Available online at www.sciencedirect.com
ScienceDirect
Int. J. Human-Computer Studies 67 (2009) 111–124
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Electronic voting eliminates hanging chads but introduces new usability challenges

Frederick G. Conrad[a,b,*], Benjamin B. Bederson[b], Brian Lewis[a], Emilia Peytcheva[a], Michael W. Traugott[a], Michael J. Hanmer[b], Paul S. Herrnson[b], Richard G. Niemi[c]

[a]*Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI, USA*
[b]*University of Maryland, USA*
[c]*University of Rochester, USA*

Received 2 March 2008; received in revised form 22 August 2008; accepted 17 September 2008
Communicated by M. Atwood
Available online 2 October 2008

Abstract

The arrival of electronic voting has generated considerable controversy, mostly about its vulnerability to fraud. By comparison, virtually no attention has been given to its usability, i.e., voters' ability to vote as they intend, which was central to the controversy surrounding the 2000 US presidential election. Yet it is hard to imagine a domain of human–computer interaction where usability has more impact on how democracy works. This article reports a laboratory investigation of the usability of six electronic voting systems chosen to represent the features of systems in current use and potentially in future use. The primary question was whether e-voting systems are sufficiently hard to use that voting accuracy and satisfaction are compromised. We observed that voters often seemed quite lost taking far more than the required number of actions to cast individual votes, especially when they ultimately voted inaccurately. Their satisfaction went down as their effort went up. And accuracy with some systems was disturbingly low. While many of these problems are easy to fix, manufacturers will need to adopt usability engineering practices that have vastly improved user interfaces throughout the software industry.
© 2008 Elsevier Ltd. All rights reserved.

Keywords: Usability of electronic voting; Usability of e-voting; Voting interfaces

1. Introduction

Hanging, dimpled, and pregnant chads achieved infamy in the contentious presidential election in the US in the year 2000. The controversy, which concerned the perforated squares of paper (chads) that voters were required to remove from ballot cards by punching with a stylus, came to symbolize ambiguity about voters' intent. Did the voter mean to push the chad all the way through or was the impression accidental? Election judges were required to decide what voters intended based on visual inspection of the punch cards. Partisans on both sides of the election found this to be deeply flawed and the world noticed usability in a way it never had before.[1] One outcome has been the widespread deployment of electronic voting systems; those with touch screens have been the most controversial. The technology is now used widely in the US, the Netherlands, India, Brazil, Japan, Venezuela, and other countries. With these systems, a vote is either registered or it is not, e.g., the check box is either selected or unselected, and there is no analogue to a hanging chad. However, electronic voting systems may introduce usability problems that threaten to undermine the credibility of voting tallies and election participation. We report a laboratory study here, which was designed to explore

*Corresponding author at: Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI, USA.
Fax: +1 734 764 8263.
E-mail address: fconrad@isr.umich.edu (F.G. Conrad).

[1]The significance of usability for election outcomes was also driven home in the same Florida contest by the infamous "butterfly ballot" (e.g., Sinclair et al., 2000).

1071-5819/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijhcs.2008.09.010

whether the usability of electronic voting systems is likely to threaten accurate and efficient voting, and if so, how usability problems affect voters' behavior and satisfaction with the experience. Based on the findings, we consider design principles for electronic voting systems in light of the characteristics of the voting task.

Electronic voting systems have generated considerable vitriol, mostly concerning the potential for inaccurate tallies due to programming errors or malice, compounded by the unverifiable nature of touch screen electronic votes (e.g., Department of Legislative Services, 2004; Brennan Center Task Force, 2006; Rubin, 2006). While such security issues are certainly a very serious concern, we believe that the consequences of hard-to-use voting interfaces are at least as serious. In a close election, even rare usability problems can distort the outcome, particularly if they lead to systematic, as opposed to randomly distributed, errors. This can occur if the usability problems are concentrated among voters who hold similar political opinions, as might be the case for minority or elderly voters on certain issues. Difficulty using the technology may mean that this group cannot vote for the candidate or position it supports and may unintentionally vote for the same opposing candidate or position. Even if voters ultimately vote the way they intend to, they may find the experience profoundly frustrating and unsatisfying. Such experiences might lead them to sit out future elections, which can itself affect the outcome of those elections. The point is that usability has proven to be of crucial importance with previous election technologies (see e.g., Roth, 1998), and is at least as important with modern electronic technologies.

These concerns are intensified by the likelihood that voting is more sensitive to usability problems than many other tasks accomplished with electronic interfaces. The user base is highly diverse and this diversity is sure to include differences in experience with technology. Because elections occur infrequently, for example in the US elections occur once a year or less often in most jurisdictions, voters are likely to remain novices with voting technology for many election cycles, experiencing usability problems that would likely be resolved with more frequent practice (see e.g., Anderson, 1983, 1989 for a discussion of learning procedures for interacting with computers in other tasks). And voting is typically done in public settings, which can create pressure on voters to move quickly and appear competent.

As a first step in exploring the usability of electronic voting, we carried out an expert review (Nielsen, 1994; Shneiderman and Plaisant, 2005, pp. 141–144) of the same voting systems and ballot designs investigated in the study reported below. Twelve internationally recognized experts, including six familiar with voting system reviews, were asked to perform a number of specific activities, including changing a vote after it was submitted, voting for a write-in candidate, and deliberately failing to vote for one or more offices. Some were asked to assume specific roles when evaluating the systems in order to simulate the experiences of novice computer users, voters with limited English language skills, including those who mostly speak another language, elderly voters, individuals who found voting stressful, or voters who made many errors using the systems. The experts were given a set of usability heuristics similar to those proposed by Nielsen (1994)[2] and instructed to apply them to the user interfaces of six electronic voting systems while carrying out the different voting tasks from different user perspectives. They reached broad consensus on potential problems with the systems.

The experts identified design flaws in the different interfaces that they believed would interfere with voters' ability to start the voting process, e.g., it is difficult to insert the access card; to read the screen and interpret the ballot, e.g., font is too small, colors are hard to read; to navigate from screen to screen, e.g., automatic advance upon selection is disorienting; to change votes, e.g., likely to be frustrating; to cast write-ins, e.g., separation of first and last name confusing; to review votes on the screen, e.g., review screen does not clearly show multiple candidates; and to verify printed paper records of their voting choices, e.g., paper record shown too fast and without instructions, among other likely problems. The expert review thus provided prima facie evidence that voters may be both frustrated and unable to vote accurately with electronic voting systems. We carried out the current study to gather detailed, empirical evidence that voters using these same systems encounter usability problems that may interfere with their ability to vote as intended and reduce their satisfaction with the experience.

2. Laboratory study

Our goals in the current study were to better understand the process of using electronic voting systems in general, i.e., irrespective of particular interface features, as well as to determine what kinds of usability problem are related to particular features of the interface in different voting systems. We developed several hypotheses about both of these issues.

Concerning the process in general, we conjectured that users would be sensitive to small differences in the amount of effort required to vote. The more effort they expend the less satisfied they will be with the experience. This was based on the finding that, for at least some tasks, users seem averse to executing even minor actions that will likely improve their performance but will lead to slightly longer task completion. In particular, individual users have been shown to opt not to invest a mouse click (Conrad et al., 2006) or even an eye movement (Gray and Fu, 2004) to obtain information that will lead to optimal performance when adequate performance is possible without these small actions. The implication is that when users are unable to take such short cuts they are likely to notice the extra effort

[2]The set of heuristics and the exact procedures followed by the experts are available from the authors.

and be unhappy about it. This may well be the case for electronic voting systems. Aside from opting not to vote in a particular race, voters have little discretion in how much effort to exert, at least in the interfaces we examined; assuming the user wants to vote in a particular race, the more work required to do this, the less satisfying the experience is likely to be.

