USENIX Journal of Election Technology and Systems (JETS)
Volume 3, Number 1 • December 2014
www.usenix.org/jets/issues/0301

DIFFUSION OF VOTER RESPONSIBILITY: POTENTIAL FAILINGS IN E2E VOTER RECEIPT CHECKING

ESTER MOHER, Children's Hospital of Eastern Ontario Research Institute
JEREMY CLARK, Concordia University
ALEKSANDER ESSEX, Western University

End-to-end verifiable (E2E) voting systems provide voters with (privacy preserving) receipts of their ballots allowing them to check that their votes were correctly included in the final tally. A number of recent studies and field tests have examined the usability of ballot casting and receipt checking. Simply checking receipts, however, is not enough to provide strong assurance that the election outcome is correct; voters must also be counted on to report any discrepancies between their receipts and the official record when they occur. In this paper we designed and ran a study examining the frequency and conditions under which voters (a) check their receipts, and (b) report discrepancies when they occur. Participants were recruited online and were asked to vote in a survey on charitable giving. Similar to previous work, we found that the proportion of voters performing a receipt check was low. More importantly, within this group, we found that the proportion of voters reporting discrepancies was also low. We did however observe that the incidence of receipt checking was significantly higher when the election outcome was unanticipated or unexpected by voters. In the condition with an adverse election result we observed that, while 7.5% of voters checked receipts, only 0.5% filed a dispute when shown an incorrect receipt. With such low reporting rates, E2E voting systems will struggle to detect fraud with high confidence, especially in elections with narrow margins of victory. We posit, therefore, that improving the usability of the receipt check component in E2E systems is an important open problem.

1. INTRODUCTION

Relative to conventional ballot casting methods, end-to-end verifiable (E2E) voting systems can be designed to provide cryptographically strong security guarantees. Depending on the specifics of the system, the correctness of the election outcome can be stated unconditionally (*i.e.,* without requiring any cryptographic or computational assumptions), even in the face of a corrupt election authority. Nearly all E2E systems share a common design element: after casting her ballot, the voter retains a privacy-preserving record of her vote called a *receipt*. After the election, the election authority publishes a list of receipts it asserts to have collected. This list is cryptographically transformed into a corresponding tally, hiding the voter-vote association and ensuring that any manipulation of this process is detectable with overwhelming probability.

If the election authority misbehaves and publishes an incorrect receipt list, the votes associated with these receipts could be modified. As one of the practical limitations of the end-to-end verification paradigm, the cryptographic checks cannot detect incorrect receipts; the burden falls to the voters to check. In order to detect maliciously modified receipts, the security guarantees of E2E systems have typically rested on the tacit assumption that a sufficient proportion of voters will (i) check that their receipts match the list produced by the election authority, (ii) report any discrepancies, and (iii) be able to convince others of the validity of an honest dispute. The system must be able distinguish between a voter mistake or lie (*e.g.,* to cast doubt on the outcome), and a malicious election authority. Recent trial elections and user studies provide some data on the first property: between 4% and 54% of voters check their receipts depending on the election. Contemporary E2E systems like Scantegrity II [Carback et al. 2010] provide a mechanism to ensure the third property, sometimes called *dispute resolution* or *accountability*. To our knowledge, however, the second property has not been explicitly considered in the literature. It is often implicitly assumed any voter checking their receipt will always uncover and report any discrepancy found. There are, however, a variety of reasons that a voter might not legitimately file a dispute: a voter might check but not notice the discrepancy, notice but assume the fault is their own, acknowledge the error should be reported but fail to do so because its too much work, or conclude the responsibility of reporting rests with another party.

In this paper, we design a study to measure how many voters are willing to report errors relative to how many check their receipts. We also examine how verification rates change according to how expected or unexpected the outcome of the election is. We find verification is higher when the result is unexpected (as might be expected in a manipulated tally), but even when voters are provided an unexpected result and wrong receipt, only 0.5% of voters actually report the discrepancy (versus 7.5% who simply check). To illustrate the consequences of this drop-off, we apply a recent US Senate race with a slim margin of victory to a fictional scenario where enough receipts were manipulated to change the election outcome. We then show that the probability of the electorate detecting such an attack drops from 99.99% to 43.25% when we factor in the low rate at which voters actually report wrong receipts.

2. RELATED WORK

2.1 End-to-end verifiable (E2E) voting systems

The literature on end-to-end verifiable (E2E) voting is vast. Staring in 1996 with a variant of FOO at Princeton, a number of E2E systems have been developed and deployed in student elections including Evox, Punchscan, Bingo Voting, Helios, and Wombat [Clark, 2011]. In 2009, Scantegrity II was used in a governmental election in Takoma Park, USA, and Prêt à Voter will be deployed in a federal election in Victoria, Australia in 2014. Although these systems differ in many details, they share a common approach to voter verifiability. Voters mark their selections, and the receipt consists of a privacy-preserving obfuscation of the selections along with a unique identifier. Different vote obfuscation methods have been proposed: an encryption of the selection, the position of the selection in a shuffled list, or short alphanumeric codes (confirmation codes) for each selection. In our study, we utilize confirmation codes.

After the election, the voter may use their receipt identifier to lookup her receipt and confirm it was recorded correctly. A second verification step (which is not a focus of this work) is performed to verify that all receipts are correctly processed to produce a final tally. Any manipulation of the association between a receipt and its corresponding vote has an overwhelming probability of detection. However the probability of detecting manipulations of receipts themselves depends on how many voters check and, importantly, how many of these actually report the discrepancy.

2.2 Reporting

The literature has focused to a great extent on the proportion of individuals who check their vote as a proxy for the number of voters who will report potential voter fraud. Two theories, however, suggest that even when a voter checks their vote and observes that an error has occurred, they may be unlikely to report a problem.

The first issue relates to the psychological phenomenon of "diffusion of responsibility", or the bystander effect [Darley and Latane 1968; Latane and Darley 1970]. For example, when a crime is committed and many people witness it, each individual witness is less likely to call the police and report the crime that when few people witness the crime. That is, individuals do not take steps to act responsibly (by reporting the crime), because they assume someone else in the crowd will do it [Darley and Latane 1968]. We hypothesize similar phenomena may arise in E2E elections, *i.e.,* individuals who are faced with a potential voting error may assume that if others have also experienced the same problem, others will have reported the problem. The result is a potentially large proportion of erroneous votes going unreported.

