

# CSD Center for the Study of Democracy

An Organized Research Unit  
University of California, Irvine  
[www.democ.uci.edu](http://www.democ.uci.edu)

Just days before the 1982 California Gubernatorial Election, black candidate Tom Bradley was expected by many to become the nation's first black Governor.<sup>i</sup> Bradley held a nearly double digit lead over his opponent Republican George Deukmejian in several polls. In spite of his large pre-election lead, Bradley was defeated by a thin margin. Seven years later, black candidate Douglas Wilder ran for governor of Virginia and held a comfortable lead over his opponent Marshal Coleman. Again there was a large discrepancy between the pre-election polls and the final result. Wilder, however, was declared the winner after an election so close that it was decided by a recount. Such large inconsistencies in the pre-election polls and the actual results have been observed for several other black candidates including the mayoral candidates David Dinkins and Harold Washington, Senate candidates Carole Moseley Braun and Harvey Gantt, and black Presidential candidate Barack Obama (NH and CA).

The phenomenon of *unusually* large discrepancies between polled pre-election support and actual electoral support for black candidates has been dubbed the Bradley Effect. It has been claimed that the Bradley Effect exists because of systematic racial biases in the electorate. While voters will tell pollsters that they intend to support the black candidate in an attempt to appear progressive, they often vote against the black candidate in the privacy of the election booth. Some question the existence of the Bradley Effect and others argue that the Bradley Effect is an outdated artifact that no longer affects polling accuracy (Keeter and Samaranayake 2007; McWhorter 2008; Walters 2008). They point to the success of polling in both the 2004 and 2006 elections of eight black candidates who ran for high-profile statewide positions. However, our analysis, to be discussed in detail below, demonstrates that this is not the case. The Bradley Effect is real and continues largely unabated.

Blacks do not seem to be the only ones that are affected by discrepancies in the polls. Both Al Gore and John Kerry lost in states that they were predicted to win just days before the election (Traugott 2001; Traugott 2005). White male candidates such as Jim Ross Lightfoot and Eddie Basha may also be considered victims of inaccuracies in pre-election polls. Although both Lightfoot and Basha led their opponents by as much as nine percent days before the election, they were both eventually defeated. This begs the question: are pre-election polls for black candidates inaccurate because of systematic biases from the public, or were the discrepancies found in the 1982 California and 1989 Virginia gubernatorial elections a result of something other than the candidates' race?

The Bradley Effect is an interesting phenomenon that has not been studied beyond one or two cases<sup>ii</sup> (Citrin et al 1990; Finkel et al 1991; Traugott and Price 1991). This paper differs from these previous studies in that we assess whether or not the Bradley Effect exists by examining nearly every election in which a black candidate ran for US

Senate or governor within the past 25 years. In addition to this, we expand on previous research by determining whether or not women, Latino, and Asian American candidates are also affected by inaccuracies in pre-election polls when compared to actual election results. As indicated above even white male candidates are not immune to polling discrepancies. Thus to isolate the Bradley Effect, which is purportedly the result of racial biases, we must include a control group (i.e., white males) to serve as a baseline of comparison. Moreover, this method will allow us to assess the claim that the Bradley Effect has become less pronounced over time.

In addition to providing the first systematic test of the Bradley Effect across minority groups, this paper also puts forth a new theoretical framework—a model of *preference falsification*—which captures the salient features of the Bradley Effect as it purportedly takes place and allows us to better identify the contexts in which the Bradley Effect might be more pronounced. Though the model is consistent with many stylized facts, we are, unfortunately, unable to provide a direct test of it. This is due primarily to the aggregated nature of the survey data, and the near-impossibility of obtaining un-adjusted exit polling data.

To accomplish these dual goals, we first review the claims in the literature regarding the causes of the Bradley Effect and other polling inaccuracies. Next, we describe our data set which was collected from several sources including the US Census, university and newspaper polls, and several election resources. We analyze this data set using the Wilcoxon match-pairs signed ranks test, the Mann-Whitney test, and multiple regression to determine what variables have an effect on poll discrepancies and which groups are most affected by them. Using pre-election polls gathered from several sources, we demonstrate that the magnitude of the Bradley Effect increases for black candidates when you control for several other variables that lead to polling inaccuracies. In addition, we find that female candidates suffer from a “reverse” Bradley Effect—in which the pre-election polls underestimate their true support—a phenomenon we name the “Hillary Effect.” Moreover, these control variables are consistent with both the literature on racial attitudes in survey research, and the model of preference falsification outlined below.

## **Literature Review**

### **Causes of Polling Inaccuracies**

Election polls have evolved over the course of a century and have become increasingly precise instruments for measuring election outcomes. This is why most pre-election polls are very successful in predicting the eventual winner (Bolstein 1991; NCPP 1997; Mitofsky 1998; Traugott 2005; Keeter and Samaranayake 2007). Several authors have shown that the best predictor of whether or not a person will vote and who they will vote for is their pre-election poll response (Mosteller 1949; Bolstein 1991; Mitofsky 1998). Despite the general accuracy of modern pre-election polls, there have been numerous instances in which the actual results do not match their predictions. Furthermore, under certain conditions related to the nature of the candidates and the race, the predictive power of these polls is further diminished.

A common problem in polling is that large numbers of respondents decline to state their voting intentions (Perry 1979; Fenwick et al. 1982; Mitofsky 1998; Berinsky 1999). With fewer people certain about their vote, it is more difficult for pollsters to forecast the results. If the “don’t know” responses are biased because the respondent is uncomfortable giving a specific response, this problem is compounded (Berinsky 1999). Another commonly cited problem is an unexpected change in the number of voters. This is a problem because polling estimates are based on assumptions regarding specific turnout rates among subgroups of the population. If the actual turnout differs from the relative rates assumed in the polls then this could lead pre-election polls to be less accurate (Mitofsky 1998; Durand et al. 2001; Traugott 2005).

Apart from these problems endemic to polling, the type of candidate and the context in which the election takes place have been shown to have an affect on the accuracy of polls. In elections with incumbents, pre election-polls tend to overestimate their support (Wright 1990; Gow and Eubank 1994). This can be explained in two ways. First, a bandwagon effect in which respondents would like to support successful candidates. Second, lack of name recognition for the challenger may lead some to support the more familiar candidate in pre-election polls. However, they may cast their vote for the challenger as they become more informed during the course of the election. Similarly, competitiveness has a large affect on the accuracy of polls (Eubank and Gow 1984; Wright 1992). In more competitive elections there is less uncertainty with respect to the candidates because voters are more informed and therefore can give a more accurate representation of their voting intention.