A related hypothesis is that users will be sensitive to how accurately they have voted, i.e., the degree to which they have voted as they intended. Of course they receive no feedback on the accuracy of their votes—this would require the systems to know their intentions!—but they may well suspect they have voted inaccurately when they have in fact voted other than they intended. Users will almost certainly be aware of their inaccuracies if inaccuracies are the outcome of effortful and frustrating interactions, as this is likely to focus their attention on the obstacles posed by the interface. In these cases, the errors would likely have the character of what Norman (1981) called *mistakes*, i.e., failure to do what the user intends such as voting for no one when the user wants to vote for a particular candidate, rather than what he called *slips*, i.e., momentary performance lapses in which the user takes an unintended action such as selecting a candidate adjacent to the one he or she wishes to vote for due to an imprecise finger action.

For each of these hypotheses about the general process, we can derive comparable hypotheses about specific interfaces. With respect to effort, we would expect interfaces that increase effort to reduce satisfaction. So if one interface requires users to exert more actions than another to carry out the same voting task, users are likely to find the lower effort design more satisfying. With respect to awareness of voting accuracy, systems that lead to longer and less-efficient interactions, especially before users fail to vote as intended, should lower users' confidence in the accuracy of their votes.

2.1. Procedure

Prior to using any voting systems, each voter indicated which candidates or ballot positions he or she intended to vote for. An experimenter confirmed that each user had selected a choice for each contest. Then the users attempted to vote for their choices on each of six voting systems. After voting on each system, the users completed a paper questionnaire about their satisfaction with the system just used, i.e., six questionnaires in all. At the end of the session each user completed another paper questionnaire, this one about his or her demographic characteristics and computer experience. One user participated at a time and his or her interactions with each system were video recorded. The order in which users interacted with the systems was randomized according to a Latin Square so that each voting system occurred in each of the six possible positions equally often across voters. While the order in which users interacted with particular voting systems was detectable in some analyses, the Latin Square design distributed such order effects uniformly across the six voting systems that we studied. Consequently, we do not mention order effects further.

2.2. Voting systems

There are numerous voting systems on the market that embody numerous design approaches. We examined the usability of six electronic voting systems chosen to represent most of the features in today's systems as well as some features that could appear in systems of the future. Two of the systems, the Diebold AccuVote-TS and Avante Vote-Trakker, were ATM-style touch screens; the second of these printed a paper record of the on-screen selections sometimes referred to as a *paper trail*; in contrast to the Diebold AccuVote-TS interface, which displayed multiple contests per screen and required the user to touch a *Next* button to advance, the Avante Vote-Trakker interface displayed one contest per screen and automatically advanced the user when a candidate was selected. A third system, the UMD Zoomable system, was an experimental prototype designed for the current research program. It used a touch screen with a zoomable interface allowing the voter to view the entire ballot or incrementally magnify (zoom in on) a portion of the ballot such as a particular race. A fourth system, the Hart Intercivic eSlate, required voters to indirectly manipulate and navigate the ballot using a physical navigation dial and buttons; upon positioning the cursor at a candidate, the voter pressed *Enter* to select the candidate. A fifth system, the Nedap LibertyVote, displayed the entire ballot at once by overlaying a translucent, physical ballot of roughly $60 \times 90\,\mathrm{cm}^2$ in size on a panel of mechanical buttons and lights visible through the ballot that users pressed to register their votes; this sort of display is sometimes referred to as a *full face ballot*. The final system, ES&S Model 100, used optical scan technology to read paper ballots on which voters manually marked their choices by filling in ovals with a pen or pencil; users received feedback about the acceptability of their ballot in a small display. All but the Zoomable system were commercially available at the time of data collection and all but one of the commercial systems were used in elections in the US or other nations at the time of the study. The Zoomable system was developed at the Human–Computer Interaction Laboratory at the University of Maryland. Its design is based on a long history of developing this kind of interface for a variety of tasks (Bederson and Meyer, 1998). This prototype, including source code, is freely available at http://www.cs.umd.edu/~bederson/voting.

2.3. Ballot design

The ballot included a mix of national and local contests as voters might encounter in a US presidential election. The ballot presented voters with 22 contests, 18 of which involved public office and four of which were ballot

questions. All candidates and ballot questions were fictional. Of the 18 contests for public office, 11 were partisan, i.e., candidates were affiliated with particular political parties, and seven were non-partisan. The ballot was either an office-bloc ballot, i.e., visually organized by the contest or office, or an office-bloc ballot with a straight-party option allowing voters to cast their votes for all candidates of a single party by making a single selection. For one system (Nedap LibertyVote), the straight-party option could not be implemented. So the alternative ballot was a party column design in which, for the partisan races, candidates affiliated with the same party appeared in the same column for all races, also a widely used ballot design. A particular participant was exposed to the same ballot design—either office bloc or office bloc with straight-party option/party column—on all six voting systems.

2.4. Participants

Forty-two members of the Ann Arbor, MI (USA) community were recruited in late July and early August, 2004 to visit our usability laboratory and vote on six electronic voting systems. Because the expert review suggested that certain types of voters, e.g., older voters or computer novices, might be at increased risk of finding the systems hard to use, and because we wished to maximize the chances of observing usability problems in a small convenience sample, we oversampled users from these subpopulations. In particular, 31 of the 42 participants had limited computer experience, that is they answered "one or two days a week" or less when asked about their frequency of computer use. In many cases these participants indicated they had never used a computer. Many used e-mail somewhat often but engaged in no other computer tasks with measurable frequency. In addition, older voters were oversampled. Twenty-nine out of the 42 voters were older than 50 years of age: 17 in the 50–64 range, nine in the 65–74 range, and three in the 75 and over range. Many of the older participants were also those without computer experience, but five of the voters older than 50 used computers 5–7 days a week. Detailed demographic information[3] about the voters appears in Table A1. Participants were paid $50 at the end of the session, which in most cases lasted between 1 and 2 h.

2.5. Voting tasks

In general, users selected one candidate or position for each contest. For two of the contests, users were required to select two candidates. In other contests, users were instructed to change an initial vote, write-in a vote, i.e., enter a name provided by the experimenters, and abstain from voting. In addition to the 22 contests on the ballot, there was a start-up task that differed by system, e.g., entering a string of four digits with the Hart Intercivic system, inserting an access card with the Diebold and Avante systems, and a task for reviewing the ballot at the end of the session with most of the systems. Because the Avante system printed a paper record of each voter's choices at the end of the voting session, a task for reviewing the paper record was also included in analyses of this system.

2.6. Measures

Users indicated their voting intentions by circling their choices in a voter guide consisting of brief descriptions of each candidate or ballot question. We defined voting accuracy as agreement between voters' intentions and their behavior, observable in the videos. An error was any discrepancy between the candidate or ballot position the user circled and the candidate or position for which the user voted. The definition of an error included not voting in a particular contest for which a choice had been indicated in the booklet. In about 0.5% of the contests the voters' choices were not visible and so we could not establish their accuracy.[4]

The video recordings provided several process measures. First, the videos were coded at the action level, e.g., a user pressing a check box on a touch screen or turning a physical navigation wheel. Seven coders classified all voter actions in the video corpus, about 60 h in all. They assigned each action to one of 76 codes comprised of four variables.[5] This enabled us to carry out two kinds of analyses. First, it enabled us to simply count the actions required to cast a vote. The median number of actions per voting contest was either one or two for each of the six systems. The maximum number of actions ranged from 25 to 101 across the systems. In addition, coding of all user actions made it possible to measure the frequency of action patterns, e.g., the number of times users pressed a "Help" button after trying unsuccessfully to select a candidate (see Sanderson and Fisher, 1994, for a discussion of sequential analyses of this sort). In addition, we measured the duration of each voting contest from immediately after the previous contest to the end of the current one. There were roughly 6000 votes cast in the data set, 22 per voter per system with additional votes coded when voters revisited contests.