A second issue relates to lack of expertise in understanding the purpose and process of E2E election verification. Voters who are unfamiliar with E2E systems may attribute an inconsistent receipt with their own error, as opposed to the error of the system. As a result, individuals may be less likely to report a problem than has been previously anticipated.

2.3 Voter Verification

The Punchscan system was deployed at U. Ottawa [Essex et al 2007] with a reported 54% of receipts queried, and no reported disputes. Helios was deployed at UC Louvain [Adida et al. 2009] with a reported 30% of receipts queried, with 7 voters filing a dispute. Scantegrity II was used in Takoma Park, MD [Carback et al. 2010] with a reported 4% of receipts queried, with 1 voter filing a dispute. In a comparative usability study of Helios, Scantegrity II and Prêt à Voter, a reported 43% and 38% of voters attempted receipt verification for Helios and Scantegrity II respectively [Acemyan et al. 2014].

In a questionnaire study, five prominent mental models of verifiability were identified [Olembo et al. 2013]: the first group references a belief that persons and/or processes are trustworthy, the second that verifiability is not possible or are of unsure it, the third that the presence of external observers can ensure integrity, the fourth that personal involvement (or the potential for it) can ensure integrity, and the final group that references auditing techniques. We do not study which groups our participants fall in, however our focus on the voters who do verify their vote likely correlate to the final group.

In later work, a study examines if textual prompts can increase a voter's intent to verify their vote [Olembo et al. 2014]. It is reported that prompts do increase intent with no measurable difference between the types of message (*e.g.,* communicating a risk vs. a social norms). Independently in this work, we chose to use a risk-based prompt to increase verification. We also measure actual verification rates as opposed to stated intentions to verify.

2.4 DRE Review Screen Verification

A user study of DRE voting machines reports that 37% of voters noticed when the review screen summarizing the voter's selections prior to casting the ballot contained manipulations (vote flips) [Everett 2007]. Similarly in our study, receipt information was manipulated (wrong code) when shown on a review page. Our study differs in a few regards that may impact the results: (i) review screens are mandatory, while receipt checking is opt-in, (ii) review screens present meaningful text (*i.e.,* candidate names), while receipt review screens have arbitrary strings (*i.e.,* confirmation codes), and (iii) their study measures if manipulations were noticed (based on self-reporting after users are told the review screen was manipulated) while we measure if a voter then files a dispute.

3. RATIONALE AND HYPOTHESES

Assuming that not all individuals who participate in an election will check their votes, checking that a vote has been cast properly may be more likely to occur when there is some level of doubt or uncertainty with the system [Andaleeb 1996; Tolin et al. 2003; van den Hout and Kindt 2003]. That is, individuals might be more likely to check whether their vote was cast properly when some level of uncertainty or discomfort with election results occurs. For example, doubt may be cast if an unexpected or undesired individual is elected. It has been observed previously that social trust is undermined when suspicion is elicited[Lee and Schwarz 2012]. In these cases, we hypothesize individuals who did not vote for the winning party may be more likely to check their receipt.

> H1: Individuals are more likely to check their receipt when the election outcome is unexpected.

A second question addresses how willing voters are to report a discrepancy in the receipt check. Individuals might be more inclined to assume that someone else in the same position would report a problem (the bystander effect), and/or that they, the voter, are at fault, rather than investigate or report that their vote was incorrectly tabulated (a knowledge gap).

> H2: Individuals may be more likely to report a problem to authorities when presented with conflicting/error messages when checking their code.

In summary, we hypothesize that the *relative* vote checking rates and reporting rates will be higher when individuals are presented with unexpected information or errors in tabulation. We are not hypothesizing a specific absolute proportion of voters that may check receipts, only that proportion of voters who do check receipts and report a discrepancy will vary by randomly assigned condition.

4. STUDY 1: EXPECTED VERSUS UNEXPECTED ELECTIONOUTCOMES

The purpose of Study 1 was to test Hypothesis 1. Here, we examine whether individual voters would be more likely to check their receipts when the outcome of the vote was unexpected. We elicited a vote from participants, and then issued them a receipt in form of a ballot ID and confirmation code. Once the outcome of the vote was shared with voters, we examined whether they checked that their receipt was correctly reflected, as well as how perseverant they were in resolving (reporting) a discrepancy. The study was split across two conditions—one where the outcome of the vote was expected, and one where the outcome was unexpected (based on proportion of actual votes obtained).

4.1 Participants

From several case studies on deployments of E2E systems in the literature, there is little consensus on what fraction of voters can be expected to check a receipt—this number has been reported as high as 50% and as low as 4%. We assumed that checking rates would fall somewhere in that range; as such, we anticipated that a sample size of 800 would suffice.

Participants were recruited from Crowdflower (http://crowdflower.com/). Crowdflower is an online crowdsourcing tool, used for recruiting samples for large-scale online surveys. A researcher is able to upload a questionnaire, which is then made available to over 50 labour partner sites. Here, participants are able to access the questionnaire, where they can complete it for a small financial incentive (<$1). The incentive is paid upon completion of the questionnaire; Crowdflower is compensated on a per-participant basis, taking an overhead of payment (33%). Previous work examining similar participant tools, such as Amazon Mechanical Turk, has suggested that the participants involved represent a wide spectrum of individuals with regard to demographics and culture [Paolacci, Chandler, and Ipeirotis 2010]. Other work has demonstrated that these online subject pools provide an inexpensive and reliable source of data [Horton, Rand, and Zeckhauser 2011]. Given that we aim to extrapolate our findings to the general voting public, we chose to use this online pool of participants as our main subject pool.

4.2 Method

In the domain of social psychology it is well understood that simply asking participants to report on their behavior in a novel setting does not always lead to them giving accurate predictions [see Orne, 1962; Seeman, 1969; Weber & Cook, 1972; Weinstein, 1980]. In order to examine actual voting behavior, therefore, we asked participants to take part in a low-stakes voting task. Participants in our study were told that the researchers were interested in charitable donation behavior, and that they would be asked about their own behaviors with regard to charitable giving.

