IN THE UNITED STATES DISTRICT COURT FOR THE NORTHERN DISTRICT OF GEORGIA ATLANTA DIVISION

DONNA CURLING, et al. **Plaintiff,** **vs.** **BRIAN P. KEMP, et al.** **Defendant.**	)	**CIVIL ACTION FILE NO.: 1:17-cv-2989-AT**

SIXTH DECLARATION OF PHILIP B. STARK

PHILIP B. STARK hereby declares as follows:

1. This statement supplements my declarations of September 9, 2018; September 30, 2018; October 22, 2019; December 16, 2019; and August 23, 2020. I stand by everything in the previous declarations.
2. In his declaration of 25 August 2020, Defendant's expert Dr. Juan Gilbert points to a peer-reviewed paper and an ArXiV manuscript about voter verification of BMD printout:
    a. Bernhard, M., A. McDonald, H. Meng, J. Hwa, N. Bajaj, K. Chang, and J.A. Halderman, 2020. Can Voters Detect Malicious Manipulation of Ballot Marking Devices? *IEEE Proc. Security & Privacy*, 1, 679-694. DOI 10.1109/SP40000.2020.00118.
    b. Kortum, P., M.D. Byrne, and J. Whitmore, 2020. Voter Verification of BMD Ballots Is a Two-Part Question: Can They? Mostly, They Can. Do They? Mostly, They Don't. https://arxiv.org/abs/2003.04997 (last visited 31 August 2020).

3. Dr. Gilbert suggests, on the basis of these papers, that reminding voters to check their printout is adequate protection against BMD malfunction, misconfiguration, and hacking. But those papers do not support that conclusion.
4. As is true for many things having to do with elections, numbers matter when considering whether a safeguard is adequate. Dr. Gilbert does not consider the numbers, only heuristics.
5. Dr. Gilbert cites the Kortum et al. (2020) manuscript: "Of the 25 voters who actually examined the printout, 19 of them detected at least one anomaly." Gilbert declaration of 25 August 2020, at 5. This is a detection rate of 19/25 = 76 percent among subjects who checked the printout. Overall, they found that only 23 percent of subjects examined the printout and only 17.6 percent of subjects noticed errors.
6. In the Kortum et al. study, the rate at which voters examined the printout and the rate at which they noticed errors depended on the number of contests on the ballot and the number of errors in the printout. For instance, for a ballot with 40 contests,[1] about 15 percent of voters reviewed the printout, of whom roughly 60 percent noticed errors. Kortum et al. (2020) at Figures 5, 6. And the rate of detecting errors among voters who inspected the printout was roughly 65 percent when there was only one error. Kortum et al. (2020) at Figure 8.
7. Taking the 76 percent number result at face value,[2] it implies that *even if some intervention could miraculously provoke every voter to check the printout*, about 24

[1] This seems closer to Georgia's elections than the other experimental condition, a ballot with only 5 contests. See paragraph 16, *infra.*
[2] This rate is an average across a number of experimental conditions involving length of the ballot, number of votes altered, and the style of the BMD printout.

percent would not notice changes to their votes. For longer ballots like those in Georgia, the Kortum et al. (2020) study finds that roughly 40 percent of voters would not notice errors, even if every voter checked the printout.

8. Consider what that means for a moderately close election. Imagine an election between Alice and Bob and suppose that every ballot has a valid vote (no undervotes or invalid ballots).
9. Suppose Alice actually won with a margin of 2 percent. If malware changed the vote from Alice to Bob on 4.2 percent of printouts and 76 percent of affected voters noticed the change and marked a new printout, the collection of BMD printouts would still erroneously show a win for Bob. If only 60 percent of voters would notice errors, malware could make Bob appear to win by changing votes on 2.5 percent of printouts.
10. If there were undervotes or invalid votes, malware could change the outcome by altering even fewer printouts. For instance, if the undervote rate were 50 percent (equivalently, if the contest is on only half the ballots in a jurisdiction), the outcome according to the printout could be flipped to a win for Bob by altering the vote on 2.1 percent of printouts if the detection rate is 76 percent, or 1.3 percent of printouts if the detection rate is 60 percent.
11. These numbers scale with the margin. For instance, if the true margin is 1 percent (rather than 2 percent) and there are no invalid votes or undervotes, malware can make Bob appear to win by altering half as many printouts, 2.1 percent for a detection rate of 76 percent or 0.8 percent for a detection rate of 60 percent. And if the undervote rate is 50 percent or the contest is on only half the ballots in the jurisdiction, malware can make Bob appear to win by altering less than 1.1 percent of the printouts if 76 percent of voters

would catch and correct errors, or by altering 0.4 percent of printouts if 60 percent of voters would catch and correct errors.

