who did not.

6) *Partisanship*: To assess the role partisanship plays in detection rates, we scored each ballot with a partisanship score, where a vote for a Democratic candidate was scored -1 and a vote for a Republican candidate was scored 1 , and we take the absolute value of the sum. There were 11 opportunities to vote in a partisan way, so a participant who voted straight-party for either major party would achieve a score of 11. Excluding **E8** and **E9**, where voters were directed how to vote, the mean partisanship score for our participants was 8.3, and the median was 11. Although our BMD did not offer an automatic “straight-party” voting option, 105 participants achieved the maximum partisanship score.

Intuitively, a voter expecting every selected candidate to be from the same party might be more likely to notice a selection from a different party. Looking at only these straight-party voters, 15 out of 105 detected the errors. Of those, nine had a partisan race swapped to a different candidate of a different party, and six of those participants wrote in the survey that they had detected the change based on party. For example, one participant wrote, “*voted GOP for governor / lieutenant governor but Libertarian was actually selected on the paper ballot.*”

This suggests that choosing a uniform set of candidates may help voters detect when something has gone wrong on their ballot, although more work is needed to establish that this is indeed the case, especially in more politically diverse populations. If this positive effect holds, it could be further promoted with ballot designs that prominently display the party, which could help voters see the information that is important to them while they review the ballot. On the other hand, BMD malware could be designed to counter this effect by focusing cheating on voters who do not cast a straight-party ballot.

7) *Slate voting*: 34 participants were assigned an intervention which asked them to vote for a preselected slate of candidates (with a partisanship score of 0). Of these, only 26 participants voted exactly as directed. Of the eight participants who did not, four voted a straight Democratic ticket (partisanship score of 11), one voted a heavily Democratic ticket (score of 9), two voted slightly Democratic tickets (scores of 3 and 5), and one voted a non-partisan ticket (score of 0), which only deviated from the slate in five positions. Of the eight participants who deviated from the slate, no participant deviated by fewer than five positions, indicating that either the deviation was deliberate or our instructions to vote the slate were unclear. Only one deviating participant managed to notice the discrepancy on their ballot, leaving participants who deviated from the slate a 13% notice rate compared to the 73% notice rate for those who did not deviate.

8) *Network effects*: One potential feature of a live polling place environment is a network effect: will a voter who is voting at the same time as a noticer be more likely to notice a problem on theirs? However, the number of people who notice in a given experiment is a confounding factor: voters are more likely to overlap with a noticer if there are more noticers. To interrogate this, we ran partial hypothesis tests for each intervention using Fisher’s exact tests with permutations of overlapping with a noticer and noticing, and then combined

using Fisher’s combining function. We found that the effect of overlapping with a noticer did not significantly impact whether a participant noticed. This suggests that our interventions were more important than overlapping.

9) *Signage*: One feature that did not correlate with improved verification performance was the signage we tested (**E4**). Our observer noted that 11 of 30 participants in the signage experiment did not notice the sign at all. Only two participants in this experiment detected the modification of their ballot and reported it, and only one accurately noted the discrepancy in their survey, suggesting that passive signage alone may be insufficient to capture voters’ attention and shape their subsequent behavior.

D. Participant Comments

Participants had two free-response sections in the exit survey. The first asked about anything “odd” they had noticed about the ballot. The second invited any additional comments. Of the 241 participants, 114 responded to at least one of these prompts. We note several features of their responses.

1) *Discrepancy reports*: In total, 44 participants (18%) noted in the free response section of the survey that they had identified some discrepancy on their paper ballot. Of these, 31 correctly identified the change, 12 gave no detail (e.g., “*At least one of my choices did not match who I picked*”), and one incorrectly identified the change (but did report that there was a mistake). We omitted this last participant from our “noticers” category where applicable.

Of the 44 participants who reported a change on their ballot in the survey, five added that they thought it could have resulted from a mistake they made. For example, one participant reported: “*I don’t remember voting for the member of Congress and there was a vote. I very well may have but just don’t remember.*”

2) *Attitudes about verification*: Twelve participants mentioned either that they would only be comfortable voting on a paper ballot or that they were comforted by the fact that a paper trail was created. Only three of these 12 participants noticed that their ballot had been modified, despite the fact that they recognized that the paper ballot was an important tool for ensuring election integrity.

Several participants seemed to realize *after* casting their vote that the evaluation of their paper ballot was important; 13 participants mentioned in the survey that they did not review or that they should have reviewed the ballot, although we did not ask them about it. This concern may have been triggered by our survey question about what they had noticed about the paper ballot, but it also might be an indication that our interventions did cause voters to think about the risk—albeit too late.

The free responses also indicate that some participants assumed that the vote was completed and submitted on the BMD, rather than the paper ballot being the official record of their vote. One participant wrote, “*I was surprised to still have a paper ballot, after using the touch system. I was expecting the results to be registered electronically.*” This assumption may discourage voters from verifying the selections on their

paper ballot. Similarly, another participant, prompted by script variant 3 (“Have you carefully reviewed each selection on your printed ballot?”), responded to a poll worker, “*I checked it on the screen, it better be right.*”

Three participants expressed concern that they would not know what to do if they noticed a problem with their paper ballot during a real election. One person wrote, “*Having the printout be incorrect was confusing and it’s not clear how that would be handled in an election environment.*”

3) *Feedback on the BMDs:* We told participants that the experiment was a study about a new kind of voting system, and many left feedback about the interface and appearance of the machines. In Michigan, where we conducted the study, BMDs are available in every precinct, but voters must request to use them. The vast majority of voters use hand-marked paper ballots, so study participants were likely unfamiliar with BMD voting. In their comments, 21 participants expressed liking the system, while only three disliked it. Although merely anecdotal, this reflects previous findings that voters like touch-screen voting equipment [22].

V. SECURITY MODEL

We are primarily motivated by the threat of undetected changes to election outcomes due to BMD misprinting attacks. Prior work has shown that such attacks cannot be reliably ruled out by pre-election or parallel testing [51], and we seek to answer whether voter verification can be an effective defense.

If a voter reports that their printed ballot does not reflect their on-screen selections, what should election officials do? Unfortunately, there is not yet a practical way to prove that the BMD misbehaved during voting. From officials’ perspective, it is also possible that the voter is mistaken, or even lying, and in a large voter population, there will always be some rate of spurious problem reports, even when BMDs are working correctly.

For these reasons, problem reports from voters can serve only as evidence that something *might* be wrong with the BMDs. If the evidence exceeds some threshold, officials could invoke contingency plans. For instance, they could remove BMDs from service to minimize further damage, perform forensic investigations in an attempt to uncover the cause, or even rerun the election if outcome-changing fraud cannot be ruled out.

Any of these responses would be costly (and none is foolproof), so the threshold for triggering them should not be too low. Moreover, attackers could exploit a low threshold by recruiting voters to fraudulently report problems, in order to disrupt or discredit the election. On the other hand, if the threshold is too high, outcome-changing fraud could be ignored.

To better understand how verification performance affects security in this setting, we construct a simple model. We assume, optimistically, that the attacker has no way to guess whether a particular voter is more likely than average to detect the alteration, and so chooses voters to attack at random. We further assume that whenever voters detect problems, they are able to remedy them and cast a correct vote by hand-marking a ballot. Except where noted, the model assumes that all voters cast their votes using BMDs.

Number of problem reports Let d be the fraction of misprinted ballots that voters detect, report, and correct. Suppose a contest had n ballots cast, and the reported fractional margin of victory was m . To have changed the outcome, the attacker would have had to successfully modify at least $n\frac{m}{2}$ cast ballots. However, since some modifications would have been corrected, the attacker would have had to induce errors in a greater number of printouts: $n\frac{m}{2(1-d)}$. Under our optimistic assumptions, if the attack changed the outcome, we would expect the fraction of voters who reported problems, a , to exceed:

$$a > m \frac{d}{2(1-d)}.$$

The model shows that the security impact of verification is non-linear, because every voter who corrects an error *both* increases the evidence that there is a problem *and* forces the attacker to cheat more in order to overcome the margin of victory. Figure 4 illustrates this effect.