The Bradley Effect could be a result of the typical context in which black candidates run. At the state level, very few blacks are incumbents and most run in non-competitive elections. In fact, between 1982 and 2006, Carole Moseley Braun was the only black gubernatorial or senate candidate to run as an incumbent. Moreover, only 9 of 27 of these black candidates competed in an election decided by 10 points or less. Both of these factors could explain why polls seem to be more inaccurate for black Senate and gubernatorial candidates when compared to their white counterparts.

In addition to the difficulties noted above, another common challenge to polling accuracy is that respondents may give misleading responses in order to appease the interviewer. This phenomenon is known as the socially desirable response and generally occurs with those respondents who have the most to gain from this deception (Hatchett and Schuman 1975; Jackman and Muha 1984; Krysan 1998). While this idea is pervasive in the survey literature, there has, to our knowledge, been no formal-theoretic model put forward to explain this phenomenon. We borrow such a model from the Public Choice literature, which is described in detail in the next section. This model, which incorporates the key features of the socially desirable response phenomenon, is also consistent with several stylized facts regarding the Bradley Effect.

### **Preference Falsification**

In the two races featuring the candidates who lend their names to the “effects” discussed in this paper—Tom Bradley and Hillary Clinton—race and gender respectively figured into the contest in important and salient ways. Because many issues regarding race and

gender are sensitive and controversial, survey respondents may—even more so than is usually the case—perceive that they are being personally judged based on their responses. Such a psychological process thus could induce the effects (the Bradley and Hillary effects) that we document below. The idea that survey respondents may misrepresent their true intentions or preferences—for reasons related to perceived social acceptance of certain views—is an established one in the field of survey research, and is generally known as the socially desirable response bias (Hatchett and Schuman 1975; Jackman and Muha 1984; Krysan 1998).

There have been several attempts to explain the causes of socially desirable response in the context of political surveys. Berinsky (1999) employs a clever technique to correct for selection bias, and is able to uncover latent opposition to government mandated desegregation of schools. In an attempt to validate his technique, he also uncovers a Bradley Effect in the 1989 New York City Mayoral election. He demonstrates that respondents who answered “don’t know” to inquiries regarding their vote choice in the upcoming election were overwhelmingly likely to be supporters of Rudy Guiliani rather than his black Democratic rival, David Dinkins. While Berinsky contributes to our understanding of the causes of polling inaccuracies, his framework nonetheless does not fully specify the contexts in which the socially desirable response is more or less likely to occur. Timur Kuran’s (1987, 1995) model of preference falsification allows us, at least in theory, to answer this question.

This theory of preference falsification hinges on a trade-off between two types of utility—*reputational* and *expressive*. In the model, individuals are allowed to hold private preferences (their true beliefs) and public preferences (their professed beliefs) that can be, and often are, at odds with one another. We first verbally sketch out the model before rendering it in formal notation. Individuals (perhaps idiosyncratically<sup>iii</sup>) are assumed to obtain utility from three sources. First, one obtains *instrumental* utility from the collective decision of society as a whole. In our case, the outcome of the collective decision is the election. Second, one can gain (or lose) *reputational* utility associated with one’s declared (public) preference. Finally, there exists *expressive* utility. This final source of utility comes from truthful self-expression on the part of the individual. When an individual falsifies his/her preference he/she foregoes some amount of expressive utility. This foregone utility increases in the magnitude of the falsification and in the strength of one’s private (“true”) preference.

If the benefits of preference falsification outweigh the costs for a given individual, then the model predicts that the individual will engage in preference falsification. First, consider *instrumental* utility. Given the individual’s nearly infinitesimal chance of directly affecting the outcome of the election, in most cases this source of utility can be safely ignored when we consider the individual calculus with respect to preference falsification. Thus, the individual’s decision to falsify his/her preference boils down to a comparison between the perceived reputational benefits of a falsification, and the cost of this action, in terms of foregone expressive utility. The following quote from Kuran (1995, p. 5) illustrates the importance of these two sources of utility—reputational and expressive—for preference falsification. “If one distinguishing characteristic of preference falsification is that it brings discomfort to the falsifier, another is that it is a response to real or imagined social pressures to convey a particular preference.” Equation

1 provides a more technical rendering of the individual's decision-making calculus with respect to preference falsification.

**Equation 1: The Individual's Preference Falsification Decision**

$$PF_i = 0 \text{ if } R_i + I_i < E_i \text{ OR } 1 \text{ if } R_i + I_i > E_i$$

Where  $PF_i$  represents the individual's decision to engage in preference falsification,  $R_i$  is the reputational benefit,  $I_i$  is the instrumental benefit, and  $E_i$  the foregone expressive utility, to individual  $i$ .

The model also contains a threshold component in which one's expected reputational utility gain is a function of one's beliefs about the distribution of preferences within one's relevant network or geographic area. Thus we should expect greater prevalence of preference falsification in areas in which it is perceived as unacceptable to be racist or sexist, e.g. areas with prevalingly progressive political outlooks. In our empirical analysis, one model is devoted to the hypothesis that social context has an effect on the aggregate level of preference falsification.

Individuals in this model may vary in the extent to which they derive disutility from falsifying their preference. One might imagine an avowed racist would feel more uncomfortable misrepresenting his true preference regarding a black candidate than would a progressive. Thus the former, *ceteris paribus*, is less likely to engage in preference falsification than the latter. The most important source of utility for our preference falsification model of the Bradley effect is reputational. According to traditional accounts of the phenomenon, the main factor contributing to the Bradley effect is socially desirable response bias, and the logic of reputational utility is present in all of its invocations in the literature.

Other individual level characteristics also affect the propensity to engage in preference falsification on political surveys. For example, those with more income, education, and who are strong partisans are most likely to over report voting (Silver et al. 1986; Berinsky 1999; Bernstein et al. 2001). In our model of preference falsification, this is the result of the fact that they have the most reputational utility at stake if they are seen as negligent in their civic duties. Silver et al. (1986) also find that those most supportive of the norm of voting are more likely to over report voting. These findings are consistent with the model. Also, minorities often over report voting when there is a fellow minority candidate (Shaw et al. 2000). They do this because they feel that they will be judged—again note the perception of social pressure, the primary driver of preference falsification—if they fail to give the answer which is expected of them. Finally, those who are more uncertain of their vote choice are also more likely to misrepresent their voting intention (Finkel et al. 1991). Those who are more uncertain must have less strongly held preferences, and thus suffer less expressive utility loss from a misrepresentation of their “true” views. If they have not yet completely made up their mind, they might as well respond in a way that they believe will please the interviewer.

White liberal Democrats are expected to vote for the Democratic candidate, but research has shown that some will cross party lines to support the white candidate over the black candidate (Bullock 1984; Giles and Buckner 1993; Reeves 1997). While they perform this act in the privacy of voting booths, they are embarrassed to admit their vote choice with an interviewer. As a result, these voters often over report voting for black

candidates in pre-election polls (Traugott and Price 1991; Reeves 1997; Wilson 2008). Moreover, Finkel et al. (1991) finds that white Democrats were significantly more likely to state a preference for Wilder (the black candidate) when the interviewer is black than when the interviewer is white.