[3]Note that we did not explicitly recruit users with sensory or motor disabilities. While universal usability of voting systems is a crucial issue it was not a focus of the current study.

[4]As an alternative to observable behavior, we also had access to the voters' automatically captured selections (ballot images). These closely matched our observations (the agreement rate was 98.3%). We did not treat them as the gold standard because unlike video-observed votes, ballot images allow us only to measure final accuracy but not initial accuracy prior to a change in votes, for example after the user consulted the "review screen."

[5]Each coder worked on a different subset of videos. Their decisions were reviewed and, in some cases, revised in two subsequent quality assurance passes through their codes. We did not compute inter-coder reliability because the coding task was too vast for us to double code the videos. Across the video corpus, the coders recorded 15,923 judgments (codes) in total.

In addition to the measures of process and accuracy, the satisfaction questionnaire, administered after the user voted on each system, provided information about users' reactions to the system in general, e.g., ease and comfort of use, readability, confidence in accuracy of vote recording, and to some specific tasks, e.g., ease of casting a write-in vote and of changing a vote. Voters indicated their satisfaction by selecting a level of agreement on a seven-point scale, where one corresponded to "strongly disagree" and seven to "strongly agree," to a statement such as "It was easy to vote with this machine."

3. Results and discussion

We first present results that cut across the individual systems, and which we believe reflect fundamental processes involved in using electronic voting technology. We next examine how the different systems affected users' performance and satisfaction in different ways that presumably derive from the different features of the six systems. Finally, we examine several specific voting tasks where usability issues were particularly vexing to users: changing a vote, casting a write-in vote, and reviewing a printed record of voting choices.

3.1. Findings across the voting systems

The clear message across these six voting systems is that voters experienced many problems, which at best increased the effort required to vote and at worst interfered with their ability to vote as intended and led to frustration. It was evident in the video recordings that many individual voters got quite lost attempting to cast particular votes with particular systems, and we discuss some specific cases below. In the aggregate this difficulty can be observed by comparing how many actions voters took in each contest on each system to the number required under ideal conditions. By "ideal conditions" we mean the minimum number of actions required to select a candidate and perform essential navigation. Typically the ideal number of actions was one (e.g., touching the on-button labeled with a candidate's name in order to select the candidate) or two (e.g., touching buttons for two candidates when the contest called for two selections, or touching a button for a candidate and touching a Next button to advance to the next screen). However, some interfaces required additional actions for each contest such as rotating the navigation dial and pressing *Enter* in order to select a candidate. Because the ideal sequence was defined in terms of its length, i.e., number of actions, and not the exact path followed by the voter, there could in principle be more than one path for casting a vote that is of minimum length without affecting our analyses of this variable. In practice we never identified more than one ideal path.

We computed the ideal sequence for each contest on each system and subtracted its length, i.e., the number of component actions, from the number of actions actually taken by each user for the corresponding contest on that system. We treated this difference as the deviation from ideal performance. If the deviation was large[6] this indicated the voter was far off the ideal path, i.e., unable to efficiently cast a vote. If users ultimately recovered from these departures from the ideal path then they are of less concern—though still not good—than if they did not recover and ultimately voted incorrectly. The latter scenario seems to have been the case relatively often. The mean deviation for accurate votes is 0.89, less than one action more than needed; the mean deviation for inaccurate votes is 2.35, nearly three times as large as for accurate votes ($t[301] = 4.52$, $p < 0.001$). The same relationship between the deviation score and voting accuracy was also evident in multivariate analyses. Specifically, in a logit model that predicts accuracy based on deviation from the ideal path controlling for voters' computer experience, the type of ballot, and the voting system, the χ^2 ($df = 1$) for the deviation score is 9.77, $p < 0.002$.

In fact, users were accurate the vast majority of the time (see below). So prolonged sequences of actions were not common. Nonetheless, it seems plausible that these longer-than-necessary episodes took a toll on users' subjective experiences. Specifically, voters might have been less satisfied the more effort they needed to expend as we proposed in our first hypothesis. We operationalized "effort" in two ways: (1) the number of voting actions and (2) the time required to cast a vote. We then computed correlations, in particular Spearman's rho, between the two effort measures and two global satisfaction measures, voters' ratings of ease of use and comfort using the system. Each correlation was computed from six pairs of values for each of the 42 voters, where each pair consisted of an average effort score and a satisfaction score for one of the six systems. There was a clear negative relationship between effort and satisfaction, reflected by the negative correlations in Table 1. The correlations range from -0.33 to -0.40, and all are significant beyond the 0.001 level. When the correlations were computed separately for voters with high and low computer experience, they were also negative and significant ($p < 0.01$ in all cases) for both groups of voters. So the relationship between effort and satisfaction is present irrespective of computer experience.

It is not surprising that voters prefer a short and quick voting experience, but their satisfaction is surprisingly sensitive to even small differences in effort. Moreover, the negative relationship between effort and satisfaction may not hold for all HCI tasks. Consider video games in which the longer the game lasts the more points the player has

[6]This difference was almost always positive. In a small number of cases voters advanced to another contest after taking fewer actions than required for ideal performance, leading to a negative deviation. This occurred primarily because voters did not complete the process for that particular contest. We include these negative deviations in the analyses reported here. However, in a parallel set of analyses we excluded negative deviations and the substantive results were very similar to what we report here.

Table 1
Correlations (Spearman's ρ) between effort (duration or number of actions) and satisfaction (ease or comfort)

	Satisfaction with voting system	
	Ease	Comfort
Average effort per contest		
Duration	−0.40	−0.37
	$n = 250$	$n = 240$
Number of actions	−0.33	−0.33
	$n = 250$	$n = 240$

$p < .001$ for all correlations.Note: Ideally n would be 252 in all cells, i.e., 42 users × 6 systems. However, a few users did not vote on all systems or did not complete all questionnaire items for a system, thus lowering the number of data points.

accumulated and the more pleasing the game. The opposite relation, i.e., the longer the session the greater the satisfaction, should be observed in this type of interaction. Of course, in our laboratory setting, users were not likely to walk away before receiving their monetary incentive at the end of the session. However it is possible that under actual voting conditions, a relatively high level of effort will lead some voters to walk away from the voting booth before submitting the entire ballot.

Finally, users seem to have some awareness that they have voted incorrectly even though there is no explicit feedback to this effect, and this detracts from their satisfaction. Inaccuracy in voting is negatively correlated with both satisfaction measures (−0.23 for ease of use, $p < 0.001$ and −0.18 for comfort, $p < 0.005$). This is consistent with our second hypothesis and suggests that the inaccuracies are the outcome of interactions that are inefficient and effortful, as implied by the earlier analysis associating off-path distance with accuracy. In sum, when voters cast inaccurate votes, substantial effort is involved and they do not like it.

3.2. Differences between the systems

Perhaps the most important measure of a voting system's usability is voters' accuracy when using it. The percent of voting error (inaccuracy) for each of the six systems is presented in Fig. 1. To produce these data we first computed the error rate for each user for each system (the total number of errors divided by the number of contests) and then computed the mean error rate across users. The error rates did not differ significantly between voting systems, even between the two systems exhibiting the highest error rates and the others. Perhaps a larger sample and more statistical power—as would be the case in an actual election—would have resulted in significantly different levels of accuracy across the systems. However, relative accuracy may be a less meaningful indication of user performance than absolute levels of accuracy.