With the 6.6% error detection rate from our non-intervention experiments and a close election with a 0.5% margin (the margin that causes an automatic recount in many states) a successful attack would cause as few as 0.018% of voters—less than 1 in 5000—to report a problem. Small changes in verification performance around our base rate cause relatively little change in the amount of evidence. More than doubling the error detection rate to 14% (the rate we found for prominent races) only increases the fraction of voters who report a problem to 0.039%. However, larger improvements have an outsized effect: with the 86% error detection rate from our most successful experiment, at least 1.5% of voters (1 in 67) would report problems.

Required detection rate Suppose election officials activate a countermeasure if the fraction of voters who report problems

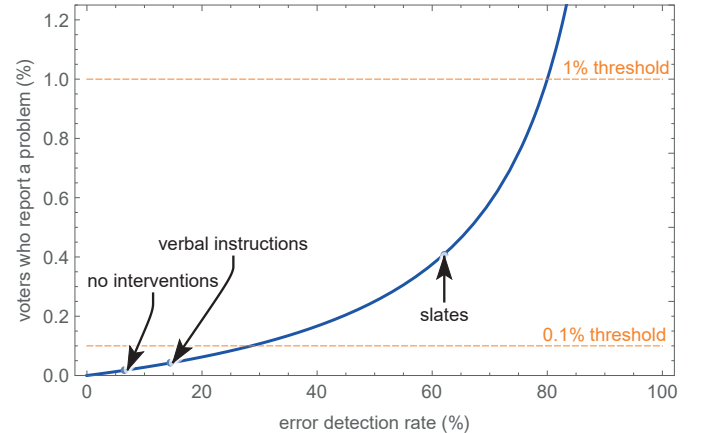


Fig. 4: *BMD security is highly sensitive to human performance.* Given a 0.5% margin of victory, we plot the percentage of voters who report a problem during the minimal outcome-changing attack as a function of the rate at which errors are detected and corrected. This model implies that using BMDs safely for all voters requires dramatically improved verification performance or very sensitive attack detection thresholds.

exceeds a threshold a^* . For a given margin, the countermeasure will be triggered by minimal outcome-changing fraud when:

$$d > \frac{2a^*}{m + 2a^*}.$$

An expensive countermeasure, like rerunning an election, will require a high trigger threshold—say, 1% of voters reporting a problem—to avoid false positives. With a 0.5% margin, reaching a 1% trigger threshold would require an error detection rate exceeding 80%. A less expensive countermeasure, such as an investigation, might be triggered by a lower threshold—say, 0.1%. Reaching this lower threshold in an election with a 0.5% margin would require an error detection rate greater than 29%. This suggests that using BMDs securely for all voters will require large improvements to verification performance or extremely low thresholds for triggering countermeasures.

Minimizing BMD voting helps dramatically Securing against misprinting attacks is far easier if only a small fraction of voters use BMDs than if all in-person voters do. This is because an attacker would be forced to cheat on a much larger fraction of BMD ballots in order to achieve the same change to the election results. Moreover, if the population of BMD voters is smaller than half the margin of victory, it is impossible for a BMD misprinting attack to change the outcome.

Let b be the fraction of voters who use BMDs. We can replace m in the expression above with $\frac{m}{b}$ and let a^* be the fraction of BMD voters that must report a problem to trigger the countermeasure. In Maryland, which uses hand-marked paper ballots but makes BMDs available to voters who request them, 1.8% of voters use BMDs [34]. With a 0.5% margin, as in the previous example, Maryland would reach a complaint threshold of 1% of BMD voters with an error detection rate of only 6.7%. If 5% of voters use BMDs, the error detection rate would need to be 17%. Our results suggest that these more modest rates of verification likely are achievable, in contrast to the far greater accuracy required when all voters use BMDs.

This model overestimates security An attacker might use any number of features (including several of the correlations we observed) to focus cheating on voters who are less likely to successfully catch errors. For instance, an attacker could preferentially modify ballots that have undervotes or a mix of selections from different parties. Attackers could also selectively target voters with visual impairments, such as those who use large text or an audio ballot. Other features, such as how long voters spend inspecting the candidate review screen, might also prove to be predictive of verification success. For these reasons, our simplified model is likely to overestimate the effectiveness of verification against sophisticated attackers.

We also note that some attackers may merely seek to cast doubt on election results by causing highly visible errors or failures—which are also possible with hand-marked paper ballots. However, in general, BMDs are vulnerable to all classes of computer-based attacks that affect hand-marked paper ballots and to others, such as the misprinting attack discussed here, to which hand-marked paper ballots are not susceptible.

VI. DISCUSSION

A. Limitations

It is challenging to capture real-world voter behavior in a mock election. However, our study followed established best practices [39], and we strived to create as realistic a polling environment as we could. It is impossible to know exactly how well we succeeded, but the effect seems to have been convincing: several people approached us to ask whether there was a real election taking place that they had not heard about. Our participants also seemed engaged in the study; many expressed strongly held political preferences in our survey (so much so that some refused to vote according to our slate), and a large majority reported voting in the 2018 midterm. On the other hand, the election used a ballot that was more than nine months old, which may have reduced participant motivation, and we had a few participants who reported that they did not vote in our state or were otherwise unfamiliar with our ballot. It is also possible that our results were skewed due to selection bias and observer effect.

Another limitation of our work is that we drew participants from a population that is locally but not nationally representative. Our participants tended to be younger, significantly better educated, more liberal, more likely to be female, and more likely to be Caucasian than the average voter in the United States [54]. Future work is needed to validate our study in more diverse and representative populations.

Although our results suggest that certain interventions can boost verification performance, the data is too sparse to provide a high-fidelity understanding of the magnitude of the improvements. In addition, due to time constraints, we were unable to test the interplay of all combinations of interventions, and some interventions appear to be sensitive to small changes (e.g., the difference in phrasing between script variants 2 and 3). Further study is needed to better characterize what makes interventions work and how they interact before we can confidently conclude that any particular set of procedures will be effective in practice.

B. Discussion of Findings

Our study provides the first concrete measurements of voter error detection performance using BMDs in a realistic voting environment. At a high level, we found that success rates without intervention are very low, around 6.6%. Some interventions that we tested did not significantly impact detection rates among participants, although others improved detection drastically and may serve as a roadmap for interventions to explore in further research. We discuss those interventions here.

1) *Verbal instructions can improve verification:* Notably, all interventions that involved poll workers verbally encouraging verification between the BMD and the scanner—those in **E6–E9**—resulted in higher ballot reviewing and error reporting rates. This, coupled with the fact that reviewing the printout was highly correlated with error detection across all of our results, suggests that interventions focused on causing the voter to review the ballot carefully may be helpful. On the

other hand, instructions at the beginning of the voting process (E5) and passive signage (E4) had no significant effect on error reporting. This pattern of effects is supported by findings from the usable security literature, which suggest that post-completion errors can be mitigated with timely interruptions that encourage individuals to take defensive steps [14].

It is worth noting that we also found that these interventions caused participants to take longer to submit their ballots, on average about twice as long. This could cause longer lines at polling places if these interventions are implemented without complementary procedural considerations, such as having adequate space for voters to stop and review their ballots.

2) *Effectiveness of slates*: Directing participants to vote for a provided slate of candidates, combined with verbally prompting them to review their printouts, resulted in strongly increased rate of error detection: 74% of participants who were given a slate and did not deviate from it noticed the errors. This finding may suggest that encouraging voters to write down their preferences in advance can boost verification.

