



Papers in this series are intended for eventual publication.

Copyrights remain with the authors and/or their publishers. Reproduction, posting to web pages, electronic bulletin boards or other electronic archives is prohibited without express consent of the copyright holders.

Comments or questions directed to the author(s) are most appreciated and will be acknowledged accordingly.

**Graduate Division, School of Arts & Sciences
Center for Organizational Dynamics
Working Paper #04-10 November 22, 2004**

Note on this revision: The thesis of the initial drafts of this paper that were widely circulated on the Internet, was that exit poll discrepancy could not have been due to chance or random error. My purpose was to raise as a legitimate question, What caused it? In this version, I begin to try to answer that question.

I'm releasing this paper in two parts, so as to replace the early drafts as soon as possible early drafts that have been posted on the Internet. Part I is to clean up the previous draft. Part II will be new analysis.

Reading note: To help non-academic readers, I have put in much more explanatory material than typical, especially in the statistical analysis sections. My apologies to those for whom these explanations are superfluous. Just skip over those sections.

The Unexplained Exit Poll Discrepancy: Part I

Steven F. Freeman, PhD

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 high, which is generally good for Democrats,¹ and as in every US Presidential election with an incumbent over the past quarter-century,² undecided voters broke heavily toward the challenger.³

1 Jack Citrin, Eric Schickler & John Sides, "What If everyone voted? Simulating the impact of increased turnout in senate elections" *American Journal of Political Science*, 2003, 47 (1) 75-90: Nonvoters are generally more Democratic than voters. Democratic party candidates generally benefit from higher turnout because the increase comes disproportionately from voters in 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; challengers exceed their final poll result by an average of 4 points.

3 Even the final "corrected" exit poll data presented on the CNN website – more on such corrected data later – indicate 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>> (Thursday Nov 14, 2004)

But then, in key state after key state, counts showed 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 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 states,⁴ the tallied margin differs from the predicted margin, and in every one, the shift favors Bush.

Table 1⁵: Predicted vs. tallied percentages in battleground states

	Bush predicted	Kerry predicted	Predicted differential	Bush tallied	Kerry tallied	Tallied differential	Tallied vs. predicted
Colorado	49.9%	48.1%	Bush 1.8	52.0%	46.8%	Bush 5.2	Bush 3.4
Florida	49.8%	49.7%	Bush 0.1	52.1%	47.1%	Bush 5.0	Bush 4.9
Iowa	48.4%	49.7%	Kerry 1.3	50.1%	49.2%	Bush 0.9	Bush 2.2
Michigan	46.5%	51.5%	Kerry 5.0	47.8%	51.2%	Kerry 3.4	Bush 1.6
Minnesota	44.5%	53.5%	Kerry 9.0	47.6%	51.1%	Kerry 3.5	Bush 5.5
Nevada	47.9%	49.2%	Kerry 1.3	50.5%	47.9%	Bush 2.6	Bush 3.9
New Hampshire	44.1%	54.9%	Kerry 10.8	49.0%	50.3%	Kerry 1.3	Bush 9.5
New Mexico	47.5%	50.1%	Kerry 2.6	50.0%	48.9%	Bush 1.1	Bush 3.7
Ohio	47.9%	52.1%	Kerry 4.2	51.0%	48.5%	Bush 2.5	Bush 6.7
Pennsylvania	45.4%	54.1%	Kerry 8.7	48.6%	50.8%	Kerry 2.2	Bush 6.5
Wisconsin	48.8%	49.2%	Kerry 0.4	49.4%	49.8%	Kerry 0.4	No dif

The media has largely ignored this discrepancy (although the Blogosphere has been abuzz), suggesting that the polls were either flawed, within normal sampling error, a statistical anomaly, or could otherwise be easily explained away. In Part I of this paper, I examine the validity of exit polls, the likelihood of sampling error, and the possibility of statistical anomaly and show that the exit poll discrepancy could not have been due to chance or random error. In Part II, I explore the further whether the count was correct.

Exit Poll Data

The data I use for this paper are those posted on the CNN website Election Night. CNN had the data by virtue of membership in the National Election Pool (NEP), a consortium of news organizations that had pooled resources to conduct a large-scale exit poll (as was done in the 2000

⁴ These eleven states are classified as battleground states based on being on at least two of three prominent lists: *Zogby*, *MSNBC*, and the *Washington Post*. (These eleven did in fact turn out to be the most competitive states in the election along with Oregon where Kerry won by 4%; in no other state was the winning margin was within 7%.)