We chose donation behavior as a voting domain because we reasoned that in general, charitable giving is positively valenced. That is, in general, most people feel positively toward charitable giving. Upon reading and signing the consent form (clicking "I Consent"), participants were asked to complete a short questionnaire about their charitable giving behaviors (see Appendix A). Following the questionnaire, participants were given the opportunity to vote for one of five charities and were told that the charity receiving the most votes would be presented with a $50 donation from the experimenters. This voting task represented the main variable of interest in this study. Since a charity vote is relatively low-stakes voting activity, we hypothesize that user trust will be high, and individuals are unlikely to suspect a nefarious outcome. As such, participants may at baseline be unlikely to check their receipt in the same way that voter apathy might impact a political election. In the instructions we decided to reference the potential for "browser compatibility issues" in order to evoke some degree of skepticism in the voting system, and, in turn, to encourage greater incidence of receipt checking. This script was provided for participants in both conditions. The voting task was described in the following text:

> "In addition to receiving payment for participating in our study, we will be donating $50 to a charitable organization. On the next page, you may select one charity from a list; the charity with the most votes will receive the donation. A donation vote is not required; you may simply select "No Thanks" to complete the study.
> After you vote, we will send you an email that will include a confirmation code for your vote, as well the results of the vote. At this point, you can use the confirmation code to check that your vote was correctly included in the result. This is optional, but we recommend it due to compatibility issues that have occurred with certain browsers."

Participants were then taken to the voting page, where they were given the choice to vote for one of the following charities (see Fig. 1).

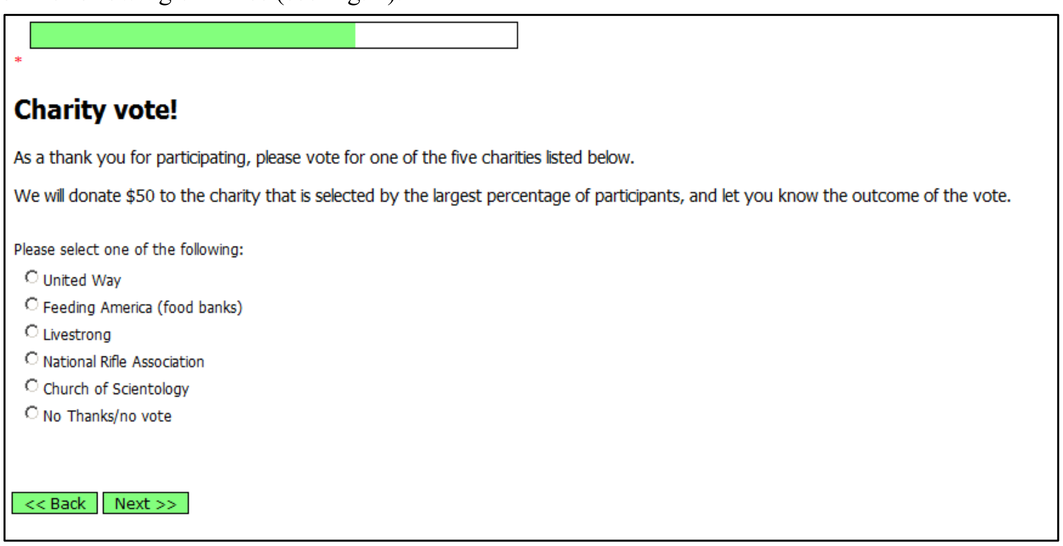
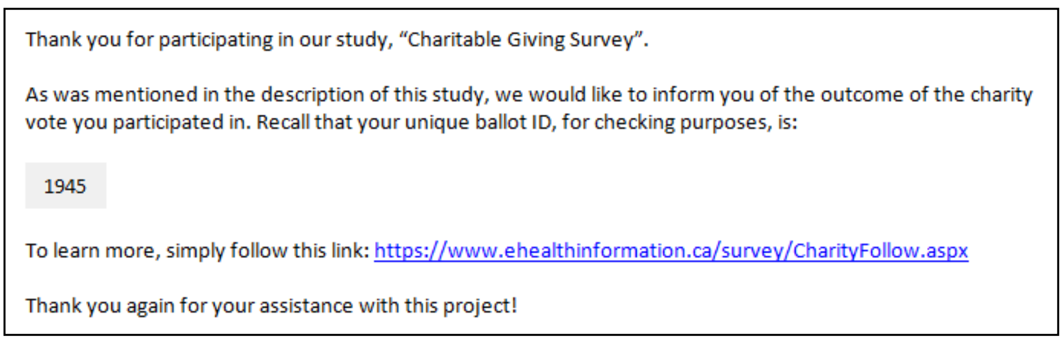


Figure 1. Screenshot of charity vote offered to participants.

We purposely selected some charities that were relatively broad in scope (*i.e.,* United Way, Feeding America/food banks), as well as some that had more narrow interests (*i.e.,* National Rifle Association; Church of Scientology), as well as one associated with recent scandal (*i.e.,* Livestrong's association with Lance Armstrong). Importantly, we did not select "expected" and "unexpected" charity outcomes subjectively; we based these decisions on objective voting behavior of our participants.

After the election closed and all votes were tallied, participants were emailed and thanked for participating. This email included a reminder of each participant's ballot ID and confirmation code, as well as a link to the outcome of the vote (see Fig. 2). For half of the participants (expected outcome condition), the link produced a web page (see Fig. 3) displaying the accurate winner of the vote (i.e., the organization that received the most votes was displayed as the winner). Thus, for this group of participants, the outcome of the vote was highly believable. For the other half of participants (unexpected outcome condition), the web page displayed the winner as the organization that received the least number of real votes. For this group of participants, the outcome of the vote was hypothesized to be less believable, which may have acted as a cue to participants that something suspicious had occurred with the vote. Following this, participants saw the following message:

> "At this point, you may check that your ballot was included in the final result by using the confirmation code we sent previously to you by email, using this link [link to receipt check]." (see Fig. 3).



Thank you for participating in our study, "Charitable Giving Survey".

