Donovan
Ex. 4

Report on Factual Political Knowledge and Voter Confusion
Todd Donovan, Ph.D. August 2010

I. Summary....1
II. Major Flaws in Research Design....3
A. Lack of Proper Experimental Controls....5
B. Failure to Test for the Effect of Top Two Ballot Design....8
III. Critical Sampling Problems....11
A. The Logic of Sampling....12
B. Critical Problems with the Manweller Sample of "New Voters"....13
C. Effect of Sample Bias on Claims About New Voters....14
D. Critical Problems with the Manweller Sample of Registered Voters....16
E. Effects of Sample Bias on Claims About Registered Voters....17
F. Critical Problems with the Sample of "Active Voters"....19
G. Effect of Sample Bias on Claims About Active Voters....20
IV. Sample Bias Problems are Compounded by Flawed Statistical Analysis....22
V. Problems with Survey Outcomes and Response Rates....23
A. Unreasonably Low Response Rates....25
B. Effects of Low Response Rates on Claims About Voters in Washington....26
C. Inability to Make Inferences to Populations Using Small, Biased Samples....27
VI. Flawed Statistical Analysis....29
A. No Information about the Number of Observations....30
B. Critical Errors in Data Analysis and Presentation....32
C. Double Counting of Individual Subjects....33
VII. Tests for Statistical Significance Cannot be Interpreted....35
VIII. Mistaken Measurement of Voter "Errors" in the Survey Design....38
A. The Need for Unambiguous Survey Questions....38
B. Measuring the Concept of "Official Party Nominee"....39
C. Failure to Measure the Concept of Official Party Nominee....41
D. Effects of Flawed Survey Wording on Measures of Voter Confusion....43
IX. Mistaken Interpretation of Voter "Errors"....44
A. Many Candidate on Top Two Ballot are Endorsed by Political Parties....45
B. Many Candidates on Top Two Ballot are Nominated by Political Parties....46
C. So-called Errors are not Errors....46
X. Conclusion....48

Report on Paper by Mathew Manweller

Todd Donovan, Ph.D. August 2010

I. Summary

The paper[1] authored by Dr. Manweller purports to show "that voters are highly confused by the new Top-Two system" in Washington State.[2] The first half of Dr. Manweller's paper contains his assessment of the status of political parties, election law as applied to parties, and how various courts have interpreted laws affecting political parties. This report focuses exclusively on the second part of the paper, where Dr. Manweller presents the results of experiments that he claims measure voter confusion associated with Washington's Top Two primary.

This report demonstrates that the results presented in the Manweller paper are flawed on several fundamental points. These fundamental flaws include, but are not limited to:

1) Problems with research design, including the lack of proper controls needed to compare voter confusion under a Top Two primary to a Partisan primary.

2) Critical sampling problems, including highly biased samples that are not representative of actual voters, nor of informed voters.

3) Sample bias problems that are compounded by flawed statistical analysis.

---

[1] Manweller, Mathew. 2010. "The Very Partisan Non-Partisan Primary: Understanding What Voters Don't Understand." Paper presented at the Western Political Science Association meeting, San Francisco, CA. April 1 - 3.

[2] Manweller, p. 2.

4) Problems with survey response rates, including non-response bias that may inflate the proportion of less informed subjects in the samples.

5) Problems with survey response rates, such that low response rates produced very small samples. Even if they were not biased, these small samples cause high levels of sampling error that prevent any inferences from the samples to the population of voters.

6) Flawed statistical analysis, where subjects are double counted in a manner that inflates the measurement of voter confusion.

7) Tests for statistical significance that cannot be interpreted, and a lack of statistically significant effects.

8) Mistaken measurement of voter "errors" due to poor survey design and flawed question wording.

9) The failure to validly measure perceptions of official party nominees, party endorsed candidates, and party association with candidates.

10) Counting responses to survey questions about candidates as 'errors' when such responses may be correct in the context of the survey instrument.

Any single flaw listed above is enough to render the results reported in the Manweller paper invalid. The cumulative effect of these flaws mean that Dr. Manweller's claims about voter confusion associated with, or caused by Washington's Top Two ballots, lack merit and are indefensible.

II. Major Flaws in Research Design

Social science research must use care in all aspects of research design when the goal of the research is to make claims about causation.[3] At issue here is whether the Top Two ballot designed by the state of Washington causes voters to be confused about whether candidates listed on the ballot are nominated by, endorsed by, or approved by the official agents of Washington's political parties. As documented in my Report on Factual Political Knowledge and Voter Confusion, widespread confusion about political matters is a constant, enduring feature of American politics.

Several key empirical questions here are: 1) Does the Manweller paper determine whether the Top Two ballot design causes voters to be *more* or less confused about the status of candidates than they already are; 2) Does the paper assess if the state's ballot design - including the disclaimer related to party status on Top Two ballot - mitigates voter confusion about a candidate's relationship with a party; and 3) are any observed results reported by Dr. Manweller representative of actual voters in Washington state? Dr. Manweller's research design is not structured to answer the first two questions, and the data reported are not representative of voters in Washington state.

In this section, I detail how Dr. Manweller's research design is flawed to the point that it prevents us from gaining any sense of whether the ballot design affects the level of confusion in the electorate.

---

[3] For a brief introduction to designing social research and causation, see Babbie, Earl. 1995. *The Practice of Social Research,* Chapter 3 and Chapter 4. Wadsworth, 7th Edition; Hoover, Kenneth and Todd Donovan. 2011. *The Elements of Social Scientific Thinking*. Wadsworth, 10th Edition.

Dr. Manweller presents the research question in terms of whether the state of Washington could craft ballots that could "be printed in such a way as to eliminate the possibility of widespread voter confusion."[4] Specifically, Dr. Manweller notes that the United States Supreme Court was concerned about whether a ballot design might cause voters to "misinterpret the candidates' party-preference designation as reflecting endorsement by the parties."[5]

The Top Two ballots clearly inform voters that candidates listed on the ballot are not endorsed by any political party. Candidates are allowed to state a party preference when they file, and that preference appears on the ballot. Thus, rather than having candidates appear with a standard party label by their name, the Top Two ballot allows candidates to indicate which party they "prefer." The ballot clearly states that "A candidate's preference does not imply that the candidate is nominated or endorsed by that party or that the party approves of or associates with the candidate." Some counties[6] enclosed the same information on a separate sheet of colored paper in ballots mailed to voters.[7] Some counties have the "does not imply" statement in a bold font as large as or larger than all or most other text on the ballot.[8] Under Washington's previous partisan primary system, candidates were simply identified with a party label listed with their name.