⁵ Source: CNN website. Wednesday November 3, 2004 12:21 am.

election). NEP, in turn, had contracted two respected firms, Edison Media Research and Mitofsky International,⁶ to conduct the polls.

Calibrated and Uncalibrated Exit Poll Data

Part of the reason the issue went away for the media – and simultaneously raised suspicion on the web – is secrecy and confusion about the data and what exactly is being characterized as the exit poll. If you go to the CNN website or any other website on which 2004 exit poll data is available, you'll see numbers very different from those that were released on Election Day. That's because the survey results originally collected and presented to subscribers were subsequently "corrected" to conform to official tallies.

The pollsters explain this as a natural process: the "uncalibrated" data were preliminary; once the counts come in, they recalibrate their original data on the assumptions that the count is correct, and that any discrepancies must have been due to imbalanced representation in their samples or some other polling error. The pollsters have taken great pains to argue that their polls were not designed to verify election results,⁷ but rather to provide election coverage support to subscribers: as one piece of data (among many) that networks could use to "call" states and to explain voting patterns, i.e., who voted for whom, and why people voted as they did.

Whatever the merits of calibrating exit poll data, it confuses the issue of *why* the uncalibrated polls were so far off and mostly in the same direction. Although this calibration process may seem perfectly natural to NEP, it confuses nearly everyone else, even sophisticated analysts intimately involved in voting issues. The MIT-Caltech Voting Project, for example, issued a report concluding that exit poll data were consistent with state tallies and that there were no discrepancies based on voting method, including electronic voting systems. But they used this adjusted data to exonerate the process! In other words, they used data in which the count is

6 Warren Mitofsky, the founder of Mitofsky International is credited with having invented the exit poll. David W. Moore, Senior Gallup Poll Editor, "New Exit Poll Consortium Vindication for Exit Poll Inventor" 10/11/2003

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

assumed correct to prove that the count *is* correct. And, sadly, this report is being used to dismiss allegations that anything might be awry.⁸

It's an awful mistake, but it is understandable how they could make it. Few of us realized that this data is corrected. Neither the CNN website, nor any other site of which I'm aware, gives any indication that the data were anything other than what nearly all of us imagine exit poll data to be – data based (solely) on subjects surveyed leaving the polling place.

Data Used in This Report

For this report, I use data that apparently are based solely on subjects surveyed leaving the polling place. These data were reportedly not meant to be released directly to the public,⁹ and were reportedly available to late evening Election Night viewers only because a computer glitch prevented NEP from making updates sometime around 8:30 p.m. that night.¹⁰ They were collected by Jonathon Simon, a former political survey research analyst, and are corroborated by saved screen shots (see Figure 1). I happened to have ten exit poll internet pages stored in my computer memory, and in each case, his figures are identical to mine. The numbers are also roughly consistent with those released elsewhere (Appendix B shows *Slate* numbers at 7:28 EST).

To derive the “predicted values” used in Tables 1 & 5, I combine the male and female vote, weighted for their percentage of the electorate. Ohio exit poll data (Figure 1) indicates that 51% of men and 53% of women voted for Kerry. Since the electorate is 47% male/53% female, Kerry's overall share of the exit poll was calculated as $(51\% \times 47\%) + (53\% \times 53\%)$ or 52.1%.¹¹ Doing the same for calculations for other battleground states and comparing these numbers with final tallies (*New York Times*, Nov. 7), I completed the columns in Tables 1 & 5.