12. If the margin were halved to 0.5 percent, the numbers in paragraph 11 would be halved as well. Even smaller margins occur in real elections, and some contests are on less than 50 percent of the ballots cast in a jurisdiction. As long as the rate at which voters detect and correct errors is less than 100 percent, there will be contests whose outcomes can be altered by BMD malware that changes an arbitrarily small number of votes.

13. The numbers in paragraphs 7–12, *supra*, are computed on the assumption that there is some magical intervention that could get every voter to check the printout. There is no reason to believe such an intervention exists. Neither paper Dr. Gilbert cites says there is. Indeed, according to Bernhard et al. (2020), reminding voters verbally to review their ballots increased the rate at which voters detected errors from less than 7 percent to less than 20 percent. Bernhard et al. (2020) at Table 1. The highest rate at which subjects noticed errors—which occurred only when voters were given a written slate to use for reference—was below 86 percent.[3] Bernhard et al. (2020) at Table 1. Evidently, details matter: "Neither signage [] nor poll worker instructions issued before the participant began voting [] yielded a statistically significant improvement to any aspect of verification performance. In contrast, poll worker instructions issued after the ballot was printed [] did have a positive effect, boosting reporting rates to 20% on the exit survey and 14% to poll workers (averaged across the experiments)." Bernhard et al. (2020) at 7–

[3] This was in a relatively small sample: only 21 subjects received the "treatment" that led to an 85.7 percent detection rate (i.e., 17 of the 21 noticed an error). If the subjects are considered a random sample of voters, a 95 percent lower confidence bound on the rate at which voters would notice errors is 67 percent. This bound was calculated by inverting binomial hypothesis tests.

8. Dr. Gilbert's claim is speculation, not science. I am not aware of any evidence to support the conclusion that reminding voters to check the printout can possibly ensure that BMD misbehavior did not change the apparent winner of one or more contests in an election.

14. The insidious gap in BMD security is that if a voter notices and complains that the BMD altered their vote, there is still no way for an election official to tell whether the BMD malfunctioned, the voter erred, or the voter is crying "wolf."[4] BMD systems do not provide any evidence whatsoever that a voter can present to election officials to demonstrate that BMDs malfunctioned. In the terminology of Appel et al. (2020), BMD-based voting systems are not *contestable*. Conversely, there is no way for an election official to demonstrate that BMD malfunctions did not alter election outcomes: BMD-based voting systems are not *defensible* in the terminology of Appel et al. (2020).

15. *Some* voters detecting problems and correcting their votes does nothing for the voters who do *not* notice and does not ensure that reported outcomes are correct, no matter how loudly the voters who notice problems complain. And the number of voters who notice and complain could be very small, even if errors, malfunctions, or hacking altered election outcomes. For instance, in the last example in paragraph 11, *supra*, only 0.6 x 0.004 = 0.24 percent of voters would request a fresh chance to mark a ballot. It is implausible that election officials would call for a new election simply because 0.24

[4] See, e.g., Appel, A.W., R. DeMillo, and P.B. Stark, 2020. Ballot-Marking Devices Cannot Ensure the Will of the Voters, *Election Law Journal*, DOI 10.1089/elj.2019.0619. Appel, A.W. and P.B. Stark, 2020. Evidence-Based Elections: Create a Meaningful Paper Trail, Then Audit, *Georgetown Law Technology Review, 4,* 523–541. Stark, P.B., and R. Xie, 2020. Testing Cannot Tell Whether Ballot-Marking Devices Alter Election Outcomes, ArXiV preprint, https://arxiv.org/abs/1908.08144 (last visited 31 August 2020).

percent of voters requested a fresh ballot. The "spoiled ballot rate" is not a usefully reliable indicator of malfeasance or malfunction. See Stark and Xie (2020).

16. I understand that in Fayette County, Georgia, ballots for the 19 May Democratic Presidential Preference Primary and Nonpartisan General Election included 29 contests. As an instructor with 32 years of experience who has taught and tested tens of thousands of undergraduate and graduate students, I am quite confident that the majority of college students would not reliably notice a change to votes nor the addition or omission of a contest from a list that long without relying on a written "slate" of selections. Human memory and human attention are not perfect. Even with a written slate, some voters will not notice changes.

I declare under penalty of perjury, in accordance with 28 U.S.C. § 1746, that the foregoing is true and correct.

Executed on this date, 31 August 2020.

Philip B. Stark