PROJECT SUMMARY

The events surrounding the 2000 presidential election called attention to a variety of seldom-discussed facts: that voting technology and ballot design can influence election outcomes; that most polling places in the United States employ outdated technology; that training in election administration is inadequate; that minorities and the poor are more likely to cast their ballots on outdated systems; and that voting procedures affect how voters feel about their ability to exercise their right to vote and their willingness to accept the results of an election as legitimate. Despite some improvements, Florida voters had another round of difficulties in their recent primary.

Since 2000, states have commissioned studies, revamped election administration, redesigned ballots, and begun to invest in new voting equipment. Existing research, however, provides little help in understanding current technology and, especially, the interface between voters and various voting systems and ballots. Thus, there is little solid information on which to base massive reforms, significant expenditures, or even the redesign of individual local ballots. The problems associated with the September 2002 primary election in Florida illustrate this.

This project brings together social and computer scientists from a number of disciplines to study voting technology and ballot design. The research team will first assess the impact of existing technology and ballot designs on a variety of factors: the ability of voters to cast their ballots accurately and efficiently, voter ease in casting complete ballots (if so desired), voter comfort using different technology and ballot interfaces, voter confidence that their ballots will be accurately recorded, and the level of voter turnout. They will also assess the effects of alternative voting systems and ballot formats on the frequency of incomplete ballots and split-ticket voting.

The project will use a variety of research designs, data collection methodologies, and analysis techniques, including: laboratory experiments, expert review, close-up observation, field tests, and "natural experiments" that occur as local jurisdictions change their voting technology and procedures. The project will also subject the researcher team's own "zoomable" voting interface design to the same rigorous evaluation applied to existing systems. Finally, the researcher team will create a generalized protocol for testing voting technology and ballot formats that will be disseminated for nationwide use.

The public benefits of this approach will be substantial: a considerable improvement in the totality of the election process, including reduced voter frustration, increased voter confidence in elections, the casting of more completed ballots, and, possibly, an increase in voter turnout. Scholarly benefits include improved understanding of human-computer interaction as it relates to voting behavior and commencement of a new approach to data collection and analysis in the study of voting technology, ballots, and voting behavior.

The project will be guided by regular interaction with practitioners and policy makers who have responsibility for the effective administration of the electoral system in the United States. Regular interaction will allow these individuals and agencies to disseminate information about the project, encourage potential beneficiaries to review the findings and use the testing protocol,

communicate project findings and recommendations to companies that manufacture voting machines, and collect and analyze data from those who use the protocol.

PROJECT DESCRIPTION

I. Background

A. *Statement of the Problem*

A major lesson derived from the 2000 presidential election is that the manner in which voters cast their ballots is important. Voting technology and ballot design can influence election outcomes, affect how voters feel about their ability to exercise their right to vote, and influence voters' willingness to accept the results of an election as legitimate. The election also showed that most polling places nationwide employ outdated technology, including unreliable punch-card ballots and mechanical lever machines; only a third of the electorate used modern computerized technology, such as optical scanning systems or direct recording electronic (DRE) systems with ATM-style touch-screen voting (Election Data Services 1998; Federal Election Commission 2000). And, because the poor and ethnic and racial minorities were more likely to cast their ballots on outdated systems, their votes were among the least likely to be counted.

States have responded to the problems associated with the 2000 elections by commissioning studies, revamping election administration, redesigning ballots, and, in some cases, by investing heavily in new voting equipment. The Florida experience in 2002, however, indicates that major problems still occur—in voting equipment, training of poll workers, and ballot design. Moreover, there is little solid information overall about the interface between voters, various voting systems, and ballots on which to base or evaluate the success of the massive reform and significant expenditures that are coming with the likely passage of the Ney/Hoyer "Help America Vote" Act.

This project brings together social and computer scientists from a number of disciplines to study these problems. Their efforts will be guided by regular interaction with a number of practitioners and policy makers who have a wide range of responsibilities for the effective administration of the electoral system in the United States. The research team will first assess the impact of existing voting technology and ballot designs on a variety of factors related to voting: the ability of voters to cast their ballots accurately and efficiently, voter ease in casting complete ballots (if so desired), voter comfort in using different technology and ballot interfaces, voter confidence that their ballots will be accurately recorded, and the level of voter turnout. We will also assess the effects of alternative voting systems and ballot formats on the frequency of incomplete ballots and split-ticket voting.

The project will use a variety of research designs, data collection methodologies, and analysis techniques. It will employ laboratory experiments to ascertain the underlying causal mechanisms that affect voters' use and evaluation of voting systems. These results will be compared to actual experiences with these systems through a series of natural experiments that occur as local jurisdictions change their voting technology and procedures. In addition, we will study voting systems and ballot formats by means of expert review, close-up observation, and field tests. We will also subject a "zoomable" voting interface of our own design to this same rigorous evaluation. Finally, as part of a set of outreach efforts, we will create a generalized

protocol for testing voting technology and ballot formats that will be disseminated for nationwide use so that problems such as those observed in 2000 and 2002 will not reoccur.

The public benefits of this approach are straightforward but powerful: nothing less than a substantial improvement in the totality of the election process, including reduced voter frustration, increased voter confidence in elections, the casting of more completed ballots, and, possibly, an increase in voter turnout. The immediate scholarly benefits include improved understanding of human-computer interaction as it relates to voting behavior. The long-term scholarly benefits include the commencement of a new approach to data collection and analysis in the study of voting technology, ballots, and voting behavior.

This proposal is supported by the Department of Labor, and [possibly others]. By partnering with these agencies, we will be able to draw on their expertise and experience, use the Bureau of Labor Statistics' usability laboratory for research, disseminate information about the project and its results, encourage potential beneficiaries to review our findings and use our protocol, communicate our findings and recommendations to companies that manufacture voting machines, and collect and analyze data from those who use our protocol. (See the appendix for letters of support.)

B. *Prior Research*

There is a long history of research in political science on the impact of procedural reform on elections. It has focused on changes in the form of ballots and shifts in the administrative procedures for casting them. Ballot reform was a frequent topic of papers and articles in the earliest days of the discipline—including prescriptions about excessive length and lack of uniformity, but also evidence of the effects of ballot design on roll-off and split-ticket voting (Allen 1906; Beard 1909).[1] Fifty years later, there was a flurry of new studies on the relationship of ballot formats to roll-off and split-ticket voting, often taking advantage of new survey data. Since then, there has been a small, but continuing series of studies on these effects along with the potential effects of candidate name order.