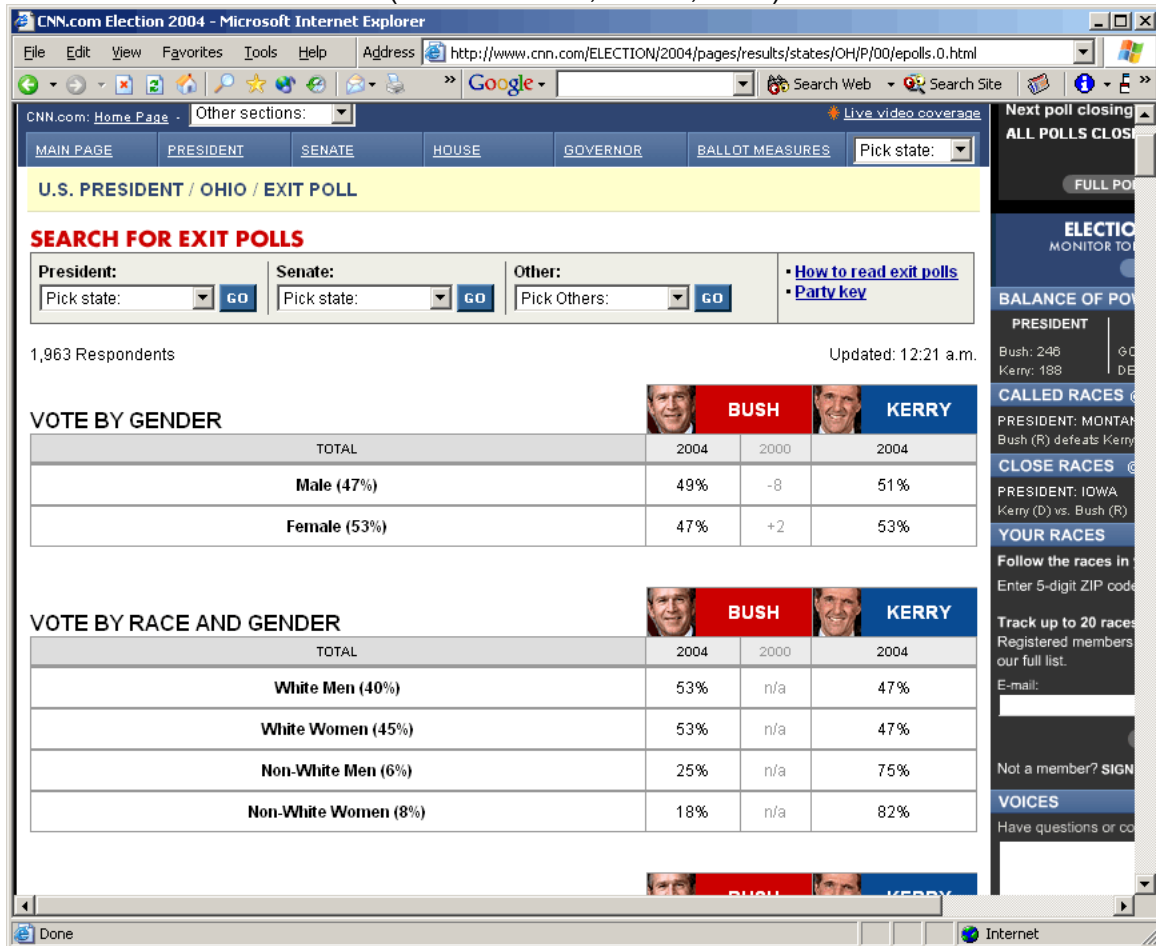
⁸ Tom Zeller, Jr. "Vote Fraud Theories, Spread by Blogs, Are Quickly Buried," *New York Times* (Front page); John Schwartz, "Mostly Good Reviews for Electronic Voting," *New York Times*; Keith Olbermann MSNBC Countdown. All three on November 12, 2004.

⁹ Martin Plissner, "In Defense of Exit Polls: You just don't know how to use them" *Slate* Thursday, Nov. 4, 2004

¹⁰ Richard Morin, "New Woes Surface in Use of Estimates," *Washington Post*, Thurs, Nov. 4, 2004; Page A29

¹¹ 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.

Figure 1. CNN web page with apparently “uncorrected” exit poll data (12:21 am Wed, Nov. 3, 2004)



Is the Data Valid?

Some commentators on an early draft of this paper rejected these data as unweighted, meaning that they have not been adjusted to appropriately weight demographic groups pollsters knowingly under- or over-sampled,¹² but it makes no sense that NEP would *ever* to distribute unweighted data to anyone, let alone publish them on the web election night. NEP’s predecessor, Voter News Service, warns in bold letters in its 2000 Methodology statement never to use unweighted data for any reason (see Appendix A). Pollsters want to get it right. Their customers are depending on it. Broadcasters want to be alerted to probable outcomes, so as to plan their coverage accordingly

¹² Ruy Teixeira produced a chart showing that by using *unweighted* data, Democrats also outpoll Republicans. This is because pollsters typically oversample minorities so that they have a sufficient sample size of important demographic groups. But then they negatively weight respondents in these groups to adjust for their actual percentage of the electorate. ([http://www.emergingdemocraticmajorityweblog.com/donkeyrising/.](http://www.emergingdemocraticmajorityweblog.com/donkeyrising/))

(e.g., pre-writing stories so they could be completed shortly after poll-closings, assigning top reporters to winners' headquarters, being prepared for when concession or victory speeches might be forthcoming, etc...). In this case, subscribers were very surprised. Anchor people were discussing who Kerry would choose for his cabinet, conservative radio hosts were warning how now we're going to see the true John and Teresa Heinz Kerry. Prominent pollster John Zogby trusted the data sufficiently to call the race for Kerry. In the end, network managers had to scramble for coverage; editors had to rewrite headlines, and journalists had to rewrite their lead stories.

