Systematic User Evaluation of a Second Device based cast-as-intended Verifiability Approach

Tobias Hilt[1][0000−0001−9267−5109], Benjamin Berens[1][0000−0002−9284−7924], Tomasz Truderung[2], Margarita Udovychenko[2], Stephan Neumann[3][0000−0003−0091−493X], and Melanie Volkamer[1][0000−0003−2674−4043]

[1] Karlsruhe Institute of Technology, Karlsruhe, Germany
{tobias.hilt,benjamin.berens,melanie.volkamer}@kit.edu
[2] POLYAS GmbH, Kassel, Germany
{t.truderung,m.udovychenko}@polyas.com
[3] stephan@stephanneumann.it

Abstract. End-to-end verifiable e-voting schemes enhance the verifiability of individual votes during the election process. Specifically, methods for cast-as-intended verifiability empower voters to confirm that their cast votes have not been manipulated by the voting client. There are mainly three approaches to implement cast-as-intended verifiability in remote e-voting systems: (1) return-code based, (2) challenge-based and (3) second-device-approach. To investigate the usability, perceived trustworthiness and manipulation effectiveness for the second-device-approach, we conducted a user study with 133 participants. The results are similar to those from related work investigating the other two approaches.

Keywords: Cast-as-intended verifiability · Second-Device approach · User Study · Manipulation Detection Efficacy

1 Introduction

Elections are the bedrock of modern democracies. In an era of increasing digitalization, governments have adopted electronic solutions in various areas, and elections are no exception. Switzerland [23] and Estonia [8] stand out as notable examples, allowing voters to exercise their right to vote in national elections through remote e-voting channels. France has recently (re)joined this trend, allowing citizens living abroad to vote online in the 2022 legislative elections, after the introduction of an online channel for the 2012 election and it being halted in the 2017 election due to security concerns [7]. Germany has also made progress in this area, introducing an online voting channel in last year's social security elections (being the third-largest nation wide German election) in addition to the traditional postal voting channel [12].

The adoption of remote e-voting offers clear advantages as a voting channel. For example, it simplifies the voting process for citizens living abroad and increases the efficiency and accuracy of the counting process. However, it is important to recognize that the integration of technology introduces the risk of

deliberate manipulation of votes [10]. To mitigate this risk and increase the likelihood of detecting such tampering, security measures similar to election audit procedures for paper-based voting systems are essential. This includes verifying that: (1) the voting client accurately encoded the vote as intended by the voter (cast-as-intended verifiability), (2) the vote recorded by the voting system corresponds to the cast vote (recorded-as-cast verifiability), and (3) the recorded vote is accurately included in the final election result (tallied-as-recorded verifiability).

Our focus is on cast-as-intended verifiability, for which there are primarily three approaches: the Benaloh Challenge, Return-Codes, and the second-device approach. However, only the first two have undergone extensive user studies, assessing their general usability and effectiveness in detecting vote manipulations. This research focuses on investigating the usability, manipulation detection efficacy and perceived trustworthiness of the third approach, commonly known as the second-device approach. This approach is for example employed in Estonian national elections and was used in the GI-Election 2023[4].
We conducted a user study using an actual system implementing this approach with 133 participants. The study had two phases, in which each participant had to cast their vote once. The first phase was used to assess usability and simultaneously served as a deception, since participants were informed that their provided usability feedback on the voting system would be (at least partly) implemented for them to reassess in the second phase of the study. In reality, the second phase examined manipulation detection efficacy for two types of manipulation in addition to perceived trustworthiness.

Our results, encompassing both general usability and manipulation effectiveness, are compared with findings from related studies on the other two approaches. Of particular note is the manipulation detection efficacy of the two manipulation types closely matching the results reported in comparable studies.Trustworthiness of the system was perceived neutral, while high usability was attributed.

2 Background and Related work

2.1 Cast-as-intended Verifiability in Remote Electronic Voting

The term "E-Voting" describes the process of casting one's vote with the help of an electronic device, which can range from automated teller machines to complex remote electronic voting systems. The focus of this research is on remote electronic voting systems . One important aspect about remote electronic voting systems is the possibility to check that one's vote was cast-as-intended. There are mainly three approaches to implement cast-as-intended verifiability in remote electronic voting systems: (1) Using the Benaloh Challenge introduced in [5], (2) Providing so called return codes after the vote is cast as required in Switzerland and e.g. proposed in [9] which voters are supposed to compare with the codes

[4] GI = Gesellschaft für Informatik; `https://gi.de`, Last accessed 15.02.24

provided on their code sheet send to them via postal mail before the election, (3) Enabling voters after having cast their vote to use a second device with an independent verifying application to check if their vote cast as intended. This approach is for instance applied in Estonia since 2013 (see e.g. [11] for a detailed description).