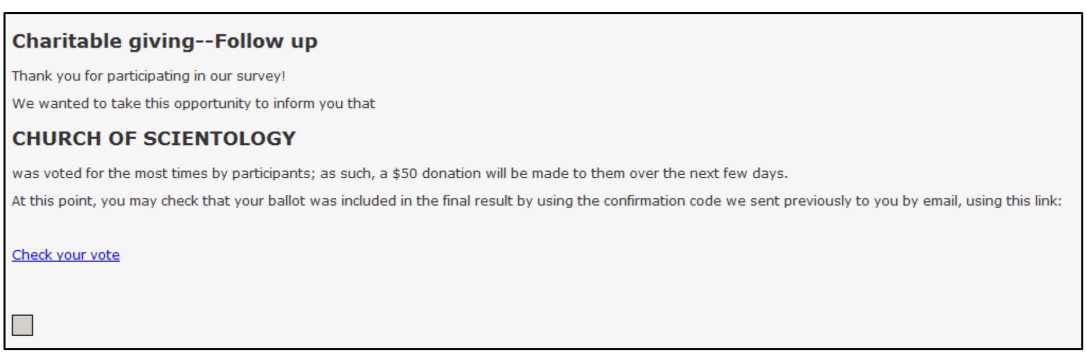
As was mentioned in the description of this study, we would like to inform you of the outcome of the charity vote you participated in. Recall that your unique ballot ID, for checking purposes, is:

1945

To learn more, simply follow this link: https://www.ehealthinformation.ca/survey/CharityFollow.aspx

Thank you again for your assistance with this project!

Figure 2. Screenshot of email sent to participants, informing them of their ballot ID and a link to learn the outcome of the vote.



Charitable giving--Follow up

Thank you for participating in our survey!

We wanted to take this opportunity to inform you that

CHURCH OF SCIENTOLOGY

was voted for the most times by participants; as such, a $50 donation will be made to them over the next few days.

At this point, you may check that your ballot was included in the final result by using the confirmation code we sent previously to you by email, using this link:

Check your vote

Figure 3. Screenshot of the follow-up survey's first page, alerting the participant to the outcome of the vote (in this case, the unexpected outcome).

After clicking the link, participants were prompted to enter their Ballot ID (see Fig. 4). After clicking to "check confirmation code", a confirmation code was then displayed on the following page—this code was either correct or incorrect (i.e., matched or did not match the code that was sent via email). Below the code display, participants saw a link to report an error (see Figs. 5, 6). We hypothesized that individuals would be more likely to check their vote in the unexpected outcome condition (versus the expected outcome condition). Further, we hypothesized that when an incorrect code (versus the correct code) is displayed when checking, participants would be more likely to report an error to authorities.

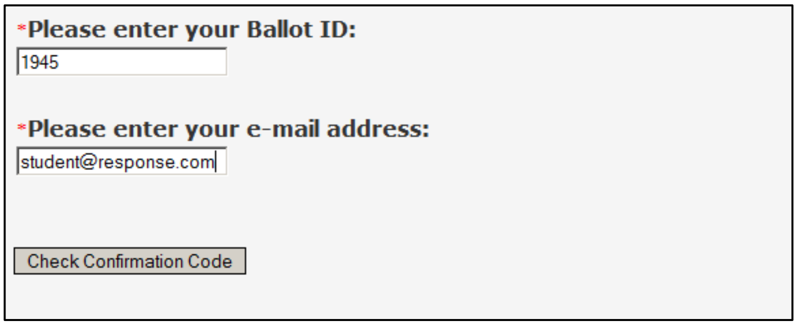


Figure 4. Second page of the follow-up survey, where participants were asked to enter their ballot ID and email address. This form was not pre-filled, but we have included a template for the reader.

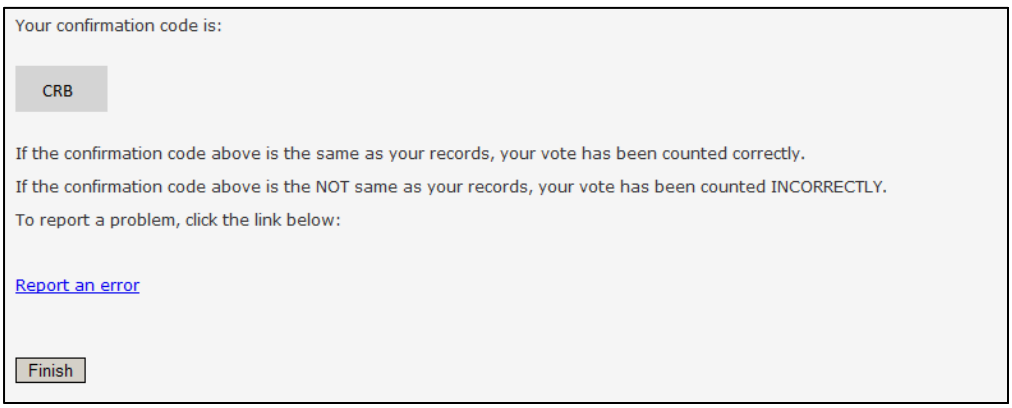


Figure 5. Third page of the follow-up survey, where participants were presented with either a correct or incorrect confirmation code, and given the option to report a problem.

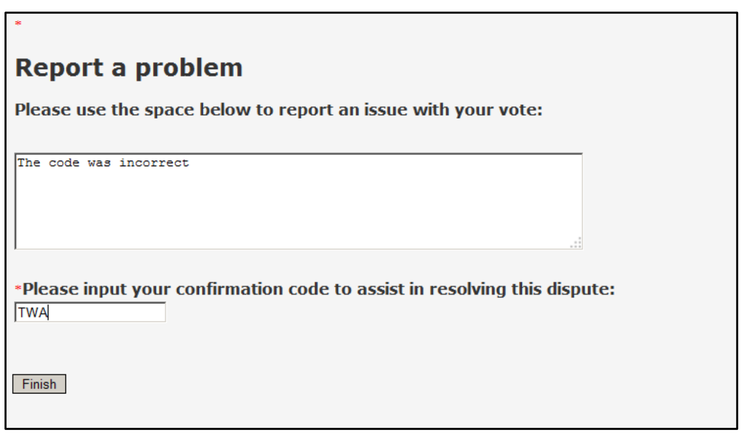


Figure 6. The fifth page of the follow-up survey, where participants were given the option to report a problem with their vote. This form was not pre-filled, but we have included a template for the reader.