Why would white liberals *not* vote for a black candidate who largely shares their political views? Liberal whites may vote against black candidates because they do not trust black candidates to protect white political interests (Bullock 1984; Williams 1991). Several studies have shown that when whites have less experience with black elected officials, they are uncertain about how they will govern and this uncertainty leads to lower levels of support (Hajnel 2001). Second, the findings of Williams (1991), Terkildson (1993), and Reeves (1997) suggest that whites believe blacks are less capable of holding such a high profile elected position. Finally, whites may believe that electing a black candidate to such a high position could lead to a loss of political power. Several authors have identified competition between racial groups as a strong predictor of racial political attitudes (Bobo and Kluegel 1993; Bobo and Hutchings 1996; Kinder and Sanders 1996). Whites who feel most threatened are more likely to support racist policies and candidates (Giles and Buckner 1993).

Similar to racial biases in the electorate, several studies have found that polls measuring attitudes on gender related issues are also subject to inaccuracies. These studies have found that the gender of the interviewer has a significant effect on the responses of both males and females (Johnson and DeLamater 1976; Reiss 1976, DiCamillo 1991). These and other studies have shown that women will be more supportive of women's issue when the interviewer is a female and less supportive when the interviewer is a male (Ballou and Del Boca 1980; Lueptow et al 1990; Pendelton 1990; Northrup 1992). Males tend to take the same approach; they are more supportive of women's issues with female interviewers but want to assert their dominance when interviewed by males (Huddy et al. 1997).

The underlying reason for misreporting their preferences regarding these issues is consistent with the model of preference falsification. It appears as if both males and females give responses that will enhance their reputation with respect to the interviewer. As Kuran (1995, p. 4) states, "(p)reference falsification aims specifically at manipulating the perceptions others hold about one's motivations or dispositions." We expect the same biases to occur for female candidates. Like race and ethnicity, we expect that systematic biases due to gender may lead to consistent problems in pre-election polls when there is a female candidate. Thus we expect some males to misreport their voting intentions with respect to the female candidate in order to appear masculine. Also, we expect some females to falsify their preference—if that preference is for the female candidate—in order to preclude any assumptions regarding voting based on gender.

## **Data & Methods**

In order to test the existence of the Bradley Effects and Hillary Effects, we analyze the differences between pre-election polls and final election results for black, Latino, Asian American, and white female candidates at the state level. While the Bradley Effect, if it exists at all, likely occurs at multiple levels of government, we concentrate only on

prominent statewide races—those for Senate and governor—for several reasons. At this level of government candidates tend to be better known than those at the local or congressional level. Also, there are many more female and minority candidates at this level than in Presidential elections. Finally, the Bradley Effect is most often and originally reported at the state level.

Our sample includes 155 cases from over 40 states. It contains 31 white female candidates, 23 black candidates, ten Latino candidates, six Asian-American Candidates, and to serve as controls, 85 races in which two white males were the major party candidates. The sample includes elections with incumbents and without. It also includes elections that were competitive and those that were not competitive. The sample includes elections in which minorities were successful and elections in which minorities were not.

To test our hypotheses, we collected data from various newspapers and other data sources. Our dependent variable is the magnitude of the Bradley Effect, or the difference between the pre-election polls' forecast of the election and the actual election results. The election results were obtained from the website [www.uselectionatlas.org](http://www.uselectionatlas.org). In addition to collecting information on the final results, we also collected information on pre-election polls. This data was obtained from newspapers from the state in which the election took place. We used four main criteria in selecting these polls. First, we used the poll that was closest to the election date. Therefore, we had the most current—thus the most accurate—estimate of the candidate's support. Second, we only included surveys that were conducted by telephone, because they are widely considered the most reliable. Studies have shown that face to face interviews as well as internet surveys are more susceptible to biases (Catania et al. 1986; Finkel et al. 1991; Zhang 2000). Third, we only used surveys that were conducted by polling firms, newspapers, or universities. Finally, we only used surveys that contacted more than 300 respondents.

Our dependent variable—the Bradley Effect—is measured in the following way. We first calculate the margin of victory as measured by pre-election polling. We then repeat this calculation for the final election results. Finally, we subtract the predicted margin of victory from the election margin of victory. If this difference is positive, then there is evidence of a Bradley Effect. That is, when compared to his/her opponent, the candidate's expected performance (measured by the pre election poll) was worse than his/her actual performance. The reverse is true if this difference is negative. Equation 2 below provides a more pithy definition of the Bradley Effect as we measure it.

### **Equation 2: Construction of the Bradley Effect measure**

$$\textit{Bradley Effect} = \textit{Predicted MOV} - \textit{Actual MOV}$$

Where *Predicted MOV* is the difference between the candidate's pre-election poll support and the opponent's pre-election poll support. *Actual MOV* is the difference between the (minority) candidate's actual electoral support and the opponent's actual electoral support. Positive values indicate a Bradley Effect, negative values indicate a Hillary Effect.

One could measure the Bradley effect simply by subtracting the pre-election polling for the candidate of interest from his/her actual support at the ballot box, but such an approach has at least two serious flaws. First, in nearly every case, unadjusted pre-election support underestimates final support in the absolute sense, because of the large numbers of undecided voters in many polls. This general problem is compounded in the present context because, as discussed above, there is evidence which indicates that the percentage of those who define themselves as undecided is a function of the degree of preference falsification in races with minority candidates. We partly sidestep this issue by operationalizing our variable this way; furthermore, we take this concern into account by adding percent of undecided voters (as measured by the pre-election poll) as a control variable in our regression. Second, there are many factors, unrelated to the race or gender of a candidate, which can influence the accuracy of political polling in unpredictable and perhaps undetectable ways. By including *both* candidates' poll-vote differential in our measure, we should be able to obviate many of these confounding factors, thus allowing us to better isolate the effect of the candidate's race or gender on their polling performance.

We also collected data on several independent variables that we expected would influence poll inaccuracies. Our main variables of interest were race/ethnicity and gender. We collected information in pre-election polls and final results for nearly every black, Asian American, and Latino statewide candidate after 1982 where the data was available.<sup>iv</sup> For white female candidates, we randomly selected 31 (from a total number of more than 150) candidates from 1982 to 2006. The races which feature only white males are of two types. Some are randomly selected from within our time period of interest (1982-2006). The remaining cases were selected according to matching criteria, so as to have one matched "white" case for every case of an election with a minority candidate in our sample. This was done in order to minimize the potential confounds introduced by ever-changing polling and weighting formulas as well as to control for certain temporal and spatial effects on polling accuracy. Thus, for each case in our sample which includes a minority candidate, we chose the most recent election—either Senatorial or gubernatorial—from the same state.

