Designing Optical Scan Ballots: 5 Simple Rules

Greg Dennis
gdennis at mit dot edu

Introduction

Approximately 1.5 million votes for president and 3.5 million votes for governor and senator are spoiled each election cycle.[1] More than 80% of these lost votes can be attributed to a poorly designed ballot, such as the infamous "butterfly ballot" used in the Florida 2000 election. As many learned during that controversy, poor ballot design has the potential to change the outcome of an election.

Researchers have examined many aspects of the ballot for its affect on voting behavior, from the ordering of the candidates' names and the typography of the text, to the physical attractiveness of the candidate. Though the existing research is somewhat limited, the evidence suggests that each of these attributes may cause spoiled ballots if not addressed properly. What follows is a survey of ballot design research, specifically the design of optical scan (or mark-sense) ballots, interspersed with some of the author's own observations. The survey is compiled into five basic rules for election officials to follow when designing such ballots.

Though they are referred to as "rules", "guidelines" is a more accurate term. It seems unlikely that any short list of rules could account for all the heterogeneity present in the highly decentralized voting administration of the United States. As pointed out by the Caltech-MIT Voting Technology Project, most states have devolved their election administration to local governments so that "there are not fifty election divisions, but over three thousand election administrators maintaining voter registration systems, choosing equipment, formatting ballots, setting up polling places, handling absentee voters, and conducting counts, audits, and recounts." Plus, much research is left to be done. So take these rules, not as strict design laws, but as a good-faith effort to summarize the current thinking about what makes a good optical scan ballot.

Rule #1: Rotate the Candidates

One of the more well-documented voting phenomena related to ballot design is the "primacy effect." The "primacy effect" refers to the inclination of a voter to choose a name listed towards the top of a vertical column of candidates. A bias towards the candidate listed first has long been assumed by political operatives, even if they haven't assigned a specific name to the phenomenon.[2] In the past, challengers have even brought lawsuits to prevent incumbents from enjoying the prized top spot.[3]

Though there is some disagreement amongst researchers as to the extent of the effect, very few doubt its existence. In 1998, Joanne Miller and Jon Krosnick found compelling evidence of candidate name-order effects in the 1992 Ohio general election. Specifically,

their results showed a statistically significant bias towards the first candidate in 57 of the 188 races they studied.[4] Importantly, they found the effect of ballot position to be stronger when the races were non-partisan, when they did not include an incumbent, and when the race was not widely covered in the media. As Krosnick and Miller note, name-order effects "were smaller when a cue was available to help people cast substantively meaningful votes."[5] That is, when information such as party affiliation and incumbency were absent, voters are more likely to be swayed by an arbitrary aspect of the ballot format such candidate name position.

Susan Banducci, et al, confirmed these findings in August 2003.[6] Their data confirm a bias towards candidates based on ballot position, as well as physical attractiveness (based on a photo on the ballot) in low-information elections. In addition to lack of media attention and absence of substantive cues on the ballot, they conclude that the primacy effect is also enhanced by the presence of a large number of candidates.

Most recently, Jonathan Koppell and Jennifer Steen studied the effect of ballot positions in the 1998 Democratic primary in New York.[7] That it was a primary election is significant because it meant the voter could not use party affiliation on the ballot as a cue for whom to vote. To mitigate the primacy effect, New York rotates the name of its candidates on the ballot across precincts. Among the 180 candidates in contested primaries that they studied, 161 received a statistical advantage when listed first. Moreover, in 7 of the 79 races, the advantage offered to the first position *exceeded* the margin of victory. That is, if the ballots had not been rotated and if the runner-up had received the top spot on the ballot, that runner-up would have likely been elected instead.

Despite the demonstrated impact of the primacy effect, most states do not rotate or randomize candidate name positions across geographic regions. One exception is California, where election law mandates rotation *and* randomization in statewide races. However, the law only requires randomization for municipal elections. This inconsistency runs counter to the research: local races are almost certainly of lower-information, and therefore will be subject to a greater primacy effect, compared to statewide races. Michael Alvarez recounts a case where this inconsistency was nearly decisive in the 2001 June mayoral runoff election in the Compton, California:

> *The Compton City Clerk . . . failed to correctly randomize the name of candidates on the runoff election ballot. Following state elections law, [he] requested and used the appropriate randomized list of candidate names in the March 2001 primary election, but he again used the same randomized list for the June 2001 runoff election (according to the court ruling in this case, the Clerk should have requested and used a second randomized list for the runoff election). The sitting incumbent mayor, Omar Bradley, was listed second on the runoff election ballot; his challenger, Eric Perrodin, was listed first and won the election by a slim 261 vote margin. The expert witness for Bradley, Jon Krosnick, testified that this incorrect ordering of candidate names on the runoff ballot could have accounted for at least the 261 vote margin, and perhaps many more ballots for Perrodin. The court was convinced by Krosnick's testimony, and on this basis alone, ruled in Bradley's favor, threw out the results of the June 2001 runoff election, and reinstated Bradley as mayor of the City of Compton.*

Due to the primacy effect, it is clear that candidates' names should be rotated so that each name is listed first an equal number of times. But is randomization—like that required by California election law—also necessary? Senior editor at *Wired*, Jessie Scanlon doesn't think so. "As almost any designer would tell you, it would be far better to simply rotate through the trusty A to Z from district to district," she writes.[8] However, she's speaking largely in reference to the California recall election and its record 133 candidates, and she's worried about voters merely being *able* to locate their candidates. But Scanlon does raise a good question: is the initial randomization an unnecessary complexity when rotation is already in place? Will simple rotation of an alphabetized list suffice?

One problem with just listing candidates A though Z is a greater likelihood that confusingly similar will be juxtaposed. In fact, alphabetical ordering is *the* way to order the names if one wants to maximize the matching prefixes of consecutive names (think candidate "Church" followed by candidate "Churchill"). Adding randomization, however, doesn't guarantee the absence of consecutive similar names either. For example, a randomized order could possibly juxtapose candidates "Hatcher" and "Thatcher."

Taking into account these factors, the following simple strategy is suggested. First, choose an order of the candidates names such that no name is similar to any consecutive name and the first is not similar to the last. Next, rotate the name order across small geographic regions such as precinct, wards, or state legislative districts. This strategy should negate the primacy effect and avoid the juxtaposition of confusingly similar candidate names.

Rule #2: One Column of Ovals

There are two main types of voting targets found on today's optically-scanned ballots: empty ovals that are to be filled-in and broken arrows that are to be connected.[9] The empty ovals seem like a natural choice given the public's familiarity with them from standardized tests and lottery tickets. The motivation for unconnected arrows is attributed to earlier optical scan machines, which detected horizontal lines better than it did vertical ones; but given that optical scan technology has progressed since then, it is unclear whether broken arrows still have a technological advantage.[10]

Regardless of any *technological* advantage broken arrows might still possess, the available data, though they are limited, suggest that the *overall* advantage goes to the ovals. In their study of residual vote rates in the Georgia 2000 general election, Bullock and Hood found that 4.63% of connect-the-arrow ballots registered as undervotes, compared to only 2.29% for the fill-in-the-oval ballots. Not surprisingly, a press release from the Georgia Secretary of State acknowledged "indications from county election officials that voters have particular difficulty with one type of optical scan ballot that requires electors to complete an arrow adjacent to a candidate's name (rather than blackening a circle)."[11] More recently, Theresa LePore, of butterfly ballot fame, was the subject of additional criticism when she opted for arrows instead of the ovals.[12]

In addition to indicating an advantage to ovals over arrows, residual vote rates indicate that listing candidates in a single column on an optical scan ballot is significantly better

than spreading the list over multiple columns. In the media consortium study of the Florida election in 2000, for example, Washington Post staff writer Dan Keating found that optical scan ballots where candidates were listed across two columns—what he terms "caterpillar" ballots—yielded a high rate of overvoting because voters would often mark one candidate in each column.[13] In fact, the rate of overvoting with the caterpillar layout was so great that polling locations with in-precinct count and caterpillar ballots actually had higher residual vote rates than those with central count and single-column ballots. This finding concurs with Doug Jones' data, which demonstrate higher overvote rates for two-column ballots in Florida.[14]

Thus, the rule is ovals, not arrows, and one column, not two.

Rule #3: Text Matters

A ballot is fundamentally a user-interface. Yet most ballots do not adhere to basic user-interface guidelines, such as those dictating proper typeface, capitalization, and plain-language instructions. In a human-factors study of voting systems, for example, Susan King Roth found fundamental usability problems related to the "legibility and readability of information displayed on the ballot."[15] Some of the more elderly participants in her study complained that the text was "too small" and some had to get "really close" to read it. Others complained that the wording of ballot questions was confusing. King points to one section of instructions on the general election ballot in Clearfield County, Pennsylvania to illustrate the potential for complex wordings:

> *To vote for an individual candidate of another party, after blackening an oval in the Straight Party Column, blacken the oval to the left of the candidate's name. For an office where more than one candidate is to be voted for, the voter, after blackening an oval in the Straight party Column, may divide his vote by blackening the oval to the left of each candidate for whom he or she desires to vote. For that office, votes shall not be counted for candidates not individually marked.*[16]

Perhaps the problem is, as Jessie Scanlon claims, that "ballots aren't designed by *designers*."[17] So what would a designer do?