These studies generally conclude that office-bloc ballots result in greater roll-off than party-column or party-row ballots (Key 1956, 212; Walker 1966),[2] but that the provision of straight-party circles or levers tends to reduce roll-off, at least in partisan contests (Robinson and Standing 1960; Walker 1966; Nichols 1998, 109-10; Kimball, Owens, and McAndrew 2002). More recently, studies have shown that electronic voting machines result in less roll-off, presumably because they alert voters to whether they have completed the ballot (Nichols and Strizek 1995; Nichols 1998), and that voting by mail does not significantly alter roll-off

[1] Roll-off is the failure to cast votes for some offices on the ballot—usually offices below those at the top of the ballot or for ballot propositions. It is typically measured as the percentage of those who vote for the top office but not for a given lower office. Split-ticket voting is the decision to cast votes for more than one political party. Where actual ballots are not available (which is typically the case), it is estimated from surveys of voters or, sometimes, by calculating the difference between the percentage of votes for a party's most and least popular candidate.

[2] An office-bloc ballot lists candidates for each contest (e.g., governor) in a bloc. Often the candidates are listed alphabetically. It is in contrast to the party-column (or party row) ballot, which lists all candidates for a given party listed under a single, party heading. Often, but not always, party-column ballots have a circle or other device that allows a voter, with one mark, to vote a straight ticket (i.e., for all members of the party).

percentages (Hanmer and Traugott 2002). Studies have also shown the party-column ballot encourages straight-ticket voting in comparison to the office-bloc ballot (Bain and Hecock 1957; Campbell, Converse, Miller, and Stokes 1960, ch. 11; Rusk 1970).

Studies of ballot order effects often report that candidates listed first on the ballot are favored, at least in nonpartisan and non-salient elections (e.g., Bain and Hecock 1957; Mueller 1970; Taebel 1975; Darcy and McAllister 1990, Darcy 1996; Miller and Krosnick 1998; Krosnick. Miller, and Tichy forthcoming). Bowler, Donovan, and Happ (1992, 564-65) also found that ballot propositions were less favored the further down they appeared on the ballot.

Research has also focused on the voting-by-mail (VBM) procedures adopted recently in Oregon. Using the behavioral histories of individual voters from before and after the introduction of VBM, a significant "selection bias" has been found; the reforms are much more likely to keep past voters in the system than to mobilize new voters (Berinsky, Burns, and Traugott 1999). A result of this bias is that the demographic and other characteristics of the electorate have not changed under VBM. Nor is there a demonstrable impact of VBM on roll-off or drop-off (Hanmer and Traugott 2002) or on straight or split-ticket voting (Traugott and Hanmer 2001).

In the aftermath of the 2000 election, a number of studies evaluated various aspects of the Florida vote, somewhat altering the focus of investigation in the process. Analyses showed, for example, that there were many more "overvotes" in Palm Beach County where the butterfly ballot design was used (Wand et al., 2001).[3] The U.S. Commission on Civil Rights (2001) concluded that poorer and minority communities more often utilized less modern equipment that is prone to overvotes and other errors. The Commission also noted a failure of the state to educate voters on the mechanics of voting, [4] a problem that seems to have reoccurred in their 2002 primary.

The largest post-2000 study was conducted by an interdisciplinary group of voting and computer specialists from the California Institute of Technology and the Massachusetts Institute of Technology. The report from this group found, surprisingly, that "residual votes"[5] were typically greater in jurisdictions using electronic voting than other kinds of machines (with the exception of punch cards), even when confounding factors were controlled (Caltech/MIT Voting Technology Project 2001). Kimball et al. (2002) also found that electronic voting machines produced a slightly higher residual vote than optical scan ballots.

Only rarely have researchers considered whether or how ballot features might confuse voters. In one noteworthy though very small-scale comparison, it was found that the labeling of rows on lever machines resulted in considerably different ballot order effects (Bain and Hecock 1957). In another study it was found that placing a salient race at the bottom of the ballot caused some voters not to cast a ballot—though a confusing straight-party device may also have contributed to this decline (Darcy and Schneider 1989), especially as Bullock and Dunn (1996) found no such

[3] Overvotes occur when individuals cast votes for more candidates than are to be elected for a given office (typically one).

[4] It should be noted that in May 2001, Florida passed legislation to purchase new voting machines and to provide $6 million for voter education. See http://www.ncsl.org/programs/legman/elect/taskfc/flbill.htm.

[5] Residual votes were defined as ballots on which no presidential vote was counted—because no vote was cast or because of human or machine error.

effect. The fact that particular demographic groups, including the elderly, poor, and uneducated, are more likely to cast incomplete ballots also suggests the possibility of confusion, though indifference and lack of knowledge about the candidates might also explain these results (Walker 1966; Darcy and Schneider 1989; Nichols and Strizek 1995; Nichols 1998; Brady et al. 2001). The only experimental studies of the voting process, while very small scale, also revealed some confusion on the part of voters (Roth 1998; Sinclair et al. 2000).

Finally, the events surrounding the 2000 presidential election highlighted an aspect of voting that has not been dealt with since the introduction of the Australian (secret) ballot at the end of the 19th century—namely, that voting technology and ballot design affect how voters feel about their ability to exercise their right to vote and influence voters' willingness to accept the legitimacy of election results. Thus, despite high quality research on turnout, roll-off, split-ticket voting, and order effects, there is little information about how to reform voting technology and ballot design in ways that will develop, encourage, and support perceptions of the voting process as an accurate and fair reflection of voters' intentions.

Because of the interdisciplinary nature of this proposal, relevant background material is also found in a collection of techniques and activities known as "usability engineering" or "user-centered system design." It has increasingly (though not universally) been recognized by the software development community that there are significant costs associated with computer systems that are difficult and frustrating to use. In response, usability engineering is becoming a standard part of the development process for new human-computer interfaces. The point of these approaches is to build systems from the perspective of the user rather than that of the system developer. Book-length introductions to the topic (e.g., Mayhew 1999; Nielsen 1993; 2000; Norman and Draper 1986) describe the rationale for usability studies as well as providing detailed information on methods for carrying them out.