If one assumes that voters should be able to vote as intended 100% of the time, then systems on which voters make any errors do not meet the criterion for this task. Thus we looked for reliable departures from perfect performance, i.e., error rate significantly greater than zero. In a logistic regression model predicting voting accuracy on the basis of voting system, controlling for duration, computer experience, and the type of ballot, the error rate was reliably more than 0% for the Hart Intercivic (9%) and Zoomable (8%) systems ($p < 0.05$ for both). As discussed below, many of the errors with the Hart system concerned difficulty using the physical navigation dial, for example overshooting the target, and coordinating this with an *Enter* button (once the user selected a candidate with the navigation dial, he or she had to then select that candidate by pressing *Enter*). For write-in votes, users had to navigate to and select each letter of the write-in candidate's name from an on-screen keyboard. In fact, the error rate was particularly high with this system (34%) for the contest (Library Board Member) requiring a write-in. For the Zoomable system, error rates were particularly high for contests occurring later in the ballot due to premature submission of the ballot by a 7% (3/42) of the users.

Thus, if one adopts the reasonable standard that voting accuracy should be indistinguishable from perfect, these two systems failed to meet that criterion.[7] It is of course possible that the error rates for some of the other voting systems would also depart reliably from zero with increased statistical power, as would be the case in an actual election, where the number of observations would be much larger than in our laboratory study. A larger and more representative sample would likely include a higher proportion of computer savvy users but because we controlled statistically for computer experience and still observed error rates reliably above zero, there is reason to be concerned about accuracy in actual elections.

Of course, a critical mass of these errors would have to lean in the same direction in order to reverse an election's outcome. But this is not inconceivable. In Florida's 13th Congressional District in the 2006 general election, about 18,000 voters in one county (Sarasota) did not cast any votes for a congressional candidate. By most accounts this was due to the ballot design, which placed the congressional race above a visually more prominent race for governor so that voters just did not notice the congressional race in the presence of the gubernatorial race. In other counties where the congressional race was on a screen by itself, many fewer instances of no votes occurred

[7]One could argue for a more relaxed standard than zero errors but then what should such a standard be? In principle it could vary as long as it is less than the difference in votes for the top two candidates but this difference is not knowable ahead of time. A zero error rate is the only rate sure to be less than the vote differential in all elections. The criterion in the *Voluntary Voting System Guidelines* developed by the US Election Assistance Commission (http://www.eac.gov/voting%20systems/voting-system-certification/2005-vvsg) is equally intolerant of error: "[v]oters should encounter no difficulty or confusion regarding the process for recording their selections" (p. 45).

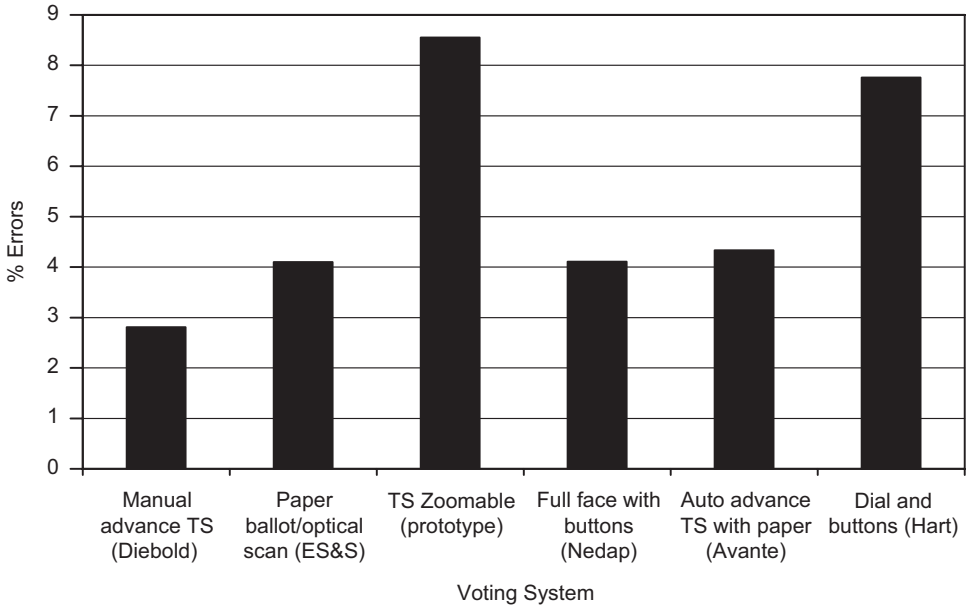

Fig. 1. Percent errors across all votes cast for each voting system. Percent for Zoomable and Hart systems reliably greater than 0. Note: TS = touch screen.

(see e.g., Frisina et al., 2008). Of the votes actually cast for a congressional candidate in Sarasota County, the Democratic candidate Christine Jennings collected slightly more than her Republican counterpart Vern Buchanan. Sarasota was Jennings's home county and the one county she won, so it stands to reason that many of the missing votes would have gone to her had voters made a selection in the contest. Nonetheless Buchanan garnered about 400 more votes than Jennings across the congressional district and was certified as the winner. The Jennings campaign contested the election on the grounds that in Sarasota County it is likely that had the missing 18,000 votes been cast as intended, i.e., for some candidate, the number going to Jennings would have been at least 400 more than to her opponent. Thus here is a case where it seems plausible that more of the errors were in one direction, i.e., would have gone to Jennings, than the other, i.e., would have gone to Buchanan.

The controversy surrounding the election in Florida's 13th congressional district concerns missing votes or "undervotes," but various other sorts of errors are also possible. We classified the errors in the video corpus into five categories and present the percent of each type of error with respect to total errors for the system—in each of the six systems in Fig. 2. The single most frequent type of error was missing votes.[8] The Zoomable system has the highest percent of missing votes, due largely to three of the 42 voters prematurely terminating the session by pressing the *Review and Cast Ballot* button instead of the *Next* button immediately to its left in a horizontal band along the bottom of the screen. All subsequent contests were treated as missing votes. So a single action led to a variable number of unrecorded votes depending on the contest in which voters made this error. Voters were next most likely to err by voting for a candidate or ballot position visually adjacent to their intended choice, the proximate candidate. This strikes us as unintentional, a slip in Norman's (1981) sense, because it could result from an imprecise hand movement that was probably not what the user intended to do; a slower movement or a larger target would likely have prevented these errors in the case of touch screens (see the discussion of Fitt's Law in Card et al., 1983). When the touch screen interface of the Avante Vote-Trakker automatically advanced the users, they may never have seen the error because the next screen was displayed before they could inspect what they had done. This was also anticipated in the expert review. If users did note their error, extra effort would have been required to repair it. In fact, in the Avante Vote-Trakker interface, the user was required to open the review screen, scroll to the contest for which the error had been made, select that contest to redisplay the ballot screen for that contest, and correct the vote. For these two reasons the auto-advance may have led to a large proportion of what we refer to as proximity errors (Herrnson et al., 2008a). The physical navigation dial in the Hart Intercivic system may well have been related to the relatively large proportion of votes for proximate candidates with that system. When users turned the dial quickly it frequently overshot the intended candidate and users could have selected the candidate on which the cursor had landed without noticing the error. Non-proximate candidate errors were very rare. Errors for write-in votes included writing-in the wrong candidate

[8]The phrase "missing votes" is neutral with respect to voters' intentions but in our study, as opposed to the Sarasota county incident, we know that users intended to cast a vote in these cases and failed to do so. Thus we treat these as errors.

