
The Unexplained Exit Poll Discrepancy

Steven F. Freeman, PhD

November 11, 2004

stfreema@sas.upenn.edu

Most Americans who had listened to radio or surfed the Internet on Election Day this year, sat down to watch election night coverage expecting that John Kerry had been elected President. Exit polls showed him ahead in nearly every battleground state, in many cases by sizable margins. Although pre-election day polls indicated the race dead even or Bush slightly ahead, two factors seemed to explain Kerry's edge: turnout was extraordinary high¹ and the exit polls indicated that, in every US Presidential election over the past quarter-century², undecided voters broke heavily toward the challenger.³

But then in key state after key state, counts were showing very different numbers than the polls predicted; and the differentials were all in the same direction. The first shaded column in Table 1 shows the differential between the major candidates' predicted (exit poll) percentages of

1 The common wisdom, expressed for example, in articles such as The Gallup Poll Monthly, Oct 1998 i397 p5(3) is that turnout will be a deciding factor in close elections, and that Democratic party candidates generally benefit from higher turnout because the increase comes disproportionately from voters in lower socio-economic groups that traditionally vote Democratic.

2 Guy Molyneux, "The Big Five-Oh", The American Prospect Online, Oct 1, 2004: There have been four incumbent presidential elections in the past quarter-century. On average, the incumbent comes in half a point below his final poll result. The challenger(s) exceed their final poll result by at least 2 points, and the average gain is 4 points. In 1980, Ronald Reagan received 51 percent, fully 6 percentage points above his final poll results.

3 Even the final "corrected" exit poll data (more on that later) presented on the CNN website, indicates that those who decided in the last three days chose Kerry over Bush 55% - 42%.
(<http://www.cnn.com/ELECTION/2004/pages/results/states/US/P/00/epolls.0.html>) Sunday Nov 14, 2004

Author's note: Given the timeliness of the subject matter, I have released this paper despite not having the opportunity to use normal academic safeguards. If you have any questions or comments, please write. Likewise, if you publish or post it to web pages, electronic bulletin boards, or other electronic archives, please let me know.

I have tried to be as rigorous as possible in my data collection, review, and analysis and I believe it compares favorably to the vast majority of commentary currently in the public domain. To hold it to an academic standard of rigor, however, requires extensive peer review; this work has barely begun to be challenged by – and improved from – this peer review process. I will revise the paper as legitimate objections are raised, as I learn more, and as more data becomes available.

the vote; the next shaded column shows the differential between their *tallied* percentages of the vote. The final shaded column reveals the “shift.” In ten of the eleven consensus battleground states⁴, the tallied margin differs from the predicted margin, and in every one, the shift favors Bush.

Table 1⁵: Predicted vs. Actual percentages in battleground states⁽¹⁾

	Bush predicted	Kerry predicted	Predicted differential	Bush tallied	Kerry tallied	Tallied differential	Tallied vs. predicted	Sample size
Colorado	49.9%	48.1%	Bush 1.8	52%	46.8%	Bush 5.2	Bush 3.4	2515
Florida	49.8%	49.7%	Bush 0.1	52.1%	47.1%	Bush 5.0	Bush 4.9	2846
Iowa	48.4%	49.7%	Kerry 1.3	50.1%	49.2%	Bush 0.9	Bush 2.2	2502
Michigan	46.5%	51.5%	Kerry 5.0	47.8%	51.2%	Kerry 3.4	Bush 1.6	2452
Minnesota	44.5%	53.5%	Kerry 9.0	47.6%	51.1%	Kerry 3.5	Bush 5.5	2178
Nevada	47.9%	49.2%	Kerry 1.3	50.5%	47.9%	Bush 2.6	Bush 3.9	2116
New Hampshire	44.1%	54.9%	Kerry 10.8	49.0%	50.3%	Kerry 1.3	Bush 9.5	1849
New Mexico	47.5%	50.1%	Kerry 2.6	50.0%	48.9%	Bush 1.1	Bush 3.7	1951
Ohio	47.9%	52.1%	Kerry 4.2	51.0%	48.5%	Bush 2.5	Bush 6.7	1963
Pennsylvania	45.4%	54.1%	Kerry 8.7	48.6%	50.8%	Kerry 2.2	Bush 6.5	1930
Wisconsin	48.8%	49.2%	Kerry 0.4	49.4%	49.8%	Kerry 0.4	No dif	2223

The media has largely ignored this discrepancy (although the Blogosphere has been abuzz), suggesting that the polls were flawed, within normal sampling error, or that it was a statistical anomaly, or have otherwise explained it away. In this paper, I examine the validity of exit polls, the likelihood sampling error, the possibility of statistical anomaly, and other explanations thus far offered.