There are a number of general design principles that are frequently used to motivate and analyze interface usability. For example, Nielsen (1993, 120) notes that "Less is More," meaning that it is important to give users no more options than necessary since more options require the user to make more complex decisions. Another principle is that of consistency (Nielsen 1993, 132). If the interface uses the same language to mean the same action from one part of the program to another, the user will come to learn that language and will be able to understand new parts of the program more easily. This is related to the well-known fact that interfaces that rely on user's memory tend to have problems (Shneiderman 1997), so minimizing the user's memory load is important as well. A set of such principles, specifically related to online survey practices, has recently been articulated by Terhanian, Bremer, and Salomon (2000).[6]

In addition to analyzing the user's task, an integral part of building user-friendly interfaces involves tests, where testing is usually understood to mean experimental sessions in which users interact with a prototype version or paper mock-up of the software (Redmond-Pyle and Moore

[6] One specific practice is closely related to ballot design. With traditional paper questionnaires, designers are often conflicted about whether or not to present a "don't know" option; including it might encourage "don't' know" responses, but omitting it makes it difficult to distinguish between "don't know" answers and simple errors of omission. In online surveys, designers can allow respondents to advance without answering but only after a message appears that calls attention to the fact that they have not answered a question (and, if desired, calls attention to the importance of respondents' opinions) (DeRouvray and Couper 2002).

1995). Testing sessions are often videotaped in order to provide a visual record of problems encountered by users. The problems are sometimes quantified with measures such as their frequency of occurrence or mean duration. Think-aloud protocols—i.e., asking users to report verbally on their thinking as they interact with the system (or immediately afterwards)—are also used to illuminate users' intentions in a way that recording actions alone cannot.[7] Likewise, a common practice is to ask users to react to alternative design proposals. Other methods include event level analyses, such as counts of user actions like backtracking, changing previous actions, pulling down several menus in search of the intended one, inactivity accompanied by some suggestion of uncertainty, confusion, or frustration such as gestures, verbal and facial expressions, and so forth (Mayhew 1999). To assess subjective experience, it is typical to administer a satisfaction questionnaire immediately after the session.

Importantly, with respect to recruiting users, there are several published arguments that a small number of users (e.g., 4-6) is enough to identify most unique usability problems (e.g. Hix and Hartson, 1993; Nielsen and Levi 1994). If the user base is particularly heterogeneous—as is the case for voters—then a number of groups of 4-6 users should be recruited to reflect the heterogeneity of the population.

C. *Recent Election Reforms*

Following the 2000 elections, numerous states began to explore the possibility of upgrading their voting systems and procedures. The National Conference of State Legislatures [NCSL] (2002) reports that as of July 19, 2002, 15 states have committed to purchasing new voting equipment before the 2002 or 2004 general elections and have appropriated more than $236 million to purchase optical scan and touch screen voting systems. Six states legislated new requirements for voting system standards and testing, and six states banned punch card voting machines. NCSL also reports that at least 16 states have tightened their procedures for counting votes, including specifying how to count ballots with hanging, dimpled or pregnant chads, or stray marks.

Based on changes that have already taken place, there have been a number of formal and informal reports of voting under the new systems. In a number of instances, the new systems seem to have been highly successful (U.S. House of Representatives 2001; Phillips 2001; Goetzl; 2002), though problems were detected in new machines shipped to Florida (Barton 2002; Wyman 2002), and the 2002 primary election experience in that state was fraught with difficulties (Canedy 2002; Merzer et al. 2002). Likewise, an examination of a system proposed for Maryland showed that considerable improvements could be made (Bederson and Herrnson 2002). In general, inadequate attention to human factors in designing new voting systems has proven to be a significant impediment to smoothly functioning systems (Mintz and Slevin 2001).

These tests are a step in the right direction. However, the number of elections using new systems is limited, and the education efforts surrounding them have generally been intense. How the new systems will work when put into place on a large scale is not fully determined. Florida's initial

[7] There is a long tradition of using think aloud techniques in psychology to illuminate how people solve problems and engage in other types of high-level cognition (e.g., Ericsson and Simon 1993); however, the method has been controversial (Crutcher 1994).

experience was disheartening; another trial comes in November 2002, in Georgia, where a statewide system of DRE machines has been adopted with considerable education efforts (www.georgiacounts.com). In addition, many jurisdictions still need to make decisions about whether or not to replace their current systems and, if so, which type of system and which particular company's product to purchase.[8] Moreover, with respect to ballot formats, almost no new research has been conducted apart from assessments of the failure of the butterfly ballot used in Palm Beach County, Florida in 2000. Given the large number of ballot formats needed for nationwide coverage, it remains to be seen whether successful designs can be implemented in virtually all jurisdictions.[9]

II. Proposed Research

We intend to produce a set of generalizations about the effects of contemporary voting technologies and ballots that is both currently applicable and that will take account of changes that will occur in the future. Though there is no standard set of evaluation criteria for assessing the quality or effectiveness of any particular reform, we will rely on a generally accepted set of normative principles to guide our efforts. For any given comparison between alternative systems, these include:

1. Accuracy—more accurate tabulation, within a required level of accuracy, is preferable;
2. Reduction in the number of spoiled ballots and residual votes—a system that minimizes spoiled ballots through overvotes and unintended undervotes is preferable[10];
3. Faithful recording of voter intent—a system that records voters' preferences as they intend them is preferable;
4. Ease of use—a system that is easier to use and less confusing is preferable;
5. Satisfaction with the voting experience—a system that makes voters more comfortable and confident that their votes are recorded is preferable;
6. Flexibility—a system that does not impede a voter's wish to cast a straight ticket or to vote for "lower" ballot offices is preferable.

In order to carry out this research, the project draws upon the expertise of an interdisciplinary team of political scientists, human factors psychologists, design engineers, computer scientists, and election officials to achieve the following goals from the perspective of human factors ("usability") considerations:

1. Test existing voting technology and ballots
2. Create and test new voting technology and ballots.
3. Develop guidelines for voting system manufacturers and election officials.
4. Develop a protocol for testing voting technology and ballots

---

[8] One might recall, for example, the surprising result of the CalTech./MIT Voting Technology Project (2001) that DRE machines had a higher residual vote than even lever machines.