[4] Manweller, p. 9, quoting *Grange* p. 13.
[5] Grange decision, p. 12.
[6] Examples include King, Spokane, Snohomish, Jefferson and Whatcom counties.
[7] Most voters in Washington state vote by mail.
[8] Examples include Chelan and

A) Lack of Proper Experimental Control

Experimental research requires that the researcher identify a potential effect of some treatment while controlling for other forces that might produce the potential effect.[9] In this case, the effect is potential voter confusion about the candidates' relationship with the political parties caused by a particular ballot design. The treatment in Dr. Manweller's experiment is variation in ballot design. In designing experiments to assess the effect of the Top Two ballot on voter confusion, we would need a control condition that represents the baseline level of confusion that exists in a random, representative selection of reasonably informed voters. Dr. Manweller's research fails to provide a proper control condition and thus it provides no baseline for assessing voter confusion. Furthermore, the research fails to use representative samples (more on this below), so even if the design had been appropriate, the results would still lack validity.

Three mock ballots are used in Dr. Manweller's experiments (see Appendix A in the Manweller paper). These mock ballots are said to mimic: 1) a partisan general election, 2) a nonpartisan Top-Two primary election, and 3) a nonpartisan Top-Two general election. There is no mock ballot representing a partisan primary election. This is a critical omission, since it prevents Dr. Manweller from having a control condition that allows us to make 'apples-to-apples' comparisons about voter perceptions and the potential for confusion about primary elections that exists independent of the Top Two ballot. Without this measure, we cannot assess a potential effect of changing between a partisan primary and a Top Two primary.

[9] Nachmias, Chava and David Nachmias. 1992. *Research Methods in the Social Sciences*. St. Martin's, 4th Edition. P. 100-102. Babbie, 1995. Chapter 9.

Dr. Manweller uses these mock ballots in an attempt to measure how subjects perceive the candidate's relationship with the political parties. He makes dubious assumptions about when a subject makes an "error" when evaluating these ballots. With the Top Two primary ballot and Top Two general election ballot, a subject who responds that a candidate is the "nominee" of a party is coded as making an error, and being confused. With the partisan general election ballot, a subject who responds that a candidate is not a "nominee" of the party is coded as making an error, and being confused.

The validity of these assumptions about what constitutes voter "error" or confusion are highly suspect, and are addressed in further detail in other sections of this report (Section XIII and Section IX). As a matter of research design, however, the omission of a mock ballot that mimics a partisan primary is critical. Consider the mock Top Two primary ballot Dr. Manweller used in the experiments. It lists multiple candidates said to "prefer" the Republican Party and multiple candidates said to "prefer" the Democratic Party. Indeed, Dr. Manweller's reported data (Table 3) suggest that subjects shown the Top Two ballot format were far less likely to perceive candidates as party nominees (30%) than subjects show the mock partisan general election ballot (80%).[10]

However, given widespread public confusion about political matters documented my Report on Factual Political Knowledge, it is not possible to conclude that 30% of these subjects shown the Top Two ballot primary perceived a candidate as a "nominee" due to confusion produced by the design of the Top Two ballot. Had these experiments

[10] Manweller, p. 28.

included a fourth mock ballot that mimicked a partisan primary, it is highly likely that the partisan primary ballot would have shown higher levels of voter confusion than the Top Two primary ballot.

The reasons for this are straightforward: With a partisan primary ballot, candidates (other than independents) are listed on the ballot with a clear, robust party label - there is nothing about "preferring" a particular party. On a partisan primary ballot, there is no disclaimer on the ballot stating that the candidates are not nominated or endorsed by the parties. Furthermore, under a partisan primary ballot, the candidates are not legally the nominee of a party until they win the primary. Using the coding logic Dr. Manweller describes in his paper, a subject would be coded as making an error, and would be coded as being confused, if she or he responded as saying *any* candidate on the partisan primary ballot was the nominee of a party.[11]

It is highly likely that given the general baseline level of voter confusion, and given the lack of any disclaimer on a partisan primary ballot, and given the use of robust party labels on partisan primary ballots, many voters would likely be confused by a partisan primary ballot (in the sense that Dr. Manweller measures confusion), because many voters would probably respond (wrongly) that candidates listed with partisan labels on the partisan primary ballot are party nominees. This control condition is required for us to understand what the baseline level of voter confusion may be, but it was not part of the research design. Indeed, I would expect that had the experiment included a partisan primary ballot, and had Dr. Manweller's own coding logic been used, the so-called 'error rates' in voter perceptions of the candidates' relationships with the parties would have

[11] Manweller, p. 13.

been greater under a partisan primary ballot than what was reported to be associated with the Top Two ballot. However, this critical experimental control condition was not part of Dr. Manweller's research design.

B) Failure to Test for the Effect of Top Two Ballot Design

As noted above, Dr. Manweller's experimental design lacks the necessary control that would allow us to make meaningful assessments of the effect of the Top Two ballot compared to a partisan primary ballot. Dr. Manweller's experimental design also lacks the capacity to test the effects of a major element of the Top Two ballot design - the disclaimer informing voters that candidates are not endorsed by or nominated by a political party. This is important, because a simple experiment would likely show, and common sense strongly suggests, that the state's use of the disclaimer reduces voter confusion regarding the candidates' relationship with the parties.

Given that voter confusion is widespread and constant, any attempt to assess whether Washington's Top Two ballot causes confusion about candidates' relationships with parties would need to establish a baseline level of voter confusion independent of key features of the state's ballot design. Specifically, we need to know how voters might perceive the candidates' relationships with parties when presented with a ballot that does not have the disclaimer stating that candidates are not endorsed or nominated by the parties. We would then need to compare voter perceptions under those conditions to how voters might perceive candidates when presented with a ballot where the disclaimer is included.

This test would simply require presenting subjects with two mock ballots. One would be something similar to a Top Two primary (or general) election ballot. This ballot would list the candidate's party preference. This ballot would include the statement (under the bold heading "**READ**") that informs voters that a candidate's party preference does not imply that the candidate is nominated or endorsed by a party. To mimic the actual voting experience, the mock Top Two ballot would also include a separate sheet of colored paper that also included the "does not imply" disclaimer. To isolate the effect of the "does not imply" disclaimer from the "prefers party" listing, a second mock ballot would be used that is identical (including the "prefers party" listing, but without the disclaimer. Subjects would be asked appropriate questions (see Section VIII, Part B of this report) to measure their perceptions of a candidate's relationship with a party. In this case, the first mock ballot (with the disclaimer) would be the treatment, and the second mock ballot (without the disclaimer condition) would be a valuable control that would establish another baseline level of voter confusion. A proper experiment would use mock ballots that replicated the actual voting experience. These would present the disclaimer as it appears on actual ballots: appearing on the ballot in bold font, or appearing in bold font larger than most other text on the ballot, with the disclaimer repeated on a separate sheet of colored paper. Dr. Manweller's mock ballots did not include any of these things.