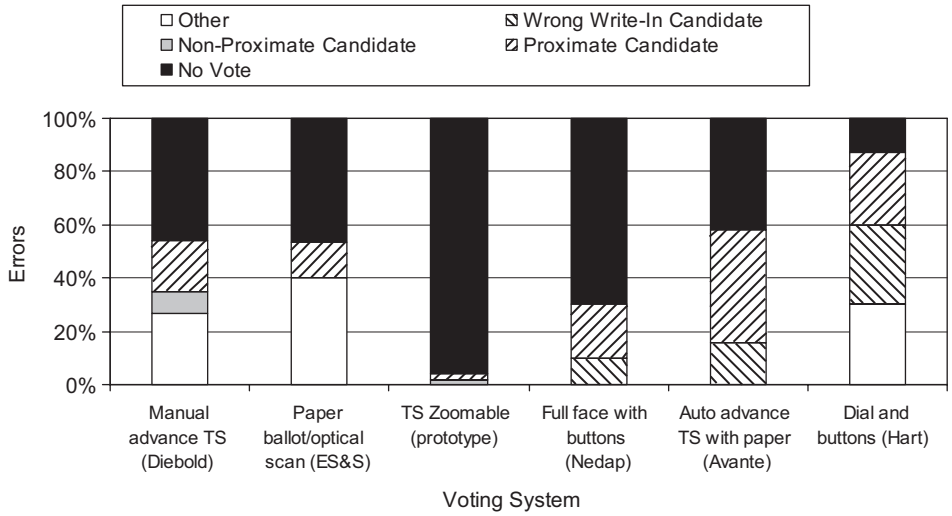

Fig. 2. Percent of errors by error type across the six voting systems. Note: TS = touch screen. Note also, "Other" errors include votes for a candidate when the user was instructed not to vote for a candidate, votes for a candidates on the ballot when the user was instructed to submit a write-in vote, failure to fill in the oval indicating a write-in vote on the paper, optical scan ballot (ES&S), and voting for the wrong party when instructed to make a straight-party selection.

name, usually a spelling error. We discuss the different sources of write-in errors on different systems in the next section. A residual category of "other" included votes for a candidate when the user was instructed not to vote for that candidate, votes for a candidate on the ballot when the user was instructed to submit a write-in vote, failure to fill in the oval indicating a write-in vote on the paper, optical scan ballot (ES&S), and voting for the wrong party when instructed to make a straight-party selection.

Users seemed more sensitive to potential errors on some systems than others based on the time spent reviewing their choices. Four of the six systems displayed the user's selections in one or two review screens, usually presented after a vote for the last contest had been registered. The other two systems, the full face Nedap system and paper ballot/optical scan ES&S system, made all choices available for inspection without any special review feature. Mean review time in minutes varied significantly across the systems, $F(5,185) = 10.62$, $p < 0.001$. Voters spent more than twice as long reviewing their choices with the auto-advance touch screen (1.6 m) and physical navigation dial and enter button (1.3 m) than they did with the other four systems: optical scan paper ballot (0.7 m), full face touch screen (0.7 m), manual advance touch screen (0.7 m), and Zoomable prototype (0.5 m). While the review time for the auto-advance touch screen system is likely to have been inflated by the non-review navigation use of the review screen discussed above, voters may nonetheless have been uncertain about the accuracy of their votes because they could not inspect their selections before the system advanced to the next screen. The time users spent reviewing their choices when voting with a navigation dial and buttons may reflect their recognition that they had in fact been relatively inaccurate with this system, perhaps due to overshooting the target. If so, these cases would support the fourth hypothesis that when users have voted inaccurately with a particular system they sense this inaccuracy or lack confidence in the accuracy of their votes. Curiously, despite the differences in review time, the number of votes changed did not vary across the six systems, $F(5,200) < 1$, n.s.

To the extent that some systems instilled confidence in users about the accuracy of their votes, this did not always serve them well. Consider the case of Voter 26, voting on the manual advance touch screen (Diebold) system, who inexplicably skipped one contest out of many on the screen. She advanced to the review screen where a red background flagged her missing vote, yet she almost immediately (2 s after the review screen was displayed) pressed the "Cast Ballot" button, terminating the session. She then commented, "That one I felt confident in that I didn't even need to go over it." A little more skepticism about her accuracy might have enabled her to detect this error.

Turning to overall duration, i.e., not just the review screen, users spent more time voting on some systems than others, $F(5,200) = 29.07$, $p < 0.001$ (see Fig. 3). They completed the ballot most quickly (4.8 m) with the manual advance touch screen (Diebold) system, marginally faster than the with second fastest system, $F(1,40]) = 3.98$, $p = 0.053$, and most slowly (10.00 m) with the dial and buttons (Hart Intercivic) interface, reliably slower than the fifth place system, $F(1,40) = 64.79$, $p < 0.001$.

The voting duration patterns more or less mirrored the number of actions required to vote on the different systems (see Fig. 4). The action count varied across the systems ($F[5,200] = 32.25$, $p < 0.001$). The single reversal in the ordering of systems relative to the pattern for duration was that voting on a paper ballot (ES&S Model 100) required fewer actions (but more time) than voting on the manual

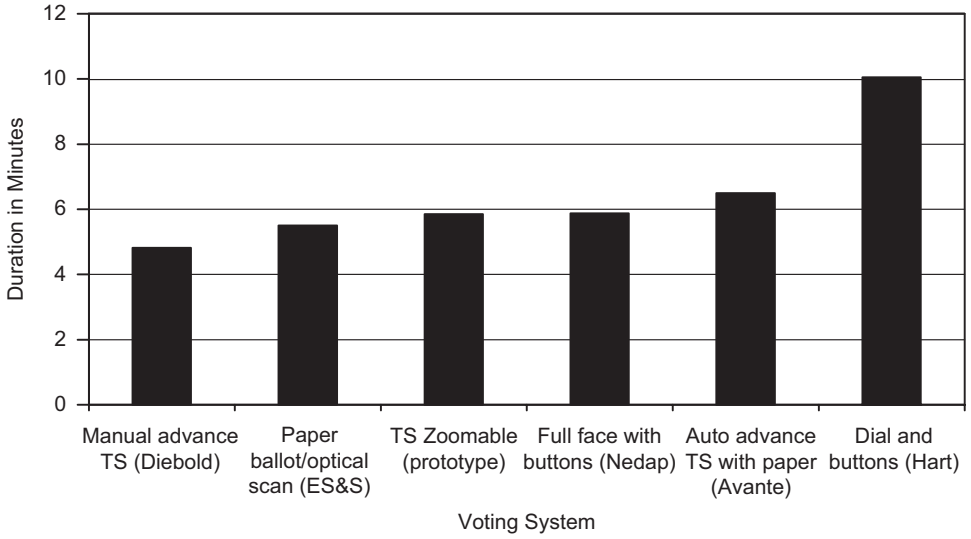

Fig. 3. Mean duration in minutes to vote (entire ballot) on each of the six systems. Note: TS = touch screen.

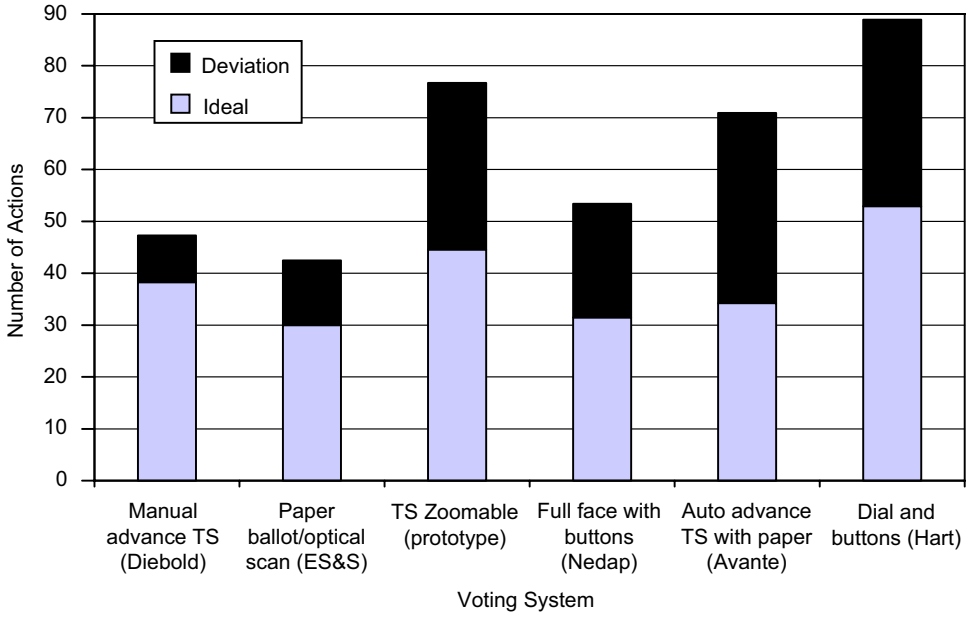

Fig. 4. Mean number of actions to vote (entire ballot) on each of the six systems, presented as sum of ideal number of actions and deviation from ideal. Note: TS = touch screen.