The standard way for user-interface designers to quantitatively judge a UI is to evaluate it against a set of usability heuristics. One popular set of heuristics in the UI community is the ten recommended by usability guru Jakob Nielsen[18]. Heuristics like these are typically applied to computer user interfaces that are highly interactive; a ballot, on the other hand, is entirely passive. Nevertheless, ballots have at least one thing in common with most computerized UI's: they both try to convey information via text. And, as we shall see, in the realm of typography, the heuristics will dictate some important guidelines.

We won't discuss every Nielsen heuristic, but let's start with his second one: "match between system and the real world." This heuristic dictates that the "system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms." In other words, ballot instructions should be in clear and simple

language and avoid jargon. Nielsen's third heuristic, "user control and freedom" recommends UIs provide an "undo" option for every task. For a ballot, this suggests clearly marked instructions that explain what the voter should do in the case he or she makes a mistake. "Consistency and standards," the fourth heuristic, emphasizes the consistent treatment of words and actions across a UI. To score well on this heuristic, a ballot should probably employ one typeface throughout with minimal variation in style and size and consistent capitalization. Finally, Neilsen's eighth heuristic, "aesthetic and minimalist design," cautions against unnecessary and redundant information. That is, keep a ballot simple and uncluttered.

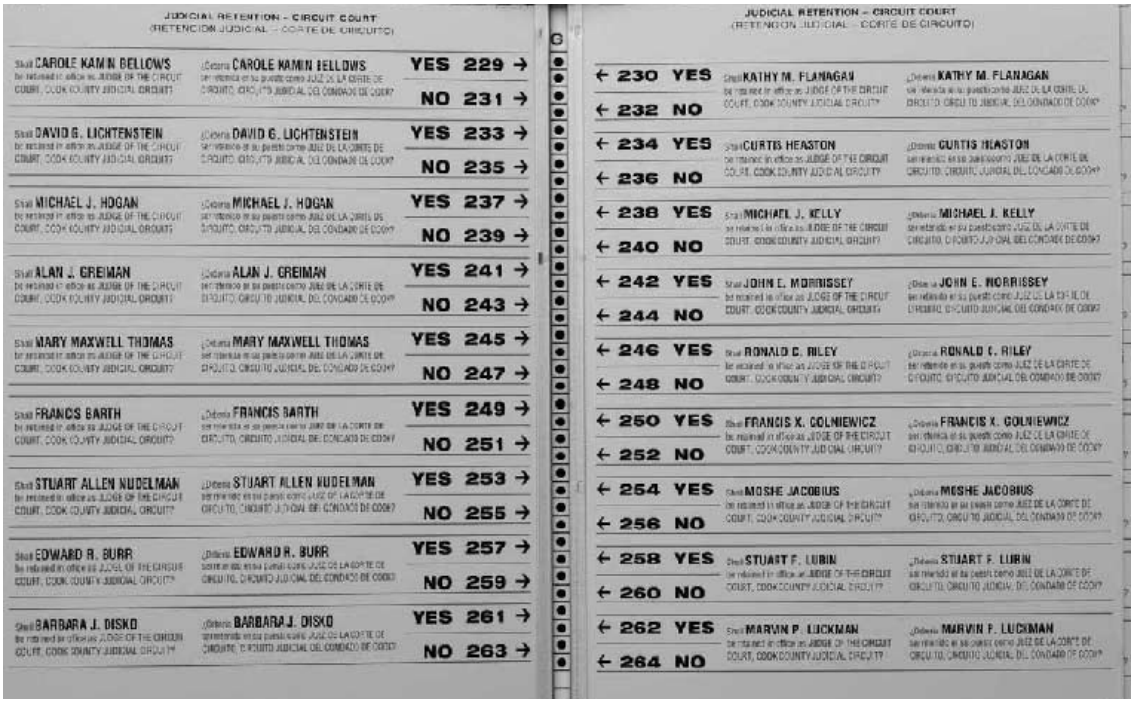

Figure 1. Cook County Ballot, November 2000

In an attempt to improve ballot design along these lines, the American Institute of Graphic Arts (AIGA) launched an initiative in 2000 called "Design for Democracy."[19] The goal of the initiative was to bring ballots, election manuals, and voting equipment up to the standards of modern graphic arts. In a revealing case study, Design for Democracy took on the tall order of redesigning the perplexing ballot that caused so much confusion in Chicago in Cook County in the 2000 general election (Figure 1).[20] For greater consistency, AIGA replaced the mismatch of font faces, styles, capitalization, and sizes on the 2000 ballot with a single, legible Univers typeface in lowercase throughout and with limited size and weight variations. To simplify the layout, they replaced every nine yes/no judicial retention questions with a single question and nine yes/no candidate listings. And if a voter accidentally spoils a ballot, instructions on the new ballots tell the voter to ask the judge for a new one. The result was a simple, clean, and aesthetically-pleasing ballot that received high marks from Chicago and Cook County voters in the 2002 election (Figure 2).[21]