Finally, all participants were emailed a copy of the feedback form.

4.3 Results

Participants (N = 841) were recruited from Crowdflower. We collected the following demographic information:

Age. We asked respondents to categorize themselves into one of four age ranges. The majority of respondents fell in the 18-30 year range (48%; non-response: 1.4%). Of the remaining respondents, 32.9% were in the 31-45 age range, 15.3% were in the 46-60 age range, and 1.8% were 60 or older.

Income. We asked respondents to categorize themselves into one of six income brackets The majority of respondents reported an annual income of less than $20,000 (24.7%), or of $20,000 to $39,999 (24.6%; non-response: 7.3%). Of the remaining respondents, 18.1% reported earning $40,000 to $59,000, 11.3% reported earning $60,000 to $79,999, 6.7% reported earning from $80,000 to $99,999, and 7.4% reported earning $100,000 or more. .

Donation behavior. The majority of respondents reported donating once or twice a year (37.6%; non-response: 3.9%). Of the remaining respondents, 5.7% reported never donating, 15.2% reported donating less than once a year, 18.4% reported donating every 2 to 3 months, and 19.3% reported donating at least once a month. Participants tended to donate between $1 and $50 (47.1%; non-response: 4.0%). The majority of participants reported that donating did improve their moods (82.4%; non-response: 4.1%). Of the remaining respondents, 6.1% reported donating nothing, 16.0% reported donating $51 to $100, 10.3% reported donating $101 to $200, 8.2% reported donating $201 to $500, 4.2% reported donating $501 to $1,000, and 4.2% reported donating $1.001 or more.

Lottery winnings. We asked participants whether they would donate lottery winnings to charity, in the event that they won. Most reported that they would donate less than half of their winnings to charity (74.2%; non-response: 7.2%). Of the remaining respondents, 5.7% reported that they would not donate any winnings to charity, 11.7% reported that they would donate more than half of their winnings to charity, and 1.2% reported that they would donate all their winnings to charity. . Participants also reported they were likely to donate this money to multiple charities versus a single charity (76.4%; non-response: 5.6%). Of the remaining respondents, 5.2% said they would not donate, and 12.8% said they would donate to only one charity.

Charity vote. Finally, we asked participants to vote for one of 5 charities for which we would donate $50. The breakdown of votes was as follows:

- United Way: 20.5%
- Feeding America/Food banks: 58.0%
- Livestrong: 6.6%
- National Rifle Association: 6.1%
- Church of Scientology: 1.3%
- No vote: 7.6%

The modal response was for Feeding America/Food banks; a chi-square test confirmed that it was chosen more frequently than all other options, χ^2 (5, 791) = 1072.21, p <.001. As such, we selected

Feeding America/Food banks as our "expected" charity outcome. Similarly, because the Church of Scientology received the fewest number of votes, we selected it as our "unexpected" charity outcome.

Follow-up participation. The last question of the charitable behaviors survey asked participants to leave their email address if they were interested in learning of the results of the vote. Of the 841 total respondents, 603 (71.7%) left their e-mail addresses.

4.4 Vote outcome follow-up

All 603 participants who left their email addresses were contacted with information on the outcome of the charity vote. Participants were emailed a reminder with their unique ballot ID and confirmation code, as well as a link to see the results of the study. The link randomly sent participants to one of two web pages, one of which informed participants of the expected vote outcome (Feeding America/Food banks), and one of which informed participants of the unexpected vote outcome (Church of Scientology). From these pages, participants could then click to check that their vote had been tallied. We measured the proportion of participants who moved on to check whether their vote has been correctly tallied.

Of the 603 participants we emailed, 84 (13.9%) clicked on the link in the email. Of this 84, half of participants received the expected outcome (N = 42; Feeding America/Food banks), and half received the unexpected outcome (N = 42; Church of Scientology). We next examined what proportion of these respondents saw the outcome of the vote and continued to check whether their ballot ID and confirmation codes matched. For each participant, we coded which page they had progressed to[1], and used this value as a continuous variable. A one-way Analysis of Variance revealed that participants in the expected outcome condition (M = 2.52 pages/22 of 42 progressing to ballot ID check) did not progress as far as did those in the unexpected outcome condition (M = 2.76 pages/32 of 42 progressing to ballot ID check), MSE = 1.19, $F(1, 83) = 5.39$, $p = .02$. Thus, seeing an unexpected (versus expected) outcome of a vote encouraged more participants to check their votes. These results support our first hypothesis, and suggest that unexpected vote outcomes are likely to produce more receipt-checking behavior, relative to expected vote outcomes. Only three participants (3.6%) clicked the through to see the "report" web page, and only one individual actually filed a report (despite the fact that half of participants who checked their ballot ID and confirmation codes received an incorrect response; $N \approx 26$). This suggests that even when participants are given overt evidence of a discrepancy, few will report a problem. Since we had designed Study 1 only to measure H1, we did not track which condition these participants came from. In response, we formulated H2 and designed Study 2 to specifically test whether individuals presented with unexpected outcomes and incorrect receipts check their receipt in sufficient numbers to detect fraud in an E2E election with high probability.

4.5 Study 1 Discussion

We observed that overall, rates of vote checking were low, though not significantly lower than in previous work[Carback et al. 2010]. Importantly, however, we observed that checking rates were significantly higher when the outcome of the vote was unexpected by voters (H1). Further, though approximately half of participants who checked their ballot IDs and confirmation codes received an incorrect response, only one participant actually filed a report regarding the issue. This suggests

[1] Page 1: Thank you/welcome page; Page 2: Expected vs. unexpected outcome; Page 3: Correct vs. incorrect confirmation code; Page 4: Reporting problem to authorities.

that even when presented with an incorrect code, reporting an error was an unlikely step taken by voters (H2). In Study 2, we specifically test the case where reporting should be highest—that is, when outcomes are unexpected and when the displayed confirmation codes are incorrect.

5. STUDY 2: TARGETED UNEXPECTED OUTCOMES AND INCORRECT CONFIRMATION CODES

In Study 1 we observed that checking rates in E2E voting were greater when the outcome of the vote was unexpected. In the second study we restrict vote outcomes to be unexpected only, in order to maximize the number of samples of voters checking their receipts. Further, we also restrict confirmation codes to be incorrect only. Together, these conditions should encourage vote checking and reporting behaviors.