It's alternatively possible that the data were already partially calibrated to tallies, but given the news story and the abrupt change at 1:30 am, that seems unlikely. If, in fact, the data already had been partially calibrated, however, it would mean that the pure exit poll numbers favored Kerry to an even greater extent.

In summary, I'd rather have NEP data; lacking that (and unless NEP has a change of heart, no one is going to see those until well into 2005), these CNN data look good, and can be used to generate some highly suggestive findings.

On (Uncorrected) Exit Polls

There are many challenges with conducting an exit poll, several of which potentially might have caused errors that would have resulted in Election Day discrepancies. I'll discuss these at length in a later section of this report (along with a discussion of potential count errors), but in general, we have reason to believe that exit polls, by which I mean *uncorrected* exit polls, are accurate survey instruments. Exit polls are surveys taken of representative respondents from the overall voting population. Both the logic behind them and experience suggest that these surveys should be able to predict overall results within statistical limits. It's relatively easy to get a representative sample, and there is no problem with figuring out who is actually going to vote or how they will vote.

In Germany, the minute the polls close, polling agencies release prognoses that have proven highly reliable. In the three most recent national elections, poll percentages diverged from official counts by only **0.26%** on average (Table 2). They have been almost as accurate for the German vote in the European Parliament Elections (Table 3), averaging **0.44%** differential from tallied results over the past three elections.

Table 2¹³: Exit poll vs Tallies in German National Elections

Parties	2002 predicted	2002 tallied	2002 diff	1998 predicted	1998 tallied	1998 diff	1994 predicted	1994 tallied	1994 diff	average dif
SPD	38.0%	38.5%	0.5%	41.0%	40.9%	0.1%	36.5%	36.4%	0.1%	
CDU/CSU	38.0%	38.5%	0.5%	35.0%	35.2%	0.2%	42.0%	41.4%	0.6%	
Green	9.0%	8.6%	0.4%	6.5%	6.7%	0.2%	7.0%	7.3%	0.3%	
FDP	7.5%	7.4%	0.1%	6.5%	6.2%	0.3%	7.0%	6.9%	0.1%	
PDS	4.0%	4.0%	0.0%	5.0%	5.1%	0.1%	4.0%	4.4%	0.4%	
Rest	9.0%	8.6%		6.0%	5.9%		3.5%	3.6%		
Average differential			0.30%			0.18%			0.30%	0.26%

Table 3: Exit poll vs Tallies in European Parliament Elections (only German part)

Parties	2004 predicted	2004 tallied	2004 diff	1999 predicted	1999 tallied	1999 diff	1994 predicted	1998 tallied	1998 diff	average dif
SPD	22.0%	21.5%	0.5%	31.0%	30.7%	0.3%	33.0%	32.2%	0.8%	
CDU/CSU	45.5%	44.5%	1.0%	48.0%	48.7%	0.7%	40.5%	38.8%	1.7%	
Green	11.5%	11.9%	0.4%	7.0%	6.4%	0.6%	10.0%	10.1%	0.1%	
FDP	6.0%	6.1%	0.1%	3.0%	3.0%	0.0%	4.0%	4.1%	0.1%	
PDS	6.0%	6.1%	0.1%	6.0%	5.8%	0.2%	4.5%	4.7%	0.2%	
REP							3.5%	3.9%	0.4%	
Rest	9.0%	9.8%		5.0%	5.4%		4.5%	6.2%		
Average differential			0.42%			0.36%			0.55%	0.44%

To make the numbers more comparable to the US Presidential election, I have grouped the parties into their coalitions in Table 4.¹⁴ As shown, the results are very accurate, in all cases predict correctly the winner, and show no systematic skew.

¹³ Source: Election data: <http://www.bundeswahlleiter.de/> (English: http://www.bundeswahlleiter.de/wahlen/e/index_e.htm)
Prognoses: www.forschungsgruppe.de Polling data also can be found at
http://de.wikipedia.org/wiki/Prognosen/Hochrechnungen_der_Bundestagswahl_2002 (gives the predictions of another polling company for the 2002 election; it's predictions are in all cases within 1% of ZDF). I'd like to thank Dr. Andreas M. Wuest, Michael Morrissey, Kurt Gloos and Lars Vinx for their help in compiling this data.