2.2 User Studies in Remote Electronic Voting

In the general context of electronic voting there have been made several user studies such as [27], examining Pret-a-Voter [22], but as the system we used is based on remote electronic voting our focus is on user studies that are comparable to our approach. From the before mentioned three approaches to implement cast-as-intended verifiability in remote electronic voting systems, the first two approaches have been extensively evaluated with respect to their usability and usability improvements have been proposed and evaluated - e.g. the Benaloh Challenge has been studied in [2, 20] and the return code approach in [15, 17]. A comparative user study of both approaches was conducted in [14]. In [18], all three cast-as-intended verifiability approaches were compared regarding their manipulation detection efficacy in a user study. The third approach has received comparatively less attention; however, recent studies, such as the one conducted by [13], have delved into this area. In their research, the authors used semi-structured interviews to explore Estonian i-Voters' understanding of the cast-as-intended verifiability implemented in the Estonian i-Voting protocol.

3 System description: Vote casting and cast-as-intended verifiability

The cast-as-intended verifiability mechanism considered in this paper is based on the use of a second device performing the corresponding cast-as-intended verifiability mechanism. There are two commonly used realizations of such a mechanism, e.g. a web application or a native mobile application. For this user study, our focus is on the web application as we consider it difficult to find participants who are going to install an app on their mobile phones just for such a study.

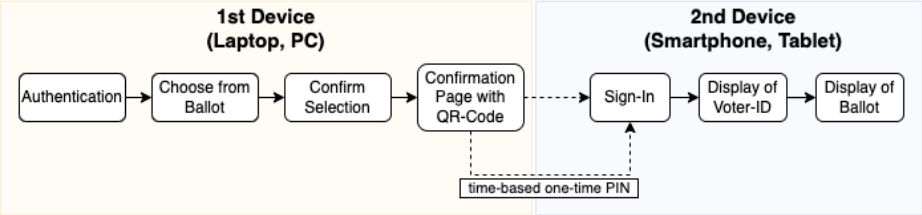

Fig. 1. Simplified Vote Casting and Verifying Process

From the voter's point of view the voting process is divided into two parts (see Fig. 1).

1st Device. The voter authenticates themselves at the voting system, using the credentials they received[5]. Upon successful authentication the voter is greeted and the ballot is displayed. The voter selects and confirms their choice. After that the vote is cast and the voter is presented a confirmation page (see Fig. 2), including a QR code that can be used to perform cast-as-intended verifiability with a second device[6]. In addition this page displays a time-based one-time PIN, which is needed for authenticating at the web application. The PIN is refreshed every 30 seconds to deter vote selling and prevent voters from easily sharing their cast vote.

Fig. 2. Confirmation page of the voting system with the QR code and time-based one-time PIN

2nd Device. In order to perform cast-as-intended verifiability, which is optional, voters must scan the QR code with a suitable device (i.e. a smartphone or tablet). After scanning the QR code the voter is transferred to the web application, hosted by an independent provider[7], where they must authenticate using the

[5] In this study, these credentials were provided in the form of a role-card, containing the voter-ID and an election invitation letter, containing the password.

[6] The interfaces are inspired by the interfaces used by the Polyas company in their verifiable voting system. We improved the language and design based on our usability expertise.

[7] In our study we simulated the host to be OSCE (Organization for Security and Co-operation in Europe), while we hosted the web application on our own servers.

time-based one-time PIN, that is currently displayed on the confirmation page on the first device.

The details of the system used are described in [19], giving information why the following cast-as-intended verifiability specific security properties are provided.

Election integrity. Election integrity wrt. cast-as-intended is ensured under the following assumptions: (1) voters actually verify their vote and check if both their displayed voter-ID and selection matches; (2) one of the devices (the primary voting device or the second device) is not corrupted; (3) either the voting-system or at least one of the verification mechanisms (if there are more than one) used by the voter to verify are not corrupt.

Ballot privacy. For ballot privacy, the voter needs to trust the second device, as it learns the voter's choice (which it displays to the voter).

4 Methodology

4.1 Recruitment, Ethics & Data Protection

Recruitment Participants needed to be 18 years or older and be fluent in German, in order to be recruited. Additionally, only participants using a PC or laptop were allowed, as we wanted to minimize the possibility of participants being unable to verify their vote with a second device due to them participating with their smartphone.