However, the slates we used functioned quite differently from slates likely to be used in practice. The choices we provided were randomly generated and had no basis in the subject's preferences—in a real election, slates would reflect who the voter intended to vote for, most likely created by the voter or their political party [29]. It is possible that the success rate we observed was primarily due to participants carefully attempting to follow our instructions and vote for unfamiliar candidates. Further study is needed with more realistic slate conditions (i.e., asking subjects to write down their preferences) in order to assess whether slates really do help voters catch errors.

C. Recommendations

Since BMDs are widely used today, we recommend several strategies for improving voter verification performance. While we are unable to conclude that these strategies will enhance error detection to the point that BMDs can be used safely in close or small elections, our findings indicate that they can help.

1) *Design polling places for verification*: Polling place layout and procedures should be designed with verification in mind. As we have discussed, voters need time and space to verify their ballots. If tables or areas to stand out of the way are provided, voters will be able to carefully verify without causing lines to form or slowing polling place throughput. The presence of such a “verification station” might also encourage verification.

Another practical concern is privacy. Several of our participants expressed discomfort with the fact that we did not provide a privacy sleeve for their ballots (a requirement in Michigan), and that the scanner accepted the ballots face-up only, with one participant stating, “*I feel like inserting the ballot face up in the scanning machine will make people uncomfortable.*” Voters may not feel comfortable reviewing their ballots in front of poll workers but may be unsure where to go to review them privately.

2) *Incorporate post-voting verbal instructions*: As all of our script-based interventions that took place after the ballot was printed (E6–E9) showed an increase in verification performance, we recommend that poll workers interrupt voters

after their ballot has printed but before it is scanned and ask them to review it. Signage with a similar message to our scripts placed at the optical scanner (E4) or instructions before the participants voted (E5) did not result in significant differences in error detection; nevertheless, further study with additional variations is prudent before ruling out such strategies.

3) *Encourage personalized slate voting*: Although our study tested randomized slates, rather than personalized slates, the effect size was so large that we tentatively recommend encouraging the use of personalized slates by voters. In our experiments (E8 and E9), participants who were directed to vote using a randomized slate (and did not deviate) reported errors at a rate of 73%. If voters prepare their own slates at home (or use a printed slate prepared, for instance, by a political party or other organization), they can use them to check each selection on the BMD printout. We note that, since we did not directly test the use of personalized slates, further research is necessary to ascertain whether large performance gains are actually achieved. Furthermore, even if personalized slates are effective, the gain will be limited to the fraction of voters who can be induced to use them.

Slates have potential downsides and should be used with care. They have the potential to compromise ballot secrecy, so we recommend providing a closed trash can, paper shredder, or other means for voters to privately dispose of them before leaving the precinct. Coercion is also a threat, but voters could be advised to prepare multiple different slates as a defense.

4) *Help voters correct errors, and carefully track problems*: Verification-promoting interventions will be of little use if action cannot be taken to remedy misbehaving BMDs—something that even our participants expressed concern about.

First, it is crucial that polling places have a procedure for voters who want to correct their printed ballots. Several subjects commented that they would not know what to do if something was wrong with their ballot in a real election, indicating that this problem is present in current election procedures.

Second, detailed records should be kept about which BMD the voter used and what the specific issue was, including the contest and candidates involved (to the extent that the voter is willing to waive ballot secrecy). Problems should be treated as potentially serious even when the voter believes they are at fault—we note that several participants in our study believed they had made a mistake even though the BMD actually was programmed to be malicious. Problem reports should be centrally reported and tracked during the election, so that issues affecting multiple precincts can be identified as rapidly as possible.

5) *Prepare contingency plans*: What to do in the event that BMDs are known or suspected to be misbehaving is a more difficult question. If an elevated number of voters have a problem with a single machine, it should be taken out of service, provided there are other BMDs available for use (especially for voters with disabilities, who may have no alternative).

If widespread problem reports occur—particularly problems focused on a tightly contested race or significantly exceeding the rate reported in past elections—officials could consider

taking most BMDs out of service and encouraging all remaining voters who can to use hand-marked ballots. This raises logistical challenges: polling place would need to have enough ballots available for hand-marking, or the ability to print ballots on demand, and votes already cast on the BMDs would be suspect.

After the election, forensic analysis of the BMDs could be performed to attempt to determine the cause of reported errors. Unfortunately, such analysis cannot in general rule out that a sophisticated attack occurred and left no digital traces. Even if programming errors or attacks are uncovered, they may be impossible to correct if officials are unable to determine whether the effects were large enough to change the election outcome. The only recourse might be to re-run the election.

Our findings show that, in the event of an actual error or attack, the rate of reported problems is likely to be only the tip of the iceberg. In our non-intervention experiments, undetected errors outnumbered reported problems by almost twenty to one. Our results further suggest that an attacker who cleverly focused cheating on voters who were less likely to verify could achieve an even higher ratio of undetected errors. An effective response requires either being very sensitive to reported problems—which increases the chances that an attacker could trigger false alarms—or achieving very high error correction rates.

6) *Educate voters about BMD operations and risks:* Like in other human-in-the-loop security contexts, greater education could boost voters’ awareness of the importance of careful verification and boost error detection and reporting rates.

To this end, we recommend educating voters that the paper, rather than what the BMD screen shows, is the official record of their votes. Several of our participants said they realized after scanning that they should have, but did not, review their printouts. Others stated that they had checked the review screen on the machine and that they trusted the paper to be correct. It is likely that many participants incorrectly assumed that the BMDs, rather than the paper and scanner, tabulated their votes.

We also recommend educating voters about the possibility of BMD malfunction. Many of our participants seem not to have even considered that the machine might have changed their votes, as indicated by the voters who blamed themselves for the misprinted ballots. Raising threat awareness could help motivate voters to carefully inspect the paper, as well as give them greater confidence to report any discrepancies they detect.

7) *Consider the needs of voters with disabilities:* Further research is needed to specifically examine verification performance among voters with disabilities, but we offer some initial recommendations here. Detecting errors in printed ballots may be especially challenging for voters with impaired vision. Designing BMD ballots for maximum legibility might help, and so might encouraging voters who use text-to-speech devices to bring them to the polls for use during verification. Jurisdictions could also provide air-gapped accessible devices to read the ballot back to voters, in case voters do not have their own text-to-speech devices. These steps would have the added benefit of reinforcing the message that the content of the paper ballots is what gets counted. If BMDs are to live up to the promise of

better and more accessible voting, enabling all voters to verify their printed ballots is a must.

8) *Require risk-limiting audits:* Even perfectly verified paper ballots are of little use for security if they are not rigorously audited to confirm the results of computer-based tabulation. Fortunately, risk-limiting audits [32] (RLAs) are gaining momentum in the United States. Colorado, Nevada, and Rhode Island mandate statewide RLAs, and states including Michigan, Virginia, Georgia, and Pennsylvania are considering implementing them soon [17]. RLAs and effective verification are both necessary in order for paper to provide a strong defense against vote-stealing attacks, and we recommend that efforts to achieve both be pursued vigorously.

VII. CONCLUSION

We conducted the first empirical study of how well voters using BMDs detect errors on their printed ballots, which is a limiting factor to the level of security that a BMD-based paper trail can provide. Based on the performance of 241 human subjects in a realistic polling place environment, we find that, absent specific interventions, error detection and reporting rates are dangerously low. Unless verification performance can be improved dramatically, BMD paper trails, particularly when used by all in-person voters, cannot be relied on to reflect voter intent if the machines are controlled by an attacker.

Nevertheless, we also find that procedural interventions can improve rates of error detection and reporting, potentially increasing the security offered by BMDs. The interventions we tested should serve as examples of what is and is not likely to be effective, and we hope they will point the way for further research and experimentation. These findings add to the broad literature of human-in-the-loop security results and recommendations, and they provide additional examples of what does and does not work in human-centric security.