advance touch screen system (Diebold). Moving the navigation dial on the Hart Intercivic eSlate and then pressing the Enter button for each vote required the most actions, about twice as many as with paper or the manual advance touch screen. This is predicted by a simple task analysis: two actions minimum per vote with this type of interface versus one action to touch a candidate's button on a touch screen, one to fill in an oval with a marker on paper, and one to press the micro-switch for a candidate on the full face system.

The action count for each system in Fig. 4 is presented in two parts, the ideal number of actions summed across all contests on the ballot and the deviation between this figure and the actual count summed across all contests on the ballot. Proportionally, the greatest deviation occurs for the auto-advance touch screen (Avante Vote-Trakker) system. This is likely due to the difficulty this interface posed to users when changing a vote. As described above, this could be done only through the review screen, rather than by pressing a *Back* button, and then scrolling to the contest requiring a change. Many voters became quite lost in this process.

Our third hypothesis suggested that when more actions are required for one system than another, satisfaction will be correspondingly lower for the system requiring more actions. This is indeed what we observed. Users'

subjective experiences varied across the voting systems ($F[5, 150] = 9.16$, $p < 0.001$ for ease of use and $F[5, 200] = 16.63$, $p < 0.001$ for comfort) in almost exactly the same order as did the time and number of actions required to vote. Users rated the manual advance touch screen (Diebold) system easiest to use and felt most comfortable using it while they judged the navigation dial and physical buttons (Hart Intercivic eSlate) the least easy to use and felt the least comfortable using it.

3.3. Individual voting tasks

Voters performed differently across the 22 different voting contests. When we compute mean performance measures for each of the contests on each of the systems, there is an effect of contest for accuracy ($F(21, 861) = 6.98$, $p < 0.001$), number of actions ($F(22, 902) = 39.48$, $p < 0.001$), and duration[9] ($F(22, 902) = 51.22$, $p < 0.001$). These differences can be attributed primarily to poor performance on two tasks: changing a vote (voters were asked to change their vote on the Probate Court Judge contest) and casting a write-in vote relative to the remaining tasks. Although users performed distinctly worse on these tasks than others, they did so for different reasons with the different voting systems. We turn now to a closer examination of the interaction in these low-performance tasks.

In the vote-changing task, voters were instructed to select a candidate (Jeanette Anderson) in the non-partisan race for Probate Court Judge and then change the vote to the other candidate (Kenneth Hager). The voters performed this task relatively inaccurately but, it seems, for different reasons on different systems. Accuracy was quite low with paper ballots (78% for ES&S) for reasons that could reflect some voters' reluctance to obtain a replacement ballot, as instructed on the printed ballot, or even erase their original mark which had been made in pencil, a problem anticipated by the expert review: six of the 42 users voted directly for Kenneth Hager without first voting for Jeanette Anderson. This was considered inaccurate in the sense that they did not follow all instructions although it was accurate in the sense that it fulfilled their final instruction for the task. The remaining two voters did initially vote for Jeanette Anderson, as instructed, but then did not change their vote. It could be that all of these voters, i.e., those who voted first for Kenneth Hager and those who never voted for him, were deterred from making any changes because to do so would have required talking to the experimenter or an election official in an actual election, and potentially reentering all their earlier selections on the replacement ballot. If this was in fact what drove the lack of successful vote change with the paper ballot, then there is a clear downside to paper-based optical scan technology.

This task was also relatively inaccurate, (82% accurate), with the auto-advance touch screen (Avante Vote-Trakker), apparently because voters had difficulty returning to the Probate Court Judge race from the subsequent race, Transit Board Member, to which they were automatically advanced upon selecting the to-be-changed candidate for Probate Court Judge. Of the eight inaccurate votes for Probate Judge with this system, five were due to selection of the wrong race from the Review screen and two resulted from not requesting the Review screen at all. Only one incorrect vote was due to initially voting for Kenneth Hager—in contrast to the paper ballot, where this was quite common.

Finally, users were highly accurate changing votes with the manual advance touch screen (Diebold) and full face (Nedap) systems (98% for both) but the few errors that did occur originated from yet another usability problem. To change a vote with the Diebold and Nedap systems it is necessary to deselect or clear the choice that has already been made by pressing the already-selected choice. One rationale for this design is to clarify users' intentions when a contest requires the user to select two or more candidates. However, the design is used in these systems for changing votes for all contests, most of which require only one selection. Two of the 42 voters could not figure this out with the Diebold system and one could not figure it out with the Nedap system, pressing the selection area next to Kenneth Hager without first deselecting Jeanette Anderson, ultimately leaving the original selection in place. Several other voters were initially stumped by de-selection but eventually succeeded. Similar difficulty with de-selection has also been observed in a field setting (Bederson et al., 2003).

In the write-in task, voters' performance was the least accurate[10] (70%) and the slowest (2.1 min) when using a navigation dial and Enter button (Hart Intercivic system). Nine of 12 incorrect write-ins involved failure to leave a space between first and last names, three involved leaving two spaces, and four involved entering either too many or two few letters in combination with space problems. This is likely due to the design of the on-screen keyboard: letters were ordered alphabetically, not in the QWERTY sequence familiar to most users; the "space" and "backspace" functions were hard to find as they were implemented as small, inconspicuous rectangles at the bottom of the keyboard; and selecting the space button does not produce any visual feedback, i.e., there is no cursor movement, leading some voters to select the space key multiple times. Voters frequently pressed the "Help" button and frequently cancelled the write-in altogether, presumably to wipe the slate clean and start again. Even voters who ultimately entered the correct name encountered similar

[9]The analyses for number of actions and duration include a start-up task before the first contest. So there are 23 tasks in all. The start-up task was not included in the accuracy analysis because the accuracy measure is defined only for actual voting contests.

[10]We adopted a strict definition of accuracy for write-ins: perfect spelling. Election law may well be more lenient in certain jurisdictions.

problems along the way. In fact, only nine of the 42 voters followed the ideal sequence of first selecting the write-in option for a particular contest, entering the appropriate characters without repairs, and submitting the write-in vote.

Voter 38 attempting to write-in "Kay Tyler" for Member of the Library Board using the navigation dial and *Enter* button illustrates many of the usability issues we have been discussing throughout the paper: prolonged duration, off-path actions, inaccuracy, and frustration. After selecting the "write-in" option, she had trouble at almost every step. She pressed the screen, which is not a touch screen, repeatedly and with increasing force. When she managed to highlight letters with the navigation wheel she could not determine how to select them; she would have had to press the physical Enter button. She sought on-line help several times. Eventually, she managed to enter the name but repeated the final "r" and could not determine how to delete unwanted letters. Eventually, she entered four "r"s at end of the name. Her unsolicited comment at this point makes plain her frustration: "Kicking the machine's probably not acceptable." Ultimately she terminated the write-in attempt without voting for anyone. The entire sequence consisted of over 80 actions, not counting the selection of individual letters from the on-screen keyboard. While this is one of the more problem-ridden interactions we observed, it is not the most extreme example.