In my opinion, it is highly likely that had such an experiment been conducted using Dr. Manweller's methods and coding scheme, it would have produced results showing the disclaimer is associated with far fewer voter 'errors,' and less confusion about the candidates' relationship with parties. The reason for this is that the disclaimer

is unambiguous. Dr. Manweller presents results (Table 1) that suggest many voters are unable to assess what a nominee is, regardless of the ballot they are presented with. There are 'errors' under every ballot condition.[12] This makes sense, given low levels of knowledge about politics documented in my Report on Factual Political Knowledge. A ballot that included an unambiguous statement that candidates are not nominated by nor endorsed by a party would certainly cause some respondents to conclude that the candidates are not nominated by nor endorsed by a party.

Indeed, although his experiments were not designed to properly test for the effect of the disclaimer,[13] Dr. Manweller's reported results illustrate that the state's ballot design corresponds with dramatically reduced perceptions that candidates are nominees of a party. Table 2 in the Manweller paper shows that 80% of people he classifies as "new voters" were exposed to the Partisan General election ballot (with no disclaimer) agreed that the candidates were party nominees. In contrast, only about 26% said this when shown the Top Two primary ballot with the disclaimer, and 57% said this when shown the Top Two general ballot with the disclaimer.[14] In the registered voter sample (Table 3) the Partisan General election ballot (without a disclaimer) corresponds with a 80% of people agreeing that the candidates are nominees of a party. This drops to 30% among people shown the Top Two ballot with the disclaimer.[15]

A key point in this section is that any experiment using Dr. Manweller's method, using any ballot design, will show that many voters fail to understand what a party

[12] Manweller, p. 26.
[13] The Manweller experiments do not isolate the effect of the disclaimer from the "prefers party" designation, nor do they present the disclaimer as it is actually presented to voters.
[14] Manweller, p. 27
[15] Manweller, p. 28

nominee is. Dr. Manweller's report makes this clear, as does the material presented in my Report on Factual Political Knowledge. The latter documents widespread levels of voter confusion, and low levels of factual knowledge about politics. Many voters lack an understanding of what a nominee is regardless of the ballot they are presented with (I elaborate on this below, in Section XIII and Section IX of this report). A proper assessment of the effects of the Washington Top Two ballot on voter confusion would need to assess levels of confusion with and without the disclaimer. And (as noted above), a proper analysis would also compare rates of so-called voter confusion associated with the Top Two primary ballot to rates of confusion associated with a partisan primary ballot. Neither test was conducted. The result of this is that - even if samples were representative (and, as I document in Section III of this report, they are not) - data on "confusion" reported in the Manweller paper lack any point of reference and are thus un-interpretable.

To summarize: The research design is critically flawed. The experimental design lacks proper controls, and it lacks any ability to place measured rates of voter confusion in the context of widespread voter confusion that is well known, and regularly observed.

III. Critical Sampling Problems

Even if Dr. Manweller had designed his experiments such that he could isolate voter confusion associated with the Top Two ballot (as Section II demonstrates - he did not) the subjects used in the experiments are so far removed from a representative sample of an informed electorate as to render results meaningless.

The description of the samples used in Dr. Manweller's paper is so brief that it fails to meet standards (discussed below in Section V of this report) used in peer reviewed scholarly journals. As a result, it is difficult to fully assess who the subjects of the experiments were. Nonetheless many elements of the samples that are described by Dr. Manweller are highly problematic.

The Manweller paper describes three distinct groups of voters who are supposedly represented by the subjects who participated in the survey experiments: 1) "New Voters"; 2) "Registered Voters"; and 3) "Active Voters." As detailed below, none of the three samples prove to be representative of the population they depict.

A) The Logic of Sampling

Scientific research - particularly research on social and behavioral phenomena - require that the researcher collect data from samples of a target population that are broadly representative of the target population.[16] Without representative samples, scholars cannot make inferences about a larger population from their observations. The essential requirement of any sample is that it is representative of the population from which it is drawn.[17]

A random probability sample is required for a sample to be representative. Such a sample is defined by the fact that any member of the population has an equal probability of being selected to be in the sample.[18] If this key quality is lacking, the sample will likely be biased such that the researcher cannot make generalizations from the sample to

[16] Nachmias and Nachmias, p. 170-172,
[17] Bohrnsteadt, George and David Knoke. 1982. *Statistics for Social Data Analysis*. F.E. Peacock. P. 19.
[18] Nachmias and Nachmias, p. 174;

the target population.[19] For example, if we want to know about new voters, registered voters, or active voters, we need samples that are representative of such populations. A biased sample is one where the subjects used in the study fail to represent the population of interest.

B) Critical Problems with the Manweller Sample of "New Voters"

The first category of voters that Dr. Manweller attempts to make claims about are "new voters." A proper, random, representative sample of new *voters* would include people who have actually voted. In order to know if the sample was representative of the population of new voters in Washington state, a study would collect basic demographic information from the sample so that it might be compared to the population.

The "new voter" sample used in Dr. Manweller's paper is highly biased. There is no information about the sample that allows us to assess if it is representative of the new voter population, nor is there information that allows us to judge if it is representative of students from the university where the data were collected. As such, nothing can be generalized from data presented from the sample.

Consider Dr. Manweller's sample that purportedly represents the population of new voters. Ideally, some random process would be used to capture the sample, yet there is no discussion in the Manweller paper about how subjects entered this sample. Nor is there any demographic information that would allow us to know if the subjects were

[19] A classic example of non-representative sample bias is the 1936 *Literary Digest* prediction that Alf Landon would defeat Franklin D. Roosevelt. The *Digest* sampled people who had telephones and magazine subscriptions, and thus failed to represent a Great Depression-era population that included many voters who lacked phones and magazine subscriptions.

representative of the new voter population or even representative of students at Central Washington University (CWU). There is no information about their age, year in school, or voting history. In short, we have no way of knowing if any subject in this sample had ever actually voted. Indeed, Dr. Manweller notes that the subjects were mostly 18-24 years of age.[20] This means that many (perhaps most) had never voted - hardly a sample of "new voters" - let alone a sample of reasonably informed voters. Dr. Manweller notes the subjects from CWU may have never been voters, and that there is no way to know if they were voters.[21]

Furthermore, Dr. Manweller himself recognizes that the failure to collect critical demographic information from this CWU sample was an "oversight" that needed to be "corrected."[22] In short, there is no way to know who these students were. No demographic information about these subjects were collected. The paper simply notes that this sample consists of 183 students at Central Washington University.

