Investigating Voter Perceptions of Printed Physical Audit Trails for Online Voting

Karola Marky[1,3], Nina Gerber[2], Henry John Krumb[2], Mohamed Khamis[3], Max Mühlhäuser[2]
[1]*Ruhr University Bochum, Germany, karola.marky@rub.de;* [2]*Technical University of Darmstadt*
name_middlename.surname@tu-darmstadt.de; [3]*University of Glasgow, mohamed.khamis@glasgow.ac.uk*

***Abstract*—Online elections come with security challenges since digital votes do not produce physical audit trails that are easily verifiable. We present and investigate a *hybrid* online voting system that combines the benefits of voting from home via the internet with those of physical ballots, such as risk-limiting audits and verifiability. After voting online, the system generates a tracking code and a physical printout – either paper or 3D-printed – of the encrypted vote that can be visually verified by the voters through live video-broadcasts. Through an online experiment (N=150), we compared hybrid voting with paper and 3D-printed votes to a baseline (digitally stored votes), investigating perceived trust, UX, usability, and security readiness. Among our results, we show that paper printouts enhance trust without negatively impacting UX. 3D-printouts enhance perceived privacy, yet impact usability and UX. We conclude with recommendations and practical considerations to inform the implementation of hybrid online voting schemes.**

1. Introduction

Elections form the cornerstone of modern democracies. Online voting could further strengthen democracies by delivering benefits compared to traditional in-person voting, such as location-independent participation or support faster announcement of election results. Further, online voting can offer more inclusive elections by accommodating the needs of voters with impairments [1] or individuals whose family or professional obligations make it difficult to visit the polling station. State-of-the-art online voting protocols offer *individual verifiability* allowing voters to verify that their voting intention is stored in the ballot box and hence considered in the tally [2]. Individual verifiability for in-person voting is possible; however, voters currently have to be physically present in polling stations to observe the handling of votes, which may take several hours.

Implementing online voting, however, requires complex cryptographic protocols to secure elections from malicious third parties and ensure democratic principles, such as secrecy of the ballot and integrity of the result [3]. Even if security is carefully considered and implemented, there can still be unforeseeable problems that result in invalid digital votes [4]. This results in two main challenges of fully digital elections:

1) Human factor challenges: Online voting systems are very different from in-person voting with paper ballots by requiring additional steps that go beyond the voting process. These extra steps can result not only in human errors by voters [5] and poll workers [6], but also in a loss of voter trust [7]–[9], and even in non-participation [10], [11].

2) Audit challenge: Manual recounts in case of failure are not possible because there are only digital votes. This might result in disenfranchising voters.

If these two challenges are not adequately considered, very basic principles of modern democracies are defeated. The two challenges are rooted in the fact that online voting systems are primarily designed to be completely digital, including ballot marking, vote casting, and tallying. Yet, the benefits of online voting can also be achieved by digitising only part of that process: ballot marking and submitting it to the ballot box.

In this paper, we propose *hybrid online voting*, a concept that leverages the benefits of voting from home via the internet while enabling a physical audit trail to strengthen understandability and security. After submitting an encrypted ballot over the internet, an analogue representation is printed in a secure facility and inserted into a physical ballot box. This allows an online voting process that is partly similar to in-person voting. Voters can observe the printing and insertion processes via live video broadcasts for verifiability purposes. The printing process addresses auditing issues by creating a physical audit trail that allows verification of electronic tallies, risk-limiting audits, and recounts. It further addresses the human factors challenges by modelling a system close to in-person voting where verifiability to similar to checking a paper ballot. Finally, the broadcasts are also available to the public allowing for universal observability. Any interested party can verify that the ballots are handled according to the voting protocol.

We investigate our proposed voting concept in an online experiment with 150 participants contributing the first evaluation of a hybrid online voting system specifically investigating the following two research questions:

RQ1: *How does hybrid online voting impact subjective trust and perceived security?* We wanted to find out if and how the physical audit trail and watching the generating of it impacts the subjective trust of voters and their security perceptions.

RQ2: *How do potential voters perceive hybrid online voting in terms of usability and user experience (UX)?* We wanted to investigate whether hybrid online voting indeed reduces the complexity of online voting and verification procedures.

For our investigation, we designed two voting systems that broadcast the manufacturing of physical audit objects to voters via a live video broadcast. The first audit trail is a paper printout of the encrypted ballot, while the second one is a physical token (3D-printed). Both audit objects have a tracking code for the voters to identify their ballots. Tracking codes are physically detached before the tally to anonymize the votes. In our online experiment, we compared paper and token-based audit trails to a baseline (no audit trail).

The results reveal that observing the paper printing process enhances voter trust. Further, the paper audit trail does not negatively impact the user experience and participants stated that they want to watch it for verification purposes. Using a token was perceived to be significantly better at preserving vote secrecy when compared to no audit trail and paper. However, using tokens impacts usability and user experience. Based on our results, we conclude that paper is the most appropriate medium for the physical ballots.

Research Contribution

1) We propose **hybrid online voting** – the first concept that leverages improved availability and inclusion of online voting with the audit and verification benefits of physical ballots. We outline requirements for the concept and discuss options as well as challenges considering security and practicability to implement them.
2) We contribute the **first investigation** of a hybrid online voting system focusing on the voters' subjective perspectives. Based on a prototype implementation of two possible realisations of hybrid online voting, we conducted a user study with 150 participants in which 3D-printed tokens are compared to paper printouts and a baseline (no audit trails).
3) Based on our results, we discuss **opportunities and challenges** of hybrid online voting, including security and scalability issues, and conclude with final recommendations for practitioners and policymakers and guidance for future work.

2. Background & Related Work

This section first summarises research on hybrid voting systems, audit trails, and user studies.

Hybrid Voting and Audit Trails. In elections, *audit trails* provide supporting documentation through which the accuracy of the integrity of the election result can be verified [12]. Such documentation consists of official documents and protocols produced by voting authorities. In paper-based elections, the ballots can be archived for a certain period as part of the audit trail. Further, audit trails can be used for conducting *risk-limiting audits* [13]. For this, a random subset of paper ballots is used to determine whether the tally was executed correctly. Risk-limiting audits can also be used to audit voting equipment. In case voting equipment malfunctions, the audit trail can be used for a recount.

Electronic voting refers to any system that uses technology in the voting process. Direct recording electronics (DREs) – the voting computers used in polling stations – are the most prominent technology. There are two main types: fully digital and printout-based DREs. Using fully digital DREs, voters fill in digital ballots which are stored on memory cards or servers. Printout DREs print the ballot producing an audit trail. DREs with printouts can be considered hybrid electronic voting systems. The specific design of DREs has a high impact on usability and comprehensibility [14]–[16]. Fully digital DREs have many limitations. Risk-limiting audits are not possible, meaning that malfunctioning DREs cannot be identified [17]. Furthermore, voters do not receive assurance that their votes are indeed cast matching their voting intention, which can impact trust [14], [18]. Printout DREs allow for risk-limiting audits [18] yet, the printed output of existing DREs has been shown to be challenging to count for poll workers [6], [19] showing that the ballot design highly impacts the counting process.

To address these issues, the printout DRE system *EasyVote* [20] was proposed. It prints two representations of the voters' choices: (1) a clear-text representation that the voters can verify and (2) a QR code that encodes the voters' choices. After visually inspecting the printout, voters insert it into the ballot box. During counting, poll-workers scan the QR code and compare the scan result with the clear-text presentation. If both match, they proceed, if not, the clear-text representation is counted since this is verified by the voter. An evaluation of EasyVote showed that this system leads to fewer errors from voters while allowing a fast tally [20]. Precinct optical scanners are another form of a hybrid system. Those are ballot boxes that digitise the voters' physical ballots after insertion [21], [22]. Some precinct optical scanners require a specific ballot format that can impact usability [14], [16], [23].

Another stream of research on hybrid voting aims to address the issue of coercion, i.e., forcing someone to vote for a specific candidate. These protocols allow voters to overwrite their coerced online vote with an analogue vote that voters cast in a polling station [24], [25]. Such protocols have not been investigated in user studies in the literature.

In sum, hybrid voting enables risk-limiting audits contributing to election integrity and addressing audit challenges. The design of the electronic system highly impacts its usability for both voters and poll workers. Usable systems result in fewer voting errors by voters and can speed up the counting process. This paper first translates hybrid voting from in-person to online voting combining the advantages of printed audit trails and online voting. Second, it presents the first user evaluation of hybrid online voting specifically focusing on the voters' perspectives.

Verifiability in Voting. A variety of online voting schemes has been proposed in the literature. Researchers have argued to provide verifiable systems allowing voters to verify their votes [2]. A comprehensive overview of verifiable online voting revealed that there are five types of verification: (1) audit-or-cast, (2) tracking codes, (3) verification devices, (4) code sheets, and (5) delegation [26]. *Audit-or-cast* schemes operate based on cut-and-choose cryptographic

protocols [27]. After preparing an encrypted vote, voters either verify the vote's ciphertext or cast it. The main idea behind this is that during the voting process, an attacker cannot know about the voter's next action. In case, the attacker manipulates the vote, the voter might verify and find the manipulation. Consequently, the scheme is probabilistic and assumes that an attacker cannot anticipate how often a voter is going to verify. Considering human factors, several studies revealed that this assumption is not realistic because voters either verify at most twice [26], [28] or consider verification as a redundant extra task because cast votes cannot be verified [28]. *Tracking codes* use personal tracking codes for verification [29]. After the election, the authorities publish all counted voting options together with the voters' codes. The codes, then, allow voters to identify their votes in officially published tally records in a privacy-preserving way, meaning that only voters know their code. Investigations show that attempts to explain cryptographic components of tracking code systems during voting resulted in higher security perception but hampered understandability [7]. Further, the tracking codes were not perceived as convincing proof [30], and some voters even considered them to be a threat to vote privacy [26], [31]. *Verification devices* use a second trusted device, e.g., a smartphone. Estonia, for instance, provides a smartphone app to download and verify the encrypted vote from the ballot box server [32]. According to log files from Estonian elections, about 4% of voters perform this kind of verification [33]. *Code sheet schemes* use paper-based materials distributed to voters via postal mail for verification [34]. Code sheets are similar to indexed transaction authentication numbers used in online banking. They contain a verification code for each possible voting option individual to each voter. Once the encrypted vote is submitted, the voting system answers with the code that the voters then have to compare to their code sheet. Further, it is possible to *delegate* verification to a trusted third party, which, however, complicates the voting process [26].

Investigating Verifiability in User Studies. A large-scale investigation of printout DREs in a realistic setting revealed that less than half of the voters verify their ballot without instruction to do so [35]. This was also shown for online systems where some study participants interacting with an audit-or-cast scheme did not even try to verify [28]. Participants who interacted with the online voting prototype used in Norway could not determine whether their votes were submitted to the electronic ballot box [36].

The interface design can also impact the voters' ability and willingness to verify as shown in a study of the Swiss online voting interface that uses code sheets [37]. If the interface does not guide voters step-by-step through verification, about one-third of voters might overlook incorrect votes. A comparative study of the verification schemes that specifically investigated (1) audit-or-cast, (2) tracking codes, (3) verification devices, and (4) code sheets revealed that even if voters are specifically instructed to verify within the voting interface, they might not be successful [26]. More specifically, in audit-or-cast schemes, over 70% of incorrect votes were not detected. Further, in tracking code and verification device schemes, the share of not found errors was around one-fourth. Using code sheets, verification was mandatory, resulting in the detection of all errors. No matter which online system was used so far, verifiability resulted in user friction since it is a concept very different from paper voting. Based on that, we leverage results about DREs [35] to create a hybrid online system that replaces the verifiability component with an observable broadcast that creates a printed ballot.