Ethics We coordinated the process alongside the ethical guidelines and received approval by the ethical committee of the KIT (Karlsruhe Institute for Technology). Participants were granted a compensation of 3€(Second phase: 4€), which was calculated using the approximate study time multiplied by the minimum wage in Germany. We offered participants to abort the study after debriefing and still receive the money, which none of them did.

Data Protection In cooperation with the data protection officer of our university, we created information about the usage of collected data, conforming to recent GDPR, which we presented to participants at the beginning of the study to inform participants about their rights and the usage of their collected data.

4.2 Research Questions

As already pointed out in Section 2 there has been limited work done to examine usability of remote electronic voting systems with cast-as-intended verifiability utilizing a second device. Thus, we try to contribute by answering the following research questions:

RQ1: *How usable do voters perceive a remote electronic voting system with cast-as-intended verifiability utilizing a second device?*

RQ2: *What is the manipulation detection efficacy of voters using a remote electronic voting system with cast-as-intended verifiability utilizing a second device?*

RQ3: *How trustworthy do voters perceive a remote electronic voting system with cast-as-intended verifiability utilizing a second device?*

To answer these questions we designed an extensive, two phase user study, which is explained in the following subsections.

4.3 Study Procedure

The study consisted of two phases. Phase one focused on perceived usability, while phase two examined manipulation detection efficacy and perceived trustworthiness. As participants were told the study was solely about perceived usability and that in the second phase they would have to reassess a reworked system, phase one also served as deception. All supplementary material used is attainable at `https://doi.org/10.35097/1934`.

Phase 1 The most important processes of phase one are illustrated in Fig. 3 and explained below.

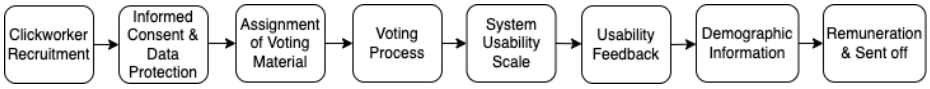

Fig. 3. Structure of Phase 1

Clickworker Recruiting Participants were exclusively recruited with the online panel "clickworker"[8], from which participants were transferred to the online questionnaire[9].

Informed Consent & Data Protection Starting the online questionnaire participants were presented the informed consent form and the data protection regulation.

[8] `https://www.clickworker.de`, Last accessed 14.12.2023
[9] `https://www.soscisurvey.de`, Last accessed 14.12.2023

Assignment of Voting Material Participants received a role card, including their personal identifier and the choice they should vote for in this election. They also received an invitation letter to the election, which contained the voting rules, the password needed for authentication, the link to a report website, in case they experience problems during the election and brief paragraph encouraging voters to verify their vote. For this research, we chose to replicate the European Parliament election scenario. In addition, we chose to simulate the letter as if it came from the Federal Returning Officer. Within this simulated letter, a brief paragraph was included to encourage voters to actively verify their votes. We could have used a straightforward informational approach, but this would likely have not motivated many participants, as seen in the German Social Election [12]. Our decision to create a unique paragraph was influenced by research demonstrating that simply providing information about this verification feature does not increase the rate at which voters engage in verification [25]. Based on research from [21], we developed the following text, which includes both an analogy and a norm cue to further motivate voters (translated from German):

> By verifying, attempts of manipulation can be detected, which are usually uncovered in a classic election with the help of independent election observers. [**Analogy**]
> Voters who want to protect democracy should therefore use their second device to check whether their vote was correctly transmitted to the digital ballot box. [**Norm**]

Voting Process. The voting process followed the logic explained in Fig. 1, from Section 3. The voting system was hosted by POLYAS[10], while the verification web application was hosted by our institute (Karlsruhe Institute for Technology).

System Usability Score. To objectively assess the perceived usability of the voting process we utilized the system usability scale [6], in the German version[11].

Usability Feedback. As part of the deception we asked participants open-ended questions about perceived usability.

Demographic Information. Participants were asked their age and gender.

Remuneration & Sent-off. Participants were thanked for participating, paid and reminded, that the second would start two weeks later.