Our results should not be read as demonstrating that BMDs can be used securely. Further work is needed to explore the potential for attackers to predict which voters will verify, and additional human-subjects testing is necessary to confirm whether sufficient rates of verification success can be achieved in practice. The cost of implementing interventions and contingency plans may also be prohibitive. Nevertheless, BMDs do offer advantages, including uniform accessibility and ease of administration. We hope our work will help election officials make better informed choices as they weigh these benefits against the security risks of using BMDs for all voters.

ACKNOWLEDGMENTS

The authors are grateful to Jackie Fleischer Best, Eli Neiburger, Emily Howard, Matt Dubay, and everyone at the Ann Arbor District Library, without whom this study would not have been possible. We also thank Philip Stark for advice about our statistical analyses; Ben Adida, Monica Childers, and Ben VanderSloot for feedback about the experimental design; and the anonymous reviewers. This material is based in part upon work supported by the National Science Foundation under Grant No. CNS-1518888, by the Facebook Fellowship Program, and by the Andrew Carnegie Fellows Program.

REFERENCES

- [1] C. Z. Acemyan, P. Kortum, and D. Payne. Do voters really fail to detect changes to their ballots? An investigation of ballot type on voter error detection. *Proceedings of the Human Factors and Ergonomics Society*, 57:1405–1409, 2013.
- [2] D. Akhawe and A. P. Felt. Alice in Warningland: A large-scale field study of browser security warning effectiveness. In *22nd USENIX Security Symposium*, pages 257–272, 2013.
- [3] H. Almuhammedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal. Your location has been shared 5,398 times! A field study on mobile app privacy nudging. In *33rd ACM Conference on Human Factors in Computing Systems*, CHI, pages 787–796, 2015.
- [4] Ann Arbor District Library. Classic shop drop! Plus, fish election results!, Aug. 2019. <https://aadl.org/node/396262>.
- [5] Ann Arbor District Library. Mock the vote, July 2019. <https://aadl.org/node/395686>.
- [6] Ann Arbor District Library. Mock voting @ AADL, Sept. 2019. <https://aadl.org/node/397364>.
- [7] A. Appel, R. DeMillo, and P. Stark. Ballot-marking devices (BMDs) cannot assure the will of the voters, 2019. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3375755.
- [8] S. Bell, J. Benaloh, M. D. Byrne, D. DeBeauvoir, B. Eakin, G. Fisher, P. Kortum, N. McBurnett, J. Montoya, M. Parker, O. Pereira, P. B. Stark, D. S. Wallach, and M. Winn. STAR-Vote: A secure, transparent, auditable, and reliable voting system. *USENIX Journal of Election Technology and Systems*, 1(1), 2013.
- [9] M. Bernhard, J. Benaloh, J. A. Halderman, R. L. Rivest, P. Y. Ryan, P. B. Stark, V. Teague, P. L. Vora, and D. S. Wallach. Public evidence from secret ballots. In *2nd International Joint Conference on Electronic Voting, E-Vote-ID*, pages 84–109, 2017.
- [10] R. Bernstein, A. Chadha, and R. Montjoy. Overreporting voting: Why it happens and why it matters. *Public Opinion Quarterly*, 65(1):22–44, 2001.
- [11] D. Bowen et al. Top-to-bottom review of voting machines certified for use in California. Technical report, California Secretary of State, 2007. <https://www.sos.ca.gov/elections/ovsta/frequently-requested-information/top-bottom-review/>.
- [12] C. Bravo-Lillo, L. F. Cranor, J. Downs, and S. Komanduri. Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, 9(2):18–26, Mar. 2011.
- [13] C. Bravo-Lillo, S. Komanduri, L. F. Cranor, R. W. Reeder, M. Sleeper, J. Downs, and S. Schechter. Your attention please: Designing security-decision UIs to make genuine risks harder to ignore. In *9th Symposium on Usable Privacy and Security*, SOUPS, 2013.
- [14] M. D. Byrne and S. Bovair. A working memory model of a common procedural error. *Cognitive Science*, 21(1):31–61, 1997.
- [15] B. A. Campbell and M. D. Byrne. Now do voters notice review screen anomalies? A look at voting system usability. In *USENIX Electronic Voting Technology Workshop/Workshop on Trustworthy Elections*, EVT/WOTE, 2009.
- [16] Clear Ballot Group. ClearAccess administrators guide, 2015. <https://www.sos.state.co.us/pubs/elections/VotingSystems/systemsDocumentation/ClearBallot/ClearAccess/ClearAccessAdministratorsGuideRev4-0-r0.pdf>.
- [17] A. Cordova, L. Howard, and L. Norden. Voting machine security: Where we stand a few months before the New Hampshire primary. Brennan Center, 2019. <https://www.brennancenter.org/analysis/voting-machine-security-where-we-stand-six-months-new-hampshire-primary>.
- [18] L. F. Cranor. A framework for reasoning about the human in the loop. In *1st Conference on Usability, Psychology, and Security*, UPSEC. USENIX, 2008.
- [19] M. De Jong, J. Van Hoof, and J. Gosselt. Voters’ perceptions of voting technology: Paper ballots versus voting machine with and without paper audit trail. *Social Science Computer Review*, 26(4):399–410, 2008.
- [20] R. DeMillo, R. Kadel, and M. Marks. What voters are asked to verify affects ballot verification: A quantitative analysis of voters’ memories of their ballots, 2018. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3292208.
- [21] S. Egelman, L. F. Cranor, and J. Hong. You’ve been warned: An empirical study of the effectiveness of web browser phishing warnings. In *26th ACM Conference on Human Factors in Computing Systems*, CHI, pages 1065–1074, 2008.
- [22] S. P. Everett, K. K. Greene, M. D. Byrne, D. S. Wallach, K. Derr, D. Sandler, and T. Torous. Electronic voting machines versus traditional methods: improved preference, similar performance. In *26th ACM Conference on Human Factors in Computing Systems*, CHI, pages 883–892, 2008.
- [23] A. J. Feldman, J. A. Halderman, and E. W. Felten. Security analysis of the Diebold AccuVote-TS voting machine. In *USENIX Electronic Voting Technology Workshop*, EVT, 2007.
- [24] M. Fitts. SC chooses new voting machines that will print paper ballots but some fear it’s not safe. The Post and Courier, June 10, 2019. https://www.postandcourier.com/article_f86632ce-8b83-11e9-8dab-5fb7858906cc.html.
- [25] S. Fowler. Georgia awards new voting machine contract to Dominion Voting Systems. Georgia Public Broadcasting, July 29, 2019. <https://www.gpbnews.org/post/georgia-awards-new-voting-machine-contract-dominion-voting-systems>.
- [26] S. N. Goggin and M. D. Byrne. An examination of the auditability of voter verified paper audit trail (VVPAT) ballots. In *USENIX Electronic Voting Technology Workshop*, EVT, 2007.
- [27] K. K. Greene, M. D. Byrne, and S. P. Everett. A comparison of usability between voting methods. In *USENIX Electronic Voting Technology Workshop*, EVT, 2006.
- [28] Indiana Fiscal Policy Institute. Vote centers and election costs: A study of the fiscal impact of vote centers in Indiana, 2010. https://www.in.gov/sos/elections/files/IFPI_Vote_Centers_and_Election_Costs_Report.pdf.
- [29] D. Jones and B. Simons. *Broken Ballots: Will Your Vote Count?* CSLI Publications, 2012.
- [30] D. Kasdan. Early voting: What works. https://www.brennancenter.org/sites/default/files/publications/VotingReport_Web.pdf.
- [31] T. Kohno, A. Stubblefield, A. D. Rubin, and D. S. Wallach. Analysis of an electronic voting system. In *25th IEEE Symposium on Security and Privacy*, 2004.
- [32] M. Lindeman and P. Stark. A gentle introduction to risk-limiting audits. *IEEE Security & Privacy*, 10:42–49, 2012.
- [33] J. Mack. Who votes in Michigan? A demographic breakdown. MLive, 2018. <https://www.mlive.com/news/erry-2018/11/340b0f9c406363/who-votes-in-michigan-a-demogr.html>.
- [34] S. Maneki and B. Jackson. Re: Comments on Ballot Marking Devices usage for the 2018 elections, 2017. Letter to Maryland State Board of Elections, citing SBE data.
- [35] P. McDaniel, M. Blaze, and G. Vigna. EVEREST: Evaluation and validation of election-related equipment, standards and testing. Technical report, Ohio Secretary of State, 2007. <https://www.eac.gov/assets/1/28/EVEREST.pdf>.
- [36] National Academies of Sciences, Engineering, and Medicine. *Securing the Vote: Protecting American Democracy*. The National Academies Press, Washington, DC, 2018.
- [37] National Conference of State Legislatures. Funding elections technology, 2019. <https://www.ncsl.org/research/elections-and-campaigns/funding-election-technology.aspx>.
- [38] R. G. Niemi and P. S. Herrmson. Beyond the butterfly: The complexity of U.S. ballots. *Perspectives on Politics*, 1(2):317–326, 2003.
- [39] M. M. Olemba and M. Volkamer. E-voting system usability: Lessons for interface design, user studies, and usability criteria. In *Human-Centered System Design for Electronic Governance*, pages 172–201. IGI Global, 2013.
- [40] S. Patil, R. Hoyle, R. Schlegel, A. Kapadia, and A. J. Lee. Interrupt now or inform later? Comparing immediate and delayed privacy feedback. In *33rd ACM Conference on Human Factors in Computing Systems*, CHI, pages 1415–1418, 2015.
- [41] J. Petelka, Y. Zou, and F. Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. In *37th ACM Conference on Human Factors in Computing Systems*, CHI, 2019.
- [42] Pew Charitable Trusts. Colorado voting reforms: Early results. https://www.pewtrusts.org/_media/assets/2016/03/coloradovoting-reformsearlyresults.pdf, 2016.
- [43] Pro V&V. Test report for EAC 2005 VVSG certification testing: Clear-Ballot Group ClearVote 1.4 voting system, 2017. <https://www.eac.gov/file.aspx?A=kOBM5qPeI8KZlJyADXYTieIXLwsxw4gYKIVroEkEBMo%3D>.
- [44] W. Quesenbery. Ballot marking devices make voting universal. Center for Civic Design, 2019. <https://civicedesign.org/ballot-marking-devices-make-voting-universal/>.