In addition to the race and/or gender of the candidate, we also expect polling inaccuracies to be tied to the circumstances of the election and the context in which the election took place. With respect to the election context, we considered whether or not the candidate of interest was a Democrat or an incumbent. In addition, we also considered whether or not the election had an uncharacteristically high turnout. This variable was constructed using voting age population turnout rates for each state of interest from George Mason's United States election project. For each state, we calculated a standard score for both Presidential and non-Presidential elections. Using standard scores will allow us to see if the election of interest had much higher or lower turnout than expected. Turnout of certain subgroups may be particularly important in elections with underrepresented minorities. To control for this we include the standardized score for black and female turnout. As mentioned above, we also include a variable for the number of undecided voters, because it is well established that large numbers of undecided voters can lead to polling inaccuracies. Next, we created a variable that measured competitiveness. We used the margin of victory in the pre-election polls as an indicator of competitiveness.

We also included variables meant to measure the social context in which each election is taking place. According to our model of preference falsification, these should affect the decision making calculus of the survey respondents. As mentioned above, we expect that voters who live in an area that is more progressive may have the most to gain by misleading pollsters. We use five variables to measure the social context of the election. We include average Leadership Conference on Civil Rights (LCCR) scores, percent of women in the state legislature, proportion of state residents who live in urban areas, percent of black population in the state, and a standardized measure of Democratic Presidential support. The LCCR measure is an average of the individual LCCR scores for that state's congressional delegation. The percent of women in the legislature variable was collected from Rutgers' Center for American Women in Politics. We expect that respondents in states with higher LCCR scores and with higher percentages of women in the legislature will be the most likely to misreport their intentions based on the preference falsification model. The black population and urbanity variables were collected through the Census. For these variables, we used the most recent census data for each state. Greenwald and Albertson (2008) in an unpublished study of the 2008 Democratic Party primaries, demonstrate that as the proportion of blacks in the population increases the magnitude of the Bradley Effect decreases. For the Democratic partisanship measure, for each state we compared its support of the most recent Democratic Presidential candidate to the national average.

## Methods

To understand the relationship between race/ethnicity and gender and the accuracy of pre election polls we use several statistical methods. To have a basic understanding of this relationship, we review the descriptive statistics. We then run two common non-parametric tests—the Wilcoxon Rank-Sum and the Mann-Whitney—to determine whether there are statistical differences in the size distribution of the Bradley effect between the comparison group (elections with two white male candidates) and each minority group. The Wilcoxon test, which is used for independent samples, gives us a basic understanding of the relative accuracy of the pre-election polls for each group. However, polling firms often use different formulas for sampling and weighting. In order to isolate the effect of the candidate's race or gender on polling accuracy, we utilize the Mann-Whitney test. The Mann-Whitney test is a dependent sample matched-pairs test for differences in distributions. For each election of interest we chose a comparison statewide election (i.e., one which involves two white males). All comparison cases are either the concurrent or most recent preceding election in the same state, and the polls are most often collected by the same polling firm as that which conducted the polls for the election of interest. Using this method, we are able to control for geographic, temporal and firm-specific effects to the maximum extent possible.

Because there are multiple factors, aside from the candidate's race or gender, which we expect to have effects on polling accuracy, we run three different OLS regressions which predict the accuracy of pre-election polls. In the first regression, we analyze the differences between white male, white female, black, Latino, and Asian American candidates without any controls. This model will allow us to recognize whether

any group is significantly different from our comparison group. The second model will test the same relationship, but in this model we control for the context of the election (Democratic party ID, competitiveness, standardized voting age population turnout, black turnout, female turnout, percent undecided, incumbent in the race, and a gubernatorial election indicator variable). Our final model will test for the effect of both election context and variables related to preference falsification on the accuracy of polls.

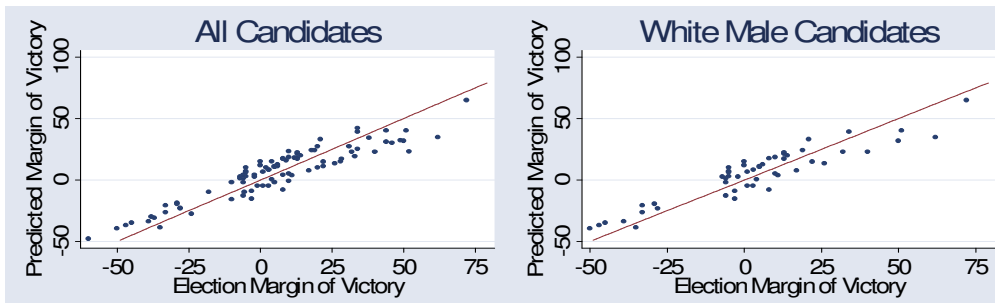
**Table 1: Summary of the Bradley Effect by Race/Ethnicity and Gender**

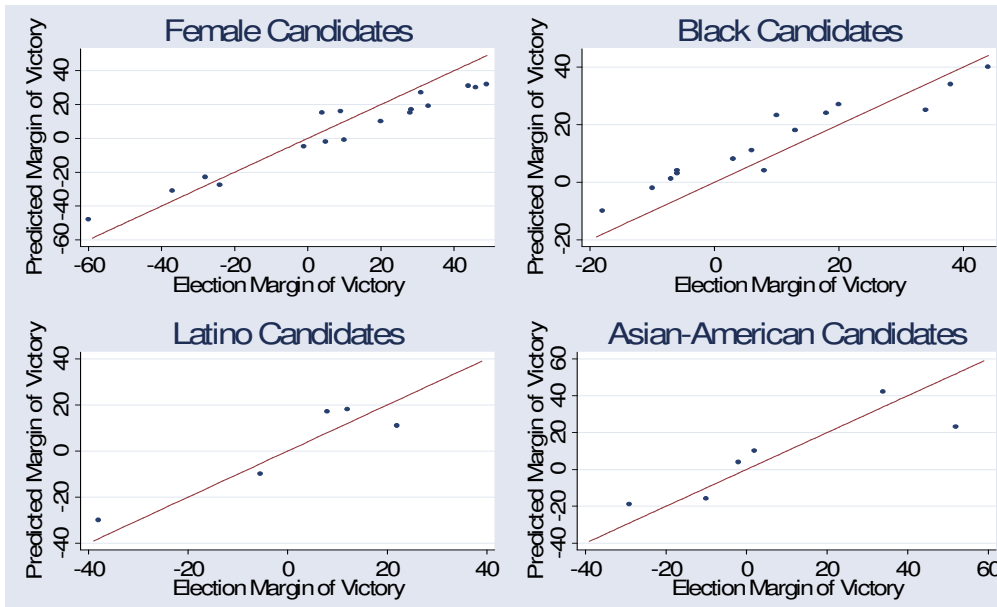
Candidate Type	N	Average	SD	% Predicted within Margin of Error	% Negative Outside MOE	% Positive Outside MOE
White Male	85	0.56	7.9	41%	41%	59%
White Female	31	-3.1	7.3	42%	72%	28%
Black	23	3.05	5.5	35%	26%	74%
Latino	10	0.71	4.8	50%	40%	60%
Asian American	6	-0.5	12	0%	34%	66%