[10] `www.polyas.de`, last accessed 20.02.2024

[11] `https://community.sap.com/t5/additional-blogs-by-sap/system-usability-scale-jetzt-auch-auf-deutsch/ba-p/13487686`, last accessed 16.02.2024

Phase 2 Two weeks after the completion of phase one the second phase started. Eligible for this phase were only the participants from phase one that actually participated in the election and verified their vote ($n = 133$)[12]. The general structure of this phase is illustrated in Fig. 4 and the processes that differ from phase one are explained below.

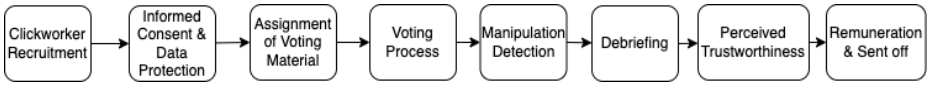

Fig. 4. Structure of Phase 2

Voting Process Participants were assigned to one of two voting systems based on the ID assigned in their role card, with each system subject to manipulation. In the event that participants detected tampering they had different reporting options, which are explained in Subsection 4.4. The voting process for the two types of manipulations and their respective viable reporting options is illustrated in Fig. 5.

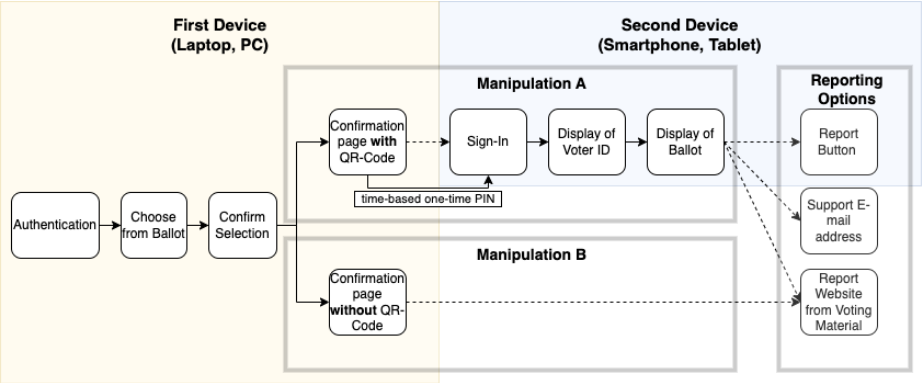

Fig. 5. Simplified Voting Process in Phase 2

Manipulation Detection Upon completion of the voting process, participants were asked an open-ended question to determine if they observed any anomalies. If they responded in the affirmative, they were then asked about the nature of the irregularities they observed and whether they took the initiative to report them.

[12] We confirmed this by providing an anonymised list of IDs to the online survey. This ensured that only Clickworkers whose IDs were on the list were able to participate.

Debriefing Based on the version of the voting system they experienced and the detection status of the manipulation, participants received a customized debriefing explaining the actual purpose of the study.

Perceived Trustworthiness. After the debriefing, participants were asked to mark their perceived trustworthiness of the remote electronic voting system on a Likert scale from one to five.

4.4 Type of Manipulation

We investigated two potential attack vectors applicable to remote electronic voting systems with cast-as-intended verifiability, inspired by those manipulation types examined by related work. First, we simulated a scenario in which a malicious voting device manipulated the vote without the voter's knowledge. Specifically, the voting application was configured to change the vote cast from "CDU" to "SPD", which are both German parties. Detection of this manipulation is possible if the voter chooses to verify his vote with a second (honest) device. In this case the voter has several secure reporting mechanisms available, including using the dedicated button within the verifying application, accessing the reporting website linked from the invitation letter, or reporting the problem using the dedicated support email addresses from the verification web application. Henceforth, this manipulation scenario is referred to as ***MT-A*** *(= Manipulation Type A: Vote Tampering Manipulation).*

We also simulated a scenario in which, the voting system itself has malicious intent, by manipulating the interface, making it impossible for the voter to verify their vote. We modified the system to withhold the display of a QR code at the end of the voting process, preventing voters from verifying their votes. The voter should refrain from using the support email address provided by the voting interface, as the platform lacks integrity. Consequently, in this scenario, the only viable option is to report using the reporting website from the election invitation letter. This form of manipulation is referred to as ***MT-B*** *(= Manipulation Type B: Verification Prevention Manipulation).*

4.5 Pre-Study

A pre-study involving 20 participants was conducted to assess system functionality and connectivity. Insights from the pre-study led to adjustments in the logic governing credential assignment, ensuring reliable payment for participants completing the online questionnaire. Minor modifications were also made to the voting material, such as aligning the voting period in the election invitation letter with the date specified in the role card.