Source of the Data

All of the 2004 exit poll data that I use here are unofficial, not meant to be released directly to the public.⁶ The data come from exit polls conducted for the National Election Pool, a consortium of the major television networks and the Associated Press, by two respected polling firms, Edison Media Research and Mitofsky International, whose founder Warren Mitofsky is credited with having invented the exit poll.⁷

4 These eleven states are classified as battleground states based on being on at least two of three prominent lists: Zogbys, MSNBC, and the Washington Post.

5 Material for this chart comes from Jonathon Simon, a former political survey research analyst, who collected and tabulated data from the CNN website before the data changed. An explanation of the how the columns were derived is presented in the “Data and Statistical Analysis” section of the paper.

6 Those who purchased the information signed agreements prohibiting the release of the data. (Martin Plissner, “In Defense of Exit Polls: You just don’t know how to use them” *Slate* Thursday, Nov. 4, 2004)

7 David W. Moore, Senior Gallup Poll Editor, “New Exit Poll Consortium Vindication for Exit Poll Inventor” 10/11/2003

The pollsters have taken great pains to argue that the polls were not designed to verify election results⁸, but rather to help subscribers explain voting patterns and as one piece of data networks could use to “call” states. The data I use for this analysis were available apparently only because a computer glitch allowed apparently “uncalibrated” data (not yet “corrected” to conform to announced vote tallies) to remain on the CNN website until approximately 1:30 AM election night.⁹ At that time, CNN substituted data “corrected” to conform to reported tallies.¹⁰

On Exit Polls

Caveats aside, the data appear to be good.¹¹ In general, we have every reason to believe that exit polls are accurate survey instruments. Exit polls are surveys taken of representative respondents from the overall voting population. Although exit polls have not been academically studied, both the logic behind them and experience suggest that we can use these surveys to predict overall results with very high degrees of certainty. It's easy to get a statistically valid representative sample; and there is no problem with figuring out who is going to actually vote -- or how they will vote.

Thom Hartmann relates that in Germany,

... people fill in hand-marked ballots, which are hand-counted by civil servants, watched over by volunteer representatives of the political parties. ... even though it takes a week or more to count the vote ... the German people know the election results the night the polls close because the news media's exit polls, for two generations, have never been more than a tenth of a percent off.¹²

Dick Morris, Republican consultant and Fox News regular, concurs:

Exit polls are almost never wrong ... So reliable are the surveys that actually tap voters as they leave the polling places that they are used as guides to the relative honesty of elections in Third

8 IBID (Plissner)

9 Richard Morin, “New Woes Surface in Use of Estimates” *Washington Post*, Thursday, November 4, 2004; Page A29

10 I have been in discussion with Warren Mitofsky about trying to get raw exit poll data from the National Election Poll; in my most recent correspondence, however, Joe Lenski wrote, “All of the exit poll data will be archived at the Roper Center at the University of Connecticut some time in early 2005. You will be able to access of the exit poll data at that point.”

11 I corresponded with Jonathan Simon, a former political survey poll analyst, who says that the 12:21 am data may have already been adjusted to match counts, but were probably still pure. (Internet correspondence Nov 6, 2004). If, in the unlikely event that the data already had been adjusted, it would mean that the pure exit poll numbers favored Kerry to an even greater extent.