The kinds of problems voters encountered en route to an incorrect vote seem to be the result of difficulty mapping their plans, e.g., write-in "Kay Tyler," to the particular interface, e.g., dial, Enter button, on-screen keyboard. The voters knew what they wanted to do but could not determine how to do it given the interface choices. This type of interaction has much the flavor of what Rasmussen (Rasmussen, 1986; Reason, 1990) calls knowledge-based control as opposed to skill- or rule-based control. Because people in this situation lack well-honed procedures for performing the task, they try to reason their way through, improvising, and adjusting based on feedback.

Users were slightly better though still relatively poor at writing-in votes on paper and with the auto-advance touch screen (81% accurate for both), but the problems affecting accuracy were quite different in the two systems. In the case of auto-advance touch screen (Avante Vote-Trakker), many of the errors could be traced to the design of the entry fields on the touch screen. For all of the other systems, voters entered the full name of the write-in candidate in a single entry field. However, for this one, the write-in interface required voters to enter first name, middle name, and last name in separate rectangular fields. The interface was designed so that after a voter entered the candidate's first name, he or she could advance to the next field only by first bringing it into focus by touching it. This is equivalent to clicking on a window not in active use in a desktop interface in order to make it active. The expert review flagged this design as likely to create problems for users and this was indeed the case. Many voters seemed unfamiliar with this convention after entering the first name. Fourteen of the 42 voters pressed the space bar at this point, several times in some cases, presumably in an attempt to shift focus to the last name field. The designers' decision to segregate text entry for different components of write-in of candidate names was presumably made to clarify the resulting data. However, this decision seems to have led to voter confusion, long durations ($M = 1.10$ m), many ($M = 13.8$) actions, and, ultimately, failure to write in the name as intended in a relatively high proportion of cases.

Users' errors with the paper ballot (ES&S) were due to the mechanics of optical scan forms. In order to alert the scanner to the presence of a write-in name with these paper ballots, voters must fill in or mark the oval to the left of the write-in field. The expert review again anticipated that this would cause users problems and this was correct. Of the five erroneous write-ins in the current study, four were due to failure to mark the oval and one was due to the opposite problem, marking the bubble but not writing in a name.

Casting a write-in ballot proved difficult on the Zoomable system because the voter was required to press a *Record Write-in* button on the on-screen keyboard, and many voters did not do this. Perhaps they did not notice the button or perhaps they were unclear of its purpose. This is clearly an obstacle to accurately write in votes but different from what was observed with the Avante, Hart Intercivic, and ES&S systems.

The final voting task that we consider is the use of the Voter-Verified Paper Audit Trail. The controversy over the security of electronic voting systems has led many to recommend, and 30 states in the US to mandate (http://www.verifiedvoting.org/), that systems print a paper record of the voting transaction. The thinking is that if an election conducted with electronic voting systems is contested, the votes printed on paper records can be tallied in the recount. Of course this would reintroduce the many problems of counting paper ballots, including the possibility that some could be stolen or added during manual counting (Campbell, 2005), but as a stopgap measure it would allow a complaint to be investigated. A critical piece of the logic is that a voter must verify the printed record or else one cannot be certain that it reflects the voter's intentions. While the approach holds promise, our study indicates that there are usability problems associated with the use of paper records, at least as implemented in the Avante system.[11]

With the Avante system, when voters submit the ballot by pressing the "Cast Ballot" button on the review screen, a paper record of the votes entered on the touch screen is printed and displayed in an enclosed case adjacent to the touch screen. The following message appears on the touch screen "Thank you for Voting! A paper record is presented

[11]Since the time we conducted our study, both the Diebold and Nedap systems have been modified to print paper records.

for your review only. When you are finished reading your record, press the 'Deposit Paper Record' button to deposit your ballot. You may not remove the paper record from the voting system.'' A button labeled *Deposit Paper Record* appears at the bottom of the screen. After 20 s an additional message appears on the screen below the original message, ''Do you need more time to review your voting record?'' and two buttons *Yes to Review* and *No to Deposit* replace the *Deposit Paper Record* button. If the voter does not press one of these buttons within 20 s, the system automatically deposits the ballot without explicit instruction from the voter.

This design presented several problems. First, voters who did inspect the paper record were often looking at the paper record and not at the touch screen when the message appeared on the touch screen asking if they needed more time. The expert review cautioned that for many of these voters the interval allotted for reviewing the paper ballot might time out, preventing them from verifying the paper record. This happened to 16 of the 42 (38%) users in the laboratory, a particularly serious problem because a recount could not unequivocally address the grievance about the original election if some of the printed records are unverified. Another 10 voters pressed the ''Deposit Paper Record'' button without looking at, or even in the direction of, the printed record. Thus, these voters verified the paper record without reviewing it. This also undermines the logic of voter verification, although it is certainly a voter's prerogative not to review the record. Only 11 voters followed the ideal sequence of looking in the direction of the paper record, possibly requesting more time, and then deliberately depositing the printed ballot with a button press. We cannot determine the direction of gaze for the remaining five voters. Thus, the logic of a voter-verified paper trail, while perhaps sound in concept, is far from perfectly implemented in the single design we examined.

3.4. Are these problems unique to e-voting or voting in general?

It is possible that the usability problems we have observed are intrinsic to the voting process, irrespective of the medium by which voters indicate their preferences. However, it seems more likely they are the result of the particular interface designs represented by the sample of voting systems we studied. Certainly if we compare the problems described above to those that have been observed with punch card voting, there is little overlap. For example, none of the problems encountered with punch card voting are caused by auto-advance navigation or voters' failure to select an already-selected candidate in order to change a vote and the best-known problem of punch card voting, ambiguity about voters' selections, is not an issue with digital voting devices because a selection is either registered or it is not. If we compare the problems of touch screen interfaces to systems requiring the voter to mark a paper ballot for optical scanning—a comparison that is possible within our own data—the problems are also relatively different. The most common problem with the paper ballot optical scan system (ES&S) was voters' failure to signal a write-in vote by filling in a write-in oval. This was not a problem with any of the touch screen voting systems because users simply cannot enter write-in votes without first selecting a write-in option of some kind. So the problems we observed seem to arise from designers' errors in creating human–computer interfaces to support voting, not from characteristics of the generic voting task.

4. Conclusion

Usability of electronic voting matters, even if it does not distort the outcome of elections. It may well reduce people's ability to vote as they intend. Based on the current results, slow, effortful performance will reduce their satisfaction with the technology and potentially the likelihood of voting in subsequent elections. If they sense they have voted inaccurately despite the personal importance of having their intentions counted, they may be further disillusioned about the process. However, the inaccuracies that we observed are certainly large enough to alter the outcome of a close election. While there is no guarantee that errors will systematically favor one candidate over another, our field study (Herrnson et al., 2008a) and the Jennings–Buchanan contested election suggest that systemic errors do occur.

Moreover, voter performance may be worse in actual election conditions than in the laboratory. When there is pressure to be quick—for example because long lines of voters are waiting to take their turn or because the appearance of having trouble with the system humiliates the voter—some voters are likely to become flustered, compounding the original interface problem. It is possible that actual voters will care more about the accuracy of their votes than did our participants, but the degree to which our participants persevered suggested they were motivated to be accurate. We oversampled participants who we believed to be at increased risk of finding the systems hard to use but we observed no evidence that their limited computer experience intensified the frequency or consequences of usability problems relative to participants with more computer experience. Our sample size was relatively small but in a much larger sample, Herrnson et al. (2008a, b) found little evidence that individual characteristics affect voting accuracy or satisfaction with electronic voting systems. In sum, there is little reason to expect substantially superior performance or a more satisfying experience in an actual election and reason to believe it will be worse.