5 Results

5.1 Demographic Information

A total of 176 participants were initially recruited, with 43 excluded due to data pruning (non-human answers or answers from participants that did not vote and just clicked through the survey[13]). The final sample for phase one consisted of 133 participants. Among them, 90 identified as male, 43 as female[14], and the majority held a university degree ($n = 64$). For phase two, 85 participants returned, of which 79 successfully participated. Exclusions were made for those who did not vote ($n = 4$), provided insufficient responses, as in blank spaces, ($n = 1$), or aborted the survey ($n = 1$). The remaining 79 participants included 50 men and 29 women.

5.2 *RQ1:* Perceived Usability

To evaluate the perceived usability, we employed the System Usability Scale (SUS), a standardized scale utilized for evaluating the perceived usability across various products [6]. The scale ranges from 0 to a maximum of 100 points, with higher values indicating better subjective usability. In order to assess the actual perceived usability, we chose to survey participants during phase one, where they experienced the unaltered system. Outliers that were 1.5 times the interquartile range over the third quartile or below the first quartile were excluded, as proposed by the IQR-method to prune data [1]. In total, three outlier were excluded. The average SUS score among the remaining participants was 78.92 ($sd = 14.26$), indicative of good usability [4]

5.3 *RQ2:* Manipulation Detection Efficiency

When assessing whether participants detected the manipulation, we included two factors: (1) the response to a question in the online questionnaire and (2) the usage of the reporting system. Regarding (1), the online questionnaire contained the question "Did you notice anything unusual during the election?". If participants responded in the affirmative and their answer contained a clear relation to one of the manipulation types, their response was categorized as "Detected", otherwise as "Not detected". For (2) all responses to the reporting system (i.e., the report website and support mail address, see Section 4.4) were inspected and categorized by two members of the research team. After a final discussion about the responses and their categorization a percentage agreement of 97.47% was reached. As before, responses were categorized as "Detected" and "Not detected". If one of the factors was categorized as "Detected", the responding participant was categorized as having detected the manipulation.

[13] Detected by comparing the IDs from the voting system with the corresponding survey ID.

[14] Note: It was possible to state "other" or "prefer not to state" but none of the participants did so.

	Detected	Not Detected
MT-A	40 (96%)	2 (4%)
MT-B	9 (24%)	28 (76%)
Total	49 (62%)	30 (38%)

Table 1. Amount of participants that detected the manipulation for both manipulation types.

Table 5.3 gives an overview of the manipulation detection efficacy, showing the overal detection rate at 62%. *MT-A* was detected far more often than *MT-B* (96% vs. 24%). A Fisher-exact confirmed the difference to be statistical significant (OR = 0.0174, 95% CI = [0.00172, 0.0877], $p < 0.001$). To determine whether gender has a statistical impact on the manipulation detection efficacy, we performed a CHI-squared test, which proves, that gender has no effect $X^2(1, N = 79) = 0.060793$, $p = 0.8052$. As not every age group had a minimum of five participants who detected and did not detect the manipulation, a Fisher-exact test was conducted. The results indicate that age also had no significant effect on manipulation detection efficiency ($p = 0.8597$).

5.4 *RQ3:* Perceived Trustworthiness

After debriefing participants about the actual purpose of the study, they were requested to assess the trustworthiness of the remote electronic voting system by indicating their level of trust on a Likert scale ranging from one (not trustworthy at all) to five (very trustworthy). On average, participants exhibited a neutral stance toward trustworthiness with a slight inclination towards positive perceived trustworthiness (3.15, $\sigma = 1.25$), as illustrated in Table 2.

Perceived trustworthiness	**Overall**		**MT-A**		**MT-B**	
	Detected	Not Detected	Detected	Not Detected	Detected	Not Detected
Not trustworthy at all (= 1)	16.3%	6.7%	17.5%	0%	11.1%	7.2%
Not trustworthy (= 2)	16.3%	20%	15%	0%	22.2%	21.4%
Neutral (= 3)	26.5%	26.7%	27.5%	0%	22.2%	28.6%
Trustworthy (= 4)	28.6%	26.7%	27.5%	50%	33.3%	25%
Very Trustworthy (= 5)	12.2%	20%	12.5%	50%	11.1%	17.8%
Average rating	3.04	3.33	3.03	4.5	3.11	3.15
	3.15		3.09		3.22	

Table 2. Distribution of participants' perceived trustworthiness based on manipulation type and detection status.