C) *Effect of Sample Bias on Claims about New Voters*

The one result in Dr. Manweller's paper that appears to show substantially more "confusion" (by Dr. Manweller's logic and flawed research design) about a candidate's relationship with a party is found with so-called "new voters." This is shown when the so-called "new voters" were exposed to the mock ballot designed to mimic a Top Two general election (see the 56.6% figure in Table 1 in the Manweller paper).

---

[20] Manweller, p. 11.

[21] Manweller, p. 11 "these students could have voted anywhere between zero and 2 times." He also notes on p. 35 (fn 17) that no information about demographics (e.g. subjects' voting history) was collected.

[22] Manweller, p. 35, note 17.

This highly biased sample[23] - where young non-voters are probably greatly overly represented - is likely to be the reason for this result. It is highly likely that this sample is biased such that it over-represents subjects who have never voted, and who thus lack information about politics generally (more so than a typical voter would lack general political information) and, more specifically, these subjects would have a much weaker understanding of the relationship between political parties, candidates, nominations, and election procedures than a typical voter or even a new voter who had actually voted. This type of sample bias (over-representation of non-voters or inexperienced, first-time voters) is particularly debilitating in this context because it inflates the chances that any experiment attempting to measuring confusion will find voter confusion.

However, as my discussion in the section above on research design demonstrates, Dr. Manweller's measured rates of voter "confusion," reported in his paper are largely meaningless, given that the experiments that produced the reported rates of voter confusion lacked needed experimental controls that would allow us to place the reported levels of confusion in a meaningful context.

Given the substantial differences between the population of new voters in Washington state and Dr. Manweller's sample of students at CWU, this sample cannot be used to make generalizations about new voters in Washington state.

[23] There is potential here also for experimenter bias (or experimenter expectancy effect). CWU students may have been exposed to discussion of matters associated with the subject of the experiments. Nachmias and Nachmias, p. 209. The Manweller paper provides no details about who the subjects were, or what courses they were drawn from.

D) Critical Problems with the Manweller Sample of "Registered Voters"

The second group of voters that Dr. Manweller attempts to make claims about are registered voters in Washington. The actual population of registered voters includes all voters in Washington who are registered to vote. There is a wide range of age and education levels among the 3,583,278 registered voters in Washington as of November 2009[24] and substantial variation in the frequency that members of the population vote. The population is also distinguished by the length of time that individuals have been registered, and by the method that the voters used to register. The vast majority of voters in Washington state registered to vote using traditional, paper application forms. A small fraction of Washington voters registered using a new on-line system that was introduced in January of 2008.[25] The Manweller study was written in early 2010. The paper does not state when data were collected, but only a very small fraction of all registered voters had registered online between 2008 and early 2010.

A proper, random, representative sample of Washington voters would be one where every member of the population has an equal probability of being drawn into the sample.[26] This would produce a sample where levels of voter experience, voter education, age and other characteristics of people in the sample would closely match the distribution of voter experience, voter education, age and other characteristics of people in the population. Such a sample would reflect the fact that the vast majority of voters in Washington registered prior to 2008, using the traditional paper application form. A

[24] http://vote.wa.gov/Elections/WEI/VoterTurnout.aspx?ElectionID=32
[25] These basic demographic traits are described in a report by Baretto, Matt, Bonnie Glasser, et al. 2010. "Online Voter Registration Systems in Arizona and Washington: Evaluating Usage, Public Confidence, and Implementation Processes." Washington Institute for the Study of Ethnicity and Race.
[26] Nachmias and Nachmias, p. 174.

proper, representative sample would thus have a rather low proportion of people who registered to vote online, given that such a low proportion of people in the population registered online.

E) Effects of Sample Bias on Claims About Registered Voters

Dr. Manweller's sample of registered voters appears to be limited - exclusively - to people who registered online between 2008 and 2009 (or early 2010, again, Dr. Manweller does not report when these data were collected). The sample is further limited to people who were willing to respond to a request to participate in an online survey. The paper provides no information about how this sample of 1500 was drawn, nor about who collected the data, beyond noting it was done "via an online survey company."[27] This means it is not at all representative of the actual registered voter population in Washington (or any state). It is simply a sample of people who registered online, who then responded to a survey that was conducted online.

This highly biased sample is problematic because the population of voters who registered online in Washington is quite distinct from the actual population of registered voters in Washington. Voters who register online in Washington are substantially younger, and thus probably less educated, and certainly less experienced with voting than actual Washington voters. This sample bias means that reasonably informed voters are grossly under-represented in this sample, while less informed, less experienced voters are grossly over-represented. My Report on Factual Political Knowledge documents that young and less educated people are most prone to giving incorrect answers to questions

[27] Manweller, p. 12.

about political matters. As such, this sample inflates the likelihood that an experiment attempting to identify voter confusion would find confusion.

A study of online voter registration in Washington conducted jointly by the University of Washington and the University of California, commissioned by the Pew Charitable Trust[28], provides data that illustrates the magnitude of bias in Dr. Manweller's sample of registered voters. Their report demonstrates that when the population of all registered voters in Washington are considered, only 24 percent are people age 18 to 34. Nearly half of all registered voters in Washington are aged 55 years or older (48.6 percent). In contrast, the demographic characteristics of the population of voters who registered online are markedly different. Over 60 percent of these voters are aged 18 to 34. Only 15% of voters registered online are over 50 years of age.[29] Dr. Manweller is thus attempting to make claims about registered voters from a sample that is drawn from a small, unrepresentative slice of the actual population.

This sample bias is compounded by the problem of selection bias that has been demonstrated to exist with Internet samples.[30] This means that even if a random list of subjects is drawn for a sample of voters who registered online, those who choose to respond to the study will not necessarily be representative of the population of voters who registered online. The propensity for certain types of people to be more likely to participate in an online survey may further weaken the ability to make generalizations from the sample about the population of online registered voters (let alone the population

[28] Baretto and Glaser et al.
[29] Barreto and Glaser et al., p. 13.
[30] Hudson, Darren, Lee-Hon Seah, Diane Hite and Tim Haab. 2004. "Telephone Presurveys, Self-selection and non-response bias to mail and Internet Surveys in Economic Research." *Applied Economic Letters.* 11(4): 237-40.

of registered voters). For example, younger people have been shown to be more likely to participate in Internet surveys than older people.[31]

Given such differences between the actual population of registered voters in Washington, and Dr. Manweller's sample of online registered voters, this sample cannot be used to make generalizations about registered voters in Washington state. There may also be bias that limits the utility of generalizing from the sample to the population of all voters who registered online.