3. Hybrid Online Voting Concept

Hybrid online voting combines benefits from online and in-person voting to offer a more intuitive and inclusive system. This section details the voting concept.

3.1. Requirements & Trust Assumptions

In this section, we outline the requirements for hybrid online voting that we derived based on the human factors and audit challenges outlined above.

Online Participation: Voters should be able to cast their votes over the internet. Voting software should assist voters in making valid voting choices while also allowing them to abstain from the election or submit invalid votes.

Physical Audit Trail: Hybrid online voting provides a physical audit trail. The audit trail consists of the ballots that were cast by eligible voters enabling risk-limiting audits and recounts. While digital logs of ballots could be used for that, it cannot be assured that each part of the digital infrastructure will function error-free (cf. [4], [17]). Physical records allow verification of the voting equipment. Once created, it is extremely challenging to tamper with a large share of the physical records because several insiders and election officials would have to work together.

Individual Verifiability: Voters should be able to verify that their ballot was cast matching their voting intention. They should be able to observe the creation of their physical ballot representation and alert authorities in case of irregularities.

Universal Observability: Any individual should be able to observe the creation of physical ballot representations and the tally. This allows observers to verify the correct execution of the audit trail creation to make sure that this matches the voting protocol.

Tally Support: The physical ballot representations should contribute to an efficient tally that accurately reflects the voting intentions. The physical representations should be designed in a way that assists poll-workers in counting.

Trust Assumptions: To ensure the security of hybrid online voting, we need the following trust assumptions about the printing facilities:

A1 The broadcast of the printing is trustworthy.
A2 The printing facility provides non-interrupted service.
A3 The operators of the printing facility are trustworthy.
A4 Unique ballots for each voter are generated.

(a) paper printing (b) token printing

Figure 1: Broadcasts of printing that were observed by the participants in our study.

3.2. Implementation & Security Aspects

This section details how we implemented the requirements outlined above, focusing on the interactions of voters and poll workers while voting. Screenshots of our implementation are in the Appendix. While the primary focus of this paper is on voter perceptions and the interaction with hybrid online voting, we also provide security considerations.

Online Participation: We implemented a voting website informed by online voting design guidelines from the literature [37]. To participate in the election, the voters authenticate. Then, the website displays instructions and an overview of the following steps. Next, the voters cast their votes by choosing a candidate from a list, or abstain. The voters review their choices and confirm them on the next screen. Finally, the encrypted vote is sent to the electronic ballot box. This refers to a typical online voting process without an audit trail and without individual verifiability.

Physical Audit Trails: The voting system creates physical ballots in a secure facility after vote casting. For this, we considered two kinds of ballots: (1) paper ballots and (2) tokens (see Figure 1). We designed a paper ballot representation that can be printed using commercial printers. The ballots have QR codes that encode the encrypted vote similar to the EasyVote in-person voting system [20]. We further tested tokens as an alternative that does not rely on paper. Our initial idea was the usage of smart cards since they can perform basic mathematical operations and improve security. However, smart cards and their functionality are difficult to verify for voters. Hence, we used something that allowed voters to *visually* verify (a) the fabrication of the token and (b) the data that is stored on it. Based on a literature search on voting tokens, we found 3D-printed ballot representations for usage at home that leverage capacitive sensing [38]. However, these tokens do not offer any verifiability features because they just simplify entering voting codes. The 3D printing process can be observed by voters, and the 3D-printed data can be visually verified. The token is printed out of two materials, PLA (i.e., plastic) and conductive PLA (i.e., PLA with graphite). The conductive structure can be sensed by commercial touchscreens. To realise the token, the conductive structure is printed in another colour and encodes the encrypted vote comparable to a QR-code-like structure that voters can visually verify. The QR-code-like structure can be decoded with the help of a touch sensor, such as a smartphone screen.

The distribution of secure facilities with printers and ballot boxes is security-critical. Many online voting systems, such as Estonia's [32], have a central ballot box server that is physically located in a secure facility. While this might be cost-efficient, it introduces a single point of failure. Nation-wide elections with potentially millions of voters are challenging using hybrid online voting if there is only one central printing facility. Hence, we instead propose modelling the system of how ballot boxes in postal voting are currently used. Consequently, there would be (at least) one printing centre in each voting district. Having this distribution of system elements also benefits security since there is no single point of failure, and it is harder for malicious third parties to compromise a big proportion of the infrastructure.

Individual Verifiability: Voters need the possibility to verify the creation of their physical ballot. To achieve this, in-person voting is modelled by letting voters visually observe the ballot creation and its insertion into the ballot box. We realised this using a live video broadcast. To make sure voters observe their vote, rather than somebody else's, the voters get notified with a personal tracking code by the software upon vote casting. This tracking code is printed visually on the physical ballot. To ensure vote secrecy, the tracking codes are physically separated from the ballot before the tally. For paper ballots, we added the tracking code to a separate part of the ballot. For the token-based ballots, we printed a structure that could be broken off. Once the printing is done, either a poll worker takes the printout and inserts it into a physical ballot box or this is done automatically. To verify the voting choice, i.e., that the vote indeed encodes the candidate chosen by the voter, we added a visual hash pattern of the encrypted ballot to both the voting software and the printouts that voters can visually compare informed by related work [39]. In

case the voters observe irregularities, e.g., the ballot is not inserted into the ballot box, they can alert the authorities. For this, we added instructions and a hotline number to the voting software to model the process used in Switzerland.

Universal Observability: Universal observability is given by access to a public website that hosts the broadcasts of all printing facilities. If an observer notices irregularities, they can alert the authorities, e.g., via a hotline.

Tally Support: Although we did not investigate tally support in our study, the paper audit trail builds upon the prior research of EasyVote, which facilitates tallying by scanning QR codes [20]. The 3D-printed tokens could be tallied similarly using touch sensors, e.g., those integrated into smartphones [38].

For our study, we adjusted our implementation: Since we could not send physical letters to the participants, we embedded the credentials in the website using text boxes highlighted in orange colour that told the participants to imagine they received these credentials before the election via postal mail. For the paper ballots, we filmed a Samsung LaserJet printer printing a ballot. Next, the paper was manually inserted into a physical ballot box. For the tokens, we used a Prusa i3 MK3S+ 3D printer to print the tokens in two materials. The tokens have dimensions of 30 $\times$ 52mm and take. We cut and modified the video to be roughly three minutes long to be as close to the paper printouts as possible in terms of duration. We acknowledge this limitation of our implementation; however, newer printer models print faster.

4. Methodology

To investigate hybrid online voting, we opted for an online study to have a realistic environment where participants use their own devices. The conditions in the study were the two audit trails (paper and token) and a baseline (no audit trail). The study used a between-subjects design, meaning that each participant interacted with one audit trail. To ensure an equal group size, participants were randomly assigned to the conditions using urns.

Task and Apparatus: We implemented one voting website for each condition. Since the participant pool was not limited to a specific country, we implemented a ballot for a generic government election that featured eight voting options given by topics that political parties could represent. To vote for a candidate, the participant clicked a round check box next to the candidate's name. We asked the participants to consider an election in their country of residence for the government or parliament when casting a vote. To preserve the participants' political opinions, we followed recommendations from related work [40] and provided written voting instructions including a party to vote for, again embedded in the voting software as orange text boxes.

Captured Data: We captured different constructs by questionnaires. To investigate RQ1 (*How does hybrid online voting impact subjective trust and perceived security?*), we used the following constructs. We assessed *security readiness* which reflects how users are psychologically prepared and willing to adopt security measures by the security readiness scale [41]. To capture further security perceptions, we added security-related statements that participants were asked to rate on a 5-point Likert scale. Even if online voting systems can offer great usability, trust was shown to be a crucial factor for adoption [10], [11], [42]. To assess *trust*, we used the refined HCTM scale [43]. To capture the *intention of use*, we again used statements that participants were asked to rate on a 5-point Likert scale. For investigating RQ2 (*How do potential voters perceive hybrid online voting in terms of usability and user experience (UX)?*), we assessed *subjective usability* using the System Usability Scale (SUS) [44] since this questionnaire has been shown to be effective to assess voting usability (cf. [5], [28], [37], [45]). We further measured *user experience* using the User Experience Questionnaire (UEQ) [46] and used the *interaction vocabulary* to measure how the interaction was perceived and assess interaction properties [47]. The *interaction vocabulary* is a semantic differential that captures how the interaction with a product is experienced on eleven different dimensions. The focus here is on a descriptive, non-judgemental recording of the interaction since different interaction characteristics can be experienced as positive or negative depending on the situation (e.g. "gentle" vs. "powerful"). Further, we assessed the participants' preference regarding online voting compared to postal voting and polling station voting. Finally, we asked the participants questions that were specific to the study condition (see Appendix A.1). Besides the specific questionnaires, we captured the participants' general attitudes towards online voting as a control variable to account for effects on the assessed constructs based on the participants' overall opinion about online voting [40].

Study Procedure: After consenting to the study, the participants were informed about their voting task and redirected to the online voting system. The participants cast a vote using the online voting system matching their assigned conditions. Then, they were redirected back to the survey provider. This part had two attention checks. After the interaction, the participants answered the questionnaires detailed above and provided demographics. Finally, they were redirected to the Prolific platform for reimbursement.

Pre-Study: The study setup was tested in a two-step pilot study. First, five researchers provided feedback to improve the clarity of instructions and questions. Then, we ran a pre-study with nine participants to clarify the procedure and questionnaires. The data from the pre-study are not included in the results. We further clarified instructions and questions. Some participants did not notice the broadcast due to a small screen. Because of that, we placed the broadcast on top of the website.

Condition	Age [years]	Gender	Experience
No audit trail (N=50)	Average: 28.8 Median: 26 Min: 18 Max: 56 SD: 9.7	20 F 30 M 0 NB 0 N/A	Average: 2.7 Median: 2 Min: 1 Max: 6 SD: 1.7
Paper audit trail (N=50)	Average: 26.2 Median: 24 Min: 18 Max: 52 SD: 6.9	30 F 18 M 1 NB 1 N/A	Average: 3.1 Median:3 Min: 1 Max: 6 SD: 1.6
3D-token audit trail (N=50)	Average: 27.9 Median: 25 Min: 18 Max: 64 SD: 9.1	31 F 19 M 0 NB 1 N/A	Average: 2.5 Median: 2 Min: 1 Max: 6 SD: 1.7
Total (N=150)	Average: 27.7 Median: 25 Min: 18 Max: 64 SD: 8.7	81 F 66 M 1 NB 2 N/A	Average: 2.8 Median: 2 Min: 1 Max: 6 SD: 1.7

TABLE 1: Overview of our sample. Voting experience ranges from 1 to 6 where 1 denotes little experience in voting.

Recruitment and Participants: We recruited 150 participants via the recruitment platform Prolific. The study was conducted in English and, thus, limited to participants with sufficient proficiency in the English language. Fifty participants interacted with each condition. The participants were, on average, 27.7 years old. Eighty-one participants identified as female, 66 as male, and one as non-binary. Two participants preferred not to state their gender. The participants came from different countries, mostly from the US and Europe, including the UK, France, Germany, Italy, and Poland. A smaller share also came from African countries, mostly South Africa, and Latin American countries. The distribution of the countries was similar in each condition (see Table 3 in the Appendix). Table 1 provides a detailed overview of our sample. None of the participants reported having online voting experience. The participants were reimbursed with an equivalent of 9 pounds per hour.