Participants that did not detect the manipulation expressed a slightly higher level of perceived trustworthiness (3.33, $sd = 1.21$) compared to those that did (3.04, $sd = 1.27$). A Fisher-exact test confirmed this marginal difference not to be statistically significant ($p = 0.7063$).

Participants from *MT-B* rated the system similar (3.22 $sd = 1.2$) to participants from *MT-A* (3.09 $sd = 1.3$), indicating that manipulation type has no statistical impact on perceived trustworthiness. A Fisher-exact test confirmed this presumption ($p = 0.7886$).

6 Discussion

6.1 Perceived Usability

In terms of usability (i.e. the SUS scores) the second device approach we used in this study performed similar to the other approaches from related work. Whereas the approaches from related work scored SUS-scores between 73 and 85, participants from our study attributed the approach a score of 79, describing "good usability". The comparison to related studies remains difficult as we used an actual voting system, while all of them used mock-up systems. Table 3 gives an overview of the archieved SUS scores.

6.2 Manipulation Detection Efficacy

As previously mentioned the comparison to the related work is difficult, especially with regards to manipulation detection efficacy due to several reasons: (1) the study design, (2) the instructions provided to the participants and (3) the note or request to participants to verify their vote, greatly differ from study to study. Another potential factor of influence is the presumption all of the other studies made in which they determined, that participants should see the voting system for the first time, when they were objected to the manipulation, as for that point in time no e-Voting was available for the participants in the real world (all these studies were conducted in Germany).

Table 3 also shows the detection rates from the various studies[15]. In the following we mainly focus on the user studies based on the return code based and second device approach, as these approaches have already been implemented in actual elections.

Overall *MT-A* was detected in 100% of cases in several studies based on the return-code based approach. The detection efficacy for *MT-B* is always much lower compared to *MT-A*, as also shown in Kulyk et al.'s study [16]. Ultimately the detection efficacy for *MT-B* is far better in return-code based approaches compared to the second device approach. This difference is potentially explained by the used approaches: Inherent to the return-code based approach is the voters expectation to receive something back after casting the vote, hence the name "return-code". Part of the voting process, as described in the voting material, is the verification. In contrast in the second device approach, the voter **can** (but does not need to), perform cast-as-intended verifiability. As the cast-as-intended

[15] Note, the comparison only provides some hints as the study designs are different in detail. Future work should investigate in a comparative study considering both manipulation types.

Source	cast-as-intended Verifiability Approach	SUS	Manipulation Type	Detection Efficacy
[26]*	Return-Code	82	MT-A	41%
[17]*	Return-Code	85	MT-A	100%
[16]*	Return-Code	81	MT-A MT-B	100% 43%
[18]	Return-Code Challenge Second Device	85 73 85	MT-A	100% 28% 64%
[14]	Return-Code Challenge	-	MT-A	100% 77%
[24]*	Return-Code	-	MT-B	71%
This study	Second Device	79	MT-A MT-B	96% 24%

Table 3. Overview of User Studies examining the detection rate of different manipulation types. *The authors made improvements to the systems/materials. We only included the improved versions.

verifiability-step is performed after casting the vote, it may be that voters are already mentally finished with the voting process, that they are more likely to not detect something missing, i.e. the QR code to perform cast-as-intended verifiability.

There is only one previous paper [18] which studied the manipulation detection rate for the second device approach (for MT-A only). Their detection rate was 64% and ours 96%. The improved detection rate may be explained by the fact that participants in our study had already used the honest system before and were therefore more likely to detect the manipulation.

6.3 Perceived Trustworthiness

The well reported usability may also have influenced perceived trustworthiness, as trustworthiness is often attributed to usable systems [3]. Although no statistical effects based on manipulation detection, gender, age, or manipulation type could be determined, the sequencing of the manipulation detection questions, debriefing, and subsequent trustworthiness assessment may have influenced participants' perceptions due to increased awareness of manipulation risks. Those who suspected manipulation may have rated trustworthiness differently than those who suspected error or did not notice anything unusual prior to the debriefing. This potential post hoc bias should be acknowledged, and future research should examine the effect of the debriefing sequence on trustworthiness ratings. It is important to assess perceived trustworthiness both before and after debriefing to understand potential changes influenced by participants' awareness of manipulation risks.