F) Critical Problems with the Sample of "Active Voters"

The third group of voters that Dr. Manweller attempts to make claims about are "highly active"[32] or "active voters"[33] (both terms are used to describe what appears to be the same set of voters). A proper, representative sample of "highly active" or "active" voters in Washington state can be obtained by contacting state or county officials for public information about the frequency that people have voted in the past. Such information is routinely used by political campaigns that seek to target their mailing and mobilization efforts to voters known to be most likely to actually vote. Public records detail each individual voter's mailing address, vote history dating back several elections, and the date the voter registered to vote. Commercial firms and political parties assemble

[31] Miller, Thomas et al 2002. "Citizen Surveys on the Web: General Population Surveys of Community Opinion." *Social Science Computer Review*. 20:124-136; Royle, Marjorie and Destiny Shellhammer. 2007. "Potential Response Bias in Internet Use for Survey Religious Research." *Review of Religious Research.* 49(1): 54-68.
[32] Manweller, p. 12.
[33] Manweller, p. 18.

this public information into lists that also include voter age, gender, phone number. Such information can be used to generate a random sample that represents active voters.[34]

Dr. Manweller appears to define a sample of "highly active" or "active" voters without having any information about the voting history of people in the sample. He states, "The sample population for the highly active voters was drawn from a population of voters who had elected to receive emails from and/or had donated to one of the two major political parties in Washington state."[35]

The people (there is nothing in the paper that indicates if these subjects were actually voters, or if they voted frequently) in this sample were drawn from a list of 6000 email addresses obtained from the state Republican and Democratic parties. The paper contains no information about how the parties obtained these email addresses, nor about how (or if) the sample of email addresses were drawn randomly from some larger population of email addresses. The paper contains no information about how the active voter data were collected other than noting subjects were contacted "per the same procedure as the registered voter population."[36]

G) Effect of Sample Bias on Claims About Active Voters

As noted in Section II, Part E of this report, surveys conducted exclusively via Internet can potentially over-represent younger voters. The paper offers no information about what proportion of the people on these email lists were donors, or what proportion were simply people who had their email addresses collected recently by campaigns or

[34] For example, campaigns use the public records to target voters who have participated in four of the last four elections, or three of the last four elections, etc.
[35] Manweller, p. 12.
[36] Manweller, p. 12.

recent voter registration efforts. Given that donating money is more costly than listing an email address on a form or clicking on a web link that can collect email addresses, it is reasonable to expect that there are a higher proportion of email only subjects than donors in this sample. This could produce a sample that is biased to include relatively newer, younger voters. Unfortunately, there is no information in Dr. Manweller's paper regarding the demographic traits of people in his "active voter" sample. Dr. Manweller assumes that providing an email address is the equivalent of voting frequently, but he provides no evidence to support this assumption. A more plausible assumption is that the parties collect email addresses several ways, including during their voter registration drives. If this is the case, this sample of so-called "active" voters may over represent newer, less experienced voters.

Given the substantial differences between the actual population of active voters in Washington state, and Dr. Manweller's sample of people "who had elected to receive emails from" one of the two major political parties, and given that there is no information about the actual voting history of people in Dr. Manweller's sample of "active" voters, observations from this sample cannot be used to make generalizations about active voters in Washington state. Indeed, in this case, it is impossible to know what target population this sample of email addresses might actually represent. As such, no inferences can be made from the sample.

IV. Sample Bias Problems are Compounded by Flawed Statistical Analysis

If Dr. Manweller's sample of online registered voters is a random draw from the population of voters who registered online (the paper provides few details about how the sample was drawn), it likely is weighted heavily toward younger, less experienced voters. This not only biases the experiments toward showing greater rates of "confusion," but it also makes analysis of the sort attempted in Table 6 and Table 7 highly problematic. In Table 6 of the Manweller paper, "error rates" are plotted across age categories.[37] A random sample of online registered voters would have very few subjects in the two categories of age for people over 51 (Dr. Manweller does not follow conventional academic protocol and report the number of cases in each category, so this is difficult to ascertain).

With so few cases in some categories, it is likely that observed "error rates" (so-called voter confusion) for those categories are far lower of far higher of what actually exists in the target population. This is a standard property of random sampling. Any random sample includes some observations (randomly) that cause the sample to deviate from the population.[38] If for example, 400 observations are made in a representative sample of people 18-34, there is a greater chance that any random variation expected in the draw of the sample is cancelled out by the large number of observations. Conversely, if only 4 people over age 60 are in a sample, it is less likely that any random variation

[37] Manweller, p. 31.

[38] This is known as sampling error. This is the reason estimates from survey research are given with a "margin of error" and "confidence intervals." A larger number of random observations reduces sampling error and makes inferences from the sample to the population more precise. A very small number of random observations increases sampling error, and makes inferences from the sample to the population less precise. For an accessible discussion, see James Levin and James Alan Fox, 2007. *Elementary Statistics in Social Research*, p. 114-115.

associated with the sample draw is cancelled out. Put differently, small random samples are less representative of the population than larger random samples. Small samples likely have more variation (statistical error) than larger samples. Problems are compounded when small, biased samples are used. The bias noted in this report act to inflate the observed rates of voter confusion.

Point estimates offered by Dr. Manweller (such as the percent of people 51-60, or 61 and over who are "confused" by various types of ballots) must be treated with great caution. Small samples make it impossible to assess how representative such estimates might be. Indeed, as demonstrated in Section VI of my report, nearly all of the results reported by Dr. Manweller fail to provide information about the number of observations used to generate his results. When the number of observations are considered, nearly all results presented in the Manweller paper appear to be generated from samples that are so small that nothing can be inferred with confidence from the samples to any larger populations.

V. Problems with Survey Outcomes and Response Rates

Studies that rely on samples to collect survey information about voter perceptions and attitudes typically report information about the outcome of efforts to get people from a sample to participate in a study. Such information on outcomes regarding how many people from the sample were contacted, how many could not be contacted, how many who were contacted refused to participate and how many who were contacted did participate are used to calculate the response rates for the survey. Response rates are

often used as a measure of survey quality.[39] The leading scholarly journal dedicated to the study of public opinion and American politics (*Public Opinion Quarterly*) requires that papers report response rates, with rates calculated using standards set forth by the American Association of Public Opinion Research (AAPOR).[40] Information about response rates is important, because they are indicators of the potential for non-response bias.[41] Ideally, researchers know something about the difference between responders and non responders. This way, they know if the sample represents the target population.