12 Thom Hartmann, “The Ultimate Felony Against Democracy” *CommonDreams.org* Thursday, November 4 2004

World countries. When I worked on Vicente Fox's campaign in Mexico, for example, I was so fearful that the governing PRI would steal the election that I had the campaign commission two U.S. firms to conduct exit polls to be released immediately after the polls closed to foreclose the possibility of finagling with the returns.¹³

Last fall, international foundations sponsored an exit poll in the former Soviet Republic of Georgia during a parliamentary election. On Election Day, the pollsters projected a victory for the main opposition party. When the sitting government announced that its own slate of candidates had won, supporters of the opposition stormed the Parliament, and the president, Eduard A. Shevardnadze, resigned under pressure from the United States and Russia.¹⁴

Students at BYU have been conducting Utah exit polls since 1982.¹⁵ They write:

... results are very precise; In the 2003 Salt Lake County mayoral race, the KBYU/Utah Colleges Exit Poll predicted 53.8 percent of the vote for Rocky Anderson and 46.2 percent for Frank Pignanelli. In the actual vote, Anderson carried 54 percent of the vote to Pignanelli's 46 percent.

True to their word, predictions in this year's contests were quite accurate. In the Utah presidential election, for example, they predicted Bush 70.8%, Kerry 26.5%. The actual was Bush 71.1%, Kerry 26.4%. Consistently accurate exit poll predictions from student volunteers, including in this presidential election, gives us good reason to presume valid data from the world's most professional exit polling enterprise.

Data and Statistical Analysis

Three critical Battleground states

The conventional wisdom going into the election was that three critical states would likely determine who would win the Presidential election -- Ohio, Pennsylvania, and Florida. Sure enough, they did, with Bush winning two of three and ascending to electoral victory. In each of these states, however, exit polls differed significantly from recorded tallies.

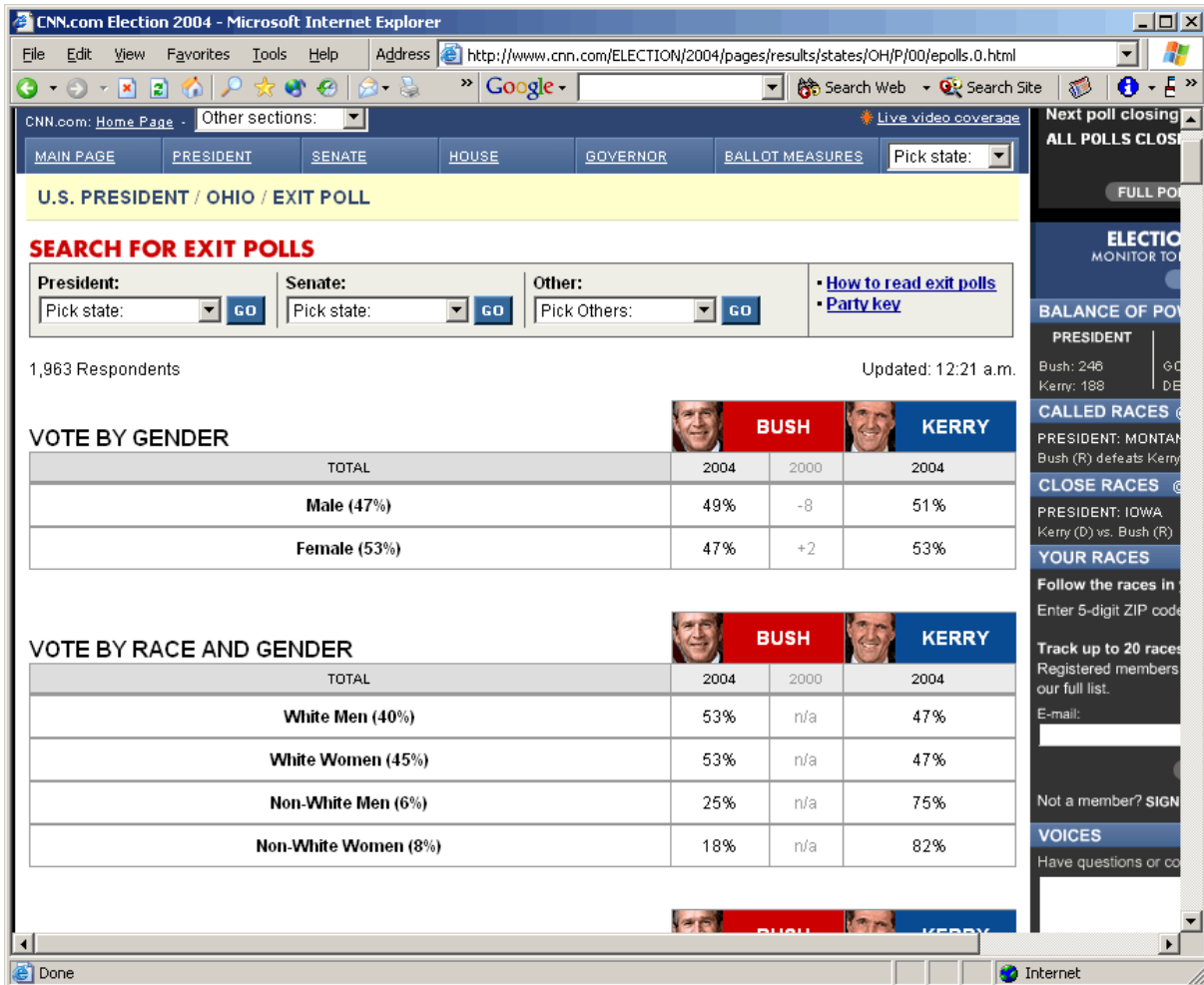
13 Dick Morris, "Those faulty exit polls were sabotage" *The Hill* Nov. 4, 2004 <http://www.thehill.com/morris/110404.aspx>