5.1 Method

As in Study 1, participants were told that the researchers were interested in charitable donation behavior. Upon reading and signing the consent form (clicking "I Consent"), participants were asked to complete a short questionnaire about their charitable giving behaviors. Also, unique to Study 2, participants reported on their online experiences, such as frequency shopping or banking online. Finally, participants were given the opportunity to vote for one of 5 charities for which the experimenters would give a $50 donation. Again, we use the potential for "browser compatibility issues" to evoke some degree of skepticism in the voting system. Participants were then taken to the voting page, where they were given the choice to vote for one of 5 charities (see Fig. 1). After the election closed, and all votes were tallied, participants were emailed and thanked for participating. Included in this email was a reminder of each participant's Ballot ID, as well as a link to learn of the outcome of the vote (see Fig. 2).

All participants were told the winner was the organization that received the least number of actual votes in Study 1 (*i.e.,* Church of Scientology; see Fig. 3). Participants were then given the opportunity to check their receipts as in Study 1 (see Fig. 4). Participants who checked their receipts were told that their confirmation codes were incorrect (See Fig. 5), and were given the opportunity to report the problem (see Fig. 5). Finally, all participants were emailed a copy of the feedback form.

5.2 Results

Participants (N = 755) were recruited from Crowdflower and completed the initial charitable giving survey. We collected the following demographic information:

Time on task. Average time spent on task was 4.12 minutes.

Age. We asked respondents to categorize themselves into one of four age ranges. The majority of respondents fell in the 18-30 year range (50,3%; non-response: 1.9%). Of the remaining respondents, 33.5% were in the 31-45 age range, 13.1% were in the 46-60 age range, and 1.2% were 60 or older.

Income. We asked respondents to categorize themselves into one of six income ranges. The majority of respondents fell across three income brackets, less than $20,000 (20.1%), $20,000 to $39,999 (22.9%), and $40,000 to $59,999 (21.9%; non-response: 9.3%). Of the remaining respondents, 13.9% reported earning $60,000 to $79,999, 6.9% reported earning $80,000 to $99,999, and 5.0% reported earning $100,000 or more.

Donation behavior. The majority of respondents reported donating to charity once or twice a year (34.0%; non-response: 3.7%). Of the remaining respondents, 18.8% reported donating at least once a month, 22.3% reported donating once every 2 to 3 months, 13.9% reported donating less than once a year, and 7.2% reported never donating. Participants reported donating between $1 and $50 (48.5%; non-response: 4.6%). Of the remaining respondents, 7.6% reported that they did not donate, 16.6% reported donating $51 to $100, 10.6% reported donating $101 to $200, 6.6% reported donating $201 to $500, 2.8% reported donating $501 to $1,000, and 2.7% reported donating $1,001 or more. The majority of participants reported that donating did improve their moods (79.5%; non-response: 5.6%).

Lottery winnings. We asked participants whether they would donate lottery winnings to charity, in the event that they won. Most reported that they would donate less than half of their winnings to charity (70.3%; non-response: 7.7%). Of the remaining respondents, 6,5% reported that they would not donate any winnings, 14.1% reported that they would donate more than half of their winnings, and 1.3% reported that they would donate all their winnings. Participants also reported they were likely to donate this money to multiple charities versus a single charity (72.6%; non-response: 5.0%). Of the remaining respondents, 6.5% reported that they would not donate, and 15.8% reported that they would donate to only one charity.

Charity vote. Finally, we asked participants to vote for one of 5 charities for which we would donate $50. The breakdown of votes was as follows:

- United Way: 22.3%
- Feeding America/Food banks: 60.0%
- Livestrong: 6.5%
- National Rifle Association: 4.4%
- Church of Scientology: 1.1%
- No vote: 5.4%

As in Study 1, Feeding America/Food banks was chosen most frequently, and Church of Scientology was chosen least frequently. Again we selected Church of Scientology as our "unexpected" charity outcome.

Follow-up participation. The last question of the charitable behaviors survey asked participants to leave their email address if they were interested in learning the results of the vote. Of the 755 total respondents, 508 (67.3%) left their e-mail addresses.

5.3 Vote outcome follow-up

All 508 participants who left their email addresses were contacted with information on the outcome of the charity vote. Participants were emailed a reminder with their unique ballot ID and confirmation code, as well as a link to see the results of the study. The link sent participants to a web page informing participants of the (unexpected) vote outcome (*i.e.,*Church of Scientology). From this screen participants could click to check their receipt, and we measured the proportion of participants who did so.

Of the 508 participants we emailed, 484 emails were delivered (undelivered emails were due to invalid addresses). Of this sample of 484, 77 clicked on the link in the email. Of this 77, 60 input their ballot ID (57 input a correct ID). Of the 57 participants, only 4 progressed to the Report page. Of the 4 who progressed to the Report page, all 4 filed a report using their confirmation code, and all 4 correctly noted in a comment box that the confirmation code they had previously been given did not match the one provided at check.

We hypothesized that by only providing an unexpected vote outcome we would increase the proportion of individuals who would check their votes. This difference was significant, $t(54) = 3.04$, $p = .004$, suggesting that by targeting participants with an incorrect conformation code, we did indeed increase the percentage of participants who would report a problem (1.8% in Study 1 to 7.3% in Study 2). This report percentage, however, is still quite small—only 0.4% of the voting population checked and reported a problem with their vote.

Lastly, we examined whether propensity to check confirmation codes and report problems was correlated with online experiences (questionnaire in Appendix). Frequency of online shopping and online banking did not correlate with either confirmation code checking ($p > .50$) or reporting a problem ($p > .60$). However, there was a significant correlation observed between number of online accounts they reported as having and propensity to report a problem, $r(754) = .09$, $p = .01$, such that individuals who reported having more online accounts were more likely to report a problem with their vote confirmation codes. This suggests that individuals who interact with more online services could perhaps be more confident in detecting errors, leading to their increased reporting rates.