6.4 Limitation

By explicitly informing participants of the study's focus on usability, there is a potential bias. Participants may have paid more attention to the usability aspects and may have overlooked the introduced manipulation. Additionally, by instructing voters how to vote their ability to detect the manipulation may have been reduced as it is more likely to detect if a personal choice (with a potential emotional connection) to a party is changed compared to a random one, that was assigned to them. The study did not comprehensively address all possible MT-Attempts. For example, we used a web-based application to reduce the burden on participants. However, this introduces an additional attack vector not covered in our study, similar to *MT-B*. In this scenario, adversaries could redirect participants to a spoofed web application that displays a fake ballot, which contains the vote the voter intended to cast. Detecting this manipulation would require voters to examine the URL of the QR code to ensure that it redirects to the legitimate web application. Another potential attack vector we did not examine is the clash attack. In this scenario, the vote is altered, but during cast-as-intended verifiability with the second machine, the voter is presented with a different ballot containing the originally intended vote. Detecting this tampering could involve asking the voter to confirm that the ballot presented on the second machine is in fact their own, typically accomplished by comparing voter IDs. However, implementing this in our study was challenging because participants were assigned new voter IDs that they had never seen before, as opposed to a real-world scenario where voter IDs are familiar (e.g., social security number or ID card number). Both of these attack vectors warrant investigation in future research. We only offered the verification application from one provider whereas in a real election voters should have the option to choose from several providers. The process of categorizing participants answers into "Detected" and "Not detected" also forms a limitation. While participants correctly identified changes in their votes or the absence of a QR code for cast-as-intended verifiability, they typically described these instances as errors, revealing a potential disconnect in perception, e.g., "My vote was wrong" (Participant 65) or "Confusingly, I did not receive a QR code" (Participant 31).

7 Conclusion

We evaluated the second device based cast-as-intended verifiability approach which has not been thoroughly investigated, yet but is used in actual elections. We conducted a two-phase user study. Phase one focused on perceived usability, while phase two examined the effectiveness of two types of manipulation detection and perceived trustworthiness. Thus, first phase was also used as deception for the second phase. Moreover with the two phases, we were able simulate that participants knew the system already when studying the manipulation detection. We found that the second device approach performed similarly in terms of both perceived usability and manipulation detection compared to existing user studies that primarily examined the return code-based approach. Furthermore, we

found a similarity between the two: in both approaches, voters showed difficulty in detecting *MT-B*, where the voting system does not allow the voter to perform cast-as-intended verifiability. In comparison to user studies from the return-code based approach, we perform slightly worse but using return-code based approach is also not feasible for all kind of elections, as the process becomes more complex with larger elections, especially visually impaired people. Consequently, in the future one should decide on a case by case basis which approach to choose.

Acknowledgments This work was funded by the Topic Engineering Secure Systems, subtopic 46.23.01 Methods for Engineering Secure Systems, of the Helmholtz Association (HGF) and supported by KASTEL Security Research Labs, Karlsruhe.

References

1. Exploratory Data Analysis, pp. 192–194. Springer New York, New York, NY (2008). https://doi.org/10.1007/978-0-387-32833-1$_1$36, https://doi.org/10.1007/978-0-387-32833-1_136
2. Acemyan, C.Z., Kortum, P., Byrne, M.D., Wallach, D.S.: Usability of voter verifiable, end-to-end voting systems: Baseline data for Helios, Prêt à Voter, and Scantegrity II. The USENIX Journal of Election Technology and Systems **2**(3), 26–56 (2014)
3. Acemyan, C.Z., Kortum, P.: The relationship between trust and usability in systems. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **56**(1), 1842–1846 (2012)
4. Bangor, A., Kortum, P.T., Miller, J.T.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies **4**(3), 114–123 (2009), publisher: Citeseer
5. Benaloh, J.: Simple Verifiable Elections. Electronic Voting Technology Workshop **EVT '06** (2006), place: Berkeley, CA, USA Publisher: USENIX Association
6. Brooke, J.: SUS-A Quick and Dirty Usability Scale. Usability Evaluation in Industry **189**(194), 4–7 (1996), publisher: CRC Press
7. Cortier, V., Gaudry, P., Glondu, S., Ruhault, S.: French 2022 legislatives elections: a verifiability experiment. In: Accepted for: International Joint Conference on Electronic Voting (E-VOTE-ID) (2023)
8. Estonian National Electoral Committee: Statistics about Internet Voting in Estonia (2015), https://www.valimised.ee/en/archive/statistics-about-internet-voting-estonia
9. Galindo, D., Guasch, S., Puiggali, J.: Neuchâtel's Cast-as-Intended Verification Mechanism. In: International Conference on E-Voting and Identity (VOTE-ID). pp. 3–18. Springer (2015)
10. Halderman, J.A., Teague, V.: The new south wales ivote system: Security failures and verification flaws in a live online election. CoRR **abs/1504.05646** (2015)
11. Heiberg, S., Willemson, J.: Verifiable Internet Voting in Estonia. In: 6th International Conference on Electronic Voting, Verifying the Vote (EVOTE). pp. 1–8. IEEE (2014)