Very low response rates are not automatically problematic, but they force the researcher to assume that the few people who choose to cooperate with the survey are representative of the many who refuse to participate. An AAPOR study of academic journals quoted a journal editor who expected at least a 60% response rate.[42] Another journal editor noted that "in most instances, 20% is too low, and 80% is a de facto standard."[43] Much of Dr. Manweller's data come from online surveys taken by people contacted by email. An online commercial survey firm (SuperSurvey Knowledge Base) notes that:

> "A high response rate is the key to legitimizing a survey's results. When a survey elicits responses from a large percentage of its target population, the findings are seen as more accurate. Low response rates, on the other hand, can damage the

---

[39] Ehrlich, Nathaniel. 2005. "Response Rate" Entry in *Polling America: An Encyclopedia of Public Opinion*. Samuel J. Best and Benjamin Radcliff, editors. Greenwood Press. p. 701.
[40] American Association of Public Opinion Research. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lenexa, KS: AAPOR.
[41] Daves. 2005. "Outcome Rates." Entry in *Polling America: An Encyclopedia of Public Opinion*. Samuel J. Best and Benjamin Radcliff, editors. Greenwood Press. p. 479-82.
[42] Johnson, Timothy and Linda Owens. 2003 (?). "Survey Response Rate Reporting in the Professional Literature." AAPOR Section on Survey Research Methods. American Statistical Association, Joint Statistical Meeting, San Francisco, CA. p. 129.
[43] Johnson and Owens, p. 130.

> credibility of a survey's results, because the sample is less likely to represent the overall target population."[44]

Dr. Manweller does not report response rates for any of the three samples used in his study. I cannot assess the response rate of the CWU sample because there is no information in the paper about how many students entered the sample, and how many were given the survey (beyond the 183 who completed it).

A) Unreasonably Low Response Rates

I can match information from the text with data in Table 1 to construct response rates for the other two samples. As noted above Dr. Manweller stated that a sample of 1500 voters who registered online were contacted by email in order to represent registered voters.[45] Data in Table 1 suggest that only 102[46] of the 1500 voters from the sample of voters registered online actually did the survey. This is a 6.8% response rate.[47] As noted above, Dr. Manweller also states that he used a sample of 6000 people with email addresses obtained from the political parties in order to represent active voters. Data in Table 1 suggest that only 549[48] of the 6000[49] people in this sample contacted by email actually did the survey. This is a 9% response rate.

---

[44] http://knowledge-base.supersurvey.com/survey-response-rate.htm
[45] Manweller, P. 12.
[46] This number is calculated by adding the smaller values in each cell in the "Registered Voters" row of Table 1.
[47] This rate is calculated with the AAPOR Response Rate 1 method.
[48] This number is calculated by adding the smaller values in each cell in the "Active Voters" row of Table 1.
[49] Manweller, P. 12.

To put this in context, a study of 31 surveys where subjects were contacted by email found an average response rate of 37%.[50] The rates of response to Dr. Manweller's surveys (6.8% and 9%) are very low.

B) Effects of Low Response Rates on Claims about Voters in Washington

Such low levels of response increase the likelihood that people who do respond to the surveys are not representative of the target population (non-response bias). For the small fraction of subjects participating in the surveys to be representative of the target populations, we would have to assume that there is no systematic bias in who chooses to participate in the study, and who does not participate. Given that the sample was contacted exclusively via email, this assumption is problematic. For example, younger people may spend more time online, and thus be more likely to respond to email queries. This would further bias the data by increasing the likelihood so-called voter confusion was being measured among a set of subjects who are less informed than the actual population of Washington voters.

One possible way to validate whether very low response rates produce observations that may be representative of the population is to compare demographic traits of people who participated in the study to the demographic traits of people in the target population. Dr. Manweller provides no such analysis. It is probably not possible to conduct such an analysis, since 1) the samples (voters registered online and people who gave their email address to the parties, respectively) fail to match the target populations (all registered voters, and registered voters who vote often); and 2) there is no

[50] Kim Sheehan. 2001. "E-mail Survey Responses: A Review." *Journal of Computer-Mediated Communications*. 6(2). January.

information about the distribution of age, education, or vote history available for the populations that Dr. Manweller actually drew his samples from.

These problems with low response rates compound several of the problems noted above. The samples used by Dr. Manweller were not representative of the populations he attempts to make claims about. Bias in the sample is increased, and the quality of inferences from the sample is further degraded, by non-response bias. As noted in Section IV of this report this bias is magnified by the use of very small samples. Sections VIII and IX of this report also show that this bias is further amplified by Dr. Manweller's methods of measurement and analysis.

C) Inability to Make Inferences to Populations Using Small, Biased Samples

Even very small samples that might be representative of a population are extremely problematic because they limit a researcher's ability to make confident inferences from the sample to the population. When random samples are observed to describe properties of a population, results vary by chance from the actual population. This is sampling error. For example, with a random sample of 1500 Americans estimating voter preferences in a close, two candidate presidential election, we could be 95% confident that the candidate preferences of the population would be +/- 2.5% of what was measured in the sample. With a sample of 1000, we would be 95% confident the population's preference was +/- 3.1% of what was observed in the sample. With a sample of 500, it would be +/- 4.4% of what was observed in the sample. As samples shrink, so does the precision of estimates from the samples. When samples are very small, sampling error becomes very high.

Table 1 of Dr. Manweller's paper reports information generated from very small samples. Dr. Manweller then attempts to make inferences from these results to the population of 3.5 million Washington voters. For example, Dr. Manweller reports that a sample of just 30 registered voter subjects were exposed to the mock Top Two ballot, and that 30% of these 30 were "confused." [51] Estimates about the actual level of such "confusion" in the Washington voter population from a representative sample of 30 registered voters would have a confidence interval of plus or minus 16.4%.[52] That is, if this were a representative sample (it is not) we might be 95% confident that "confusion" about the Top Two general election, as Dr. Manweller defines it, might occur in somewhere between 13.6 to 46.4% of the registered voter population.

Dr. Manweller also reports a sample of 31 "registered voter" subjects were exposed to a mock Partisan general election ballot, and that 19% of these were "confused."[53] Estimates about the actual level of such "confusion" in the Washington voter population from a representative sample of 31 registered voters would have a confidence interval of plus or minus 13.9%. That is, if this were a representative sample (it is not) we might be 95% confident that "confusion" about the partisan general election, as Dr. Manweller defines it, might occur in somewhere between 5.4 to 33.2% of the population.[54] In other words, even if the samples had been representative (they are

[51] Manweller, p. 26.
[52] That is, if the sample had been representative even of the actual population of people who registered online, we could be 95% confident that somewhere between 13.6% to 46.4% of that population would display voter confusion about the Top Two general election ballot (using Dr. Manweller's definition of "voter error").
[53] Manweller, p. 26.
[54] That is, if the sample had been representative even of the actual population of people who registered online, we could be 95% confident that somewhere between 5.4% to

not), there small size makes it impossible to make inferences about whether these ballots produce different rates of confusion in the actual population. In this case, the data presented by Dr. Manweller show no evidence that the mock Top Two general ballot might create more or less confusion in the Washington voter population than the mock Partisan general election ballot.