- [45] W. Quesenbery, J. Cugini, D. Chisnell, B. Killam, and G. Reddish. Letter to the editor: Comments on “A methodology for testing voting systems”. *Journal of Usability Studies*, 2(2):96–98, 2007.
- [46] R. W. Reeder, A. P. Felt, S. Consolvo, N. Malkin, C. Thompson, and S. Egelman. An experience sampling study of user reactions to browser warnings in the field. In *36th ACM Conference on Human Factors in Computing Systems*, CHI, 2018.
- [47] R. Rivest. On the notion of ‘software independence’ in voting systems. *Philos. Trans. Royal Soc. A*, 366(1881):3759–3767, October 2008.
- [48] T. Selker, E. Rosenzweig, and A. Pandolfo. A methodology for testing voting systems. *Journal of Usability Studies*, 2(1):7–21, 2006.
- [49] T. Selker, E. Rosenzweig, and A. Pandolfo. Reply to comment on: The Methodology for Testing Voting Systems by Whitney Quesenbery, John Cugini, Dana Chisnell, Bill Killam, and Ginny Redish. *Journal of Usability Studies*, 2(2):99–101, 2007.
- [50] P. Stark. Conservative statistical post-election audits. *Annals of Applied Statistics*, 2(2):550–581, 2008.
- [51] P. B. Stark. There is no reliable way to detect hacked ballot-marking devices, 2019. <https://arxiv.org/abs/1908.08144>.
- [52] J. Sunshine, S. Egelman, H. Almuhiemedi, N. Atri, and L. F. Cranor. Crying wolf: An empirical study of SSL warning effectiveness. In *18th USENIX Security Symposium*, pages 399–416, 2009.
- [53] United States Senate Select Committee on Intelligence. Report on Russian active measures campaigns and interference in the 2016 U.S. election, 2019. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume1.pdf.
- [54] U.S. Census Bureau. QuickFacts: Ann Arbor, 2019. <https://www.census.gov/quickfacts/annarborcitymichigan>.
- [55] U.S. Election Assistance Commission. Designing polling place materials. <https://www.eac.gov/election-officials/designing-polling-place-materials/>.
- [56] Verified Voting. Ballot marking devices. <https://www.verifiedvoting.org/ballot-marking-devices/>.
- [57] Verified Voting. The Verifier: Polling place equipment. <https://www.verifiedvoting.org/verifier/>.
- [58] VSAP. Voting system for all people. <https://vsap.lavote.net/>.
- [59] Wall Street Journal. Election 2018: How we voted in the 2018 midterms, November 6, 2018. <https://www.wsj.com/graphics/election-2018-votecast-poll/>.
- [60] D. S. Wallach. On the security of ballot marking devices, 2019. <https://arxiv.org/abs/1908.01897>.
- [61] M. S. Wogalter. Communication-human information processing (C-HIP) model. In M. S. Wogalter, editor, *Handbook of Warnings*, chapter 5, pages 51–61. Lawrence Erlbaum Associates, Mahwah, NJ, 2006.
- [62] M. S. Wogalter and K. R. Laughery. Warning! sign and label effectiveness. *Current Directions in Psychological Science*, 5(2):33–37, 1996.

APPENDIX A
POLL WORKER SCRIPT

Our poll workers followed four versions of the script below: a baseline version, and three variants that each add one line.

VARIANT 1: Before the voter begins using the BMD, a poll worker asks them to check their ballot before it is scanned.

VARIANT 2: Before the voter deposits the ballot, a poll worker informs them that it is the official record of the vote.

VARIANT 3: Before the voter deposits the ballot, a poll worker asks whether they have carefully reviewed each selection.

When Subject Arrives (POLL WORKER A)

Hello! Before you begin, please fill out this Institutional Review Board consent form. [Point to form and pen.] If you have any questions, feel free to ask.

You are about to participate in a study about the usability of a new type of voting machine. You will be using one of these voting machines to make selections on your ballot, which will be a truncated version of the Ann Arbor 2018 midterm ballot. Once you are finished, your ballot will be printed from the printer beneath the machine, and you can review your ballot and deposit it in the ballot box over there. [Point out ballot box.] Feel free to vote your political preference or not; no identifying information will be collected that could match you with your votes. If you would like to quit at any time during the study, just say so.

VARIANT 1: *Please remember to check your ballot carefully before depositing it into the scanner.*

You may begin at any time.

Before Subject Deposits Ballot (POLL WORKER B)

VARIANT 2: *Please keep in mind that the paper ballot is the official record of your vote.*

VARIANT 3: *Have you carefully reviewed each selection on your printed ballot?*

After Subject Deposits Ballot (POLL WORKER B)

Thank you for participating! You are now finished with the study, and should fill out the exit survey. [Point to debrief survey computers.]

After Subject Completes Exit Survey (POLL WORKER B)

Thank you for your participation! You are now finished. If you have any questions about this study, you may ask them now, although I am unable to answer some questions due to the nature of the research. Here is a debrief form. [Hand subject a debrief form.] If you think of anything after you leave, you can reach [me/the principle investigators] through the information on the debrief form.

If you know anyone who might like to participate, please refer them here; we will be here [remaining time].

Thank you again for participating!