12. Hilt, T., Kulyk, O., Volkamer, M.: German social elections 2023: An overview and first analysis. In: Accepted for: International Joint Conference on Electronic Voting (E-VOTE-ID) (2023)
13. Hilt, T., Sein, K., Mällo, T., Villemson, J., Volkamer, M.: Voter perception of cast-as-intended verifiability in the estonian i-vote protocol. In: Accepted for: International Joint Conference on Electronic Voting (E-VOTE-ID) (2023)
14. Kulyk, O., Henzel, J., Renaud, K., Volkamer, M.: Comparing "Challenge-Based" and "Code-Based" Internet Voting Verification Implementations. In: IFIP Conference on Human-Computer Interaction. pp. 519–538. Springer (2019)
15. Kulyk, O., Ludwig, J., Volkamer, M., Koenig, R.E., Locher, P.: Usable Verifiable Secrecy-Preserving E-Voting. In: 6th International Joint Conference on Electronic Voting (E-Vote-ID). University of Tartu Press (2021)
16. Kulyk, O., Volkamer, M., Müller, M., Renaud, K.: Towards Improving the Efficacy of Code-Based Verification in Internet Voting. In: VOTING. Springer (2020)
17. Marky, K., Zimmermann, V., Funk, M., Daubert, J., Bleck, K., Mühlhäuser, M.: Improving the Usability and UX of the Swiss Internet Voting Interface. In: ACM CHI (2020)
18. Marky, K., Zollinger, M.L., Roenne, P., Ryan, P.Y., Grube, T., Kunze, K.: Investigating Usability and User Experience of Individually Verifiable Internet Voting Schemes. ACM Trans. Comput.-Hum. Interact **28**(5) (2021)
19. Müller, J., Truderung, T.: Caised: A protocol for cast-as-intended verifiability with a second device. In: Volkamer, M., Duenas-Cid, D., Rønne, P., Ryan, P.Y.A., Budurushi, J., Kulyk, O., Rodriguez Pérez, A., Spycher-Krivonosova, I. (eds.) Electronic Voting. pp. 123–139. Springer Nature Switzerland, Cham (2023)
20. Neumann, S., Olembo, M.M., Renaud, K., Volkamer, M.: Helios Verification: To Alleviate, or to Nominate: Is That the Question, or Shall we Have Both? In: International Conference on Electronic Government and the Information Systems Perspective. pp. 246–260. Springer (2014)
21. Olembo, M.M., Renaud, K., Bartsch, S., Volkamer, M.: Voter, what message will motivate you to verify your vote. In: USEC. Internet Society (2014)
22. Ryan, P.Y., Teague, V.: Pretty Good Democracy. In: Security Protocols Workshop. vol. 17, pp. 111–130. Springer (2009)
23. Serdult, U., Germann, M., Mendez, F., Portenier, A., Wellig, C.: Fifteen Years of Internet Voting in Switzerland. In: ICEDEG. pp. 126–132. IEEE (2015)
24. Thürwächter, P.T., Volkamer, M., Kulyk, O.: Individual verifiability with return codes: Manipulation detection efficacy. In: Krimmer, R., Volkamer, M., Duenas-Cid, D., Rønne, P., Germann, M. (eds.) Electronic Voting. pp. 139–156. Springer International Publishing, Cham (2022)
25. Thürwächter, P.T., Volkamer, M., Kulyk, O.: Individual verifiability with return codes: Manipulation detection efficacy. In: 7th International Conference on Electronic Voting (E-Vote-ID). vol. 13553, p. 139–156. Springer LNCS (2022)
26. Volkamer, M., Kulyk, O., Ludwig, J., Fuhrberg, N.: Increasing security without decreasing usability: Comparison of various verifiable voting systems. In: Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022). USENIX Association, Boston, MA (Aug 2022)
27. Winckler, M., Bernhaupt, R., Palanque, P., Lundin, D., Leach, K., Ryan, P., Alberdi, E., Strigini, L.: Assessing the Usability of Open Verifiable E-Voting Systems: a Trial with the System Prêt à Voter. In: ICE-GOV. pp. 281–296 (2009)