¹⁴ FDP Freie Demokratische Partei (liberals) are aligned with Christian Democrats, and Social Democrats (SPD) are aligned with the Greens. PDS (socialists) and Republikaner (extreme right) are not in any coalition.

Table 4: Predicted vs. tallied percentages in the three critical battleground states

Election	Coalition	SPD/ Green predicted	CDU/CSU FDP predicted	Predicted differential	SPD/Gree n tallied	CU/FDP tallied	Tallied differential	Tallied vs. predicted
2004 European Parliament		33.5%	51.5%	CF 18.0	33.4%	50.6%	CF 17.2	SG 0.8
2002 National election		47.0%	45.5%	SG 1.5	47.1%	45.9%	SG 1.2	CF 0.3
1999 European Parliament		38.0%	51.0%	CF 13.0	37.1%	51.7%	CF 14.6	CF 1.6
1998 National election		47.5%	41.5%	SG 6.0	47.6%	41.4%	SG 6.2	SG 0.2
1994 European Parliament		43.0%	44.5%	CF 1.5	42.3%	42.9%	CF 0.6	SG 0.9
1994 National election		43.5%	49.0%	CF 5.5	43.7%	48.3%	CF 4.6	SG 0.9

In the US, exit polls have also been quite precise. Students at BYU have been conducting Utah exit polls since 1982.¹⁵ They write:

... results are very precise; In the 2003 Salt Lake [City] 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 vote, 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, suggest we should expect accuracy, within statistical limits, from the world's most professional exit polling enterprise.

Not only can exit polls accurately predict actual voting, they have been widely used to verify elections. When Mexico sought legitimacy as a modernizing democracy in 1994, Carlos Salinas instituted reforms designed to ensure fair elections, and central among these were exit polls.¹⁶ Exit pollsters were hired again for the next Presidential election in 2000,¹⁷ and not coincidentally, it was the first loss for the Institutional Revolutionary Party (PRI) in its 72 year history.

In Russia, and throughout the former Soviet block, exit polls have been used to verify elections. Last fall, international foundations sponsored an exit poll in the former Soviet Republic

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, aside from an LA Times poll, for which I could not determine whether or not the data was corrected.

¹⁶ Paul B. Carroll, Dianne Solis. "Zedillo's apparently clean win at polls diminishes threat of Mexican unrest." *The Wall Street Journal* August 23, 1994 pA2

¹⁷ Rebeca Rodriguez U.S. political consultants signed to conduct exit poll in Mexico. *Knight Ridder Newspapers*, June 16, 2000

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.¹⁸

Statistical Analysis of the Three Critical Battleground States: Ruling Out Random Error

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. Typical analyst comments included:

Since Election 2000, Republicans and Democrats have banked their aspirations on an electoral trinity: Florida, Pennsylvania and Ohio. As the Big Three goes, so goes the nation.

- David Paul Kuhn, CBS News: "High-Stakes Battle for the Big 3" Oct. 26, 2004

Conventional wisdom for months, including RCP's, had been that whoever won two of the "big three" Ohio, Pennsylvania and Florida would almost certainly become President.

- Real Clear Politics: [posting 10/28/04](#)

The accepted wisdom is that whoever wins two out of the three states of Ohio, Pennsylvania and Florida will win the election.

- Rob Watson, BBC News: October 28, 2004

The numbers and logic were straightforward. Of the other battleground states: Michigan and Minnesota always leaned Democratic, Colorado and Nevada Republican. Iowa, New Hampshire and New Mexico don't have many electoral votes. Wisconsin has a long tradition as a liberal state, and only 10 electoral votes compared to 20, 21, and 29 for the big three.

Campaign activities were also consistent with this logic. Ohio, Pennsylvania, and Florida were the three states the candidates visited most, and in which they spent the most money.¹⁹

¹⁸ Martin Plissner, Exit Polls to Protect the Vote, *New York Times* 10/17/04

¹⁹ See stories cited above.

The conventional wisdom proved correct. Bush won two of the three and ascended to electoral victory by virtue of that. In each of these states, however, exit polls differed considerably from recorded tallies (Table 5).

Table 5: Predicted vs. tallied percentages in the three critical battleground states

	Sample size	Bush predicted	Kerry predicted	Predicted differential	Bush tallied	Kerry tallied	Tallied differential	Tallied vs. predicted
Florida	2846	49.8%	49.7%	Bush 0.1 ²⁰	52.1%	47.1%	Bush 5.0	Bush 4.9
Ohio	1963	47.9%	52.1%	Kerry 4.2	51.0%	48.5%	Bush 2.5	Bush 6.7
Pennsylvania	1930	45.4%	54.1%	Kerry 8.7	48.6%	50.8%	Kerry 2.2	Bush 6.5

A Statistical Anomaly?