5.4 Study 2 Discussion

With Study 2, we aimed to encourage vote checking and reporting behaviors. Specifically, we primed participants to be suspicious of a potential error in vote tabulation by informing them that certain browsers had been known to cause trouble for the voting system; we reported the outcome of the charity vote to be the unexpected outcome; and we showed all participants incorrect confirmation codes when they checked their receipt. As such, rates of receipt checking and reporting of errors are likely to be inflated relative to the general population. Compared to Study 1, we observed that more participants filed a report when a discrepancy was present and made explicit (supporting H2). Yet again, however, we observed that rates of receipt checking were low—indeed, likely too low to detect an attack with a high degree of confidence. We observed that 57 of 755 (7.5%) voters checked their receipt and only 4 of 755 (0.5%) reported an incorrect receipt. We posit two possible explanations borrowed from the psychology literature (bystander effect and knowledge gap), which may explain why these receipt checking and reporting rates are lower than one might expect, given previous findings. These results suggest that even extreme cases of potential vote fraud could go undetected by the electorate.

6. GENERAL DISCUSSION

Overall we observed that rates of receipt checking were low, though not significantly lower than in previous work [Carback et al. 2010]. Critically, we observed that checking rates were significantly higher when the outcome of the vote was unanticipated or unexpected by voters in Study 1 (supporting H1). Further, more participants in Study 2 reported an error than in Study 1, suggesting that when incorrect confirmation codes are provided, individuals are more likely to raise an alarm (supporting H2).

Despite this, though approximately half of participants in Study 1, and all participants in Study 2 who checked their ballot IDs and confirmation codes received an incorrect response, only 7 total participants clicked to "report a problem", and only 5 reported an actual problem, suggesting that even when presented with an incorrect code, reporting an error was an unlikely step taken by voters.

The magnitude of the observed difference between voters who check receipts and those who report errors has significant consequences for E2E elections. Consider a recent close election:

the 2008 United States Senate election in Minnesota, where Sen. Al Franken was certified as the winner by a margin of 225 votes out of 2,887,337. In an election where 7.5% randomly sampled voters confirm their receipts are correct (our observed rate of receipt *checking*), a manipulation capable of altering the result would be detected with 99.99% probability.[2] However, if we reduce the receipt confirmation rate to 0.5% (our observed rate of *error reporting*), the probability of detecting the manipulation is reduced to 43.25%. In other words, it would be more likely to escape detection than not.

Two potential explanations for these inconsistencies in error information and reporting can be drawn from the psychology literature. First, it is possible that voters in this study, though from an online sample and relatively savvy (compared to a population drawn offline), may have had a knowledge gap. Voters may have been confused by the error message, and may have attributed the error to themselves (i.e., thinking they incorrectly remembered or wrote down codes, misunderstanding meaning of error messages, etc.). Alternatively, voters may have experienced a diffusion of responsibility, or the bystander effect [Darley and Latane 1968], such that when faced with a vote tabulation error, they assumed that someone else in the same predicament would report the problem, and so felt as though their reporting was not required. Both of these explanations warrant exploration in future research, as they predict different ways to attenuate the problem: a knowledge gap can be reduced by providing more clear and succinct information; a bystander effect can be reduced by reminding the voter how important their individual vote is, for example.

6.1 Limitations to Study

There are two main limitations to the present studies. First, the task of online voting for a charity donation was lower-stakes relative to the task of voting for a representative in government. As such, rates of reporting errors in the present studies may be lower than in an actual election, and it is possible that receipt checking rates and error reporting would *increase* as the stakes increase.

The second limitation of this work is that participants in this sample were recruited from an online pool and may, therefore, be more technologically adept than the general voting population. As a result, these individuals may be feel more confident in (or less confused by) the receipt checking process, which may have encouraged more checking and error reporting than a typical population. As such, the prevalence of receipt-checking in this work may be *inflated* relative to the general population. Additionally, we made receipt checking relatively easy (by prompting users through email), and error reporting was as simple as filling out a webform (while real elections may require documentation, signed forms, or in-person reporting). These factors may also cause our rates to be inflated.

6.2 Replication Studies

For our observations to be considered reliable when generalized to different elections and voting populations, our hypotheses should be retested in replication studies. We suggest a few extensions and modifications future studies might consider to address the limitations of our own study. To determine to what extent, if any, our results are due to voter apathy concerning the relatively low stakes of our election, follow-up work could implement a similar study with much larger rewards or utilize a poll concerning a real political issue. A follow-up study may also be implemented in a

[2] Assuming an adversary fraudulently changes $F = \lceil 225/2 \rceil$ receipts from Franken to his closest competitor, the probability of detecting fraud with $B = 2887337$ ballots and $\mathrm{E}[R]$ expected reported receipt confirmations is:

$$\mathbf{Pr}_{Detection}[R, B, F] = 1 - \binom{B-F}{\mathrm{E}[R]} \Big/ \binom{B}{\mathrm{E}[R]}$$

real election, however the use of deception makes this course of action questionable both ethically and legally. Future studies should strive to test different voting populations, including ones more representative of the voting-age population.

Future studies may also reexamine all aspects of the study design, and in particular the wording used to inform voters of the option to check their receipt and report an error. We know from other domains in usable security that wording is important (*e.g.,* browser warnings about HTTPS connections [Akhawe and Porter Felt 2013])—different wordings could be tested to see what, if any, measurable impact it has on user behavior. This has been studied to some extent with an emphasis on the communicated message [Olembo et al. 2014].

Finally future studies might examine which voters verify their receipts and report discrepancies, perhaps with a follow-up questionnaire to understand their motivations. The study could also track how these users voted to see if certain voters are more likely to check/report (*e.g.,* supporters of the losing candidates).

6.3 Future Directions for E2E Designers

Future work should examine how voting procedures might be altered to improve usability. In this study, individuals (especially those who are less computer-savvy) may have had difficulty understanding what vote confirmation error messages meant, as well as how to appropriately handle them (i.e., reporting the problem). As such, development of more straightforward communication with regard to receipt checking could increase checking rates, especially among those individuals who would otherwise not check. For example, creating a simple infographic about how to use the system could increase overall vote checking rates. Alternatively, priming individual voters with security concerns could also increase check rates (although likely at the cost of trust). Finally, inconsistent findings in code checking could automatically be reported without requiring the voter to click through to an additional page, or reporting buttons could be made larger and more obvious to voters.