Of course, Section III of this report documents that these samples are already heavily biased by the fact they do not represent their target populations, and Section V of this report documents that the results are further biased by non response bias. As noted, voter error rates reported by Dr. Manweller are biased in a manner that inflates observed error rates. Thus, we gain little by placing confidence intervals on his estimates. My point in doing this is to illustrate the general weakness of small sample sizes. These small, biased samples mean that Dr. Manweller cannot use these data to make generalizations about voters in Washington state. As I note in the Section VI in this report, individual respondents also appear to have been observed multiple times in order to produce an inflated measurement of "error" in assessment of a candidate's relationship with a political party.

VI. Flawed Statistical Analysis

It is nearly impossible to assess the statistical significance of the results presented by Dr. Manweller. The paper provides very little information about the number of observations used to calculate the percentages in each table, and the paper provides very little useable information about statistical significance.

33.2% of that population would display voter confusion about the Partisan general election ballot (using Dr. Manweller's definition of "voter error").

Information about sample size used in any analysis is critical information for assessing the validity of the research. It is standard practice for scholars to report the number of observations used in their analysis when presenting tables and figures that display their results. Yet that information is almost totally absent in the Manweller paper. The text of the paper does mention that the "new voter" sample included 183 subjects. As noted above in the section on response rates (Section V), there is no concrete statement in the text of the paper, nor in any footnotes, that states the number of subjects in the "registered voter" or "active voter" samples.

I can glean some ideas about sample size from the one statement about the "new voter" group and from Table 1 in the paper. Given that there are 183 "new voters," information in Table 1 suggests "N=59" of these subjects were shown the mock "Partisan General" ballot, "N=64" were shown the mock "Top Two Primary" ballot, and "N=60" were shown the mock "Top Two General" ballot (59+64+60=183). Using this same logic, it would appear that the registered voter sample contained 102 subjects (N=31+N=41+N=30= 102), and that the "active voters" sample had 549 subjects (N=180+N=138+N=231 = 549).

A) No Information about the Number of Observations

However, Table 2, Table 3, Table 4, Table 5, Table 6 and Table 7 in the Manweller paper provide no information about the number of subjects used to generate the information in these tables. This makes it impossible to understand how representative the results might be of any target population.

The matter becomes even more confusing when information from footnote 17 is considered. Here, Dr. Manweller notes that "the overall N for Tables 5-7 is less than the N reported in Tables 1-4. Tables 5-7 exclude data collected on the newer voters." [55] This is confusing on several levels. First, contrary to this claim, there is no N (number of subjects) reported by Dr. Manweller in any table other than Table 1. Second, the statements suggests that "newer voters" were included in the data presented in Tables 1 - 4, yet the heading for Table 3 states that table describes "Registered Voters" and the heading for Table 4 states that table describes "Active Voters." Third, information in Table 1 suggests that the sample of registered voters totaled 102 subjects. Thus, it would seem the total number of observations used to generate the information in Table 3 ("Registered Voters") was 102.

Information in footnote 17 thus suggests that information presented in Table 6 and Table 7 were generated with less than 102 subjects. This is particularly problematic, because Dr. Manweller divides the data into 5 categories in each table. Given the discussion of the substantial bias associated with "registered voter" sample (discussed above), we might expect that fewer than 8 subjects were in each of the two older categories of age in Table 6.[56] This means the levels of voter "error" reported for these age categories in the sample is unrepresentative of any population.

Unfortunately, Dr. Manweller failed to provide basic information about the number of cases used in his analysis so it is impossible to assess how his results might be

---

[55] Manweller, p. 35 note 17.

[56] The Barreto and Glasser et al Pew Study (p. 6) reports that only 15% of voters who registered online were over 50. Dr. Manweller's paper suggests the sample used to generate Table 6 was less than 102. 102*.15 = 15.3. There were two categories for age over 50 in Table 6 (51-60, and 61+). 15.3 /2 = 7.65, or the potential for fewer than 8 subjects in each cell.

representative of the population of voters who registered online. Such information would be expected in any academic paper in order to pass peer review.

B) Critical Errors in Data Analysis and Presentation

Many of Dr. Manweller's results are summarized in Table 1, yet the paper presents little useable information about how these data were generated, nor about how "error rates" were calculated, nor about how statistical analysis was conducted.

Dr. Manweller paper states, "respondents are confused if they identify candidates on the Top Two ballots as official party nominees…Voter confusion, therefore, can be assessed by measuring how many respondents incorrectly identify candidates as the nominee of a political party. Table 1 summarizes the error rates for each type of ballot delineated by demographic group."[57] Yet it is not clear how (or if) this is actually done.

The mock Top Two primary ballot used by Dr. Manweller includes the names of multiple candidates said to "prefer" the Democratic Party and multiple candidates said to "prefer" the Republican Party. The survey instrument asks subjects if they consider one particular Democrat to be the nominee of the Democratic Party. It then asks subjects if they consider one particular candidate to be the nominee of the Republican Party (see Appendix B of the Manweller paper). This means there are three possible outcomes built into the experiment given Dr. Manweller's definition of voter error: No errors (the subject responds "No" to the question about the Democratic candidate, and "No" to the question about the Republican candidate); one error (the subject responds "No" when asked if one candidate is a nominee, but responds "Yes" when asked about the other

[57] Manweller, p. 13.

candidate); or two errors (the subject responds "Yes" to the question about the Democratic candidate, and "Yes" to the question about the Republican candidate.

C) Double Counting Individual Subject

However, Table 1, and subsequent analysis treats "error" as a dichotomous trait. The subject is correct, or wrong. Dr. Manweller provides no information about how the three category outcome built into his experiment is reduced to a dichotomous measure. There are several reasons why this is problematic.

First, this undefined method makes results in Table 1 very difficult to interpret. Consider the middle cell in the top row of Table 1, where we read that 25.7% of "new voters" are "confused" by the Top Two ballot type. I detail elsewhere in this report that sample bias, and a lack of a meaningful baseline of "voter confusion" make it nearly impossible to assign a meaning to this percentage. The matter is further complicated by the method used analyze and present this result. There are 64 subjects in this category ("new voters" exposed to the mock Top Two ballot). Yet the 25.7 percent appears to be calculated from an N of 128, not from an N of 64. Dr. Manweller provides no information about why this is the case.

Second, it would appear that the reported rates of voter confusion are based on repeated observations of the same subjects. That may be how we get 128 observations from a group of 64 subjects. That is, each respondent appears to have had two chances at making an error - if they are "correct" in responding about one candidate's status as a nominee, they are measured again and given another opportunity to make an error. Dr.

Manweller provides no information about how many respondents made one error, or two errors.