A basic question to ask on looking at such a discrepancy is whether it is just a statistical anomaly. It can happen, for example, that a fair coin tossed ten times will land heads each time, but it doesn't happen often (1 out of 1024 times). If we witness this, we will at least suspect that the coin might be adulterated, especially if the stakes are high and we are not permitted to inspect the coin carefully.

Statistical significance, which means that the discrepancy is such that it is unlikely to occur by chance, depends on four factors – the size of the discrepancy, the sample size, sample characteristics, and the level of significance (just how unlikely does it have to be?). Table 5 provides sample size and discrepancy. For statistical purposes, these samples are quite large. Two thousand or so respondents is roughly the size of most national polls.

Without access to the data and methodology, we cannot model the sample characteristics precisely. But we do know the general procedures by which exit polls are conducted. Appendix A provides the 2000 Presidential election exit poll methodology and a bibliography of articles on the process from that and other elections. Based on these we can make a reasonable approximation.

²⁰ 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% (Appendix B)

A random sample of a population can be modeled as a normal distribution curve. Exit polls, however, are not random samples. To avoid prohibitive expense, exit poll samples are clustered, which means that precincts, rather than individuals, are randomly selected. This increases variance and thus the margin of error because of the possibility that precinct voters share similar characteristics that differentiate them from the rest of the state in ways that past voting behavior would not predict. An analysis of the 1996 exit polls estimated that the cluster sample design adds "a 30 percent increase in the sampling error computed under the assumption of simple random sampling" (Merkle and Edelman, 2000, p. 72). That study is particularly apt because the 1996 state exit polls involved roughly the same number of precincts (1,468) as this year's polls (1,480).²¹ Pollsters also have techniques to decrease variance, most notably stratification, which helps ensure that the sample is representative of the overall population.²² In the analysis below, however, I conservatively assume no counterbalancing effects.

Figure 2 depicts the resulting distribution curve for samples of 1,936 randomly selected respondents from approximately 40 randomly selected precincts in a state in which 48.5% of the vote went for Kerry. The thin blue density curve is that of a simple random sample; the wider purple curve is of a clustered sample with no stratification. The horizontal double arrow below the curve indicates the poll's statistical margin of error, the corresponding 95% confidence