APPENDIX B
RECRUITMENT SCRIPT

An investigator used the following script to recruit library patrons to participate in the study:

Hello, do you have 10 minutes to participate in a study about a new kind of voting machine that is used in elections across the United States? This study will consist of voting using our voting machine and depositing a printed paper ballot into a ballot box, and then filling out a survey about the experience. If you would like to participate, we will need you to first sign a consent form. We will provide a flyer at the end of your participation with information about the study. We cannot make all details available at this time, but full details and research results will be made available within six months of the conclusion of this study. We thank you for your consideration and hope you choose to participate!

APPENDIX C

SLATE OF CANDIDATES FOR DIRECTED VOTING CONDITION

We randomly generated a slate of candidates and provided a printed copy to voters in certain experiments. The handout voters received is reproduced below:

Race	Candidate(s)
Governor and Lieutenant Governor	Bill Gelineau and Angelique Chaiser Thomas
Secretary of State	Mary Treder Lang
Attorney General	Lisa Lane Gioia
United States Senator	Debbie Stabenow
Representative in Congress 12th District	Jeff Jones
Member of State Board of Education (Vote for 2)	Tiffany Tilley Mary Anne Hering
Regent of the University of Michigan (Vote for 2)	Jordan Acker Joe Sanger
Trustee of Michigan State University (Vote for 2)	Mike Miller Bruce Campbell
Justice of the Supreme Court (Vote for 2)	Megan Kathleen Cavanagh Kerry Lee Morgan
Judge of Court of Appeals 3rd District Incumbent Position (Vote for 2)	Jane Marie Beckering Douglas B. Shapiro
Judge of Circuit Court 22nd Circuit Incumbent Position (Vote for 2)	Timothy Patrick Connors Carol Kuhnke
Judge of Probate Court Incumbent Position	Darlene A. O'Brien
Judge of District Court 14A District Incumbent Position	Thomas B. Bourque

EXHIBIT 5

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333245177>

Ballot-marking devices (BMDs) cannot assure the will of the voters

Preprint · April 2019

DOI: 10.13140/RG.2.2.29128.98566

CITATIONS

0

READS

96

3 authors, including:



[Richard A. Demillo](#)

Georgia Institute of Technology

112 PUBLICATIONS 6,562 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Center for 21st Century Universities [View project](#)

Ballot-marking devices (BMDs) cannot assure the will of the voters

Andrew W. Appel[†]
Princeton University

Richard A. DeMillo[†]
Georgia Tech

Philip B. Stark[†]
Univ. of California, Berkeley

April 21, 2019

Abstract

Computers, including all modern voting systems, can be hacked and misprogrammed. The scale and complexity of U.S. elections may require the use of computers to count ballots, but election integrity requires a paper-ballot voting system in which, regardless of how they are initially counted, ballots can be recounted by hand to check whether election outcomes have been altered by buggy or hacked software. Furthermore, secure voting systems must be able to recover from any errors that might have occurred.

However, paper ballots provide no assurance unless they accurately record the vote as the voter *expresses* it. Voters can express their intent by hand-marking a ballot with a pen, or using a computer called a ballot-marking device (BMD), which generally has a touchscreen and assistive interfaces. Voters can make mistakes in *expressing* their intent in either technology, but only the BMD is *also* subject to systematic error from computer hacking or bugs in the process of recording the vote on paper, after the voter has expressed it. A hacked BMD can print a vote on the paper ballot that differs from what the voter expressed, or can omit a vote that the voter expressed.

It is not easy to check whether BMD output accurately reflects how one voted in every contest. Research shows that most voters do not review paper ballots

[†]Authors are listed alphabetically; they contributed equally to this work.

printed by BMDs, even when clearly instructed to check for errors. Furthermore, most voters who do review their ballots do not check carefully enough to notice errors that would change how their votes were counted. Finally, voters who detect BMD errors before casting their ballots, can correct only their own ballots, not systematic errors, bugs, or hacking. There is no action that a voter can take to demonstrate to election officials that a BMD altered their expressed votes, and thus no way voters can help deter, detect, contain, and correct computer hacking in elections. That is, not only is it inappropriate to rely on voters to check whether BMDs alter expressed votes, *it doesn't work*.

Risk-limiting audits of a trustworthy paper trail can check whether errors in tabulating the votes *as recorded* altered election outcomes, but there is no way to check whether errors in how BMDs record *expressed* votes altered election outcomes. The outcomes of elections conducted on current BMDs therefore cannot be confirmed by audits. This paper identifies two properties of voting systems, *contestability* and *defensibility*, that are necessary conditions for any audit to confirm election outcomes. No commercially available EAC-certified BMD is contestable or defensible.

To reduce the risk that computers undetectably alter election results by printing erroneous votes on the official paper audit trail, the use of BMDs should be limited to voters who require assistive technology to vote independently.

Elections for public office and on public questions in the United States or any democracy must produce outcomes based on the votes that voters *express* when they indicate their choices on a paper ballot or on a machine. Computers have become indispensable to conducting elections, but computers are vulnerable. They can be hacked—compromised by insiders or external adversaries who can replace their software with fraudulent software that deliberately miscounts votes—and they can contain design errors and bugs—hardware or software flaws or configuration errors that result in mis-recording or mis-tabulating votes. Therefore there must be some way, *independent* of any software in any computers, to ensure that reported election outcomes are correct, i.e., consistent with the expressed votes as intended by the voters.

Voting systems should be *software independent*, meaning that “an undetected change or error in its software cannot cause an undetectable change or error in an election outcome” [23]. Indeed, version 2.0 of the Voluntary Voting System Guidelines (VVSG 2.0) incorporates this principle [7].

Software independence is similar to tamper-evident packaging: if somebody opens the container and disturbs the contents, it will leave a trace.

While software independence is crucial, it is not enough: *who* can detect errors and *what happens* when errors are detected are just as important. Even if individual voters in principle could detect changes to their votes on the BMD-generated ballot, unless voters can provide convincing evidence of problems to the public and unless election officials take appropriate remedies when presented with such evidence, software independence alone does not guarantee that outcome-changing problems—accidental or malicious—can be caught, much less corrected.

To be acceptable, a voting system also must be *contestable*: We say that voting system is contestable if any change or error in its software that results in a change or error in a reported election outcome can generate public evidence that the reported outcome is not trustworthy. Evidence available only to individual voters¹ does not suffice: “trust me” is not evidence. If a voting system is contestable, it is software independent, but the converse is not necessarily true. If a voting system is not contestable, then problems might never see the light of day, much less be corrected.

Voting systems must also be *defensible*. We say that a voting system is defensible if, when it reports the correct outcome, it can also generate convincing public evidence that the reported outcome is correct. Evidence available only to an election official or voting system vendor does not suffice: in other words, “trust me” is not evidence. If a voting system is not defensible, then it is vulnerable to “crying wolf”: malicious actors could claim that the system malfunctioned when in fact it did not, and election officials will have no way to prove otherwise.

Rivest and Wack [23] also define a voting system to be *strongly software independent* if it is software independent and moreover, a detected change or error in an election outcome (due to change or error in the software) can be corrected using only the ballots and ballot records of the current election.² Strong software independence combines tamper evidence with a kind of resilience: there’s a way to tell whether faulty software caused a problem, and a way to recover from the problem if it did.

The only known practical technology for a contestable, defensible, strongly software independent voting system is *hand-marked paper ballots*, kept physically secure,

¹Specifically, if the voter is selected candidate A on the touchscreen of a BMD, but the BMD prints candidate B on the paper ballot, then this A-vs-B evidence is available to the individual voter, but the voter cannot demonstrate this evidence to anyone else, since nobody else saw—nor should have seen—where the voter touched the screen. Thus, the voting system cannot generate public evidence of errors recording expressed votes, even if those errors altered the reported outcome.