Ethical Considerations: The authors' institutions did not require formal IRB approval for this work because our institution had neither an institutional review board (IRB) nor an ethics review board (ERB) that applied to our study. However, we adhered to the strict (inter)national privacy laws and followed best practices for research conduct and transparency. The study procedure matches our institution's ethics guidelines and is GDPR-compliant. The consent form informed the participants about their rights as a participant. It mentioned that the study could be aborted at any time without any negative consequences. The data of participants that abort the study is deleted. Participants were informed about data processing, storage, privacy, and that the collected data could not be used to identify individuals.

Data Analysis: We collected 154 complete questionnaire responses. Four of them were filtered because the participants failed the attention checks. We analysed each construct using statistical methods. First, we checked whether all required assumptions for a one-way ANOVA were met. We used non-parametric tests in case the results were non-normally distributed. In case the Levene test indicated that the assumption of variance homogeneity was not met, we adjusted the degrees of freedom by calculating Welch's ANOVA. For all post-hoc tests, we used Bonferroni-Holm-corrected alpha-levels[1]. We calculated the size of our sample using G*Power considering an α-level of 0.05, power of 0.8 and considered a large effect size[2]. This resulted in a required total sample size of 66 for the ANOVA calculations and 42 for the pairwise comparisons indicating that our sample size of 150 is sufficient.

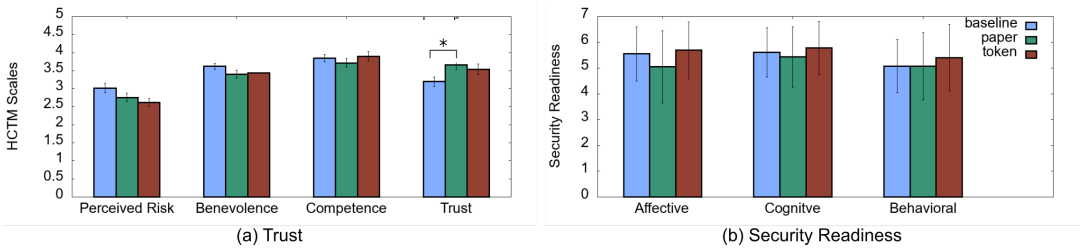
To analyse the free-text responses, we first familiarised ourselves with them by reading all answers. The responses overall consisted of comments that reflect the constructs evaluated in the study as well as accessibility and design elements as additional aspects. This resulted in a codebook with the following eight codes: usability and user experience, trust, security, design elements, verifiability, accessibility, availability, and other comments not related to the voting experience. Next, one researcher coded all answers by applying the codebook. A second researcher then verified the coding, disagreements were discussed, and final code allocations were agreed upon. During the coding, we could link the free-text answers to specific constructs that we evaluated. Hence, we provide quotes throughout the following sections.

Limitations: First, our sample was drawn via the Prolific recruitment platform. Hence, it might not represent the whole population of potential online voters as it might be biased toward relatively tech-savvy and younger people. The study, therefore, serves as a first insight into this topic and should be repeated with demographically representative samples drawn from different cultural groups.

Second, although we instructed our participants to imagine they would elect the government or parliament in their country, their perceptions of the voting process might differ in a real election setting. In a scenario where the outcome of the election plays a more decisive role for the participants, they may place more value on the verifiability of their vote [48]. Further, we simulated the printing process with pre-recorded videos. Overall, participants took 10.5 min to answer the questionnaires, yet the token printing videos were a bit longer than the other conditions which might have impacted the results. Future work should investigate a realistic scenario with real-time printing and individual voting choices of participants. However, conducting such a study during an actual government or parliament election is challenging as it would have ethical implications.

1. Since we have three conditions, the Bonferroni-Holm-corrected alpha-levels are .05, .025 and .0167 respectively

2. With 0.4 for the ANOVA calculations and 0.8 for the pairwise comparisons using the Mann Whitney U test.



(a) Trust (b) Security Readiness

Figure 2: Results of the HCTM and security readiness scales. The asterisk * indicates statistically significant differences. The error bars depict the standard deviation.

Third, although we solicited responses to open-ended questions, we mainly gathered quantitative feedback. We chose this setting to keep the duration of the online study within a time frame in which the participants can still concentrate well. However, future studies should also rely on qualitative data, e.g., by interviews, to capture people's perceptions of the concept more comprehensively.

Fourth, there are several directions in which a user investigation of online voting can go. We intentionally focused on the voters' subjective perceptions in this paper. However, from the security perspective, it is also essential to investigate to which extent the assumption that voters indeed reliably verify holds. In our investigation, all votes were correct, hence no errors could be detected. As shown by previous work, voters might not detect errors [26], [35]. Consequently, future work should carefully investigate to which extent this assumption is true, what is needed such that voters indeed verify, and whether voters indeed detect incorrect votes using hybrid online voting.

5. Results

The study lasted on average 20 minutes; participants needed, on average, five minutes to cast their vote. Considering voting, those who interacted with the baseline needed on average three minutes, the paper audit trail needed four minutes, and the 3D-printed tokens eight minutes. Detailed descriptive statistics (including SD and 95% CIs) are provided in Appendix A.3.

5.1. RQ1: Perceived Security and Trust

In this section, we detail our results on security and trust perceptions to answer RQ1 (*How does hybrid online voting impact subjective trust and security perceptions of voters?*).

Security Perceptions (SRS [41]): The SRS assesses whether users are psychologically prepared and willing to use security mechanisms on an affective, cognitive, and behavioural level [41]. Figure 2b) depicts the results of the individual levels. We calculated Kruskal-Wallis tests and found no statistically significant differences for cognitive ($\chi^2(2)$ =3.37, p = .185) and behavioural ($\chi^2(2)$ =2.63, p = .268), while the differences in affective security readiness were not significant ($\chi^2(2)$ =5.98, p = .050).

Participants were also specifically asked in the final questionnaire (Q6, see Figure 5) whether they consider vote secrecy to be protected. We analysed the results by Welch's ANOVA and revealed significant differences ($F(2, 94.242) = 3.838$, $p = .025$, η^2 =.043) The *post-hoc* tests showed a significant differences between the baseline and tokens ($p = .039$); the token audit trail was perceived as more vote-secrecy-preserving. Participant comments in the free-text answers reflect these results. P128 positively commented on vote secrecy given by the token: *"I really enjoyed the simplicity of the internet voting system. I also found the broadcast of the printing of the voting item very useful as it reassured me that my vote is safe and private and it cannot be faked."*, P128 (token). Analysis of the free-text answers that mention security aspects, in general, indicated that participants who interacted with the baseline primarily wrote negative comments about security:

"I would like a clear explanation of the operation of electronic elections, how votes are handled and counted, and making it clear how data is protected and guarantees the security of your vote.", P39 (baseline) or *"It is easy to cast a vote online but it is also risky as people can hack the system and add in fake votes which is not safe either."*, P38 (baseline). Security-related comments from participants who interacted with an audit trail condition were shifted towards other more general security-related aspects that should be considered in any voting system. P69 (paper) commented on authentication: *"I'm not sure how this voting system would check my personal information that is actually needed in political elections, i.e. if I can even take a part in election, which caused my concerns about safety."* P82 (paper) suggested a specific authentication: *"I would like to log in with my bank account to the voting system."*

Even more generic, in *all* conditions, participants equally expressed the fear of hackers that could break into any online voting system:

"My biggest concern with this system is being susceptible to hacking attacks. Although election frauds may also happen with regular voting systems, I am still not 100% sure about an online method.", P5 (baseline).

"[...] the system developers need to be wary of hackers and therefore ensure that they install strong security on the system.", P142 (token). *"I think anything on the internet can be hacked so I don't think it's secure. I would be concerned about that."*, P74 (paper).

Perceived Trust (HCTM [43]): The HCTM scale considers four subscales of perceived risk (inverted items 1-3), benevolence (items 4-6), competence (items 7-9), and trust (items 10-12). First, we calculated each subscale per participant by adding all items and dividing them by three. Each subscale ranges from 0 to 5. Figure 2a) depicts the four subscales.

We analysed each scale with a one-way ANOVA. The perceived risk of the baseline was highest with $M = 3.02$ ($SD = 0.96$), followed by the paper with $M = 2.76$ ($SD = 0.87$). The perceived risk of token audit trails was lowest with $M = 2.62$ ($SD = 0.80$). Differences between perceived risk were not found to be significant ($F(2, 147) = 2.652$, $p = .074$). The benevolence of the baseline was again rated highest with $M = 3.73$ ($SD = 0.66$), while the paper received on average 3.66 ($SD = 0.75$) and token 3.71 ($SD = 0.74$). These differences were not significant ($F(2, 147) = 1.404$, $p = .249$).

Considering competence, the tokens were rated highest ($M = 3.90$, $SD = 0.95$), followed by the baseline ($M = 3.71$, $SD = 0.75$) and paper (M=3.71, SD=0.88). Again, we found no significant differences ($F(2, 147) = 0.591$, $p = .555$). Considering the subscale trust, paper was rated highest with $M = 3.66$ ($SD = 0.95$), followed by tokens (M=3.54, SD=1.01) and the baseline ($M = 3.20$, $SD = 0.95$). Here, we found statistically significant differences ($F(2, 147) = 0.591$, $p = .043$, $\eta^2 = .039$). The *post hoc* analysis revealed that paper was found to be more trustworthy than the baseline ($t(97) = 2.019$, $p = .047$, Cohen's $d = .406$)

A number of participants mentioned trust in the free-text answers. Participants in baseline condition mentioned that missing verifiability options weakened their trust:

"Regarding the security of the voters, this needs to be stated a little more reassuring to the voter but I am quite confident about it already. Extra precautions should be taken to ensure people's confidence in the site.", P16 (baseline).

"I would love internet voting over any other but it's impossible to know what the program does and actually saves as answer.", P17 (baseline).

On the one hand, participants in the conditions with audit trails welcomed the video broadcast since it contributed to their trust:

"I also found the broadcast of the printing of the voting item very useful as it reassured me that my vote is safe and private and it cannot be faked.", P126 (token)

"The video watching should not be a compulsory aspect of the online voting, however, when I witness it being inserted in the box, it did being a sense of relief. Therefore, equally I can understand why it may be useful watching the video.", P109 (token).

"It worked well. Simple to use. The barcode helps ease the tension of maybe it is not my paper.", P62 (paper).

On the other hand, the printing of the tokens was criticised based on the duration and composition of the video:

"Everything before the printer section was easy to use and seemed secure enough. The printing part threw my confidence of the system into question. I don't know how long it takes to make a 3D printed item, but it seemed to take a long time (I also started to think it might just be looping 20 seconds of video over and over), and having a human take it at the end and put it in a box also made me wary. I would have rather seen a fully automated system from creation to putting it in the ballot box.", P126 (token).

"I did not have a sense of how long the video was going to take. When I re-read the instructions and realised that the token would be put in the ballot box I then trusted that this would happen." P111 (token).

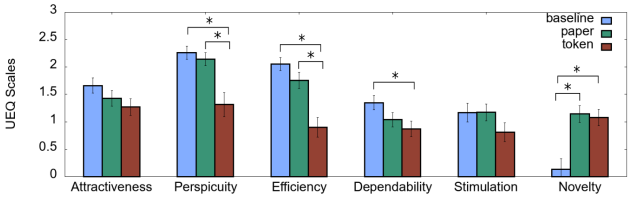


Figure 3: User Experience Scales. The asterisk * indicates statistically significant differences. The error bars indicate the standard deviation.