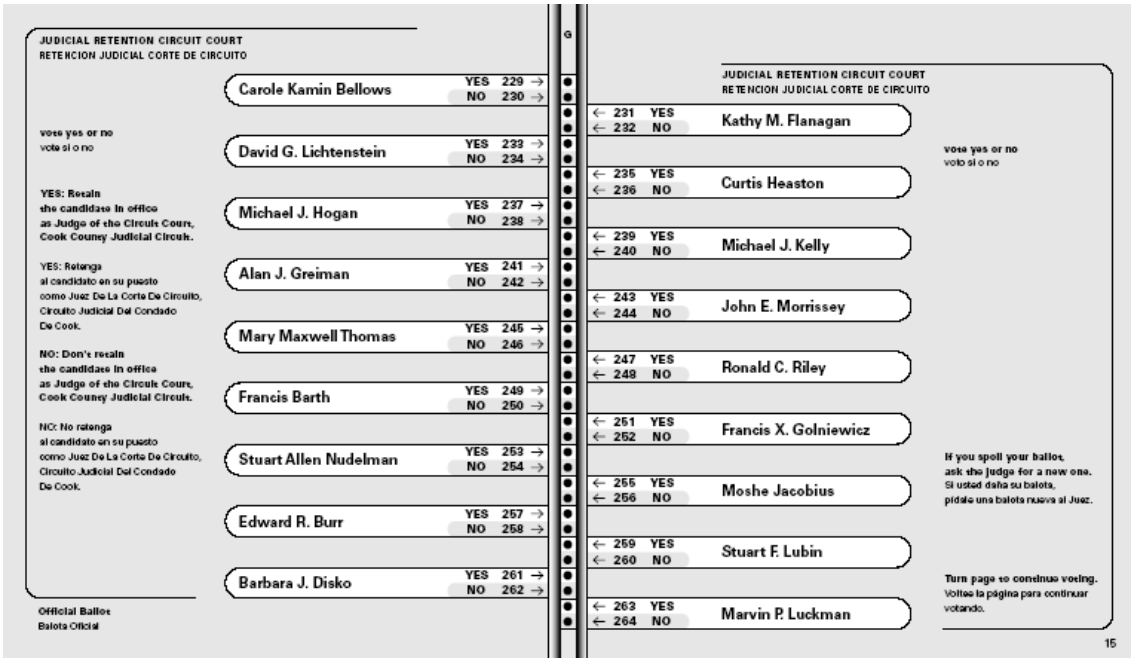

Figure 2. Cook County Ballot Design, November 2002

Although the ballot in this case study was a punch card and not an optical scan, the lesson is the same: text matters. It matters not only in what it says, but in how it looks and how much of it there is.

Rule #4: None of the above

Thought there is admittedly no data to support this claim, it would seem that many ballot-related problems could be mitigated with the addition of an explicit "None of the above" (NOTA) option in each race.

First and foremost, NOTA would distinguish between intentionally and unintentionally unmarked ballots. When combined with an in-precinct optical scan that is configured to reject any ballot without exactly one mark per race, NOTA would likely go far towards eliminating the unintentional undervote. Note that this undervote protection could help further mitigate any technological advantage that broken arrows still have over ovals.

In addition, NOTA could help diminish the effect that arbitrary cues, such as the position of a candidate's name on a ballot, have over an election. As discussed above, such incidentals have the strongest effect in low-information elections where the voter is offered few sophisticated cues, such as party affiliation or incumbency, to aid their decision. Voters who are swayed by these cues may be uninformed, undecided, or simply dissatisfied with all choices; but clearly they feel compelled to vote for *someone*, perhaps because they worry their entire ballot may be invalid it they do not. By providing an explicit NOTA option, a ballot can offer these voters with a way out of a decision they may be disinclined to make.

Though further research into NOTA's effect is needed, it seems to this author that it could only have a positive impact on the voting experience.