14 Martin Plissner, Exit Polls to Protect the Vote, *New York Times* 10/17/04

15 <http://exitpoll.byu.edu/2004results.asp>. As far as I have been able to determine, this was the only other exit poll conducted on the 2004 presidential election. I had thought that Zogby also had an exit poll, but haven't been able to verify this; they may have been using the same National Election Pool data, when they declared Kerry the winner in Ohio on Election Night. See William Douglas, "Early exit polls come under fire" *The Mercury News* (11/3/2004)

Data (Ohio). CNN reported the exit poll as illustrated in Figure 1. Combining the male and female vote, weighted for their percentage of the electorate (47% male), Kerry’s predicted share

Figure 1. CNN graphic with apparently “uncorrected” exit poll data



of the total Ohio vote was 52.1%¹⁶. Doing the same for Florida and Pennsylvania, and adding in final tallies (NY Times, Sunday evening), we derive Table 2.

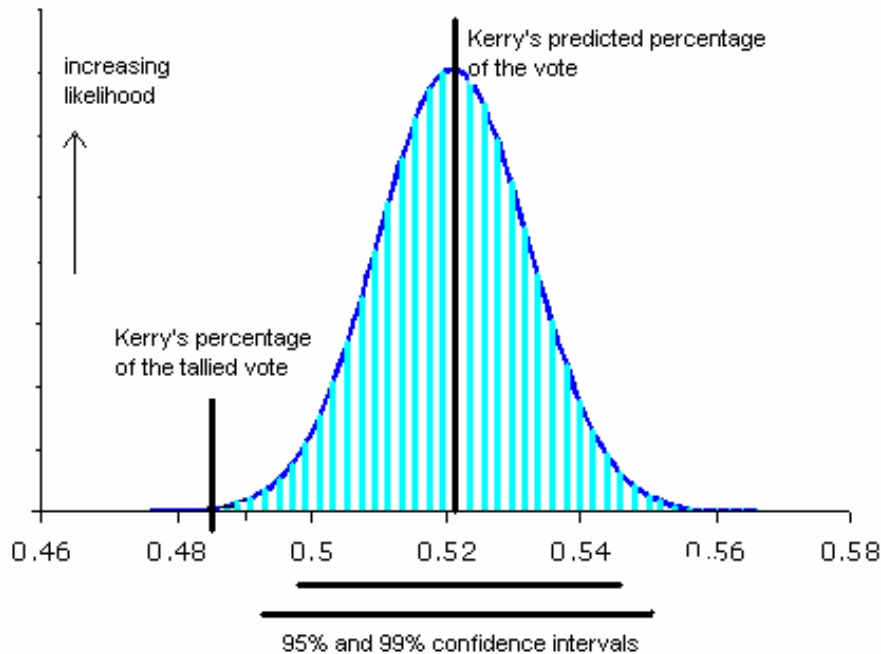
16 Among the limitations of the CNN exit poll data is the lack of significant digits. Rounding errors mean that exit poll numbers for individual state analyses could be off by up to .5. This is unlikely because it comes from two groups, male and female, and it’s unlikely that they are both rounded that much in the same direction. Regardless, the strength of the finding is such that even if all numbers had been rounded the full .5 in an unfavorable direction, the basic finding would still hold.