5.2. RQ2: Usability and User Experience

The results detailed in this section serve as a basis to answer RQ2 (*How do potential voters perceive hybrid online voting in terms of usability and user experience?*).

User Experience (UEQ [46]): The UEQ assesses the scales (1) attractiveness, (2) perspicuity, (3) efficiency, (4) dependability, (5) stimulation, and (6) novelty. Descriptive statistics of the six scales are depicted in Figure 3. Overall, (Welch's) ANOVA did not reveal statistically significant differences for the scales attractiveness, which refers to the overall impression of a system ($F(2, 147) = 1.740$, $p = .179$), and stimulation, which refers to excitement and motivation to use the system ($F(2, 147) = 1.542$, $p = .217$).

For all other scales, we found statistically significant differences, namely perspicuity ($F(2, 93.791) = 7.213$, $p = .001$, $\eta^2 = .121$), efficiency ($F(2, 147) = 13.121$, $p < .001$, $\eta^2 = .168$), dependability ($F(2, 147) = 3.122$, $p = .047$, $\eta^2 = .041$), and novelty ($F(2, 147) = 10.872$, $p < .001$, $\eta^2 = .129$).

Perspicuity refers to the ease of learning to use the system. In the Bonferroni-Holm-corrected *post-hoc* analysis of perspicuity, the tokens were perceived as more difficult to learn than the baseline ($t(100) = 3.842$, $p < .001$, Cohen's $d = .761$). Further, paper audit trails were perceived to be easier to learn than tokens ($t(97) = 3.384$, $p = .001$, Cohen's $d = .681$). This aspect is also reflected in the free-text answers: *"The voting system was simple and fast for me. I understood it quickly."*, P42 (baseline).

We found similar results for efficiency, i.e. the perceived effort to solve a task with a system. The baseline of no audit trail was perceived as significantly faster than tokens ($t(100) = 5.284$, $p < .001$, Cohen's $d = 1.04$). Paper audit trails were perceived as more efficient than tokens ($t(97) = 3.912$, $p < .001$, Cohen's $d = .787$).

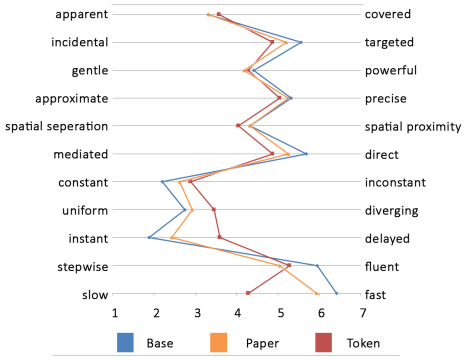


Figure 4: Interaction profiles from the interaction vocabulary.

The duration of printing the token was mentioned by a number of participants in the free-text answers, some participants even suggested using paper instead of 3D-printing: *"The live broadcast video was too long. Not too many people would bother to watch it until the end."*, P149 (token) or *"It was a very reliable system. The transmission of the printing of the ballot works, but I think that with a conventional printer it would be faster."*, P151 (token). Dependability reflects the degree to which users feel in control when interacting with a system. The tokens were perceived to offer less control compared to the baseline ($t(100) = 2.349$, $p = .045$, Cohen's $d = .465$). In the free-text answers, participants using tokens expressed this aspect, such as P101: *"Although there is the item that helps tallying the votes, it doesn't mean that votes can't be manipulated in some way, at least in my perception."* This indicates that tokens may not be perceived as a good solution for ballot representation by voters. Finally, when considering novelty, both tokens ($t(100) = -3.713$, $p<.001$, Cohen's $d = -.735$) and paper ($t(97) = -4.013$, $p<.001$, Cohen's $d = -.807$) were perceived as significantly more novel than the baseline.

Interaction Vocabulary (IV [47]): The IV assesses how users perceive the interaction with a system or product on a descriptive, non-judgemental level. The interaction is described in 11 interaction characteristics using semantic opposites such as "apparent" vs. "covered". The interaction profiles for all conditions are depicted by Figure 4. A Kruskall-Wallis test revealed significant differences in six interaction characteristics ($p<.05$). Full statistical test outputs can be found in the Appendix A.2. Pairwise comparisons using Mann-Whitney-U-tests showed that the baseline was considered to be significantly faster compared to the token (i.e., more efficient; $Z = -5.60$, $p<.001$, $r^2 = .209$) and the paper audit trail ($Z = -2.32$, $p = .02$, $r^2 = .036$), which matches our results from the user experience questionnaire. Furthermore, the baseline was considered to be more fluent (this refers to a fluent integration in the running process, which allows users to exert continuous influence on the system; $Z = -2.57$, $p = .010$, $r^2 = .044$), instant (this refers to instant feedback on user actions, which allows users to experience their impact and increases their feelings of security and competence; $Z = -4.70$, $p<.001$, $r^2 = .147$), constant (this denotes system reliability, which fosters user adaptation and feeling of security; $Z = -2.69$, $p = .007$, $r^2 = .048$), direct (this refers to experiencing the consequences of interacting with the system directly instead of mediated, which fosters a feeling of constant control; $Z = 2.94$, $p = .003$, $r^2 = .058$), and targeted (this refers to a non-incidental interaction with the system, in which users are highly concentrated on the interaction, which highlights the significance of the interaction process; $Z = -2.57$, $p = .010$, $r^2 = .044$) compared to the token audit trail; while it was considered to be significantly more fluent ($Z = -3.19$, $p = .001$, $r^2 = .068$) and instant ($Z = -2.64$, $p = .008$, $r^2 = .046$) compared to the paper audit trail. The two audit trail conditions differed significantly in terms of delay, i.e., the paper condition was considered to be significantly more instant ($Z = -2.99$, $p = .003$, $r^2 = .06$) and faster ($Z = -4.11$, $p<.001$, $r^2 = .113$).

Perceived Usability (SUS [44]): Perceived Usability was assessed by the SUS score (range between 0 and 100). The baseline received the highest average (78.86, min=45, max=100, SD=13.38). This is followed by the paper audit trails (78.26, min=33, max=100, SD=13.64) and tokens (69.44, min=15, max=100, SD=16.38). The scores of roughly 78 points refer to a "good" perceived usability or "C" on the grade scale while the 69 refer to a "D" [49].

A one-way ANOVA revealed significant differences between the three conditions ($F(2, 147) = 6.582$, $p = .002$, $\eta^2 = .082$). The post-hoc analysis revealed significant differences between the SUS score of the baseline and the tokens ($t(100) = 3.180$, $p = .004$, Cohen's $d = .93$). Also, the paper was rated to be significantly better than the tokens ($t(97) = 0.516$, $p = .009$, Cohen's $d = .104$). The differences between the paper and the baseline ($p = 1.000$) were not significant.

Comments from the free-text answers that consider usability mainly mentioned that the baseline and paper conditions were considered easy to use. P38 commented on the baseline saying: *"I think it was very easy to use."* P61 commented on the usability of the paper condition: *"The instructions given were well written, easy to understand, and clear. I changed my vote a few times to check if the voting system allows me to do so and it did, I'm impressed. The double checking to ensure that the user has made the correct vote was a brilliant touch."*

Participants in the token condition also commented on the broadcast, such as P137: *"The printing broadcast was unnecessarily lengthy, and it was not clear when it was finished. I re-started the survey as I thought the video was broken, and I was stuck on the page. Very confusing."*

5.3. Usage Intention

To capture the participants' usage intention of the three conditions, we asked eight questions that participants an-

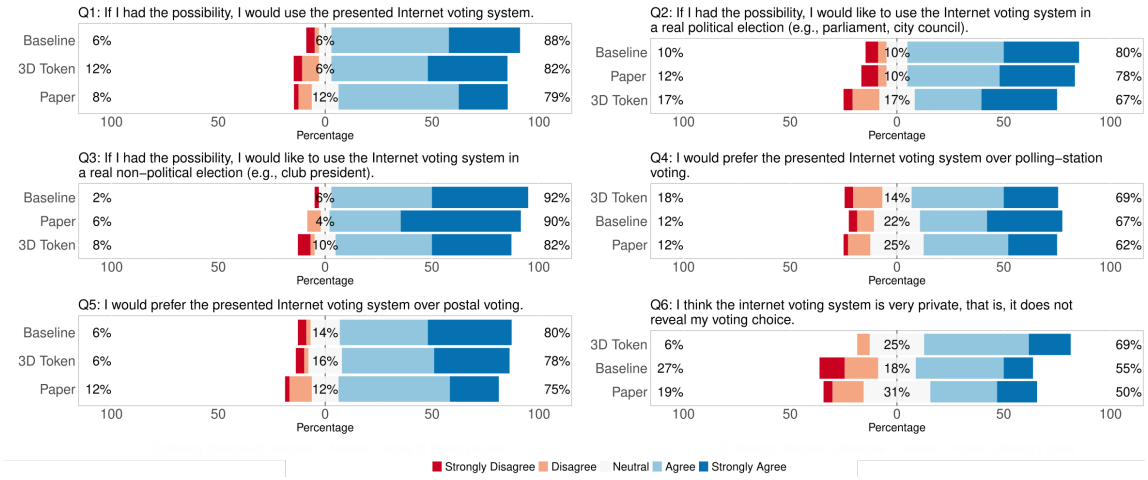


Figure 5: Likert items from the final questionnaire. The questions we asked the participants are provided above each plot. The asterisk * indicates statistically significant differences in Q6 and Q7.

swered on a 5-point Likert scale. Figure 5 provides an overview of the results to these questions. First, we asked them if they would like to use the presented online voting system in general (Q1). Across all conditions, participants agreed to use the presented system in a real election ($M = 4.03$, $SD = 0.955$). The analysis did not reveal significant differences between the conditions ($F(2, 147) = 0.551$, $p = .578$).

Next, we captured the usage intention considering political elections (e.g., for parliament, city council) (Q2) while the third question (Q3) considered non-political elections (e.g., club president). Both questions were answered similarly to the overall question with an average of 3.92 ($SD = 1.13$) for political elections, and 4.26 ($SD = 0.90$) for non-political elections. We could not find significant differences for the political elections ($F(2, 147) = 0.349$, $p = .706$) and for the non-political ones ($F(2, 147) = 2.014$, $p = .137$).

Participants were then asked whether they preferred the presented online voting system over in-person and postal voting (Q4). Overall, the participants tend to slightly favour online voting over in-person voting for all conditions. The baseline was rated with $M = 3.35$ ($SD = 1.03$), paper with $M = 3.69$ ($SD = 0.83$), and the tokens with $M = 3.35$ ($SD = 1.02$). However, differences in the preference were not significant between the conditions ($F(2, 147) = 0.308$, $p = .736$). We observed a similar trend for postal voting (Q5). The baseline was rated with $M = 3.29$ ($SD = 1.15$), paper with $M = 3.59$ ($SD = 1.00$), and the tokens with $M = 3.25$ ($SD = 1.04$). While the participants overall slightly tended to favour online voting over postal voting, we could not determine significant differences between the conditions ($F(2, 147) = 0.987$, $p = .375$).