[9] The fact that any given ballot is used in a relatively small area (and that the effects of poor design are likely to be subtle) also means that success or failure of individual ballot formats will be harder to detect than that of voting systems. Thus, it is unlikely that we will see many studies that purport to test voter response to the ballot itself.

[10] Undervoting occurs when individuals—perhaps intentionally—fail to vote for a given office or proposition. Undervoting, for whatever reason, does not invalidate the ballot. See note 3 for overvoting.

5. Disseminate findings and archive data on voting technology, ballot designs, and election results.

We will employ several methods of data collection and testing, as well as several new analytical approaches to data analysis, in order to deal effectively with issues of internal and external validity. In this section we outline the variety of methods of study along with the goals of the research project.

Test Existing Voting Technology and Ballots

A basic principle underlying our work is that no system should be introduced until it has undergone thorough testing. Thus, we propose to test both existing and newly designed systems. These tests will be directed toward Principles 1-3 (accuracy, reduction in the number of spoiled ballots, and faithful recording of voter intent). The same testing procedures will apply to voting machines and ballots, as each aspect of voting needs to be evaluated independently. Four distinct types of operations will be used in the testing phase.

Expert Review. The PIs, along with consultants who have experience in computer interface design, voting technology, and ballot construction, will do a usability study by walking through the system in detail, performing representative tasks, and recording where weaknesses occur. This approach takes advantage of experts' knowledge of scientific research about what works and what doesn't, along with practical understanding of existing designs derived from experience with many different kinds of user interfaces (e.g., ballots, questionnaires, electronic forms, instructional systems, computer games, and other computer interfaces). In addition to pointing out problems, an advantage of this approach is that experts can make concrete suggestions about how to improve problematic aspects of design, including:

- Clarity and consistency of terminology
- Graphics quality
- Quality of instructions and help commands
- Feedback or warnings for undervoting or overvoting
- Presence of on-screen distractions
- Ease of inserting or removing voting cards (used with DREs, after a voter is judged eligible to vote)
- Adequacy of review screens
- Quality of paper or on-screen ballots
- Privacy afforded to voters
- Ability to accommodate disabled voters

Experience with expert review was gained in a small-scale study carried out in cooperation with officials from several counties in Maryland (Bederson and Herrnson 2002) that were considering the purchase of a DRE system. Experts were able to point out numerous shortcomings, many of which could be corrected by small changes in the equipment tested. The observations applied both to the equipment itself and to the facsimile ballot created by the equipment manufacturer.

Laboratory Testing. Testing voting processes in a human factors laboratory produces detailed observations of voters' interactions and reactions to a variety of voting equipment and ballots. This work is directed toward Principles 3-5 (faithful recording of voter intent, ease of use, and voter satisfaction). The PIs will measure voters' intentions by having voters think aloud as they interact with voting technology, thereby providing a continuous verbal account of their intentions, goals, plans for achieving those goals, and responses to the process more overt. Voters' intentions also will be measured by collecting converging evidence, such as recording the location and amount of force used for each action on a touch screen (e.g., whether the point of contact is in the center or the periphery of a button), or directing the voters to write their intended action (e.g., to vote for candidate X) before exposing them to the user interface.

Recording voter intent, which was at the center of the debate about hanging and dimpled chads in the 2000 Florida vote count, will enable us to determine whether voter error was due to "slips" or "mistakes" (Norman 1981). A slip is an unintended action like a stray finger movement grazing a touch screen; a mistake is an intended but erroneous action like deliberately pressing the "submit" button after choosing a candidate for the first race because the user believes that each vote must be submitted separately.

Of particular concern with the introduction of computers to the voting process is whether voters believe their votes will be securely, accurately, and privately recorded. Our research will include laboratory experiments to address this concern directly. In one such approach, participants will interact with voting machines in four different "vote recording" scenarios that mimic the way in which votes are transmitted from precincts to the county level: 1) vote totals for each machine are printed out, placed in an envelope, sealed, and sent by courier to a central facility to be tallied; 2) voting machines in a precinct are linked and all the votes are sent to a local server where they are printed out and placed in the same kind of envelope as in the first case; 3) the votes at that polling station are sent electronically to a central facility to be tallied; and 4) their vote is immediately recorded on an Internet website.

In addition to the above we will use the laboratory setting to:

- Collect early feedback on the voting technology we develop, enabling us to revise the design prior to field testing. We plan to iterate between the lab and the field during the developmental process.
- Learn about areas we need to examine using other research techniques and develop the research protocols used in the usability and field tests.

Usability and Field Tests. Apart from the highly structured setting of a laboratory, we also will videotape non-experts as they proceed through the voting process in a more real-world-like scenario. This effort is also devoted to Principles 3-5. Usability testing is a close-up observational approach that does not require a specialized laboratory, but it is useful in addressing some of the same issues as in the laboratory tests. We will videotape individuals as they proceed through the voting process, asking them to describe what they are thinking while they vote so we can record their responses to different aspects of the process and identify shortcomings in voting machines and ballot technology, including:

- Time spent reading instructions
- Voter responses to paper or on-screen ballot
- Voter responses to the reporting of undervoting or overvoting
- Difficulties in changing a vote
- Complications and malfunctions of DRE or of optical scan readers

While many of the problems associated with voting technology and ballots will be apparent from expert testing, laboratory studies, and usability tests, field tests are required to be certain that our observations can be extended to ordinary individuals in circumstances similar to those faced by voters on election day. The field tests will involve a larger and more heterogeneous group in more natural settings. They will involve three components: observation of voters using systems, administration of post-voting questionnaires, and data analysis.

We expect to set up voting booths in varied locations, most likely in Maryland, upstate New York, and Michigan. A variety of venues will be chosen. For example, urban indoor shopping malls attract varied populations. Likewise, certain malls (that have on-going research operations) have locations that attract high-SES, suburban whites and other locations that attract lower-SES, urban and suburban whites and minorities. In addition, young, inexperienced populations are readily available (at universities) as are older populations (at adult care facilities). Libraries attract individuals of many ages.