### Descriptive Statistics and Non-Parametric Tests

In the aggregate there is essentially no difference between the pre-election polls and the final results for white male, Asian American, and Latino candidates. For all these groups, the average difference between the pre-election poll and the final result is under one percent. This randomness in poll inaccuracies does not seem to happen for black and white female candidates suggesting that there may be some systematic biases. Pre-election polls tend to underestimate the final results for white female candidates and overestimate the final results for black candidates. The latter result seems to be consistent with others who examine the accuracy of pre-election polls for black candidates (e.g. Traugott and Price 1991; Krysnan 1998; Berinsky 1999). The former, however, seems to be a new finding.

**Figure 1: Polling Inaccuracies: Scatter plot of White Male, White Female, Black, Latino and Asian American Candidates outside the Margin of Error W/ Equality Line.**





Only candidates whose Bradley Effect score ((pre-election poll margin of victory)-(election margin of victory)) was greater than the margin of error were included in this graph. The equality line is a 45 degree line, if candidates fall on this line they do as well as they are predicted to do. If they fall above the line the pre-election poll overestimated their support in the final election, if they fall below the pre-election poll underestimated their performance.

While polls may be quite accurate in predicting the winning candidate, most polls do not predict the margin of victory within the margin of error. Pre-election polls appear to be the most biased for white female and black candidates. In about three-fourths of the female cases, the pre-election polls under predict women’s performance in the election. The opposite is true for blacks; in about three-fourths of the cases outside of the margin of victory pre-election polls overestimated black support. These results are displayed graphically in Figure 1. For blacks, most candidates fall above the 45 degree line, showing that most blacks perform better in pre-election polls than in the election. For women the opposite appears to be true. Most women candidates fall below the 45 degree line.

**Table 2: Wilcoxon Test Predicting Bradley Effect**

Candidate Type	N	Z
White Males	85	1.5
White Female	31	-2.25*
Black	23	2.37*
Latino	10	0.35
Asian American	6	0.62

\*Significant at .05. Comparison Variables: Predicted Margin of Victory & Election Margin of Victory. Positive results show a Bradley Effect because the candidate performed better in pre election polls than in the election. Negative results indicate a Hillary effect.

The Wilcoxon test corroborates these findings. White female and black candidates’ final election margin of victory is significantly different than their predicted margin of victory. There does appear to be a Bradley Effect for black candidates and a Hillary Effect for female candidates. The differences between the predicted margin of

victory and the actual election margin of victory are not significantly different for white male, Latino, or Asian American candidates. This cursory analysis suggests that there are systematic biases for white women and black candidates; however, further analysis should be done to assess the robustness of this finding.

**Table 3: Mann-Whitney Tests Predicting the Bradley Effect**

Candidate Type	N	Z-Score
<b>Baseline</b>		
Black Cand. Vs. White Male Cand.	42	0.16
Female Cand. Vs. White Male Cand.	60	-2.09*
Latino Cand Vs. White Male Cand.	20	0
Asian American Cand. Vs. White Male Cand.	12	-0.16
<b>No Incumbent In Race</b>		
Black Cand. Vs. White Male Cand.	18	2.04*
Female Cand. Vs. White Male Cand.	28	-1.12
Latino Cand. Vs. White Male Cand.	8	0.29
Asian American Vs. White	6	0.21
<b>Republican Candidates</b>		
Black Cand. Vs. White Male Cand.	10	-1.06
Female Cand. Vs. White Male Cand.	26	-1.9*
Latino Cand Vs. White Male Cand.	4	0
Asian American Cand. Vs. White Male Cand.	8	-0.29
<b>Democratic Candidates</b>		
Black Cand. Vs. White Male Cand.	27	-1.69+
Female Cand. Vs. White Male Cand.	36	1.73
Latino Cand Vs. White Male Cand.	11	0
Asian American Vs. White Male Cand.	4	0

+Approaching Significance at .103 \*Significant at .05. DV: Bradley Effect (Pre-Election Polls-Final Results). Positive results show a Bradley Effect because the candidate performed better in pre election polls than in the election. Negative results indicate a Hillary effect.

As we mentioned in the literature review, certain types of elections and candidates are more susceptible to polling inaccuracies. Therefore, we ran several Mann-Whitney tests examining different types of candidates and contexts. Comparing the total number of the non-traditional candidates (elections with blacks, white females, Latinos, and Asian Americans) to their traditional comparisons (elections with two white male candidates), it appears as if the Bradley Effect for blacks disappears. This supports the findings of those who argue that the Bradley Effect may not be real phenomenon (Keeter and Samaranayake 2007; McWhorter 2008; Walters 2008). In contrast to the other groups, female candidates perform significantly better in the final results than they are predicted to do in pre-election polls. Women who are Republicans are the most prone to a Hillary Effect. This fits our preference falsification model. Republicans, who are known to be less progressive on women’s issues (Carroll 1984; Freeman 1993), may have the most reputation to lose by admitting a preference for a woman. Black candidates who are in elections with no incumbent and black Democrats are the most likely to fall victim to the Bradley Effect, although the former context is only marginally significant. This also fits our preference falsification model, as Democrats have the most to gain reputationally by

admitting to vote for a black Democrat and hence more incentive to be misleading (Finkel et al. 1991; Traugott and Price 1999). All other groups were not significantly different from their comparison white male candidates. The Mann-Whitney test shows that polling for black and female candidates may be problematic under certain circumstances. Multiple regression will allow us to determine whether or not these pre-election polls are significantly different from whites controlling for a host of different variables beyond just a few dichotomous variables.

## **Regression Results**

The first model (Baseline Model) shows that black, Latino, and Asian American candidates do not encounter a distinct pre-election bias when compared to white candidates. Female candidates, on the other hand, perform significantly worse—by 3.6 points—in pre-election polls than the final results when compared to white male candidates. This finding supports the relationship found in previous tests. Surprisingly, there is no significant difference between white male and black candidates with regards to polling inaccuracies, without controlling for other variables. There is a similar result for Latino and Asian American candidates.