A number of participants gave similar comments regarding their usage intention in all conditions showing that participants interested in online voting in general might not make differences based on security properties:

"I found it very simple and convenient to use, I would definitely choose to use this style of voting over postal in the future.", P148 (token) or *"I think it's very useful as you can save a lot of time. You don't need to drive to the voting station, neither do you have to go to a postal office to send your letter."*, P34 (baseline).However, some participants would like to limit the usage of online voting to low-stake elections. This is again not linked to our specific concept, but to the attitude towards online voting in general. E.g., P71 (paper) wrote: *"I think that the internet voting system could be very useful with small scale elections. Due to recent controversy around the world about voting protections and election interference, I do worry about trusting this system with an election as serious as one that determines a country's leader."*, P71 (paper).Finally, we asked condition-specific questions in the audit trail conditions, namely whether participants would indeed observe the broadcast in an actual election (Q7). Participants in the paper condition tended to agree with an average of 3.96 ($SD = 1.02$). Participants in the token condition tended to neither agree nor disagree with, on average, 3.05 ($SD = 1.28$). A t-test revealed significant differences between paper and token ($t(99) = 32.200$, $p<.001$, Cohen's $d = 1.17$). However, when asked whether they consider the broadcast to be unnecessary (Q8), participants in both audit trail conditions tended to disagree with on average 2.57 ($SD = 1.19$) for paper and 2.64 ($SD = 1.21$) for tokens. These differences were not significant ($t(99) = 21.809$, $p = .754$).

6. Discussion and Future Work

In this section, we first discuss the *results of the online study*. Since the study focused on interaction aspects, we detail *challenges* for implementing hybrid online voting. We further outline *opportunities* of hybrid online voting and how it could improve online voting. We discuss the *limitations* of our investigation and motivate *future work*. Finally, we conclude with *final recommendations* for researchers, designers, policymakers and practitioners.

6.1. User Study Results

This section discusses *broadcast perceptions*, *the security perceptions of voters*, *verifiability and trust* throughout this section, we also discuss *limitations* of our investigation.

Broadcast Perceptions & Limitations: We investigated online voting with neither audit trail nor verification in comparison to hybrid online voting. According to prior studies, the baseline reflects how most voters intuitively expect online voting to be (cf. [28], [50]). Early implementations of online voting used in Estonia were similar to that before individual verifiability was introduced in 2015 [32]. Hence, it is not surprising that some participants considered the broadcast to be unnecessary.

In our study, the token broadcast was considered redundant since printing took very long. Accordingly, our participants perceived the tokens as slower and less efficient than baseline and paper. 3D printing is a technology that most participants were unfamiliar with. Two participants even feared that we just show a looped video. Hence, better feedback about the printing progress might be necessary if 3D-printed tokens are used, e.g., by a progress bar, adjusting the camera angle, or providing multiple angles. We used a top-down view which is limited in conveying the token's height. While the interaction with paper ballots was, on average, one minute longer compared to the baseline, it was not perceived as less efficient indicating that the duration of the printing process has to be considered when implementing hybrid online voting.

Based on the perceptions of the printing process and further issues in understanding how the token works, we conclude that paper would be a better choice than tokens for physical audit trails. Using paper solves many issues associated with the tokens and was even suggested by participants in the token condition. Further, paper is something well-known in the context of elections and more cost-efficient.

For our implementation, we used video broadcasts as an intuitive and easy-to-understand way to offer verifiability. However, there might be alternatives to this that are more efficient than watching a video that should be explored in future work. Another essential task of future work is determining what should actually be seen in a broadcast. We opted to show the printing and insertion into the ballot box as a non-interrupted experience during which voters do not have to wait. However, one might argue that observing the insertion into the ballot box is sufficient. Another limitation of the broadcast in our study is the usage of pre-recorded videos to make sure that voters have an uninterrupted experience without failure. All votes cast in the study were error-free. The core purpose of verification, however, is discovering and reporting incorrect votes. Consequently, a crucial next step is investigating (a) whether voters indeed verify by watching the broadcast and (b) the detection of incorrect votes ideally using a live broadcast instead of pre-recorded videos.

Security Perceptions: Overall, we could not determine statistically significant differences between the conditions. Hence, there is no evidence that the introduction of audit trails will decrease security perceptions. Similar to other comparative studies [28], participants seemed to focus on overall risks that might be present in *any* online voting system, such as the risk of hacking. While hybrid online voting aims to address such risks, not all participants considered it beneficial to security. When analysing the free-text answers, we observed that participants who interacted with the baseline expressed more generic concerns and concerns related to missing assurance. Interestingly, the tokens were considered to be more vote-secrecy-preserving than the baseline which might be linked to their rather non-intuitive design of encoding. Our research shows that voters must be informed about the security properties of the voting system. While we did not provide such information in the voting software during the study, related work recommended specific information practices that authorities should follow to properly inform voters [42], [51].

Verifiability and Trust: Some participants in the baseline condition missed assurance about the whereabouts of their ballots. The audit trail conditions aimed at providing such assurance, yet some participants were not convinced by the broadcast functionality, especially in the token condition. This might be linked to the complexity of the token.

Paper audit trails were perceived to be significantly more trustworthy compared to the baseline. While we used a generic high-stake election to investigate a large and diverse group of participants, the specific election might impact the voters' perceptions of a voting system [48].

Trust is highly individual [52] meaning that there is no one-size-fits-all solution that each and every voter will trust equally. The construct trust can be decomposed into dispositional, learned, and situational trust [52]. Dispositional trust is a person's tendency to trust based on their psychological characteristics, it has been linked to the overall willingness to use online voting [53]. Learned trust is based on experience. It was shown that people who have experience with online voting are likely to use it again [54]. Finally, situational trust is based on the specific circumstances of interaction. Consequently, the HCTM scale [43] that we used in our study focuses on situational trust. Participants in our study had no experience with online voting but all of them had experience with in-person voting. Consequently, they might base overall trust perceptions of online voting on media reports that mainly focus on risks of specific online voting technology [55]. As a consequence, the level of trust imposed by the participants in our study might not match reality.

Furthermore, researchers have argued that in case online voting is introduced, policymakers and authorities should provide additional resources to voters, such as informative materials [42], [51], such that they can determine their trust towards the used technology over time [56]. Considering the prior research on trust explained above, our study reflected a scenario in which online voting is introduced as new a voting channel for the first time.

Regarding the overall scores of the HCTM scale, participants only expressed medium trust levels. Nevertheless, our results show that paper audit trails, which are closest to the solutions that people know from their daily lives, are preferred over novel solutions. Hence, if online voting is introduced as a new vote-casting channel, this preference for familiarity should be considered.

6.2. Challenges of Hybrid Online Voting

Even though hybrid online voting combines benefits from the analogue and digital worlds, it also introduces new security challenges. The security considerations provided by us are by no means exhaustive, demanding the need for a formal voting protocol to secure hybrid online voting. Future work should focus on developing and formally proving such a security protocol. Yet, our work serves as a first exploratory evaluation of the voters' perspective. In the remainder of this section, we discuss security-related aspects, understandability for voters and poll workers, a ballot design for hybrid online voting, and scalability aspects. In each section, we use the discussion to also motivate future research on hybrid online voting.

Secure Facilities Might be Difficult. Realising hybrid online voting requires specific secure facilities to be available in the physical world for setting up the printers. Similar facilities are already in use in several countries to store postal votes until tallying. Since such facilities are already available, e.g., in administration buildings, those could be used for hybrid online voting. However, setting up the printers comes with additional requirements compared to storing postal votes. Requirements for such facilities in terms of technology, infrastructure and human resources form an integral task for future work that specifically considers the realisation of hybrid online voting. For instance, the printers need an energy supply. Since the printers use normal power outlets, this is likely available. However, it must be made sure that the printers indeed run and that the votes are inserted into the ballot box which in turn might require the presence of maintenance personell. In our study, a poll worker inserted the vote into the ballot box. If this is implemented, poll workers with training to administer the printers who work in shifts could insert the ballots. This would require financial resources to pay the poll workers. Furthermore, even if the printers run and deliver service, the broadcasts form another point of failure either by not running properly or by being hijacked by an attacker. Consequently, broadcast availability and security need to be assured.

Vote Secrecy Might Impact Verifiability. One specific requirement for hybrid online voting is individual verifiability. Depending on the verification scheme, there is a tension between vote secrecy and individual verifiability. Using our study prototype, participants could verify that their vote is printed by a tracking code. However, it is more challenging to verify that the printed vote indeed encodes the voter's actual choice. We tried showing optical codes to the voters, but those were not well recognised. There are several possibilities to address this issue. First, similar to the EasyVote system [20], the voting choice could be printed in clear text in addition to the encoding. In doing so, the voters could visually verify that the printed choice matches their voting intention. This solution, however, comes at the price of vote secrecy. Since the broadcast is universally observable, anyone could make a protocol of tracking codes and voting options. Such a protocol would allow for intermediate results which might bias voters and hence maliciously influence the election outcome. Based on that, the voting choice cannot be included in clear text on the printed ballots.

A possibility to offer verifiability without violating vote secrecy is to use a code voting protocol [57], [58]. Before the election, voters receive an individual list of voting options and corresponding voting codes via postal mail. To cast a vote, the voters enter the voting code that belongs to the candidate of their choice. This has two advantages: First, malware on the voter's computer cannot eavesdrop on the voting choice. Second, the voting code can be printed on the ballot representation and verified by the voter. Since the vote codes are different for each voter, vote privacy cannot be broken. Code voting has been investigated in studies already that demonstrated its usability [38], [59]. However, code voting has the drawback of resulting in a more complicated voting system. While it has been shown that voters are willing to sacrifice usability for the sake of security [59], this has yet to be investigated for hybrid online voting.

The verifiability features implemented in our prototype do not target voters with visual impairments. Such voters could be assisted by an app that localises the QR code or token in the video and verbally expresses the content to them. While this is one option to assist such voters, future work should work on audit trail verification that also works for voters with visual impairments. Finally, in our study, all votes were correct meaning that voters could observe incorrect votes being inserted in the ballot box. Similar to other verifiable online voting schemes (see Section 2), our concept relies on the assumption that voters indeed verify. Yet, several investigations (cf. [26], [35], [60]) showed that this is not always the case. While we followed guidelines from related work by integrating instructions, it is crucial to further investigate this assumption.

Understandability for Voters & Poll Workers. With hybrid online voting, we aimed to propose a concept that is easy-to-understand and easy-to-use for both voters and poll workers. In our investigation, we focused on the voters' perspective. There are several tasks executed by poll workers during the electoral process. First, poll workers could be responsible for observing the printers and making sure that there are no technical issues. This task is similar to administering DREs in polling stations. Next, poll workers might insert the ballot into the ballot box, however, this process could be automated. One participant in our study mentioned that the presence of a human negatively impacted their trust. Future work should determine how the insertion

task should be done. Finally, there are several ways to conduct the tally. By tallying the stored electronic votes, the physical representations serve as an audit trail. The poll workers perform a risk-limiting audit by tallying a random subset of the physical representations indicating whether a tally of all physical representations is needed. This is similar to using a print-out DRE in in-person voting. This process is more efficient compared to hand-counting ballots and allows a faster announcement of the election result. Another option is tallying the physical representations which has the benefit that the tally is easier-to-understand. Once created, the physical representations can no longer be compromised by an attacker. This, however, impacts the tally duration and could delay the announcement of the election result.