For ballots, it is important to simulate the situation faced by actual voters. Thus we do not propose (as in one manufacturer's simulation) to use a mock ballot asking people to vote for their favorite sports figure (from three or four on the ballot) or for their favorite actress. Rather, we intend to simulate voting for a typical mix of real public offices (using an introduction similar to that used in many surveys, such as "if the election were held today"). For example, we will ask individuals to vote for president (with several choices rather than simply the two major parties), Congress, a local race, and so forth. Variations in format will be tested, including office-bloc and party-column formats, paper ballots (simulating voting by mail), multimember district elections, offices for which more than one candidate is selected (such as school board elections), and ballots with referenda. As in a real election, it is expected that voters will be familiar with some candidates (especially at the top of the ticket, along with incumbents) and unfamiliar with other candidates and even the office for which they are running. Also as in real elections, voters will not be forced to complete the entire ballot. Likewise, we will employ randomization to avoid order effects.

The outcome measures will be more limited than in a laboratory setting. In some cases we will ask voters to tell us immediately (by interrupting their simulated vote) if they encounter problems. In other cases, we will ask them to fill out a brief questionnaire or to take part in a brief interview after the simulated vote. We will look for any signs of confusion or perceived difficulty in voting. We will keep track of how long it takes individuals to vote. In addition to identifying shortcomings similar to those observed with the previous approaches, field tests will help us to record a voting system's ability to hold up under repeated use and learn about voters' impressions, including:

- Comfort level using the system

- Ability to read characters on screen or paper ballot
- Understanding of instructions and terminology
- Ability to change a vote or correct mistakes
- Trust in the system
- Overall impressions of voting
- Confidence that one cast one's vote as intended
- Confidence that one's vote will be accurately recorded

The information collected will indicate specific areas where general and targeted outreach and educational efforts may be needed to introduce new technology and ballot designs to both voters and election officials, identify elements of the population that may need special assistance, and help determine the number, if any, of voting machines that should be made available for practice at individual polling sites.

Natural Experiments. Natural experiments involve the observation of changes in the real world and the consequences of those changes. This effort is devoted to Principles 2 and 6 (reduction in spoiled ballots and maintaining flexibility). The introduction of new voting systems creates a situation akin to a laboratory experiment in which prior behavior can be observed (voting under the old system), an intervention is made (the new system), and new behavior is observed (voting under the new system). Because it is a "natural" experiment, lacking in randomization and other features of the laboratory, data are often analyzed using so-called interrupted time series models (e.g., Ostrom 1990; Hamilton 1994). However, the use of matching control groups (similar jurisdictions where no change in technology or administrative procedures took place) is also critical to establishing confidence in the causal nature of observed change. This adds a cross-sectional component to the research (Beck and Katz 1995). Collecting information about the citizenry that resides in different voting precincts enables us to control for voter demographics, different registration systems, and other elements of political context.

Federalism and the decentralization of election administration make natural experiments possible and useful. They can be used to analyze the extraordinary variation in voting technology, ballot formats, and election procedures (including voting by mail). Moreover, some states and jurisdictions have approved a variety of electoral reforms in response to the 2000 election, while others are waiting to assess the effectiveness and costs of the solutions others implement. With advance knowledge of the location and types of reforms planned, we can conduct pre- and post-reform analyses controlling for the characteristics of the electorate and other aspects of the electoral setting.

An additional analysis strategy is based on the availability of appropriate data resources for analysis over time, including access to individual vote histories (allowing us to take individual behavior into account and control for heterogeneity in the electorate), ballot images for a set of elections (allowing us to investigate behavioral consequences such as roll-off, drop-off,[11] and straight- and split-ticket voting), and demographic and contextual information available at the precinct and county levels to permit ecological analysis with some confidence (Thomsen 1987; Thomsen, Berglund, and Worlund 1991; King 1997).

[11] Drop-off is defined as the decline in turnout between high and low salience elections—especially between presidential elections and subsequent off-year congressional elections. See note 1 for roll-off.

States where individual vote histories are available include Colorado, Georgia, Kentucky, Louisiana, Minnesota, North Carolina, and South Carolina. While several studies using voter history files have been conducted, this resource has not been fully tapped (Sigelman et al.1985; Sigelman and Jewell 1986; Tomz and Van Houweling 2001). During the first year of the study we will collect data from several states where legislation has banned certain voting systems or made funds available for purchasing new systems, including Arizona, California, Florida, Georgia, Indiana, Louisiana, Maryland, Minnesota, Missouri, North Carolina, and Texas. We will collect data for the 1996 through 2000 elections and then continue the data collection activity throughout the project life as new procedures and technologies are employed. For example, we have a copy of all of the ballot images from the 2000 presidential election in Florida. This will form the basis of an analysis of how changes in technology introduced in 2002 affect voting behavior in that state. We will be able to look at how the adoption of different technologies affects straight- and split-ticket voting patterns, roll-off, and drop-off. It is insufficient to look only at the 2002 election because novelty and media coverage will affect behavior, and an off-year election (without a presidential race at the top of the ballot) will make a difference as well. Thus, we will use the 2000 results as a base from which to trace change forward through the 2004 election and eventually farther into the future.

Natural experiments will enable us to assess:

- The impact of new voting technologies, ballot designs, and procedures (voting by mail) on spoiled votes, residual votes, roll-off, split-ticket voting, drop-off, and turnout.
- How the impact of new voting technologies, ballot designs, and procedures are conditioned by voter demographics and other electoral circumstances.
- How voter education efforts contribute to understanding of new technologies.
- Whether there is a "novelty" effect that comes from heavy media coverage of the initial implementation of reforms that wears off as new procedures and technology become fully implemented over time.
- Whether there are implementation issues that can be categorized with an eye toward developing programs to alert and train election administrators in how to deal with them.

Create and Test New Voting Technology and Ballots

There is a trade-off between control and realism in any kind of testing. In understanding voting system usability, the issue is between understanding general principles of interface design effectiveness, on the one hand, and evaluating specific commercially implemented interfaces on the other.