²¹ <http://www.exit-poll.net/faq.html#a7>

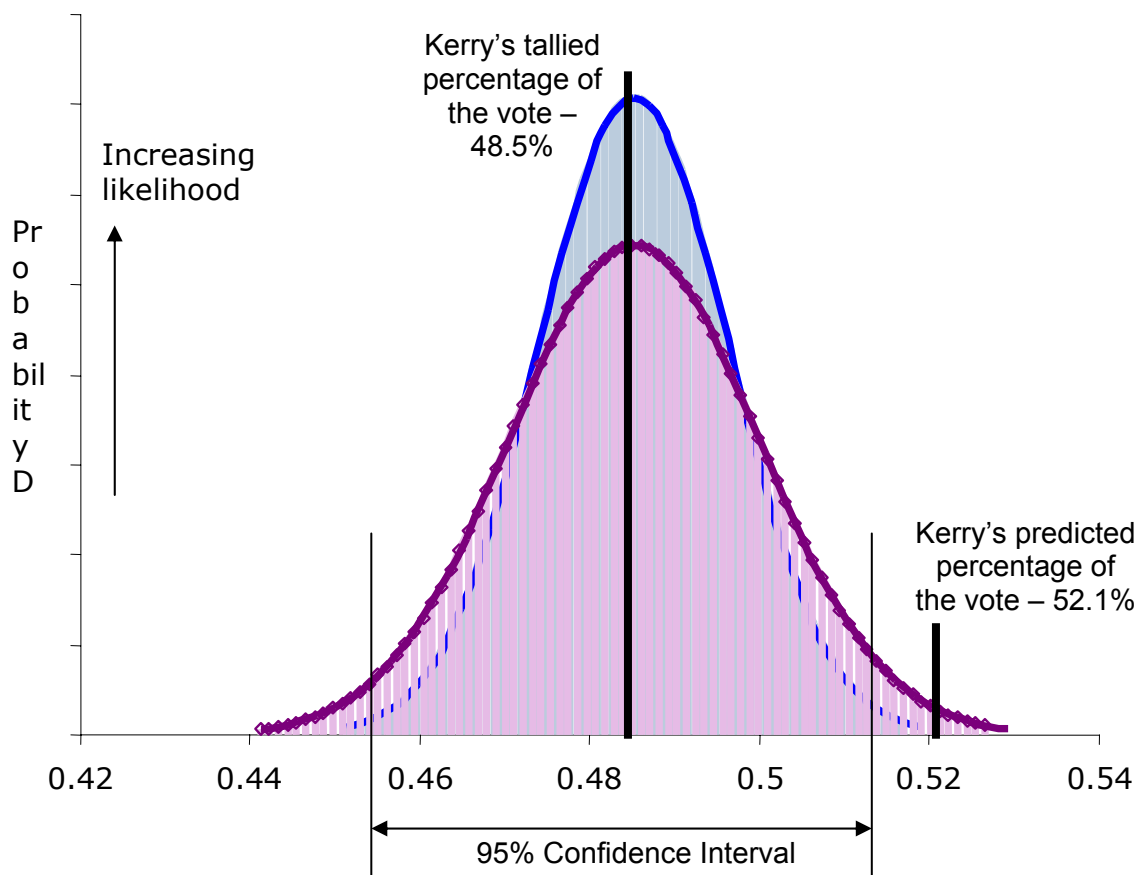
²² Pollsters also use a counterbalancing process that decreases variance – stratification. Identifying voters by key characteristics that predict voting behavior (race, sex, age, income, ethnicity, religion, party affiliation, etc...) ensures that the sample is representative of the overall population, either by seeking out subjects with specific demographic characteristics and/or weighting groups depending on their representation in the sample compared with that of the overall voting population. By getting samples in which minorities are over-represented (but subsequently negatively weighted), pollsters can ensure adequate sample sizes of each of these representative subgroups. Knowing exactly how much to weight over- or under-represented population depends on an accurate knowledge of overall demographics of the electorate. Historical data, census data, and registration roles, can be used to complement sampling site counts to try to accurately weight the sample. Assumptions

An early draft of this paper based on an assumption that the effects stratification of could balance the effects of clustering generated headline grabbing probability of 250,000,000 to odds. In this analysis, I use more conservative estimates, and assume no counterbalancing effects at all due to stratification. Although in principle, pollsters weight over- and under-sampled groups thereby ensuring a more representative sample than chance alone would dictate challenges in knowing exactly what weight to assign a group. The only measure of the demographics of actual voters on Election Day is the exit poll itself.

interval.²³ If one hundred unbiased samples were drawn from this population, we would expect that in 95 (on average), Kerry would poll between 45.6% and 51.4%. And because half of the 1-in-20 cases that fall outside the interval would be low rather than high, 97.5% of the time we would expect Kerry to poll no more than 51.4%. It turns out that the likelihood that Kerry would poll 52.1% from a population in which he receives only 48.5% of the vote is less than one-in-one-hundred (.0073).

Figure 2. Normal distribution curve for sample predictions based on Kerry's tally in Ohio

If you were to conduct this poll 20 times, 19 times you should get a result between 45.6% and 51.4%. The likelihood of the NEP poll prediction exceeding the actual vote by as much as 3.6% is less than 1 in 100.



23 To determine the margin of error, calculate the standard error of a random sample using the formula $\sqrt{\frac{p(1-p)}{N}}$ where p = Kerry percentage of the vote and N is the sample size. (.0113). To adjust for the fact that this is a clustered sample, add 30% (.01466 or 1.47%). Sixty-eight percent of the time, a prediction from a sample this size would be within one standard error, then . Ninety-five percent of the time, it will be within 1.96 standard errors (2.87% in this case).

Conducting the same analysis for Florida, we find that Kerry's poll prediction of 49.7% of the vote is likewise outside the 95% confidence interval. Given a population in which he receives only 47.1% of the vote, the chances that he would poll 49.7% out of 2846 respondent in an exit poll with no systematic error is less than two-in-one-hundred (.0164). Kerry's poll numbers are outside the 95% confidence interval as well in the third critical battleground state, Pennsylvania. Although he did carry the state, the likelihood that an exit poll would predict 54.1%, given 50.8% support of the electorate is just slightly more than one-in-one-hundred (.0126).

Assuming independent state polls with no systematic bias, the odds against any two of these statistical anomalies occurring together are more than 5,000:1 (five times more improbable than ten straight heads from a fair coin). The odds against all three occurring together are 662,000-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.