Scalability Considerations. In this paper, we evaluated hybrid online voting from an HCI perspective. Voters in our study immediately saw the broadcast without delay. In an actual election, multiple voters may want to cast their votes at the same time. which might cause printing resources to be exhausted, meaning voters have to wait. While this aspect can be improved by using multiple printers in one facility, it is unlikely to have a sufficient number of printers for the entire voting period. When visiting a polling station, people simply wait in line until it is their turn to vote. Similar queuing systems can be implemented for hybrid online voting. For this, it is crucial to consider the placements of the queue within the voting process. The queue could be after submitting the vote. In addition to the broadcast, voters should also receive feedback about the printing queue and an estimation of when their ballot will be printed. This might lead voters to abandon the broadcast and to not verify their vote. Another option is similar to in-person voting; the queue would be before vote casting. Again, voters should receive a time estimate when it is their turn to vote. Both queuing options should be investigated by future work to determine optimal queuing time and further aspects to improve scalability. Another option to improve scalability is printing several tokens on one 3D printer. This also complicates and, hence, partly addresses the timing attack on privacy mentioned above, because it is harder to link broadcast events to individuals. Since multiple prints at once impact the printing duration, determining an optimal number of prints forms an important task for future work.

6.3. Final Recommendations

When considering our study results, the question arises of whether we should use audit trails in the first place. Overall, the security benefits of audit trails are obvious: there are physical items that once created can no longer be manipulated. Yet, they add complexity to the system and new security challenges as outlined above. Overall, we argue for using a paper audit trail if possible due to the trust enhancements, and security benefits and participants that interacted with the paper condition did not experience significant UX shortcomings. However, a more in-depth investigation comparing hybrid voting against fully digital solutions in terms of human factors but also security is needed to fully answer the question of whether we should use audit trails for online voting in general. Based on our investigation and the discussion above, we conclude with five core takeaways that need to be considered when implementing hybrid online voting.

1) Verification does not automatically lead to trust: Some participants in our study mentioned they trust that the infrastructure functions correctly no matter which system they use since this is an official system used for an election. Further, the primary task of a voter is casting a vote.
2) Audit trails should be close to known solutions: Paper-based audit trails were perceived similarly as the baseline and rated as more trustworthy. This is linked to the participants' familiarity with paper-based procedures.
3) Automation might impact trust: The presence of poll workers in the broadcast might impact the voters' trust. Hence, the benefits of human presence should be weighed against those of a fully automated system.
4) Duration of printing is crucial: Participants considered 3D printing as too time-consuming. Watching the paper printer was considered efficient. Hence, we recommend keeping the duration of printing to a minimum to ensure that a large share of voters watch the broadcast. Further, it is important to communicate the printing progress to voters.
5) Simplicity can impact vote-secrecy perceptions: The tokens were perceived as most secrecy-preserving but also most complicated. Hence, authorities might want to provide additional information on why a simple system is still preserving vote-secrecy.

7. Conclusion

This paper presented a hybrid online voting which allows vote casting over the internet. Cast votes are printed and inserted into a physical ballot box in a secure facility. We investigated two kinds of physical ballots - paper and 3D-printed tokens - as well as a baseline in an online study with 150 participants. Based on our investigation, we show that hybrid online voting has to potential to enhance voter trust. Further, the provision of an audit trail did not impact the user experience – neither positively nor negatively. Yet, not all physical representations perform equally well. Paper, as it is time-efficient and already known by individuals, should be used as a physical ballot. We conclude the paper by outlining challenges introduced by audit trails and discussing core considerations for the future implementation of hybrid online voting.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2092 CASA - 390781972 and grant number 251805230/GRK 2050.

References

[1] F. Bastien, R. Koop, T. A. Small, T. Giasson, and H. Jansen, "The role of online technologies and digital skills in the political participation of citizens with disabilities," *Journal of Information Technology & Politics*, vol. 17, no. 3, pp. 218–231, 2020.

[2] B. Adida, "Advances in Cryptographic Voting Systems," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.

[3] D. A. Gritzalis, "Principles and requirements for a secure e-voting system," *Computers & Security*, vol. 21, no. 6, pp. 539–556, 2002.

[4] S. Heiberg, P. Laud, and J. Willemson, "The application of i-voting for estonian parliamentary elections of 2011," in *Proceedings of the International Conference on E-Voting and Identity (VoteID)*. Springer, 2011, pp. 208–223.

[5] C. Z. Acemyan, P. Kortum, M. D. Byrne, and D. S. Wallach, "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*, vol. 2, no. 3, pp. 26–56, 2014.

[6] T. Milic, M. McArdle, and U. Serdült, "Analysis of the antrim county, Michigan November 2020 election incident," 2021. [Online]. Available: https://content.govdelivery.com/attachments/MISOS/2021/03/26/file_attachments/1736734/Antrim.pdf

[7] V. Distler, M.-L. Zollinger, C. Lallemand, P. B. Roenne, P. Y. A. Ryan, and V. Koenig, "Security - visible, yet unseen?" in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: ACM, 2019, pp. 605:1–605:13.

[8] T. Milic, M. McArdle, and U. Serdült, "Attitudes of Swiss citizens towards the generalisation of e-voting," 2016. [Online]. Available: https://doi.org/10.5167/uzh-127938

[9] M. Warkentin, S. Sharma, D. Gefen, G. M. Rose, and P. Pavlou, "Social identity and trust in internet-based voting adoption," *Government Information Quarterly*, vol. 35, no. 2, pp. 195–209, 2018.

[10] L. Nestas and K. Hole, "Building and maintaining trust in internet voting," *Computer*, vol. 45, no. 5, pp. 74–80, 2012.

[11] B. Randell and P. Y. A. Ryan, "Voting technologies and trust," *IEEE Security & Privacy*, vol. 4, no. 5, pp. 50–56, 2006.

[12] R. G. Saltman, "Accuracy, integrity and security in computerized vote-tallying," *Commun. ACM*, vol. 31, no. 10, p. 1184–1191, Oct. 1988.

[13] M. Lindeman and P. B. Stark, "A gentle introduction to risk-limiting audits," *IEEE Security & Privacy*, vol. 10, no. 5, pp. 42–49, 2012.

[14] K. K. Greene, M. D. Byrne, and S. P. Everett, "A comparison of usability between voting methods," in *Proceedings of the Electronic Voting Technology Workshop (EVT)*. Berkeley, CA, USA: USENIX Association, 2006.

[15] F. G. Conrad, B. B. Bederson, B. Lewis, E. Peytcheva, M. W. Traugott, M. J. Hanmer, P. S. Herrnson, and R. G. Niemi, "Electronic Voting Eliminates Hanging Chads but Introduces New Usability Challenges," *International Journal of Human-Computer Studies*, vol. 67, no. 1, pp. 111–124, 2009.

[16] S. P. Everett, M. D. Byrne, and K. K. Greene, "Measuring the usability of paper ballots: Efficiency, effectiveness, and satisfaction," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (HFES)*, vol. 50, no. 24. SAGE Publications Sage CA: Los Angeles, CA, 2006, pp. 2547–2551.

[17] A. J. Feldman, J. A. Halderman, and E. W. Felten, "Security analysis of the Diebold AccuVote-TS voting machine," 2006.

[18] J. Hsu and G. Bronson, "E-voting technologies usability: A critical element for enabling successful elections," in *Emerging Challenges in Business, Optimization, Technology, and Industry*. Cham, Switzerland: Springer, 2018, pp. 61–78.

[19] S. N. Goggin and M. D. Byrne, "An examination of the auditability of voter verified paper audit trail (VVPAT) ballots." *EVT*, vol. 7, pp. 10–10, 2007.

[20] J. Budurushi, "Usable security evaluation of EasyVote in the context of complex elections," Ph.D. dissertation, Technische Universität Darmstadt, 2016.

[21] D. Chaum, A. Essex, R. Carback, J. Clark, S. Popoveniuc, A. Sherman, and P. Vora, "Scantegrity: End-to-end voter-verifiable optical-scan voting," *IEEE Security & Privacy*, vol. 6, no. 3, pp. 40–46, 2008.

[22] D. Chaum, R. Carback, J. Clark, A. Essex, S. Popoveniuc, R. L. Rivest, P. Y. Ryan, E. Shen, and A. T. Sherman, "Scantegrity ii: End-to-end verifiability for optical scan election systems using invisible ink confirmation codes." *Proceedings of the Electronic Voting Technology Workshop EVT*, vol. 8, pp. 1–13, 2008.

[23] S. P. Everett, K. K. Greene, M. D. Byrne, D. S. Wallach, K. Derr, D. Sandler, and T. Torous, "Electronic voting machines versus traditional methods: Improved preference, similar performance," in *Proceedings of the Conference on Human Factors in Computing Systems (SIGCHI)*. New York, NY, USA: ACM, 2008, pp. 883–892.

[24] O. Spycher, R. Haenni, and E. Dubuis, "Coercion-resistant hybrid voting systems," 2010. [Online]. Available: https://arbor.bfh.ch/8292/1/SHD10.pdf

[25] O. Spycher and R. Haenni, "A novel protocol to allow revocation of votes a hybrid voting system," in *2010 Information Security for South Africa*. IEEE, 2010, pp. 1–8.

[26] K. Marky, M.-L. Zollinger, P. B. Roenne, P. Y. Ryan, T. Grube, and K. Kunze, "Investigating usability and user experience of individually verifiable internet voting schemes," *ACM Transactions on Computer-Human Interaction*, vol. 28, no. 5, 2021.

[27] J. Benaloh, "Simple verifiable elections," in *Proceedings of the Electronic Voting Technology Workshop (EVT)*. Berkeley, CA, USA: USENIX Association, 2006.

[28] K. Marky, O. Kulyk, K. Renaud, and M. Volkamer, "What did I really vote for? On the usability of verifiable e-voting schemes," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: ACM, 2018, pp. 176:1–176:13.

[29] P. Y. A. Ryan, P. B. Rønne, and V. Iovino, "Selene: Voting with transparent verifiability and coercion-mitigation," in *Proceedings of the International Conference on Financial Cryptography and Data Security (FC)*. Cham, Switzerland: Springer, 2016, pp. 176–192.

[30] M.-L. Zollinger, V. Distler, P. B. Roenne, P. Y. A. Ryan, C. Lallemand, and V. Koenig, "User experience design for e-voting: How mental models align with security mechanisms," in *Proceedings of the International Joint Conference on Electronic Voting (E-Vote-ID)*. TalTech, 2019, pp. 187–202.

[31] M.-L. Zollinger, E. Estaji, P. Y. Ryan, and K. Marky, ""just for the sake of transparency": Exploring voter mental models of verifiability," in *International Joint Conference on Electronic Voting (E-Vote-ID)*. Springer, 2021, pp. 155–170.

[32] S. Heiberg and J. Willemson, "Verifiable internet voting in Estonia," in *Proceedings of the 6th International Conference on Electronic Voting, Verifying the Vote (EVOTE)*. Piscataway, NJ, USA: IEEE, 2014, pp. 1–8.

[33] S. Heiberg, A. Parsovs, and J. Willemson, "Log analysis of estonian internet voting 2013–2014," in *Proceedings of the International Conference on E-Voting and Identity (Vote-ID)*. Springer, 2015, pp. 19–34.

[34] D. Galindo, S. Guasch, and J. Puiggalí, "2015 Neuchâtel's cast-as-intended verification mechanism," in *Proceedings of the International Conference on E-Voting and Identity (VoteID)*. Cham, Switzerland: Springer, 2015, pp. 3–18.

[35] M. Bernhard, A. McDonald, H. Meng, J. Hwa, N. Bajaj, K. Chang, and J. Halderman, "Can voters detect malicious manipulation of ballot marking devices?" in *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*. Los Alamitos, CA, USA: IEEE Computer Society, may 2020, pp. 1402–1417.