²The only alternative remedy would be to void the results of the entire election and conduct a new one.

counted by machine, audited manually, and recountable by hand.³

Over 40 states now use some form of paper ballot for most voters [14]. Most of the remaining states are taking steps to adopt paper ballots. But *not all voting systems that use paper ballots are equally secure*. Some are not software independent. Some are software independent but not contestable or defensible. In this report we explain:

- *Hand-marked paper ballot* systems are the only practical technology for contestable, defensible voting systems.
- *Some ballot-marking devices (BMDs)* can be software independent, but they are neither contestable nor defensible. Hacked or misprogrammed BMDs can alter election outcomes undetectably, and elections conducted using BMDs do not provide public evidence that reported outcomes are correct. Therefore BMDs should not be used by voters who are able to mark an optical-scan ballot with a pen.
- *All-in-one BMD or DRE+VVPAT voting machines* are not software independent, contestable, or defensible. They should not be used in public elections.

Terminology

Although a voter may form an intention to vote for a candidate or issue days, minutes, or seconds before actually casting a ballot, that intention is a psychological state that cannot be directly observed by anyone else. Others can have access to that intention through what the voter (privately) *expresses* to the voting technology by interacting with it, e.g., by making selections on a BMD or marking a ballot by hand.⁴ Voting systems must accurately record the vote as the voter *expressed* it.

With a *hand-marked paper ballot optical-scan* system, the voter is given a paper ballot on which all choices (candidates) in each contest are listed; next to each candidate

³The election must also generate convincing evidence that physical security of the ballots was not compromised, and the audit must generate convincing public evidence that the audit itself was conducted correctly.

⁴We recognize that voters make mistakes in expressing their intentions. For example, they may misunderstand the layout of a ballot or through a perceptual error or lapse of attention make an unintended choice. The use of touchscreen technology does not necessarily correct for such user errors, as every smartphone user who has mistyped an important text message knows. Poorly designed ballots, poorly designed touchscreen interfaces, and poorly designed assistive interfaces increase the rate of error in voters' expressions of their votes. For the purposes of this report, we assume that properly engineered systems seek to minimize such usability errors.

is a *target* (typically an oval or other shape) which the voter marks with a pen to indicate a vote. Ballots may be either preprinted or printed (unvoted) at the polling place using *ballot on demand* printers. In either case, the voter creates a tamper-evident record of intent by marking the printed paper ballot with a pen.

Such hand-marked paper ballots may be scanned and tabulated at the polling place using a *precinct-count optical scanner* (PCOS), or may be brought to a central place to be scanned and tabulated by a *central-count optical scanner* (CCOS). Mail-in ballots are typically counted by CCOS machines.

After scanning a ballot, a PCOS machine deposits the ballot in a secure, sealed ballot box for later use in recounts or audits; this is *ballot retention*. Ballots counted by CCOS are also retained for recounts or audits.⁵

Paper ballots can also be hand counted, but in most jurisdictions (especially where there are many contests on the ballot) this is hard to do quickly; Americans expect election-night reporting of unofficial totals. Hand counting—i.e., manually determining votes directly from the paper ballots—is appropriate for audits and recounts.

A *ballot-marking device* (BMD) provides a computerized user interface that presents the ballot to voters and captures their expressed selections, for instance, a touchscreen interface or an assistive interface that enables voters with disabilities to vote independently. Voter inputs (expressed votes) are recorded electronically. When a voter indicates that the ballot is complete and ready to be cast, the BMD prints a paper version of the electronically marked ballot. We generally use the term *BMD* for devices that mark ballots but do not tabulate or retain them, and *all-in-one* for devices that combine ballot marking, tabulation, and retention into the same paper path.

The paper ballot printed by a BMD may be in the same format as an optical-scan form (e.g., with ovals filled as if by hand) or it may list just the names of the candidate(s) selected in each contest. The BMD may also encode these selections into barcodes or QR codes for optical scanning. We discuss issues with barcodes later in this report.

An *all-in-one touchscreen voting machine* combines computerized ballot marking, tabulation, and retention in the same paper path. All-in-one machines come in several configurations:

- DRE+VVPAT machines—direct-recording electronic (DRE) voting machines with a voter-verifiable paper audit trail (VVPAT)—provide the voter a touchscreen (or

⁵Regulations and procedures governing custody and physical security of ballots are uneven and in many cases inadequate, but simple to correct because of decades of development of best practices.

other) interface, then print a paper ballot that is displayed to the voter under glass. The voter is expected to review this ballot and approve it, after which the machine deposits it into a ballot box. DRE+VVPAT machines do not contain optical scanners; that is, they do not read what is marked on the paper ballot; instead, they tabulate the vote directly from inputs to the touchscreen or other interface.

- BMD+Scanner all-in-one machines⁶ provide the voter a touchscreen (or other) interface to input ballot choices and print a paper ballot that is ejected from a slot for the voter to inspect. The voter then reinserts the ballot into the slot, after which the all-in-one BMD+scanner scans it and deposits it into a ballot box.

Opscan+BMD with separate paper paths. At least one model of voting machine (the Dominion ICP320) contains an optical scanner and a BMD in the same cabinet,⁷ so that the optical scanner and BMD-printer are not in the same paper path; no possible configuration of the software could cause a BMD-marked ballot to be deposited in the ballot box without human handling of the ballot. We do not classify this as an *all-in-one* machine.

Hacking

There are many forms of computer hacking. In this analysis of voting machines we focus on the alteration of voting machine software so that it miscounts votes or mis-marks ballots to alter election outcomes. There are many ways to alter the software of a voting machine: a person with physical access to the computer can open it and directly access the memory; one can plug in a special USB thumbdrive that exploits bugs and vulnerabilities in the computer's USB drivers; one can connect to its WiFi port or Bluetooth port or telephone modem (if any) and exploit bugs in those drivers, or in the operating system.

“Air-gapping” a system (which is to say, disconnecting it from a wired network) does not automatically protect it. Before each election, election administrators must transfer a *ballot definition* into the voting machine by inserting a *ballot definition cartridge* that was programmed on election-administration computers that may have been connected previously to various networks; it has been demonstrated that vote-changing viruses can propagate via these ballot-definition cartridges [13].

Hackers might be corrupt insiders with access to a voting-machine warehouse; cor-

⁶The ES&S ExpressVote can be configured as either a BMD or a BMD+Scanner all-in-one.

⁷More precisely, the ICP320 optical scanner and the BMD audio+buttons interface are in the same cabinet, but the printer is a separate box.

rupt insiders with access to a county’s election-administration computers; outsiders who can gain remote access to election-administration computers; outsiders who can gain remote access to voting-machine manufacturers’ computers (and “hack” the firmware installed in new machines, or the firmware updates supplied for existing machines), and so on. Supply-chain hacks are also possible: the hardware installed by a voting system vendor may have malware pre-installed by the vendor’s component suppliers.⁸

Computer systems (including voting machines) have so many layers of software that it is impossible to make them perfectly secure [18, pp. 89–91]. When manufacturers of voting machines use the best known security practices, adversaries may find it more difficult to hack a BMD or optical scanner—but not impossible. Every computer in every critical system is vulnerable to compromise through hacking, insider attacks or exploiting design flaws.

Election assurance through risk-limiting audits.

To ensure that the reported outcome of each contest is that outcome that would have been found by accurately tabulating the voters’ intent as recorded, the most practical known technology is a *risk-limiting audit* (RLA) of paper ballots [25, 26, 17]. The National Academies of Science, Engineering, and Medicine, recommend routine RLAs after every election [18], as do many other organizations and entities concerned with election integrity.⁹

A RLA involves manually inspecting randomly selected paper ballots following a rigorous protocol. The audit stops if and when the sample provides convincing evidence that the reported outcome is correct; otherwise, the audit continues until every ballot has been inspected manually and the correct electoral outcome is known.