What can be done? Several lessons for designers of voting systems can be gleaned from the current study. First, it would be wise for designers to minimize voter effort. Voters may be more tolerant of effort than the participants in the Conrad et al. (2006) and Gray and Fu

Table A1
User characteristics

Race	%	Sex	%	Age	%	Education	%	Annual income ($)	%
White	90.5	Female	64.3	18–24	2.4	High school diploma or GED	14.3	0–14,999	14.3
Black	7.1	Male	35.7	25–34	4.8	Some college, no degree	28.6	15,000–34,999	21.4
Hispanic	2.4			35–49	23.8	4-year degree	19.0	35,000–49,999	19.0
				50–64	40.5	Some post-graduate work	11.9	50,000–64,999	11.9
				65–74	21.4	Master's degree	19.0	65,000–84,999	11.9
				⩾75	7.1	Doctoral degree	7.1	85,000 or more	14.3
								Non-response	7.1

(2004) studies, who were reluctant to click for clarification or move their eyes to obtain needed information. Yet, the users in the current study were very sensitive to the amount of effort. Second, designers would be wise to provide voters with an escape route so that when they are caught in a prolonged and fruitless sequence of actions, they can restore the original state with a single action. This is actually a standard and well-known recommendation in the interface design community (Shneiderman and Plaisant, 2005, p. 75,) but apparently neither well known nor followed by the designers of all of these voting systems.

Finally, at least some manufacturers of voting systems have not yet incorporated usability testing and usability engineering into their development process. If they had, we would not have observed so many problems, many of which were anticipated by the expert review. None of the problems we have observed are particularly difficult to fix. Most if not all can be addressed through the kind of usability engineering techniques that are now common throughout the software industry (e.g., Hix and Hartson, 1993; Nielsen, 1993) and quite similar to our approach in the current study. Improving the usability of voting systems is particularly important because the nature of the task leads people to persevere almost heroically in some instances to cast a vote despite substantial usability obstacles. Compared to other structurally similar tasks in which users are less motivated to finish, e.g., completing on-line survey questionnaires, designers of voting systems are at an advantage: users are relatively likely to forgive their usability oversights. Nonetheless, the great diversity of the voting population and the imperative for all eligible voters to be able to vote their intentions place an extra burden on designers to make the systems usable on the first try by anyone. It is hard to think of another human–computer interaction domain in which all members of an extremely diverse user base need to be able to perform at such high levels. Certainly there is none more central to the workings of democracy.

In the software industry it was not until it became apparent that more usable designs could increase profit that usability engineering really became part of the development process (e.g., Landauer, 1995; Bias and Mayhew, 2005). With electronic voting, some of this incentive exists—election administrators can cancel contracts with manufacturers if they are disappointed with a system's performance, including voters' ability to use the system. But in the end, usability of electronic voting is more than a business issue. It is about enabling all citizens to exercise a fundamental right without error, anxiety, or cynicism. The current research makes it apparent that this threshold has not yet been cleared.

Acknowledgments

This work was supported by National Science Foundation Grant number IIS0306698. We thank the following people for their assistance and advice: Alex Carrick, Aaron Clamage, Wil Dijkstra, Peter Francia, Ralph Franklin, Shweta Jayaprakash, Allison Negrinelli, Rachel Orlowski, Esther Park, Won-ho Park, Whitney Quesenbury, Randy Roberson, Roma Sharma, Mike Toomey, and Dale Vieriegge. In addition we thank the following manufacturers of voting systems for their partnership, and for making available their systems for testing: Hart InterCivic, ES&S, Nedap, and Avante; we thank the Maryland State Board of Elections for making available the Diebold system for testing.

Appendix A

See Table A1.

References

Anderson, J.R., 1983. The Architecture of Cognition. Harvard University Press, Cambridge, MA.

Anderson, J.R., Conrad, F.G., Corbett, A.T., 1989. Skill acquisition and the LISP tutor. Cognitive Science 13, 467–505.

Bederson, B.B., Meyer, J., 1998. Implementing a zooming user interface: experience building Pad++. Software: Practice and Experience 28, 1101–1135.

Bederson, B.B., Lee, B., Sherman, R., Herrnson, P.S., Niemi, R.G., 2003. Electronic voting system usability issues. In: Proceedings of CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters 5(1), 145–152.

Bias, R.G., Mayhew, D.J. (Eds.), 2005. Cost-justifying Usability. An Update for the Internet Age, second ed. Morgan Kauffman, San Francisco, CA.

Brennan Center Task Force on Voting System Security, 2006. The machinery of democracy: protecting elections in an electronic world. Brennan Center for Justice at NYU Law School. Also available at 〈www.brennancenter.org〉.

Campbell, T., 2005. Deliver the Vote: A History of Election Fraud, An American Political Tradition, 1742–2004. Carroll & Graff, New York.

Card, S.K., Moran, T.P., Newell, A., 1983. The Psychology of Human–Computer Interaction. Erlbaum, Hillsdale, NJ.

Conrad, F.G., Couper, M.P., Tourangeau, R., Peytchev, A., 2006. Use and non-use of clarification features in web surveys. Journal of Official Statistics 22, 245–269.

Department of Legislative Services, January 2004. A review of issues related to the Diebold AccuVote-TS Voting System in Maryland. Report presented to Senate Education, Health and Environmental Affairs Committee and House Ways and Means Committee, Maryland General Assembly.

Frisina, L., Herron, M.C., Honaker, J., Lewis, J.B., 2008. Ballot formats, touchscreens, and undervotes: a study of the 2006 midterm elections in Florida. Election Law Journal 7, 25–47.

Gray, W.D., Fu, W., 2004. Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive Science 28, 359–382.

Herrnson, P.S., Niemi, R.G., Hanmer, M.J., Bederson, B.B., Conrad, F.C., Traugott, M.W., 2008a. Voting Technology: The Not-so-simple Act of Casting a Ballot. Brookings, Washington, DC.

Herrnson, P.S., Niemi, R.G., Hanmer, M.J., Francia, P.J., Bederson, B.B., Conrad, F.G., Traugott, M.T., 2008b. Voters' evaluations of electronic voting systems: results from a usability field test. American Politics Research 36, 580–611.

Hix, D., Hartson, H.R., 1993. Developing User Interfaces: Ensuring Usability through Product and Process. Wiley, New York.

Landauer, T.K., 1995. The Trouble with Computers: Usefulness, Usability and Productivity. MIT Press, Cambridge, MA.

Nielsen, J., 1993. Usability Engineering. AP Professional, Boston, MA.

Nielsen, J., 1994. Heuristic evaluation. In: Nielsen, J., Mack, R.L. (Eds.), Usability Inspection Methods. Wiley, New York, pp. 25–62.

Norman, D.A., 1981. Categorization of action slips. Psychological Review 88, 1–15.

Rasmussen, J., 1986. Information Processing and Human–Machine Interaction. North-Holland, Amsterdam.

Reason, J., 1990. Human Error. Cambridge University Press, Cambridge, UK.

Roth, S.K., 1998. Disenfranchised by design: voting systems and the election process. Information Design Journal 9, 1–8.

Rubin, A.V., 2006. Brave New Ballot: The Battle to Safeguard Democracy in the Age of Electronic Voting. Morgan Road Books, New York.

Sanderson, P.M., Fisher, C., 1994. Introduction to this special issue on exploratory sequential data analysis. Human–Computer Interaction 9, 247–250.

Shneiderman, B., Plaisant, C., 2005. Designing the User Interface, fourth ed. Addison Wesley, Boston, MA.

Sinclair, R.C., Mark, M.M., Moore, S.E., Lavis, C.A., Soldat, A.S., 2000. Psychology: an electoral butterfly effect. Nature 408, 665–666.