----- End of Part I -----

Appendix A: How Exit Polls are Conducted

2000 Voter News Service Exit Poll Methodology²⁴

METHODOLOGY STATEMENT

The VNS exit poll was developed and conducted by Voter News Service. The exit poll results are based on interviews with a probability sample of voters exiting polling places on Election Day, 2002.

Sampling

The samples were selected in two stages. **First**, a probability sample of voting precincts within each state was selected that represents the different geographic areas across the state and the vote by party. Precincts were selected with a probability proportionate to the number of voters in each precinct. Each voter in a state had the same chance to have his or her precinct selected. There is one exception. In some states, precincts that have large minority populations were sampled at a higher rate than other precincts. The sample weighting (described below) adjusts the representation of these precincts to their correct share of the total vote. **Second**, within each precinct, voters were sampled systematically throughout the voting day at a rate that gives all voters in a precinct the same chance of being interviewed.

The National sample is a subsample of the state sample precincts. The probability of selecting these precincts was the same as if the sample had been selected at a uniform rate nationwide, with the exception that minority precincts were again selected at a higher rate.

24 <http://webapp.icpsr.umich.edu/cocoon/ICPSR-STUDY/03527.xml> or http://www.ropercenter.uconn.edu/usvns2002_2.html

Weighting

The exit poll results are weighted to reflect the complexity of the sampling design. That is, the weighting takes into account the different probabilities of selecting a precinct and of selecting a voter within each precinct. For example, states that were selected at a higher rate receive a smaller weight than other precincts of the same size. There is also an adjustment for voters who were missed or refused to be interviewed, which is based on their observed age, race and sex.

NOTE: THE EXIT POLL RESULTS ARE ONLY MEANINGFUL IF THEY ARE WEIGHTED. UNWEIGHTED TABULATIONS MAY BE SERIOUSLY MISLEADING AND SHOULD NOT BE USED FOR ANY REASON.

Bibliography on How Election Exit Polls are Conducted

- Konner, Joan (2003). "The Case for Caution." *Public Opinion Quarterly* 67(1):5-18.
- Merkle, Daniel M. and Murray Edelman (2000). "A Review of the 1996 Voter New Service Exit Polls from a Total Survey Error Perspective." In *Election Polls, the News Media and Democracy*, ed. P.J. Lavrakas, M.W. Traugott, pp. 68-92. New York: Chatam House.
- Merkle, Daniel M. and Murray Edelman (2002). "Nonresponse in Exit Polls: A Comprehensive Analysis." In *Survey Nonresponse*, ed. R. M. Groves, D. A. Dillman, J. L. Eltinge, and R. J. A. Little, pp. 243-58. New York: Wiley.
- Mitofsky, Warren J. (2003). "Voter News Service After the Fall." *Public Opinion Quarterly* 67(1):45-58.
- Mitofsky, Warren J. (1991). "A Short History of Exit Polls. In *Polling and Presidential Election Coverage*, Lavrakas, Paul J, and Jack K. Holley, eds, Newbury Park, CA: Sage. pp. 83-99

Website

The single best source that I have found for information on exit polling and polling in general has been "**Mystery Pollster: Demystifying the Science and Art of Political Polling**" - By Mark Blumenthal <http://www.mysterypollster.com/>. The comments are extensive, knowledgable, and usually fair. (The one notable exception is his unfairly harsh critique of the first draft of this paper.)

Appendix B: Slate Election Day Exit Poll Numbers

Updated Late Afternoon Numbers

Mucho flattering to Kerry; plus Nader makes an appearance.

By Jack Shafer

Updated Tuesday, Nov. 2, 2004, at 4:28 PM PT

Florida	Pennsylvania	Minnesota	New Mexico	Missouri	New Jersey
Kerry 51	Kerry 53	Kerry 52	Kerry 50	Kerry 47	Kerry 54
Bush 49	Bush 46	Bush 46	Bush 48	Bush 52	Bush 44
		Nader 2	Nader 1		Nader 1
Ohio	Iowa	New Hampshire	Colorado	New York	West Virginia
Kerry 51	Kerry 50	Kerry 54	Kerry 49	Kerry 62	Kerry 45
Bush 49	Bush 49	Bush 44	Bush 50	Bush 36	Bush 54
		Nader 1	Nader 1	Nader 2	Nader 1
Michigan	Wisconsin		Arkansas	Nevada	
Kerry 52	Kerry 51		Kerry 45	Kerry 49	
Bush 46	Bush 48		Bush 54	Bush 48	
Nader 1	Nader 1		Nader 1	Nader 1	