7. CONCLUSIONS

The present work suggests that much more work is required to maximize effectiveness of E2E voting systems, particularly with reference to how people approach vote checking and error reporting. E2E systems may struggle to reliably uncover fraud if the error reporting rates from our studies generalize to real world elections. However, we believe simple design changes could improve usability of E2E voting, thereby improving fraud detection, and thus validity, of the system overall.

8. ACKNOWLEDGEMENTS

We thank the Editorial Board of JETS for useful and actionable feedback that improved the paper. The second author acknowledges NSERC Discovery for partial funding of this work.

9. REFERENCES

Acemyan, C., P.Kortum, M.D. Byrne, and D.S. Wallach. 2014. “Usability of Voter Verifiable, End-to-end Voting Systems: Baseline Data for Helios, Prêt à Voter, and Scantegrity II.” *JETS* 2(3).

Adida, B., O. de Marneffe, O. Pereira, and J. J. Quisquater. 2009. “Electing a University President Using Open-Audit Voting: Analysis of Real-World Use of Helios.” EVT/WOTE.

Andaleeb, S. S. 1996. “An Experimental Investigation of Satisfaction and Commitment in Marketing Channels: The Role of Trust and Dependence.” *Journal of Retailing* 72 (1): 77–93.

Akhawe, D. and A. Porter Felt. 2013. “Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness.” USENIX Security.

Ben-Nun, J., N. Farhi, M. Llewellyn, B. Riva, A. Rosen, A. Ta-Shma, and D. Wikström. 2012. “Wombat: An Implementation and Field Testing of Dual Paper-Cryptographic Election System.” EVOTE.

Carback, R., D. Chaum, J. Clark, J. Conway, A. Essex, P. S. Herrnson, T. Mayberry, et al. 2010. “Scantegrity II Municipal Election at Takoma Park: The First E2E Binding Governmental Election with Ballot Privacy.” USENIX Security.

Clark, J. 2011 “Democracy Enhancing Technologies: Toward deployable and incoercible E2E elections.” PhD Thesis, Waterloo.

Darley, J. M. and B. Latane. 1968. “Bystander Intervention in Emergencies: Diffusion of Responsibility”. *Journal of Personality and Social Psychology* 8 (4, Pt. 1): 377-383.

Essex, A., J. Clark, R. Carback, and S. Popoveniuc. 2007. “Punchscan in Practice: an E2E Election Case Study.” WOTE.

Everett, S. P. 2007. “The Usability of Electronic Voting Machines and How Votes Can Be Changed Without Detection” Dissertation, Rice University.

Horton, J. J., D. G. Rand, and R. J. Zeckhauser. 2011. “The Online Laboratory: Conducting Experiments in a Real Labor Market.” *Experimental Economics* 14 (3): 399–425.

Karayumak, F., M. M. Olembo, M. Kauer, and M. Volkamer. 2011. “Usability Analysis of Helios - An Open Source Verifiable Remote Electronic Voting System.” EVT/WOTE.

Latane, B. and J. M. Darley. 1970. “The Unresponsive Bystander: Why Doesn’t He Help?”.New York, 131 pp.

Lee, S. W. S., and N. Schwarz. 2012. “Bidirectionality, Mediation, and Moderation of Metaphorical Effects: The Embodiment of Social Suspicion and Fishy Smells.” *Journal of Personality and Social Psychology* 103 (5): 737–49.

Olembo M. M., S. Bartsch and M. Volkamer. 2013. “Mental Models of Verifiability in Voting.” VoteID.

Olembo M. M., K. Renaud, S. Bartsch and M. Volkamer. 2014. “Voter, What Message Will Motivate You to Verify Your Vote?” USEC, 2014.

Orne, M. T. 1962. “On The Social Psychology Of The Psychological Experiment: With Particular Reference To Demand Characteristics And Their Implications.” *American Psychologist* 17 (11): 776–83.

Paolacci, G., J. Chandler, and P. G. Ipeirotis. 2010. “Running Experiments on Amazon Mechanical Turk.” *Judgment and Decision Making* 5 (5): 411–19.

Seeman, J. 1969. “Deception In Psychological Research.” *American Psychologist* 24 (11): 1025–28.

Tolin, D. F., J. S. Abramowitz, B. D. Brigidi, and E. B. Foa. 2003. “Intolerance of Uncertainty in Obsessive-Compulsive Disorder.” *Journal of Anxiety Disorders* 17: 233–42.

Van den Hout, M., and M. Kindt. 2003. “Repeated Checking Causes Memory Distrust.” *Behaviour Research and Therapy* 41: 301–16.

Weber, S. J., and T. D. Cook. 1972. “Subject Effects In Laboratory Research: An Examination Of Subject Roles, Demand Characteristics, And Valid Inference.” *Psychological Bulletin* 77 (4): 273–95.

Weinstein, N. D. 1989. “Optimistic Biases about Personal Risks.” *Science* 246 (4935): 1232–34

APPENDIX

A. Charity questionnaire

General demographics

1. Please select an age-range that describes you:
    18-30
    31-45
    46-60
    61-75
    76+
    Prefer not to say

2. In what income bracket do you fall?
    <$20,000
    $20,000 to $39,999
    $40,000 to $59,999
    $60,000 to $79,999
    $80,000 to $99,999
    $100,000+
    Prefer not to say

The following questions ask you about your charitable giving behavior. Please choose the response that best fits your behavior.

1. How frequently do you make a charitable donation (to a registered charity)?
    At least once a month
    Once every 2 or 3 months
    Once or twice a year
    Less than once a year
    Never
    Prefer not to say

2. If you do make a charitable donation, how much do you give on average per year?
    $0—I do not give to charity
    $1—$50
    $51—$100
    $101—$200
    $201—$500
    $501—$1000
    More than $1001
    Prefer not to say

3. If you do make a charitable donation, do you feel better about yourself after donating to charity?
    Yes
    No
    Prefer not to say

4. If you won the lottery, how much of it would you give to charity?
   - None
   - Less than half
   - More than half
   - All of it
   - Prefer not to say

5. If you won the lottery and said you would give at least some of it to charity, would you give it all to one charity, or to several?
   - I said I would not give any of my winnings to charity
   - I would give to only one charity
   - I would give to several charities
   - Prefer not to say