While it is useful to evaluate existing systems, we believe it is just as useful to understand general principles, especially for computer-based DRE systems. Existing systems are a combination of many different design elements, and they change from version to version, with different versions sharing some elements but not others. Any evaluation we could make of a specific system would quickly become obsolete and have limited lasting value. If we test individual interface components, we can develop an understanding of general principles and specific features that will enable us to make more precise recommendations to manufacturers and

purchasers. In order to accomplish this we will perform all DRE tests using software systems that we build from scratch. This will enable us not only to have complete control over the interface, but also to perform electronic logging that will enable us to analyze the use of the interfaces more precisely.

Existing DRE system interfaces are simple enough that we will be able to duplicate the interfaces in commercial systems without difficulty. One of the PIs has significant experience in software development and construction of user interfaces (Bederson 2001; Bederson et al. 2000). This experience will enable us to go one step further by developing and testing innovative interfaces, which can hopefully influence future voting system manufacturers.

We have already started experimenting with new interface designs, and one based on "zooming" seems particularly promising. One of the reasons that using computers to vote is challenging is because some elections involve many offices, and the information cannot all fit on the screen at once. Most existing DREs have several "pages" of election races that the user must step through sequentially, and then look at a summary page at the end. This is problematic because users never see an overview of the entire election, and thus use their short-term memory to relate races on one page to those on another (contrary to one of the design principles noted above). By using a "zoomable user interface," users can start with an overview of the entire election, and zoom in to specific races in context. The user can relate the individual races to each other by simply zooming out at any point.

We created an interactive prototype of a zoomable voting system, viewable on the web at http://www.cs.umd.edu/~bederson/voting/. We have built similar zoomable user interfaces in many contexts, and have shown them to have demonstrable advantages over more traditional interfaces (Bederson and Boltman 1999; Hornbaek et al. 2002).

Develop Guidelines for Voting System Manufacturers and Election Officials

We intend to develop and to disseminate widely a number of guidelines for voting equipment manufacturers and ballot designers. These will of necessity take the form of general statements. For example, we may find that confusion is generally minimized if voters can check on whom they have already voted for at any time during the voting process (on a DRE machine) or if ballots (for optical-scanning systems) are shaped like an ordinary 8½ x 11 or legal-sized paper (rather than odd shapes with cut-off corners, etc., as on current ballots). We anticipate that general principles of this sort will be derived from our laboratory and field testing. The closest analogy, perhaps, is with the construction of questionnaires for public opinion surveys, for which there are certain rules of thumb or suggested formats, such as the procedures recommended to avoid various kinds of "response sets" (Sudman and Bradburn 1986).

Develop a Protocol for Testing Voting Technology and Ballots

In the present environment, new voting technologies (especially new variants of DRE machines) are likely to be developed over the next 2 to 5 years or more. And ballot formats change in every election—the specific candidates to be sure, but also the number and type of ballot propositions, and often even the offices to be filled (because of staggered state and U.S. senate candidates).

Thus, even with the development of general guidelines, there is a need for continuing testing of voting equipment and ballot formats.

It would be impractical to establish a center whose mission is to test all new voting machines, ballots, and combinations thereof for all the nation's election officials. As an alternative, we intend to provide these officials with a protocol for such testing. The protocol will mirror the test procedures described above—i.e., a combination of expert review, laboratory testing, close-up observation, and field tests, recognizing that not every scholar, manufacturer, or election official could carry out all forms of research. We will make recommendations for the amount of and kind of testing that should be undertaken, and for the way in which each procedure should be carried out. The protocol will enable election officials to quickly assess the technology and ballots they intend to use on election day. The recommendations will be based on our own testing experience but will be designed so that the procedures can be carried out by persons (especially election officials) who do not have a background in test procedures. We will advise election officials that some tests, such as those for new technology, are best carried out when they involve individuals from many jurisdictions and involve the assistance of experts available at local universities and agencies; other tests, such as those for a slightly modified ballot, may involve testing that can take place at shopping malls, libraries, and senior citizen centers.

Disseminate Findings and Archive Data on Voting Technology, Ballot Designs, and Election Results

The nature of the subject and the events surrounding the 2000 election guarantee that the proposed research will be of interest to elected officials, the media, political parties, interest groups, election administrators, voting systems manufacturers, citizens, and researchers. We recognize that significant outreach will be needed to attract the attention of some of our intended audiences. We propose to present our work at a variety of professional and public forums. In addition to public presentations, publications in trade and academic journals, and the posting of results on websites, we will use the partnerships we have created with federal agencies and nongovernmental organizations to help us disseminate information about the project to election officials and to companies that manufacture voting machines. We plan to encourage these groups to review our findings, to use our protocol for continuing evaluation of voting systems and ballot formats, and to provide us with their recommendations and the data they collect so we can refine our results.

We will archive all of the data collected for this project at the Inter-university Consortium for Political and Social Research (ICPSR) in Ann Arbor, Michigan, and make it publicly available on CD-ROM and through zipped files that can be downloaded from a website.

Time Table

June 1-Dec 31, 2003

- acquire voting equipment and a nationwide sample of ballots
- develop prototype for electronic ballot for DRE machines
- develop prototype for ballot for optical scan machines
- plan laboratory and field tests

- collect election data on roll-off, straight/split ticket voting, spoiled ballots (from natural experiments)

Jan 1-May 31, 2004
- laboratory testing for DRE and optical scan voting machines and ballots
- close-up observation for DRE and optical scan voting machines and ballots
- plan field tests and natural experiments
- begin analyzing data from natural experiments

- June 1-Dec 31, 2004
- field test DRE and optical scan voting machines and ballots
- continue collection of data on roll-off, straight/split ticket voting, spoiled ballots (from natural experiments)
- collect survey data on voter satisfaction with new voting procedures

Jan 1-May 31, 2005
- analyze data from field tests and surveys
- analyzing data from natural experiments
- work with partners to develop the testing protocol
- distribute protocol to selected election officials

June 1-Dec 31, 2005
- continue analyzing data from natural experiments
- broad distribution of testing protocol to state and local election officials
- collect and compile feedback on testing protocol from state and local election officials
- field tests of protocol and of ballots/machines in states holding odd-year elections

Jan 1-May 31, 2006
- disseminate information on results to state and local election officials and to the public
- write scholarly articles

III. Significance of the Project

As the election of 2000 demonstrated, voting technology, ballot designs, and election procedures are subjects of concern to citizens, elected officials, election administrators, and scholars. The principal investigators for the proposed project are an interdisciplinary team drawn from political science, computer science, psychology, and usability studies at a number of universities. The project will involve educating and training graduate students in these diverse fields; creating an infrastructure for research in several areas; collecting, archiving, and disseminating new data; developing design guidelines and a testing protocol that will be broadly distributed to election administrators and voting systems manufacturers both directly by us and by the governmental agencies and nongovernmental organizations that are partners on the project. Because we will ask those using our protocol to share their results with us, the project also involves meta-research and the collection and analysis of meta-data.