Rule #5: Test, Test, Test

The last, but arguably the most important rule, is to test any ballot design before introducing it in an election. As any designer knows, the best way to evaluate an interface is through substantial user testing. Both Michael Alvarez and Michael Traugott suggest the tests take the form of controlled experiments.[22,23] Alvarez offers a helpful template study of the butterfly ballot to illustrate the form these experiments might take:

> *Two small groups of randomly selected registered voters from an election jurisdiction could be contacted about participation in a research project on election technologies . . . One group would be the control group, and they would be asked to participate in a hypothetical election, using the same ballot style used in that election jurisdiction in the most recent election. The experimental group would be asked to participate in a hypothetical election, "voting" for the same set of candidates as the control group, but using a "butterfly" ballot. Both groups would be interviewed, either using a quantitative or qualitative approach, before and after their "voting" about their experience and opinions about the ballot. The ballots cast by both groups could then be examined for errors, and for deviations in the vote cast from the voter's stated "vote" in the follow-up interview. Such a study design could produce a powerful, and clear, analysis of the precise effects of the butterfly ballot on voter behavior in the ballot booth.*

In addition to controlled experiments, Traugott rightly advises designers to deploy "mobile units" of their ballots to bustling locales such as malls and shopping centers. A mobile unit allow for a pool of subjects that is not only larger, but also more diverse, from a variety of backgrounds and socioeconomic statuses. Plus, the units would help showcase and familiarize a wider audience with a ballot they may use in the near future.

Beyond conducting good tests, the development of any ballot, or voting system in general, must be approached as a usability problem from start to finish. As the Federal Election Commission recommends, a voting system should be developed through a "user-centered" design process.[24] The "common theme," the FEC continues, "is ensuring that user needs and preferences are addressed in the design process, rather than shaping a system chiefly according to technological considerations." Though the rules outline in this document ought to help, the key is to recognize that the voting system is not so much about votes, as it is about voters.

[1] MIT-Caltech Voting Technology Project. *Voting: What Is, What Could Be*. July 2001.
[2] Jonathan GS Koppell and Jennifer A. Steen. "The Effects of Ballot Position on Election Outcomes" *The Journal of Politics*. Vol 66, No 1. February 2004. pp 267-281.
[3] Ibid.
[4] Joanne M. Miller and Jon A. Krosnick.. "The Impact of Candidate Name Order on Election Outcomes." *Public Opinion Quarterly*, 1998. 62(2): 291-330.
[5] Ibid.
[6] Susan Banducci, Michael Thrasher, Colin Rallings, and Jeffrey A. Karp. "Candidate Appearance Cues in Low-Information Elections." American Political Science Association. August, 2003.
[7] Koppell and Steen.
[8] Jessie Scanlon. "Wanted: A Legible Voting Ballot." *Slate*. 6 October 2003.
[9] Eric A. Fisher. "Voting Technologies in the United States: Overview and Issues for Congress" CRS Report for Congress RL30773. March 2001.
[10] Douglas W. Jones. "Counting Mark-Sense Ballots." http://www.cs.uiowa.edu/~jones/voting/optical/ . 2002.
[11] Cathy Cox, Georgia Secretary of State. "Secretary Cox: 2000 Election Data Shows Undervote Rates Consistent Among All Voting Equipment." Press Release. 28 December 2000.
[12] Jane Musgrave. "Ballot arrows point out new controversy." *Palm Beach Post*. 23 August 2004.
[13] Dan Keating. "Democracy Counts: The Media Consortium Florida Ballot Project." August 2002.
[14] Douglas W. Jones. "An Ethical Response." http://www.cs.uiowa.edu/~jones/voting/cogel/
[15] Susan King Roth. "Disenfranchised by design: voting systems and the electoral process." Information Design Journal. Vol 9(1), 1998.
[16] Susan King Roth. "Human Factors research on Voting Machines and Ballot Designs: An Exploratory Study." http://www.capc.umd.edu/rpts/MD_EVote_Roth.pdf
[17] Scanlon.
[18] Jakob Nielsen. "Ten Usability Heuristics." http://www.useit.com/papers/heuristic/heuristic_list.html
[19] Design for Democracy Case Studies. http://www.electiondesign.org/pdf/design_for_democracy.pdf
[20] Blair Kamin. "The (design) fix is in." *Washington Post*. 16 March 2004.
[21] Ibid.
[22] Alvarez.
[23] Michael Traugott. "Testing Alternative Hardware and Ballot Forms." Working Group on Voting Technology and Balloting. March 2002.
[24] Federal Election Commission. "Designing a User-Centered Voting System." October 2003.