[36] K. S. Fuglerud and T. H. Røssvoll, "An evaluation of web-based voting usability and accessibility," *Universal Access in the Information Society*, vol. 11, no. 4, pp. 359–373, 2012.

[37] K. Marky, V. Zimmermann, M. Funk, J. Daubert, K. Bleck, and M. Mühlhäuser, "Improving the usability and ux of the swiss internet voting interface," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13.

[38] K. Marky, M. Schmitz, F. Lange, and M. Mühlhäuser, "Usability of code voting modalities," in *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA)*. New York, NY, USA: Association for Computing Machinery, 2019, p. 1–6.

[39] J. Tan, L. Bauer, J. Bonneau, L. F. Cranor, J. Thomas, and B. Ur, "Can unicorns help users compare crypto key fingerprints?" in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. New York, NY, USA: ACM, 2017, pp. 3787–3798.

[40] K. Marky, M.-L. Zollinger, M. Funk, P. Ryan, and M. Mühlhäuser, "How to assess the usability metrics of e-voting schemes," in *Proceedings of Financial Cryptography and Data Security*. Cham, Switzerland: Springer, 2019, pp. 257–271.

[41] J. Sun and P. Ahluwalia, "How users respond to authentication methods: A study of security readiness," in *Proceedings of the 14th Americas Conference on Information Systems (AMCIS'08)*, 2008.

[42] K. Marky, P. Gerber, S. Günther, M. Khamis, M. Fries, and M. Mühlhäuser, "Investigating state-of-the-art practices for fostering subjective trust in online voting through interviews," in *Proceedings of the 31st USENIX Security Symposium (USENIX Security 22)*, 2022, pp. 4059–4076.

[43] S. Gulati, S. Sousa, and D. Lamas, "Design, development and evaluation of a human-computer trust scale," *Behaviour & Information Technology*, vol. 38, no. 10, pp. 1004–1015, Oct. 2019.

[44] J. Brooke, "SUS-a quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, no. 194, pp. 4–7, 1996.

[45] C. Z. Acemyan, P. Kortum, M. D. Byrne, and D. S. Wallach, "Users' mental models for three end-to-end voting systems: Helios, prêt à voter, and scantegrity ii," in *Proceedings of the International Conference on Human Aspects of Information Security, Privacy, and Trust (HAS)*. Cham, Switzerland: Springer, 2015, pp. 463–474.

[46] B. Laugwitz, T. Held, and M. Schrepp, "Construction and evaluation of a user experience questionnaire," in *Proceedings of the Symposium of the Austrian HCI and Usability Engineering Group (USAB)*. Cham, Switzerland: Springer, 2008, pp. 63–76.

[47] S. Diefenbach, E. Lenz, and M. Hassenzahl, "An interaction vocabulary. describing the how of interaction." in *Extended Abstracts on Human Factors in Computing Systems (CHI EA)*. New York, NY, USA: Association for Computing Machinery, 2013, p. 607–612.

[48] T. Selker, E. Rosenzweig, and A. Pandolfo, "A methodology for testing voting systems," *Journal of Usability Studies*, vol. 2, no. 1, pp. 7–21, Nov. 2006.

[49] A. Bangor, P. Kortum, and J. Miller, "Determining what individual sus scores mean: Adding an adjective rating scale," *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.

[50] F. Karayumak, M. M. Olembo, M. Kauer, and M. Volkamer, "Usability analysis of Helios - an open source verifiable remote electronic voting system," in *Proceedings of the Conference on Electronic Voting Technology/Workshop on Trustworthy Elections (EVT/WOTE)*. Berkeley, CA, USA: USENIX Association, 2011.

[51] M. Volkamer, O. Spycher, and E. Dubuis, "Measures to establish trust in internet voting," in *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 1–10.

[52] S. Marsh and M. R. Dibben, "The role of trust in information science and technology," *Annual Review of Information Science and Technology (ARIST)*, vol. 37, pp. 465–98, 2003.

[53] L. Carter and R. Campbell, "The impact of trust and relative advantage on internet voting diffusion," *Journal of theoretical and applied electronic commerce research*, vol. 6, no. 3, pp. 28–42, 2011.

[54] K. Veseli, "Voting as a habit? quantitative analyses of voting costs and turnout in direct democracy," Ph.D. dissertation, University of Zurich, 2016.

[55] A. S. Patrick, P. Briggs, and S. Marsh, "Designing systems that people will trust," *Security and Usability*, vol. 1, no. 1, pp. 75–99, 2005.

[56] E. Sillence, P. Briggs, L. Fishwick, and P. Harris, *Trust and Mistrust of Online Health Sites*. New York, NY, USA: Association for Computing Machinery, 2004, p. 663–670.

[57] D. Chaum, "Surevote: Technical overview," in *Proceedings of the Workshop on Trustworthy Elections (WOTE)*, 2001.

[58] P. Y. A. Ryan and V. Teague, "Pretty good democracy," in *Proceedings of the International Workshop on Security Protocols (SPW)*. Cham, Switzerland: Springer, 2009, pp. 111–130.

[59] O. Kulyk, S. Neumann, J. Budurushi, and M. Volkamer, "Nothing comes for free: How much usability can you sacrifice for security?" *IEEE Security & Privacy*, vol. 15, no. 3, pp. 24–29, 2017.

[60] O. Kulyk, J. Henzel, K. Renaud, and M. Volkamer, "Comparing "challenge-based" and "code-based" internet voting verification implementations," in *Proceedings of the IFIP Conference on Human-Computer Interaction (INTERACT)*. Cham, Switzerland: Springer, 2019, pp. 519–538.

Appendix A. User Study

A.1. Final Questionnaire

Custom items:

1) If I had the possibility, I would use the presented Internet voting system. (5-point Likert scale)
2) If I had the possibility, I would like to use the Internet voting system in a real political election (e.g., parliament, city council). (5-point Likert scale)
3) If I had the possibility, I would like to use the Internet voting system in a real non-political election (e.g., club president). (5-point Likert scale)
4) I would prefer the presented Internet voting system over polling-station voting. (5-point Likert scale)
5) I would prefer the presented Internet voting system over postal voting.
6) I think the Internet voting system is very private, that is, it does not reveal my voting choice. (5-point Likert scale)
7) (paper/token condition) I would observe the broadcast that shows the printing of the voting item. (5-point Likert scale)
8) (paper/token condition) I think the printing broadcast is unnecessary. (5-point Likert scale)
9) Please describe your experience with the online voting system. List positive and negative aspects, the more details we got, the better (free-text)

For the 5-point Likert scales, we used the labels "I completely disagree", "I disagree", "I neither agree nor disagree", "I agree", and "I completely agree". Besides the standardised questionnaires and our custom items, we added two attention checks asking participants to choose a specific answer, such as *please mark "I disagree"*.

A.2. Detailed User Study Results

TABLE 2: Detailed statistics of the IV

	χ^2	df	p	η^2
slow - fast	36.253	2	0.000	0.233
stepwise - fluent	11.458	2	0.003	0.064
instant - delayed	25.378	2	0.000	0.159
uniform - diverging	5.608	2	0.061	0.025
constant - inconstant	8.248	2	0.016	0.043
mediated - direct	8.932	2	0.011	0.047
spatial separation - spatial proximity	2.391	2	0.303	0.003
approximate - precise	0.847	2	0.655	-0.008
gentle - powerful	0.753	2	0.686	-0.009
incidental - targeted	6.702	2	0.035	0.032
apparent - covered	0.967	2	0.617	-0.007

TABLE 3: Participants' country of residence.

Condition	Country
No audit trail (N=50)	USA: 15 South Africa: 9 Europe: 21 Other: 5
Paper audit trail (N=50)	USA: 16 South Africa: 11 Europe: 17 Other: 6
3D-token audit trail (N=50)	USA: 13 South Africa: 12 Europe: 20 Other: 5

A.3. Descriptive Statistics

TABLE 4: Descriptive statistics of the IV.

	No audit trial		Paper audit trial		3D-token audit trial	
	M	SD	M	SD	M	SD
slow/fast	6.41 95% CI [6.18, 6.63]	0.85	5.94 95%CI [5.58, 6.27]	1.23	4.27 95% CI [3.65, 4.86]	2.13
stepwise/fluent	5,94 95% CI [5.55, 6.29]	1.39	5.04 95% CI [4.60, 5.50]	1.62	5.27 95% CI [4.78, 5.67]	1.55
instant/delayed	1,88 95% CI [1.57, 2.25]	1.24	2.44 95% CI [2.10, 2.81]	1.37	3.59 95% CI [3.06, 4.12]	1.95
uniform/diverging	2,75 95% CI [2.37, 3.12]	1.31	2.94 95% CI [2.58, 3.29]	1.23	3.45 95% CI [3.00, 3.88]	1.60
constant/inconstant	2,20 95% CI [1.90, 2.51]	1.11	2.63 95% CI [2.33, 2.92]	1.06	2.88 95% CI [2.55, 3.27]	1.34
mediated/direct	5,69 95% CI [5.27, 6.06]	1.46	5.25 95% CI [4.81, 5.67]	1.51	4.86 95% CI [4.37, 5.25]	1.56
spatial seperation/spatial proximity	4,31 95% CI [3.86, 4.75]	1.63	4.33 95% CI [3.98, 4.69]	1.23	4.04 95% CI [3.67, 4.41]	1.36
approximate/ precise	5,31 95% CI [4.86, 5.75]	1.63	5.25 95% CI [4.79, 5.65]	1.54	5.04 95% CI [4.61, 5.51]	1.71
gentle/powerful	4,41 95% CI [3.90, 4.88]	1.69	4.19 95% CI [3.75, 4.65]	1.55	4.27 95% CI [3.82, 4.73]	1.61
incidental/targeted	5,55 95% CI [5.20, 5.88]	1.22	5.21 95% CI [4.85, 5.54]	1.17	4.86 95% CI [4.51, 5.22]	1.34
apparent/covered	3,31 95% CI [2.86, 3.76]	1.64	3.33 95% CI [2.90, 3.79]	1.66	3.57 95% CI [3.14, 4.00]	1.53

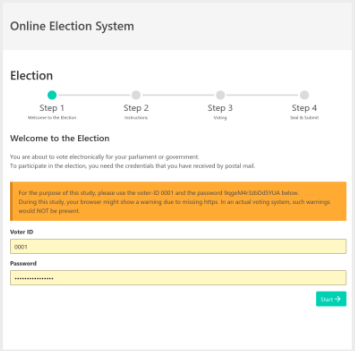
TABLE 5: Descriptive statistics of the SRS, HCTM, custom items, SUS, and UEQ scales.