Table 2: Predicted vs. Actual percentages in the three critical battleground states

	Bush predicted	Kerry predicted	Predicted differential	Bush tallied	Kerry tallied	Tallied differential	Tallied vs. predicted	Sample size
Florida ⁽¹³⁾	49.8%	49.7%	Bush 0.1 ¹⁷	52.1%	47.1%	Bush 5.0	Bush 4.9	2846
Ohio	47.9%	52.1%	Kerry 4.2	51.0%	48.5%	Bush 2.5	Bush 6.7	1963
Pennsylvania	45.4%	54.1%	Kerry 8.7	48.6%	50.8%	Kerry 2.2	Bush 6.5	1930

Statistical significance, which means that the discrepancy is such that it is unlikely to occur by chance, depends on three factors – the size of the discrepancy, the sample size, and the level of significance (just how unlikely does it have to be?) For analysis purposes, we could choose any measure: Bush’s differential, Kerry’s differential or the differential between them; it all works out the same. Based on the analysis that will follow, I’m going to examine Kerry’s percentage of the vote.

Figure 2. Statistical prediction of Kerry’s true percentage of the vote in Ohio



¹⁷ Earlier exit polls, including one released by Slate at 7:28 EST, 28 minutes after the Florida polls closed showed Kerry leading 50% to 49%.

Figure 2 depicts a normal distribution curve¹⁸ probability density showing the relative likelihood, given this poll result, of the actual percentage of the vote he would be expected to receive in the state. The black lines below the curve indicate the poll's statistical margin of error, the corresponding zone of 95 and 99 percent confidence. In this case, given that the exit poll indicated Kerry received 52.1% of the vote, we are 95 percent sure that the true percentage he received was between 49.8% and 54.4%. And because half of the 1 in 20 cases that fall outside the interval would be high rather than low, we're 97.5 percent sure that the true percentage he received was at least 49.8%. We are 99.5% sure that the true percentage he received was at least 49.2%. It turns out that the likelihood that he would have received only 48.5% of the vote is less than one in one thousand (.0008).

Conducting the same analysis for Florida, we find that Kerry's 47.1% of the vote is likewise outside the 99% confidence interval. The likelihood of his receiving only 47.1%, given that the exit polls indicated 49.7%, is less than three in one thousand (.0028). Kerry's count is also outside the 99% confidence interval in the third critical battleground state, Pennsylvania. Although he did carry the state, the likelihood of his receiving only 50.8% given that the exit polls indicated 54.1% is less than two in one thousand (.0018).

The likelihood of any two of these statistical anomalies occurring together is on the order of one-in-a-million. The odds against all three occurring together are 250 million to one. As much as we can say in social science that something is impossible, it is impossible that the discrepancies between predicted and actual vote counts in the three critical battleground states of the 2004 election could have been due to chance or random error.

18 This analysis assumes a simple random sample. Again, the strength of the finding is such that any modification of this assumption would not change the basic finding, but it might be somewhat stronger or slightly weaker depending on exactly how the exit polling was done. If the pollsters broke states into strata (e.g., separating counties into two or more groups by income, age, racial composition, etc..., and then randomly sampled within each strata, then the variances would be reduced and an even stronger case can be made. If on the other hand, states were broken into clusters (e.g., precincts) and then clusters (precincts) were randomly selected (sampling individuals within those selected precincts), the variances would increase. Most survey sampling uses a combination of clusters and strata, but we do not know the details of how this sample was conducted.