The proposed project focuses on a public policy issue of major importance. Voting is the defining feature of a democracy, and the 2000 election suggests the need to develop ways to assess and improve voting technology and ballot designs in the United States. We must see to it that new voting procedures conform to the high standards included in legislation that recently has passed or is under consideration in state legislatures and Congress. We must also instill confidence in citizens that their votes are cast, transmitted, and counted properly, in secret, and accurately. Moreover, to ensure that no group is disenfranchised, voting technology and ballot formats must take into consideration the capabilities and needs of individuals from diverse backgrounds, including less educated voters, non-native English speakers, individuals with disabilities, and members of traditionally underrepresented groups. This research promises to test existing processes and to develop, test, and disseminate design principles that will lead to these kinds of improvements in the way Americans vote.

REFERENCES

Allen, Philip Loring. 1906. "Ballot Laws and Their Workings." *Political Science Quarterly* 1:38-58.

Bain, Henry M., and Donald S. Hecock. 1957. *Ballot Position and Voters Choice*. Detroit: Wayne State University Press.

Barton, Jill. 2002. "Expert: Palm Beach's New Voting Machines Have Problems." *Tampa Bay Online*, Wed., Sept. 11. http://ap.tbo.com/ap/florida/IGAIFTWBQ30.html.

Beard, Charles A. 1909. "The Ballot's Burden." *Political Science Quarterly* 24:589-614.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time Series-Cross Section Data." *American Political Science Review* 89:634-47.

Bederson, B.B. 2001. PhotoMesa: A Zoomable Image Browser Using Quantum Treemaps and Bubblemaps. UIST 2001, ACM Symposium on User Interface Software and Technology, CHI Letters, 3(2), pp. 71-80.

Bederson, B.B., & Boltman, A. 1999. Does Animation Help Users Build Mental Maps of Spatial Information? In Proceedings of Information Visualization Symposium (InfoVis 99) New York: IEEE, pp. 28-35.

Bederson, Benjamin B., and Paul S. Herrnson. 2002. "Usability Review of the Diebold DRE System for Four Counties in the State of Maryland." http://www.capc.umd.edu/rpts/MD_EVoteMach.pdf

Bederson, B. B., Meyer, J., & Good, L. 2000. Jazz: An Extensible Zoomable User Interface Graphics Toolkit in Java. UIST 2000, ACM Symposium on User Interface Software and Technology, CHI Letters, 2(2), pp. 171-180.

Berinsky, Adam J., Nancy Burns, and Michael W. Traugott. 2001. "Who Votes by Mail? A Dynamic Model of the Individual-level Consequences of Voting-by-Mail Systems." *Public Opinion Quarterly* 65:178-97.

Bowler, Shaun, Todd Donovan, and Trudi Happ. 1992. "Ballot Propositions and Information Costs: Direct Democracy and the Fatigued Voter." *Western Political Quarterly* 45:559-68.

Bullock, Charles S. III, and Richard E. Dunn. 1996. "Election Roll-off: A Test of Three Explanations." *Urban Affairs Review* 32:71-86.

Caltech/MIT Voting Technology Project. 2001. "Residual Votes Attributable to Technology: An Assessment of the Reliability of Existing Voting Equipment." http://www.vote.caltech.edu/Reports/.

Campbell, Angus, Philip E, Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.

Canedy, Dana. 2002. "Despite Overhaul of Voting System, Chaos Triumphs Again in Florida Election." *New York Times*, September 11, A28.

Crutcher, R.J. 1994. "Telling What We Know: The Use of Verbal Report Methodologies in Psychological Research." *Psychological Science* 5:241-44.

Darcy, Robert. 1986. "Position Effects with Party Column Ballots." *Western Political Quarterly* 39:648-62.

Darcy, R., and Ian McAllister. 1990. "Ballot Position Effects." *Electoral Studies* 9:5-17.

Darcy, R., and Anne Schneider. 1989. "Confusing Ballots, Roll-off, and the Black Vote." *Western Political Quarterly* 42:347-64.

DeRouvray, C. & Couper, M.P. 2002. "Designing a Strategy for Reducing "No Opinion" Responses in Web-based Surveys." *Social Science Computing Review* 28(1): 3-9.

Election Data Services, Inc. 1998. *1998 Voting Equipment Study Report*. http://electiondataservices.com/content/vote_equip.htm.

Ericsson, K.A. & Simon, H.A. 1993. *Protocol Analysis: Verbal Reports as Data,* rev. ed. Cambridge: MIT Press.

Federal Election Commission. 2000. The Administrative Structure of State Election Offices: Voting Systems. http://www.fec.gov/elections.html.

Goetzl, Aron. 2002. "Philadelphia Makes Successful Change from Level to Electronic Machines." Electionline Weekly, http://www.electionline.org/article.jsp?id=electionline_Weekly_for_5-23.

Hamilton, John. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hanmer, Michael J., and Michael W. Traugott. 2002. "The Impact of Voting by Mail on Voter Behavior." Paper presentated at the annual meeting of the Midwest Political Science Association, April 25-28, Chicago, IL.

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

Hornbaek, K., Bederson, B.B., & Plaisant, C. 2002. Navigation Patterns and Usability of Zoomable User Interfaces With and Without an Overview. ACM Transactions on Computer Human Interaction, (in press).

Key, V.O., Jr. 1956. *American State Politics*. New York: Knopf.

Kimball, David C., Chris Owens, and Katherine McAndrew. 2002. "Unrecorded Votes in the 2000 Presidential Election." Unpublished manuscript, University of Missouri-St. Louis.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.