The Bradley Effect reappears and the Hillary Effect becomes stronger both in terms of significance and magnitude when you control for several election context variables (Model 2, Election Context Model). Black candidates now perform 3.3 points better in pre-election polls and women candidates perform 4.1 percent worse in their respective polls when compared to white male candidates. Asian American candidates also appear to face some sort of a Bradley Effect. While only significant at .10, Asian American candidates perform 5.3 percent better in the pre-election poll than in the final result. We are more cautious about this finding however, due to the small number of cases. Non competitiveness was the only other variable that was significant in this specification. As an election becomes less competitive, polls underestimate a candidate's support by .16 points for every predicted percentage point difference between the two candidates. Other variables, including turnout and percent undecided, had no significant bearing on pre-election poll inaccuracies<sup>v</sup>.

Our final model (Social Pressure Model) seems to reify the findings about black and women candidates in Model 2 (Election Context Model). In the social pressure model, pre-election polls continue to overestimate support for black candidates and underestimate support for women candidates. Again competitiveness appears to be a factor in the accuracy of polls. Voting age population turnout becomes significant in this model when you drop black and women turnout and include other controls. As turnout increases, pre-election polls become less successful at predicting the final result. Greater turnout leads polls to overestimate the candidate's election performance by 1.21 percent for each standard deviation above the mean in turnout.

**Table 4: Predicting the Bradley Effect: OLS Regression Predicting the Difference between the Latest Pre-Election Poll and the Final Results for Black, Latino, Asian American, and Women Candidates from 1982-2006.**

DV=Bradley Effect	Baseline Model	Election Context Model	Preference Falsification Model
<u>Candidate Type</u>			
Black	2.57 (1.82)	3.29* (1.56)	3.51* (1.65)
Female	-3.57* (1.63)	-4.12* (1.47)	-3.33* (1.39)
Latino	.24 (2.59)	1.20 (2.86)	1.81 (2.39)
Asian American	-.97 (3.28)	5.28+ (3.06)	.83 (2.93)
<u>Election Context</u>			
Non Competitive		-.16* (.025)	-.18* (.024)
Democrat		-0.036 (1.14)	1.00 (1.13)
VAP Turnout		.49 (.66)	1.21* (.58)
Black Turnout		-0.02 (0.57)	X
Women Turnout		.75 (.70)	X
Incumbent in Race		-.134 (1.18)	-.24 (1.14)
% Undecided		-.109 (.128)	.002 (.124)
Gubernatorial Election		-1.09 (1.92)	.64 (1.12)
<u>Falsification Variables</u>			
LCCR Score			.009 (.036)
Black Population			-0.05 (0.07)
% Urban			-0.005 (0.006)
Women in Legislature			-.008 (.07)
Democratic Partisanship			9.37 (9.89)
Constant	.46 (.84)	1.06 (4.47)	.37 (4.47)
N	155	115	151
Adj R	0.03	0.33	0.32

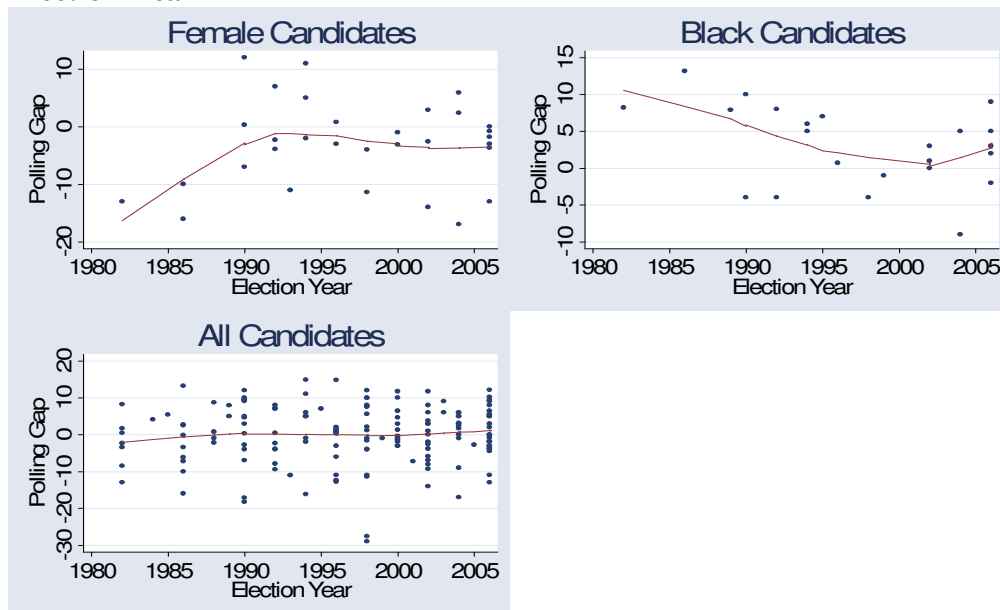
\*Significant at .05

DV: (For all Models) Latest Pre-Election Poll Margin of Victory-Final Result Margin of Victory (Positive direction signals that polls overestimated margin of victory). Positive results show a Bradley Effect because the candidate performed better in pre election polls than in the election. Negative results indicate a Hillary effect. Standard Errors are presented within parenthesis.

Omitted Comparison: White Male Candidates.

Unfortunately, none of the social pressure variables are significant. The Social Pressure Model also dropped the Adjusted  $R^2$  from .33 in the 2<sup>nd</sup> model (Election Context Model) to .32. These findings do not confirm our hypothesis about preference falsification, but they also do not rule it out. Preference falsification occurs at a more local context, so these aggregate measures are not the ideal way to test our hypothesis.

**Figure 2: Scatter plot with Smoothing Line: Bradley Effect and Hillary Effect by Election Year**



## Conclusion

Those who claim that the Bradley Effect is an artifact of the past are partially correct. Though the smoothed line in Figure 2 suggests that the Bradley Effect may be disappearing for black candidates, this apparent increase in the accuracy of polls is, according to our findings, due largely to changes in the context in which black candidates are running. This seeming decrease in the Bradley Effect could be due to that fact that the black gubernatorial and senate candidates are less competitive after the year 2000. Before 2000, 33 percent of black gubernatorial and Senate candidates ran in races that were decided by 4 points or less; only seven percent were in such elections after 2000. Furthermore, before 2000 almost 50 percent ran in elections that were decided by 10 percent or less; after 2000 only 16 percent were in such elections. Our OLS regression showed that predicted margin of victory was a strong predictor of poll inaccuracies, with the less competitive candidates performing better in the final result. Also, more black Republicans (31 percent after 2000 compared to 16 percent before 2000) ran for gubernatorial and Senate positions. This may explain why a cursory analysis would conclude that the Bradley Effect has diminished over time.

When controlling for these and other variables in our models, the Bradley Effect becomes significant. Our models suggest that the decrease in the Bradley Effect observed

by several social scientists and political pundits (e.g., Keeter and Samaranayake 2007; McWhorter 2008; Walters 2008) could be the result of the context in which the recent crop of black candidates ran. Our findings suggest that black candidates are truly susceptible to problems with pre-election polling.