Official Explanations

The New York Times tells us that they obtained a report issued by the pollsters that debunked the possibility that the exit polls are right and the vote count is wrong¹⁹, but does not explain beyond that declaration how the possibility was “debunked.” In fact, no evidence at all is presented of skewed survey data or any problem at all with the polls except that “uncorrected” data were released to the public. *Slate* reports that Mitofsky and Lenski insist that the polls were perfectly fine.^{20 21} Mitofsky himself only says, “We are still investigating the source of the exit poll problem.”²²

One of the few people close to the pollsters to offer an explanation early on was Martin Plissner, former executive political director of CBS News (and self-described close friend of the pollsters), who identifies three problems with the polls:

The pollsters who work outside the polling stations often have problems with officials who want to limit access to voters. Unless the interviews have sampled the entire day's voters, the results can be demographically and hence politically skewed. Finally, it is of course a poll, not a set of actual recorded votes like those in the precinct samples collected after the polls close.²³

Regarding the first problem, Jack Shafer of *Slate* observes:

... exit pollsters always encounter overzealous election officials enforcing electioneering laws. Can we really believe that a significant number of the 1,480 exit poll precincts in 50 states and the District of Columbia that Edison/Mitofsky surveyed on Election Day were so affected? And in sufficient numbers to bend state-by-state exit polls in Kerry's favor?²⁴

If in fact, pollsters have a limited sample of voters from a given precinct, they can weight the voters who are polled from that precinct or demographically similar precincts accordingly.

Regarding time of day variation, this paper does not refer to mid-day reports, but rather end-of-day data, which happened to still be available at midnight. But even if there were an early voter bias, is there any reason to believe that early votes would be skewed Democrat?

Stereotypically, Republicans are early risers.

19 Jim Rutenberg, “Report Says Problems Led to Skewed Surveying Data” *New York Times*, Nov. 5, 2004

20 Martin Plissner “In Defense of Exit Polls: You just don't know how to use them. *Slate* Thursday, Nov. 4, 2004

21 Jack Shafer, “The official excuses for the bad exit poll numbers don't cut it.” *Slate* Friday, Nov. 5, 2004, 9:23 PM PT

22 Warren Mitofsky, personal communication, November 10, 2004

23 Martin Plissner “In Defense of Exit Polls: You just don't know how to use them. *Slate* Thursday, Nov. 4, 2004

24 Jack Shafer, “The official excuses for the bad exit poll numbers don't cut it.” *Slate* Friday, Nov. 5, 2004, 9:23 PM PT

Regarding the last ditch argument that it's just a poll, true (of course), but, as I have documented, the evidence and logic on exit polls suggest that we have every reason to believe they are accurate within statistical limits.

Under-representation?

Other explanations put forth by the *Washington Post* charge that samples may have included too many women, too few Westerners, not enough Republicans, etc ...” Regarding the first part of this critique, Morris writes:

The very first thing a pollster does is weight or quota for gender. Once the female vote reaches 52 percent of the sample, one either refuses additional female respondents or weights down the ones one subsequently counted. This is, dear Watson, elementary.

Moreover, the issue of male/female ratio is irrelevant. CNN and others released data presenting male and female preferences separately, thus automatically weighting sex appropriately.

Other potential imbalances are part of normal sampling error. A random sample would result in the poll precision and confidence intervals that I reported. Under such conditions, Republicans, westerners, etc., are equally (un)likely to be over- or under-represented. Imprecise representation is incorporated within the margin of error. (That's why we have the concept of probability densities, margin of error, etc.... If you could choose a perfectly representative sample, you could predict outcomes precisely.) In theory, techniques to ensure sample representativeness²⁵ make the exit polls be even more accurate than my analysis indicates, thus making the observed discrepancies even more unlikely.

Bush voter unwillingness to participate and other “explanations”

Most recently, Senior *Gallup Poll* Editor David W. Moore, report that Mitofsky and Lenski say that,

Kerry voters apparently were much more willing to participate in the exit poll than were Bush voters. The interviewers at each of the sample voting locations are instructed to choose voters as they exit from the voting booth -- every third, or fifth, or eighth voter -- some sequence of this sort

25 Pollsters normally either choose precincts so as to ensure representative samples, or weight respondents by key demographic categories. The Utah Colleges Exit Poll website gives a fairly good basic explanation of polling practices: http://exitpoll.byu.edu/about/survey_sampling_faq.asp and http://exitpoll.byu.edu/about/sample_design.asp.

that is determined ahead of time. Many voters simply refuse to participate in the poll. If the refusers are disproportionately for one candidate or another, then the poll will be biased....²⁶

OK, true enough. If Republicans disproportionately refuse to participate, that *could* explain exit poll error. But do we have any reason to suspect that?