Krosnick, Jon A., Joanne M. Miller, and Michel P. Tichy. Forthcoming. "An Unrecognized Need for Ballot Reform: The Effects of Candidate Name Order on Election Outcomes." In *Rethinking the Vote: The Politics and Prospects of American Election Reform*, ed. A.N. Crigler and M.R. Just. New York: Oxford University Press.

Mayhew, D. J. 1999. *The Usability Engineering Lifecycle*: *A Practitioner's Handbook for User Interface Design*. San Francisco: Morgan Kaufman.

Mercer, Martin, Joni James, and Alfonso Chardy. 2002. "Here We Go Again: Confusion Reigns in Sequel to 2000 Election." *Miami Herald* (online), September 11. http://www.miami.com/mld/miami/archives/.

Miller, Joanne M., and Jon A. Krosnick. 1998. "The Impact of Candidate Name Order on Election Outcomes." *Public Opinion Quarterly* 62:291-330.

Mintz, John and Peter Slevin. 2001. "Human Factor at Core of Vote Fiasco." *Washington Post*, June 1.

Mueller, John E. 1970. "Choosing among 133 Candidates." *Public Opinion Quarterly*. 34:395-402.

National Conference of State Legislatures. 2002. "The States Tackle Election Reform." http://www.ncsl.org/programs/legman/elect/taskfc/TackleElectRef.htm.

Nichols, Stephen M. 1998. "State Referendum Voting, Ballot Roll-off, and the New Electoral Technology." *State and Local Government Review* 30:106-17.

Nichols, Stephen M., and Gregory A. Strizek. 1995. "Electronic Voting Machines and Ballot Roll-off." *American Politics Quarterly* 23:300-18.

Nielsen, J. 1993. *Usability Engineering.* Cambridge AP Professional.

Nielsen, J. 2000. *Designing Web Usability: The Practice of Simp*licity. Indianapolis: New Riders Publishing.

Nielsen, J. & Levi, J. 1994. "Measuring Usability: Preference vs. Performance." *Communications of the ACM* 37:66-75.

Norman, D. 1981. "Categorization of Action Slips." *Psychological Review* 88:1-15.

Norman, D. A. & Draper, S. W. (eds.). 1986. *User Centered System Design*. Hillsdale, NJ: Lawrence Erlbaum.

Ostrom., Charles W. 1990. Time Series Analysis: Regression Technique, 2nd ed. Beverly Hills: Sage.

Phillips, Carly. 2001. "Burnettown Election Tests Machines." Augusta (Georgia) Chronicle, August 22. http://www.augustachronicle.com/ns-search/stories/082201/met_201-2690.000.shtml?NS-search-set=/3d3c6/aaaa123083c635b&NS-doc-offset=0&

Redmond-Pyle, D., & Moore, A. 1995. *Graphical User Interface Design and Evaluation*. London: Prentice Hall.

Robinson, James A., and William H. Standing. 1960. "Some Correlates of Participation: The Case of Indiana." *Journal of Politics* 22:96-111.

Roth, Susan, King. 1998. "Disenfranchised by Design: Voting Systems and the Election Process." *Information Design Journal* 9(1):1-8

Rusk, Jerrold. 1970. "The Effect of the Australian Ballot Reform on Split Ticket Voting: 1876-1908." *American Political Science Review* 64:1220-38.

Shneiderman, B. 1997. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 3rd ed. Massachusetts: Addison-Wesley.

Sigelman, Lee, and Malcolm E. Jewell. 1986. "From Core to Periphery: A Note on the Imagery of Concentric Electorates." *Journal of Politics* 48:440-49.

Sigelman, Lee, Philip W. Roeder, Malcolm E. Jewell, and Michael A. Baer. 1985. "Voting and Nonvoting: A Multi-Election Perspective." *American Journal of Political Science* 29:749-65.

Sinclair, Robert C., Melvin M. Mark, Sean E. Moore, Carrie A. Lavis, and Alexander S. Soldat. 2000. "An Electoral Butterfly Effect." *Nature* 408 (Dec. 7):665-66.

Sudman, Seymour, and Norman M. Bradburn. 1986. *Asking Questions*. San Francisco: Jossey-Bass.

Taebel, Delbert A. 1975. "The Effect of Ballot Position on Electoral Success." *American Journal of Political Science* 19:519-26.

Terhanian, George, John Bremer, and Amanda Salomon. 2000. "Advice on Reducing Error in Internet Research." *CASRO Journal*, pp. 55-59.

Thomsen, Søren R. 1987. *Danish Elections 1920-79: A Logit Approach to Ecological Analysis and Inference*. Aarhus: Politica.

Thomsen, Søren R., Sten Berglund, and Ingemar Worlund. 1991. "Assessing the Validity of the Logit Method for Ecological Inference." *European Journal of Political Research* 19:441-77.

Tomz, Michael, and Robert Van Houweling. 2001. "Who Spoils Their Ballots? Voting Equipment and Race in the 2000 Presidential Election." Unpublished Manuscript.

Traugott, Michael W., and Michael J. Hanmer. 2001. "Oregon Vote-by-Mail Project: Report to the League of Conservation Voters Education Fund." University of Michigan.

U.S. Commission on Civil Rights. 2001. "Voting Irregularities in Florida During the 2000 Presidential Election." June. http://www.usccr.gov/.

U.S. House of Representatives. 2001. "Election Reform in Detroit: New Voting Technology and Increased Voter Education Significantly Reduced Uncounted Votes." Prepared for Representative Henry A. Waxman by the Minority Staff, Special Investigations Division, Committee on Government Reform, April 5. http://www.house.gov/reform/min/pdfs/pdf_inves/pdf_elec_detroit_rep.pdf

Wand, Jonathan N., Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Michael C. Herron, and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95:793-810.

Walker, Jack L. 1996. "Ballot Forms and Voter Fatigue: An Analysis of the Office Block and Party Column Ballots." *Midwest Journal of Political Science* 10:448-63.

Wyman, Scott. 2002. "County: New Machines Have Flaws: Official Wants Replacements for Devices with Dim Monitors that May Be Hard To Read." South Florida Sun-Sentinel, Feb. 12.