Most pre-election polls predicted Barack Obama to win in the 2008 New Hampshire Democratic Presidential Primary. On election night, Hillary Clinton surprised everyone and defeated Obama. This was déjà vu for many political commentators and some began to argue that the Bradley Effect had reappeared. While several political pundits called this the Bradley Effect, our study provides new insight into this case. Perhaps it was not that the polling was not only overestimating Obama's support, but also underestimating Clinton's. This seems to be consistent with the findings from this election. One of the major findings in this paper seems to be that female candidates are the victims of the Hillary Effect. White women candidates like Hillary Clinton tend to do worse in pre-election polls than in actual elections.

Poll discrepancies for women could be due to preference falsification. Under reporting the vote for women could come from either or both men and women. First, it could be that male respondents want to show their masculinity by telling the pollster that they are not going to vote for a women candidate. Second, women respondents may be embarrassed about admitting to vote for a women candidate. Women who truly support a female candidate may feel that they are being judged for voting on superficial grounds, and thus may falsify their response to appease the interviewer. Unfortunately, we do not have the data to test these hypotheses so these are just conjectures.

In 2008, Geraldine Ferraro argued that sexism is still pervasive in US culture and may be more acceptable when compared to racism. These findings suggest that there may be a stigma attached to supporting a female candidate, much like there is for those who do not vote for a black candidate (assuming that the candidate is viable and the member of the respondent's party). This problem of gender bias is nothing new. As several authors have shown, polls often are inaccurate when measuring gender related issues (Ballou and Del Boca 1980; Lueptow et al. 1990; Northrup 1992). Female candidates do not appear to be immune to this gender bias.

While our data show that, on average, these discrepancies are "within" the common margin of error for most polls, the operationalization of our variables and our research design, render this concern moot. Margins of error are probabilistic, thus, in expectation they will cancel each other out. Because we find significant and substantial differences in the accuracy of polling for black and female candidates, *when compared to* white male candidates, our findings are nonetheless substantive. Though these findings suggest that a small proportion of survey respondents are misrepresenting their voting intentions, even a small proportion of such respondents may alter the accuracy of predictions in close elections.

We were unable to adequately test our preference falsification hypothesis, but not for lack of effort.<sup>vi</sup> That being said we believe that there is some evidence that can be drawn from this paper to support our hypothesis. The strongest evidence can be taken from our Mann-Whitney Tests where female Republican and black Democratic candidates are the most susceptible to polling inaccuracies. Due to the fact that most pre-election polls are not available to the public and that exit polls are already adjusted to mirror the actual results, to fully understand the causes of the Bradley Effect we suggest

experimental studies such as those conducted by Krysnan (1998) to determine whether preference falsification can explain poll inaccuracies for blacks and women candidates

## References

- Ballou, Janice, and Francis K. Del Boca. (1980). "Gender Interaction Effects on Survey Measures in Telephone Interviews." Paper presented at the 1980 meetings of the American Association for Public Opinion Research, Mason, OH.
- Berinsky, Adam J. (1999). "The Two Faces of Public Opinion." American Journal of Political Science 43(4): 1209-30.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. (2001). "Overreporting Voting: Why It Happens and Why It Matters." Public Opinion Quarterly 65(1): 22-44.
- Bobo, Lawrence and Vincent L. Hutchings (1996). "Perceptions of Racial Group Competition: Extending Blumer's Theory of Group Position to a Multiracial Social Context." American Sociological Review 61(6): 951-972.
- Bobo, Lawrence and James R. Kluegel (1993). "Opposition to Race-Targeting: Self-Interest, Stratification Ideology, or Racial Attitudes?" American Sociological Review 58(4): 443-464.
- Bolstein, Richard (1991). "Predicting the Likelihood to Vote in Pre-Election Polls." The Statistician 40(3): 277-283.
- Bullock Iii, Charles S. (1984). "Racial Crossover Voting and the Election of Black Officials." Journal of Politics 46(1): 238-251.
- Carroll, Susan, Women Candidates and Support for Feminist Concerns: The Closet Feminist Syndrome. Western Political Quarterly, 1984. 37(2): p. 307-323.
- Catania, Joseph A., Lois J. McDermott, and Lance M. Pollack, (1986). "Questionnaire response bias and face-to-face interview sample bias in sexuality research." Journal of Sex Research 22(1): 52-72.
- Citrin, Jack, Donald P. Green, and David Sears. (1990). "White Reactions to Black Candidates: When Does Race Matter?" The Public Opinion Quarterly 54(1): 74-96.
- Crespi, Irving. (1988). Polling Sources of Accuracy and Error, New-York: Russell Sage.
- DiCamillo, Mark.( 1991). "Exploring Gender Effects in the June 1990 California Gubernatorial Primary Election." Paper presented at the 1991 meetings of the American Association for Public Opinion Research, Phoenix.
- Durand, Claire., Andre Blais, and Sebastien Vachon. (2001). Review: A Late Campaign Swing or a Failure of the Polls? The Case of the 1998 Quebec Election, AAPOR. 65: 108-123.

- Fenwick, Ian, Federick Wiseman, John F. Becker, and James R. Heiman. (1982). "Classifying Undecided Voters in Pre-Election Polls." The Public Opinion Quarterly 46(3): 383-391.
- Finkel, Steven. E., Thomas M. Guterbock, and Marian J. Borg. (1991). "Race-of-Interviewer Effects in a Preelection Poll: Virginia 1989." The Public Opinion Quarterly 55(3): 313-330.
- Freeman, Jo, Feminism vs. Family Values: Women at the 1992 Democratic and Republican Conventions. PS: Political Science & Politics, 1993. 26: p. 21-28.
- Giles, Michael W. and Melanie A. Buckner (1993). "David Duke and Black Threat: An Old Hypothesis Revisited." Journal of Politics 55(3): 702-713.
- Gow, John. and Eubank, Robert B., 1984. The pro-incumbent bias in the 1982 national election study. American Journal of Political Science 28, pp. 224–230
- Hajnal, Zoltan L. (2002). "White Residents, Black Incumbents, and a Declining Racial Divide." American Political Science Review 95(03): 603-617.
- Hatchett, Shirley, and Howard Schuman. 1975. "White Respondents and Race-of-Interviewer Effects." Public Opinion Quarterly 39:523–28.
- Hopkins, Daniel (2008) No More Wilder Effect, Never a Whitman Effect: When and Why Polls Misperceive Black and Female Candidates. <http://people.iq.harvard.edu/~dhopkins/wilder13.pdf>
- Huddy, Leonie, Joshua. Billig, John Bracciodieta, Lois Hoeffler, Patrick Moynihan, and Patricia Pugliani. (1997). "The Effect of Interviewer Gender on the Survey Response." Political Behavior 19(3): 197-220.
- Jackman, Mary R., and Michael J. Muha. 1984. "Education and Intergroup Attitudes: Moral Enlightenment, Superficial Democratic Commitment, or Ideological Refinement?" American Sociological Review 49:751–69.
- Johnson, Weldon T., and John D. DeLamater. 1976. "Response Effects in Sex Surveys." Public Opinion Quarterly 40:165-81.
- Keeter, Scott, & Samaranayake, Nilanthi. (2007). Can you trust what polls say about Obama's electoral prospects. Retrieved March 12, 2008, from <http://pewresearch.org/pubs/408/can-you-trustwhat-polls-say-about-obamas-electoral-prospects>
- Kinder, Donald. R. and Lynn. M. Sanders (1996). Divided by Color: Racial Politics and Democratic Ideals, University Of Chicago Press.