It is *conceivable* that Kerry voters were much more willing to participate in the exit poll than were Bush voters, but although it's not difficult to imagine why some Bush voters might not want to participate, it's also not difficult to imagine why some Kerry voters might not want to participate either.

The problem with this "explanation" or even one that would have considerably more "face validity" (which means that it makes sense on the face of it), such as the possibility that absentee/early voters disproportionately favored Bush²⁷, is that it is not an explanation, but rather a hypothesis. It's *apparent* that "Kerry voters were much more willing to participate in the exit poll than Bush voters" *only given several questionable assumptions*. An explanation would require independent evidence.²⁸

The Role of the Exit Poll

The pollsters have made clear that the purpose of their poll was not to verify the integrity of the election; they were hired by the AP-Network consortium to provide supporting data for election coverage. In this election, concerns about this exit poll-count discrepancy are compounded by concerns about voting technologies, both new (especially electronic voting machines without paper trails) and old (punch card ballots still widely in use). Allegations about miscount and worse have been lighting up the Internet. In at least two cases, vote count errors have been acknowledged and corrected.²⁹ Additional sources of concern include mistabulation through "spoilage," (as we saw in Florida in 2000, large numbers of votes are commonly lost due

26 David W. Moore, Senior Gallup Poll Editor, "Conspiracies Galore" Gallup News Service: November 9, 2004.

27 To the best of my knowledge, the pollsters have not offered absentee/early voters as an "explanation," presumably because they were able to predict any disproportionate support based on previous elections.

28 I could imagine various types of supportive evidence. One possibility would be verifying sampled results versus actual voting patterns in random sample precincts where counts are unimpeachable.

29 "Glitch gave Bush extra votes in Ohio" *cnn.com* 11/05/04. <http://www.cnn.com/2004/ALLPOLITICS/11/05/voting.problems.ap/>

to imperfections in the voting process), overuse of provisional ballots, and limited access by observers to some vote tallies.³⁰

Verifying the integrity of elections is important not only in Mexico, Venezuela, and (the former Soviet Union Republic of) Georgia. Whatever the original purpose of this particular exit poll, it could be used to help verify the integrity of this election if its data and methodology were released.

Summary and Implications

My purpose in this paper has not been to allege election theft, let alone explain it. Rather, I have tried to demonstrate that exit poll data are fundamentally sound, that the deviations between exit poll predictions and vote tallies in the three critical battleground states could not have occurred strictly by chance or random error, and that no explanations of the discrepancy have yet been provided.

The unexplained discrepancy leaves us with two broad categories of hypotheses: the exit poll data are wrong (or misleading) in ways that have not yet been adequately explained, or the count is off. The most important investigations concern verification of the tallies and allegations of fraud on one side; and the exit poll data and methodology on the other. Some useful statistical analyses would compare the “shift” in (a) battleground states vs. non-battleground states; and (b) states, counties and precincts where safeguards are strong vs. those where they are suspect. Obviously, if we had the polling consortium’s raw data, we could do more definitive analyses.

Given that neither the pollsters nor their media clients have provided solid explanations to the public, suspicion of fraud, or among the less accusatory, “mistabulation,” is running rampant and unchecked. That so many people suspect misplay undermines not only the legitimacy of the President, but faith in the foundations of the democracy.

30 Erica Solvig, “Warren's [Warren County, Ohio] vote tally walled off” *Cincinnati Enquirer* Friday, November 5, 2004

Systematic fraud or mistabulation is a premature conclusion, but the election's unexplained exit poll discrepancies make it an unavoidable hypothesis, one that is the responsibility of the media, academia, polling agencies, and the public to investigate.

----- END -----

Dr. Freeman is lecturer and visiting scholar at the Center for Organizational Dynamics, School of Arts & Sciences, University of Pennsylvania; his areas of expertise include resilience, innovation, and research methods. He earned his Ph.D. from the Massachusetts Institute of Technology. Contact him at stfreema@sas.upenn.edu.