RLAs can check whether errors in tabulating recorded votes altered election outcomes, but cannot check whether errors in recording expressed votes altered election outcomes. Properly preserved hand-marked paper ballots ensure that expressed votes are identical to recorded votes. On the other hand, BMDs might not record expressed

⁸Given that many chips and other components are manufactured in China and elsewhere, this is a serious concern. Carsten Schürmann has found Chinese pop songs on the internal memory of voting machines (C. Schürmann, personal communication, 2018). Presumably those files were left there accidentally—but this shows that malicious code *could* have been pre-installed deliberately, and that neither the vendor’s nor the election official’s security and quality control measures discovered and removed the extraneous files.

⁹Among them are the Presidential Commission on Election Administration, the American Statistical Association, the League of Women Voters, and Verified Voting Foundation.

votes accurately, for instance, if BMD software has bugs, was misconfigured, or was hacked. Thus, RLAs that rely on BMD output cannot ensure that election outcomes are correct.

RLAs protect against vote-tabulation errors, whether those errors are caused by failures to follow procedures, misconfiguration, miscalibration, faulty engineering, bugs, or malicious hacking.¹⁰ The *risk limit* of a risk-limiting audit is the maximum chance that an outcome that is incorrect because of tabulation errors will pass the audit without being corrected. The risk limit should be determined as a matter of policy or law. For instance, a 5% risk limit means that, if a reported outcome is wrong because of tabulation errors, there is at least a 95% chance that the post-election audit will correct it. Smaller risk limits give higher confidence in election outcomes, but require inspecting more ballots, other things being equal. RLAs never revise a correct outcome.

RLAs can be very efficient, depending in part on how the voting system is designed. If the computer results are accurate, an efficient RLA with a risk limit of 5% requires examining about (7 divided by the margin) ballots selected randomly from the contest.¹¹ For instance, if the margin of victory is 10% and the results are correct, the RLA would need to examine about $7/10\% = 70$ ballots to confirm the outcome at 5% risk. For a 1% margin, the RLA would need to examine about $7/1\% = 700$ ballots. The sample size does not depend (much) on the total number of ballot cast in the contest, only on the margin of the winning candidate's victory.

A paper-based voting system (such as one that uses optical scanners) is systematically more secure than a paperless system (such as DREs) only if the paper trail is trustworthy and the results are audited against the paper trail using a rigorous method such as an RLA.

But what if the paper ballots are not a trustworthy record of the votes expressed by the voters? If it is possible that error, hacking, bugs, or miscalibration caused the recorded votes to differ from the expressed votes, an RLA or even a full hand recount does not provide convincing public evidence that election outcomes are correct: such a system cannot be *defensible*. In short, paper ballots provide little assurance against hacking if they are never examined or if the paper might not accurately record the vote expressed by the voter.

¹⁰RLAs do not protect against problems that cause BMDs to print something other than what was shown to the voter on the screen, nor do they protect against problems with ballot custody.

¹¹Technically, it is the *diluted margin* that enters the calculation. The diluted margin is the number of votes that separate the winner with the fewest votes from the loser with the most votes, divided by the number of ballots cast, including undervotes and invalid votes.

Security Flaws

A BMD-generated paper trail is not a reliable record of the vote expressed by the voter. Like any computer, a BMD (or a DRE+VVPAT) is vulnerable to hacking, installation of unauthorized (fraudulent) software, and alteration of installed software.¹²

If a hacker sought to steal an election by altering BMD software, what would the hacker program the BMD to do? In cybersecurity practice, we call this the *threat model*.

The simplest threat model is this one: In some contests, not necessarily top-of-the-ticket, change a small percentage of the votes (such as 5%).

In recent national elections, analysts have considered a candidate who received 60% of the vote to have won by a landslide. Many contests are decided by less than a 10% margin. Changing 5% of the votes can change the margin by 10%, because “flipping” a vote for one candidate into a vote for a different candidate changes the difference in their tallies—i.e., the margin—by 2 votes. If hacking or bugs or misconfiguration could change 5% of the votes, that would be a very significant threat.

Although public and media interest often focus on top-of-the-ticket races such as President and Governor, elections for lower offices such as state representatives, who control legislative agendas and redistricting, and county officials, who manage elections and assess taxes, are just as important in our democracy. But most voters are not as familiar with the names of the candidates for those offices.

Research by one of us [9], in a real polling place in Tennessee during the 2018 election, found that half the voters *didn't look at all* at the paper ballot printed by a BMD, even when they were holding it in their hand and directed to do so while carrying it from the BMD to the optical scanner. Those voters who did look at the BMD-printed ballot spent *an average of 4 seconds* examining it to verify that the eighteen or more choices they made were correctly recorded. That amounts to 222 milliseconds per contest, barely enough time for the human eye to move and refocus under perfect conditions and not nearly enough time for perception, comprehension, and recall [22].^{13 14}

¹²It is also vulnerable to bugs and misconfiguration.

¹³You might think, “the voter really *should* carefully review their BMD-printed ballot.” But because the scientific evidence shows that voters *do not* [9] and cognitively *cannot* [12] perform this task well, legislators and election administrators should provide a voting system that counts the votes *as voters express them*.

¹⁴Studies of voter confidence about their ability to verify their ballots are not relevant: in typical situations, subjective confidence and objective accuracy are at best weakly correlated. The relationship between confidence and accuracy has been studied in contexts ranging from eyewitness accuracy [6, 8,

The same study found that among voters who examined their hand-marked ballots, half were unable to recall key features of ballots cast moments before, a prerequisite step for being able to recall their own ballot choices.

Suppose, then, that 10% of voters examine their paper ballots carefully enough to even *see* the candidate's name recorded as their vote for legislator or county commissioner. Of those, perhaps only half will remember the name of the candidate they intended to vote for.¹⁵

Of those who notice that the vote printed is not the candidate they intended to vote for, what are they supposed to think, and what are they supposed to do? Do they think, "Oh, I must have made a mistake on the touchscreen," or do they think, "Hey, the machine is cheating or malfunctioning!" There's no way for the voter to know for sure—voters do make mistakes—and there's *absolutely* no way for the voter to prove to a pollworker or election official that a BMD printed something other than what the voter entered on the screen.¹⁶

Either way, polling place procedures generally advise voters to ask a pollworker for a new ballot if theirs does not show what they intended. Pollworkers should void that BMD-printed ballot, and the voter should get another chance to mark a ballot. Anecdotal evidence suggests that many voters are too timid to ask, or don't know that they have the right to ask, or are not sure whom to ask. Even if a voter asks for a new ballot, training for pollworkers is uneven, and we are aware of no formal procedure for resolving disputes if a request for a new ballot is refused. Moreover, there is no sensible protocol for ensuring that BMDs that misbehave are investigated—nor can there be, as we argue below.

Let's summarize. If a machine alters votes on 5% of the ballots (enabling it to change the margin by 10%), then optimistically we might expect $\frac{1}{20} \times \frac{1}{10} \times \frac{1}{2}$ or 0.25% of the voters to request a new ballot and correct their vote. This means that the machine will change the margin by 9.75% and get away with it.

In this scenario, 0.25% of the voters, one in every 400 voters, has requested a new ballot. You might think, "that's a form of *detection* of the hacking." But is isn't, as a

28] to confidence in psychological clinical assessments [10] and social predictions [11]. The disconnect is particularly severe at high confidence. Indeed, this is known as "the overconfidence effect." For a lay discussion, see *Thinking, Fast and Slow* by Nobel economist Daniel Kahnemann [15].

¹⁵We ask the reader, "do you know the name of the most recent losing candidate for county commissioner?" We recognize that some readers of this document *are* county commissioners, so we ask those readers to imagine the frame of mind of their constituents.

¹⁶Voters should *certainly* not videorecord themselves voting! That would defeat the privacy of the secret ballot and is illegal in most jurisdictions.