- Krysan, Maria. (1998). "Privacy and the Expression of White Racial Attitudes: A Comparison across Three Contexts." Public Opinion Quarterly 62(4): 506-544.
- Kuran, Timur. (1987). "Chameleon voters and public choice." Public Choice 53(1): 53-78.
- Kuran, Timur. (1987). "Preference Falsification, Policy Continuity and Collective Conservatism." Economic Journal 97(387): 642-665.
- Kuran, Timur. (1995). "Private Truths." Public Lies: The Social Consequences of Preference. Harvard University Press. Cambridge, MA; Harvard University Press
- Lueptow, Lloyd B., Susan L. Moser, and Brian F. Pendleton. 1990. "Gender and Response Effects in Telephone Interviews about Gender Characteristics." Sex Roles 22:29-42.
- McWhorter, John. "Bradley Effect R.I.P. *New York Sun* Jan 17, 2008 E9
- Mitofsky, Warren. J. (1998). "Review: Was 1996 a Worse Year for Polls Than 1948?" The Public Opinion Quarterly 62(2): 230-249.
- Mosteller, Frederick. 1949. "Measuring the Error." In *The Pre-election Polls of 1948*, by Frederick Mosteller et al. New York: Social Science Research Council.
- National Council on Public Polls (NCPP). 1997. "Polling Council Analysis Concludes Criticisms of 1996 Presidential Poll Accuracy Are Unfounded." Press release, February 13. National Council on Public Polls, Fairfield, CT.
- Northrup, David A. (1992). "Gender, Gender, Gender: The Effect of Interviewer's Gender on Respondent's Answers to Affirmative Action Items." Working Paper. North York, Ontario: Institute for Social Research.
- Perry, Paul. (1979). "Certain Problems in Election Survey Methodology." The Public Opinion Quarterly 43(3): 312-325.
- Reeves, Keith. (1997). Voting Hopes Or Fears?: White Voters, Black Candidates & Racial Politics in America, Oxford University Press, USA.
- Schuman, Howard, Charlotte Steeh, Lawrence Bobo, and Maria Krysnan. (1997). Racial Attitudes in America: Trends and Interpretations, Harvard University Press. Cambridge, MA; Harvard University Press
- Shaw, Daron, Rodolfo. de la Garza, and Jongho Lee. (2000). "Examining Latino Turnout in 1996: A Three-State, Validated Survey Approach." American Journal of Political Science 44(2): 338-346.

Silver, Brian D., Barbara. A. Anderson, and Paul Abramson (1986). "Who Overreports Voting?" American Political Science Review 80(2): 613-624.

Terkildsen, Nayda. (1993). "When White Voters Evaluate Black Candidates: The Processing Implications of Candidate Skin Color, Prejudice, and Self-Monitoring." American Journal of Political Science 37(4): 1032-53.

Traugott, Michael W. (2001). "Trends: Assessing Poll Performance in the 2000 Campaign." The Public Opinion Quarterly 65(3): 389-419.

Traugott, Michael W. (2005). "The Accuracy of the National Preelection Polls in the 2004 Presidential Election." Public Opin Q 69(5): 642-654.

Traugott, Michael W. and Vincent Price (1992). "A Review: Exit Polls in the 1989 Virginia Gubernatorial Race: Where Did They Go Wrong?" The Public Opinion Quarterly 56(2): 245-253.

Wilson, John. (2008). Barack Obama: The Improbable Quest. Boulder, CO: Paradigm Publishers

Williams, Linda F. (1990). "White/Black Perceptions of the Electability of Black Political Candidates." Black Electoral Politics: 45-64.

Wright, Gerald C., 1990. Misreports of vote choice in the 1988 NES senate election study. Legislative Studies Quarterly 15, pp. 543–563

Wright, Gerald C., 1992. Reported versus actual vote: there is a difference and it matters. Legislative Studies Quarterly 17, pp. 131–142

Zhang, Yin (2000). "Using the Internet for survey research: A case study." Journal of the American Society for Information Science 51(1): 57-68.

---

<sup>i</sup> We are grateful for helpful comments on early drafts of this paper from Bernard Grofman, Louis DeSipio, Carole Uhlaner, David Lublin, Tasha Philpot, Peter Wielhouwer and participants of the 2008 PREIC summer colloquium in Irvine, CA and the 2008 APSA annual meeting.

<sup>ii</sup> Hopkins 2008 does, in fact, systematically analyze the Bradley effect (the author uses an alternate definition of the phenomenon—the Wilder Effect). We are happy to find that others are interested in this phenomenon. Our analysis, however, differs from his in several important ways, including the operationalization of the dependent variable, use of statistical methods, and the inclusion of other underrepresented groups (Asian Americans and Latinos). This paper was written and presented before the authors discovered his paper.

<sup>iii</sup> This is consistent with research in social psychology. For example, Snyder (1987) argues that individuals differ to the extent that they engage in “self-monitoring” and thus in the importance they accord to the views of others.

<sup>iv</sup> Illinois 2004, Mississippi 2000 and 2006, and New Mexico 2002 were excluded. In the 2004 Illinois senate election there were 2 black candidates—thus making the relevant comparison impossible. In the 2002 New Mexico gubernatorial election there were two Latino candidates. No information was found for Mississippi 2000 or 2006. For the same reasons of comparison, elections with two female candidates were also excluded.

<sup>v</sup> We tested many other variables which are not included in our final models shown below. These variables include a period effect (in order to capture the effect of racial re-districting in 1992), a polling firm effect, numerous interactions between the “falsification” variables, a regional effect, an indicator for female minorities, and an interaction between female candidacies and gubernatorial elections.

<sup>vi</sup> Numerous attempts were made to analyze different data sets in order to make inferences with respect to preference falsification. These included attempts to use ecological inference and polynomial regression. Unfortunately, no data was available. Exit polls generally make ex-post adjustments in order to better their data to the actual result. Other data sets considered include the American National Election Study, Senate Election Survey, and National Election Pool General Election Exit Polls among others.