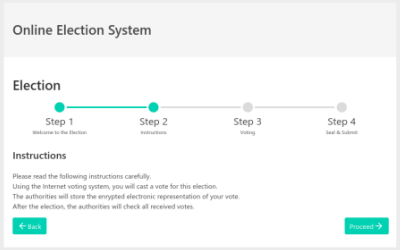
	No audit trail		3D-token audit trail		Paper audit trail	
	M	SD	M	SD	M	SD
Sec_read_affective	5.69 95% CI [5.34, 5.99]	1.15	5.06 95% CI [4.62, 5.45]	1.43	5.56 95% CI [5.26, 5.85]	1.06
Sec_read_cognitive	5.78 95% CI [5.45, 6.05]	1.03	5.43 95% CI [5.09, 5.73]	1.17	5.61 95% CI [5.33, 5.88]	0.97
Sec_read_behavioral	5.40 95% CI [5.05, 5.75]	1.29	5.08 95% CI [4.69, 5.42]	1.35	5.08 95% CI [4.78, 5.36]	1.04
HCTM_perceived_risk	3.02 95% CI [2.74, 3.28]	0.96	2.76 95% CI [2.51, 3.01]	0.87	2.62 95% CI [2.45, 2.88]	0.80
HCTM_competence	3.71 95% CI [3.62, 4.05]	0.75	5.43 95% CI [3.46, 3.98]	0.88	3.90 95% CI [3.61, 4.15]	0.95
HCTM_benevolence	3.73 95% CI [3.54 , 3.91]	0.66	3.66 95% CI [3.44, 3.88]	0.75	3.71 95% CI [3.50, 3.92]	0.74
HCTM_trust	3.20 95% CI [3.09, 3.64]	0.95	3.66 95% CI [3.30, 3.85]	0.95	3.54 95% CI [3.20, 3.34]	1.01
FQ01_02	4.00 95% CI [3.71, 4.27]	1.08	3.81 95% CI [3.46, 4.10]	1.18	3.94 95% CI [3.65, 4.25]	1.16
FQ01_03	4.33 95% CI [4.10, 4.53]	0.766	4.40 95% CI [4.13, 4.63]	0.84	4.06 95% CI [3.78, 4.33]	1.05
FQ01_04	3.86 95% CI [3.57, 4.16]	1.11	3.71 95% CI [3.42, 4.00]	1.01	3.73 95% CI [3.43, 4.02]	1.12
FQ01_05	4.10 95% CI [3.82, 4.35]	0.99	3.83 95% CI [3.54, 4.08]	0.96	4.04 95% CI [3.78, 4.29]	0.98
FQ01_10	2.06 95% CI [2.00, 2.18]	0.42	2.06 95% CI [2.00, 2.17]	0.32	2.00 95% CI [2.00, 2.00]	0.00
FQ01_06	3.35 95% CI [3.08, 3.65]	1.04	3.35 95% CI [3.08, 3.63]	1.02	3.69 95% CI [3.47, 3.92]	0.84
FQ01_07	3.29 95% CI [2.96, 3.63]	1.15	3.25 95% CI [2.96, 3.52]	1.04	3.59 95% CI [3.33, 3.86]	1.00
FQ01_08	3.29 95% CI [2.96, 3.63]	1.24	3.46 95% CI [3.15, 3.75]	1.09	3.82 95% CI [3.61, 4.04]	0.82
FQ01_09	3.43 95% CI [3.12, 3.73]	1.12	3.40 95% CI [3.06, 3.69]		3.76 95% CI [3.51, 4.04]	0.91
SUS	78.26 95% CI [75.39, 82.95]	13.38	78.26 95% CI [73.79, 81.74]	13.64	69.44 95% CI [65.14, 74.34]	16.38
Attractiveness	1.66 95% CI [1.37, 1.96]	1.05	1.49 95% CI [1.22, 1.76]	0.92	1.21 95% CI [0.87, 1.54]	1.18
Perspicuity	2.23 95% CI [2.02, 2.51]	0.88	2.17 95% CI [1.91, 2.41]	0.86	1.30 95% CI [0.87, 1.74]	1.54
Efficiency	2.05 95% CI [1.80, 2.30]	0.90	1.80 95% CI [1.50, 2.09]	1.01	0.87 95% CI [0.50, 1.24]	1.32
Dependability	1.35 95% CI [1.07, 1.62]	0.99	1.03 95% CI [0.76, 1.30]	0.91	0.88 95% CI [0.59, 1.16]	1.01
Stimulation	1.66 95% CI [0.81, 1.51]	1.23	1.21 95% CI [0.92, 1.51]	1.01	0.77 95% CI [0.41, 1.12]	1.26
Novelty	1.40 95% CI [-0.26, 0.53]	1.40	1.15 95% CI [0.83, 1.47]	1.09	1.06 95% CI [0.75, 1.38]	1.11

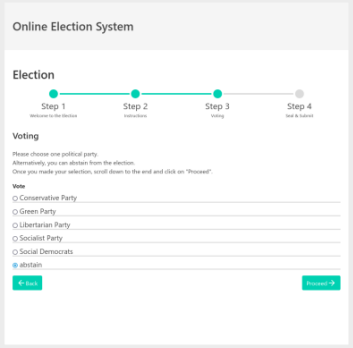
A.4. Prototype Screenshots

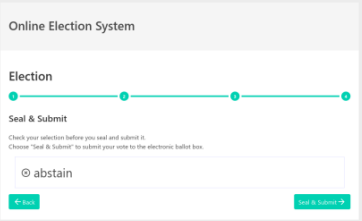
In this section, we provide screenshots of the implemented prototypes that we used in the user study. For higher resolution screenshots, the reader is referred to the paper's supplementary material available under https://www.gla.ac.uk/tangiblevoting.

Baseline

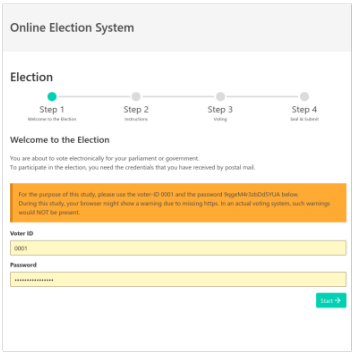


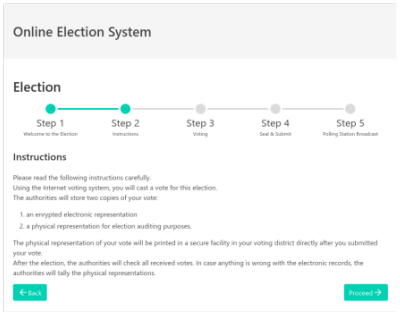


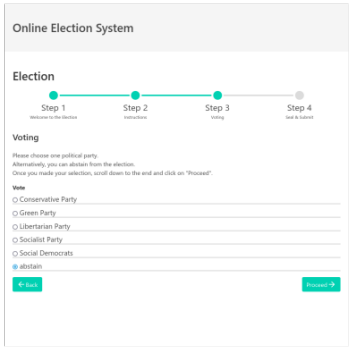


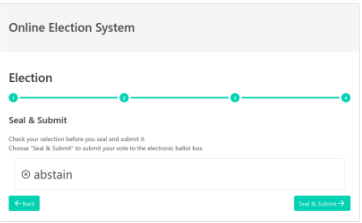


Paper

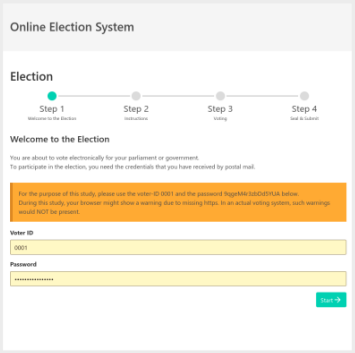


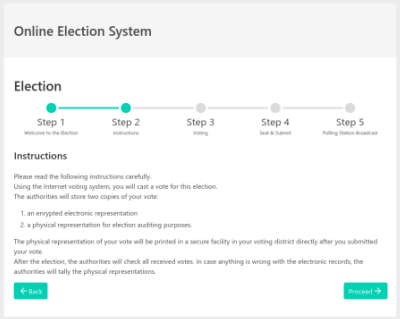


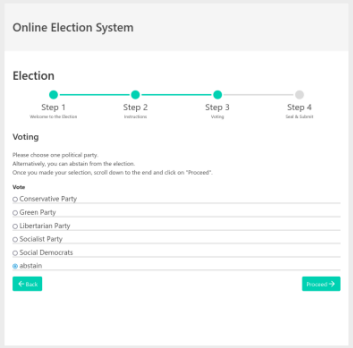


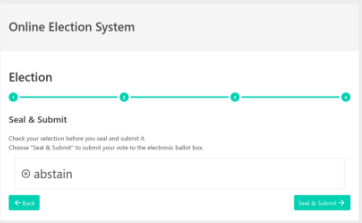


Token

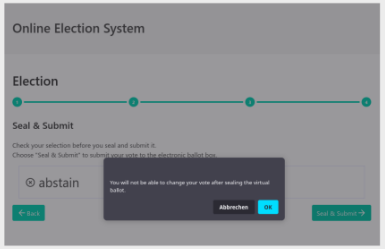


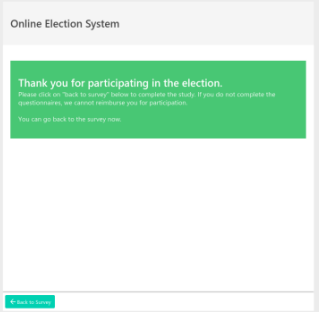




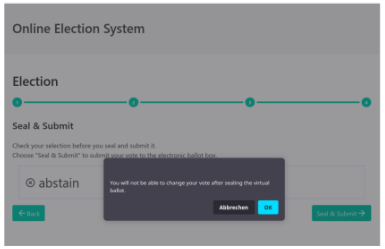


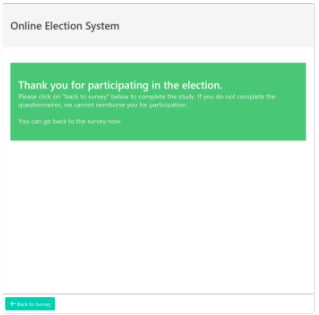
Baseline



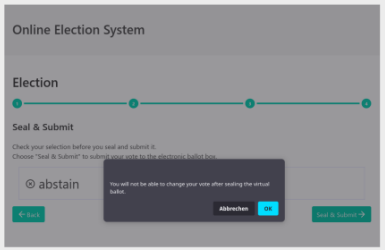


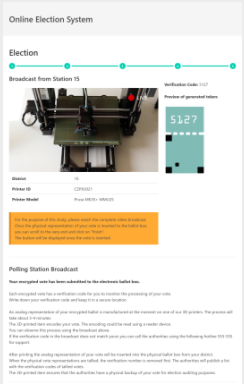
Paper

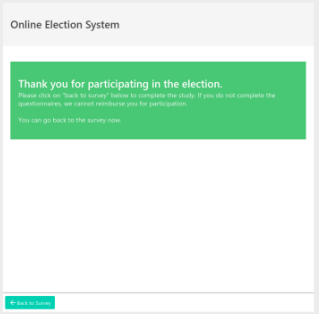




Token







Appendix B.
Meta-Review

The following meta-review was prepared by the program committee for the 2024 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

Summary

This paper evaluates a potential hybrid online voting system where encrypted versions of cast ballots are printed and observable by the voter and the general public on a live video stream. They tested the usability of this scheme with 150 participants. Of these, 50 participants were asked to vote electronically without any verification method (control), 50 were shown a live feed where their vote was printed on paper, and 50 were shown a live feed where a token 3-D was printed to reflect their vote. Their results show that the paper-printed option did the best to improve participant trust in the vote without impacting usability, though participants remained wary of online voting generally.

Scientific Contributions

- Provides a Valuable Step Forward in an Established Field
- Establishes a New Research Direction
- Other

Reasons for Acceptance

1) This paper provides a valuable step forward in an established field and establishes a new research direction. This paper builds a system to support hybrid online/in-person voting, demonstrating an approach that improves online voting security, as well as a new class of approaches (i.e., hybrid) that should be investigated further in later work. The paper also shows how people actually interact with hybrid systems and shows the usability of these voting schemes.