Third, this practice of using repeated observations of the same subjects in each category likely further biases the results such that they further inflate any reported rates of voter error. This method increases the likelihood that a voter shown the Top Two primary ballot will be coded as making an "error" about a candidate's relationship with a party. Consider a clinical drug trial as an analogy, where the research question is whether a 10 unit dose of a drug produced a particular effect. Experiments would be designed to treat some subjects with a single 10 unit dosage, then make an observation. Measures of the potential effects of a 10 unit dosage would be in error if these same subjects were quickly given a second 10 unit dose, and observed a second time.

Fourth, the reported number of subjects and the reported observations in Table 1 simply do not add up. In the cell where 41 "registered voters" are exposed to the Top Two primary ballot, 29.4 percent are reported to have made "errors." It appears that this percentage is calculated from 78 observations from the 41 subjects in this group. If each of the 41 subjects was observed twice (as appears to be the case with some categories in Table 1), we would expect this percentage to be out of 82 observations, rather than 78. This suggests that some subjects were counted twice, and some not. This discrepancy is also apparent in the "active voters" exposed to the "Partisan General" ballot and the "active voters" exposed to the "Top Two" ballot. In the former case, there are 351 observations drawn from 180 subjects. We would expect 360 observations if all subjects were counted twice. In the latter case, there are 266 observations from 138 subjects. We

would expect 276 observations if all subjects were observed twice. In short, it is not at all obvious what is being observed and reported here.

Dr. Manweller provides no explanation in the paper for why the number of observations in each cell of Table 1 exceed the number of subjects in each cell. Dr. Manweller provides no explanation in the paper for the discrepancies across cells, where some cells appear to have counted each subject twice, and others appear to count a fraction of subjects twice.

Fifth, using multiple observations of the same subject to create multiple data points presents problems for statistical analysis of the experimental data. Standard statistical tests are based on the assumption that one observation - one data point - is independent of others. Statistical tests such as Chi-square "assume that observations in the data sample are independent of one another - that no observation within a sample in any way influences another observation in the sample."[58] By counting subjects from the same sample twice, this fundamental statistical assumption of independent observations is violated. Statistical tests are thus biased.

VII. Test for Statistical Significance Cannot be Interpreted

Tests of statistical significance are designed so that we can understand if results being observed are significantly different than what might be observed due to simple random chance.[59] The "null" hypothesis in statistical tests is that observations are random (in this case, that there is no difference in rates of voter confusion across groups

[58] Evans, James D. 1996. *Straightforward Statistics for the Behavioral Sciences*. Brooks/Cole. P. 445; Levin and Fox, p. 223.
[59]Mansfield, Edwin. 1986. *Basic Statistics with Applications*. WW Norton. P. 322.

of subjects shown different ballot formats. Statistical significance is simply the probability that the null hypothesis is wrong. A conventional level of statistical significance is $p < .05$ (that is, there is less than a 5 in 100 chance that it is wrong to reject the null hypothesis.[60]

Dr. Manweller does report statistical tests (Chi-Square and Fisher's Exact). However, the conflation of the number of subjects and the number of observations reported in Table 1 makes it impossible to assess the statistical tests reported with Table 1. Moreover, Chi Square and Fischer's do not tell us how confident we can be when making inferences from values observed in the tests to the larger population (see Section V, Part C of this report for a discussion of inference from random samples).

Several important points should be considered in assessing the validity of these tests. First, most of the statistical tests Dr. Manweller reports show there are no significant differences across the groups being compared. Of the nine tests comparing pairs of groups shown different ballots, only four tests suggest there were more "errors" (as Dr. Manweller defines errors) among subjects shown a Top Two ballot. However, two of these tests showing "significant" effects make "apples-to-oranges" assessments by comparing subjects who were shown different versions of a Top Two ballot (and thus fail to compare "confusion" under the Top Two ballot to confusion under a partisan ballot).

Second, given the discussion of research design above in Section II, the most relevant statistics would be those making "apples-to-apples" comparisons. That is, comparing people exposed to the Top Two General Election ballot with those exposed to the Partisan General ballot; or comparing people exposed to the Top Two primary ballot

[60] Levin and Fox. p. 156.

with those shown the Partisan primary ballot (this experiment was not conducted). There is no significant difference in "confusion" across groups of "registered voters" shown the Top Two general and partisan general ballots.

The tests suggest statistically significant differences between subjects in the highly biased sample of "new voters" (see Section III, Part D and Part E of this report) who were shown the Top Two General election ballot, and those shown the Partisan General election ballot. The tests also suggest a significant difference between subjects from the biased and ill-defined sample of "active" voters shown the Top Two General ballot and subjects from this sample shown the partisan ballot. However, as noted in Section VI of this report, these results appear to be driven by double counting (some) individual subjects, making these suggestions of statistical significance suspect.

Third, and related to the point above, tests that show statistical significance are largely meaningless if the data being analyzed are drawn from flawed samples (see Section III and IV of this report); particularly if the instrument used to measure the concepts being compared were also biased (see Section VIII and Section IX of this report).

Fourth, statistical tests should be based on the number of subjects (observed one time each) in an experiment (see Section VI, Part D of this report). Dr. Manweller fails to report the degrees of freedom or numbers of cases used to generate the tests of statistical significance, so it is not possible to determine how the statistics were calculated. As a result, one cannot assess whether or not the tests were calculated with numbers that are inflated by the double counting of some subjects that is reported in Table 1.

VIII. Mistaken Measurement of Voter "Errors" in the Survey Design

As noted in my Report on Factual Political Knowledge, confusion about matters of politics and government is widespread in the American electorate. This is one reason that survey researchers regularly include clarifying information in their questions when asking subjects about political matters. In cognitive survey experiments of the sort Dr. Manweller attempts, "respondents must comprehend key elements of the questions."[61] Some concepts that are measured by opinion surveys are rather unambiguous, and might not require a clarifying introductory statement prior to administering the survey question. For example, a standard question asking about the President is worded: "Do you approve or disapprove of the way Barack Obama is handling his job as president?"[62] With this question, there is little need for an introduction to clarify and explain the meaning of key concepts to subjects who are surveyed.

A) The Need for Unambiguous Survey Questions

Conversely, when concepts in a survey question are likely to be more ambiguous, or less clearly understood by the survey respondent, an introduction is used in order to make sure the respondent understands the concept that the question is attempting to measure. For example, a recent Pew survey asked people about how states should balance their budgets. We might expect that many people do not know that state laws require budgets to be balanced. This is something that is probably more difficult for

---

[61] Willis, Gordon. 2005. "Cognitive Interviewing." Entry in *Polling America: An Encyclopedia of Public Opinion.* Samuel J. Best and Benjamin Radcliff, editors. Greenwood Press. P. 93. Also see Willis, Gordon et al. 1991. "The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires" *Applied Cognitive Psychology.* 5: 251-67.

[62] Pew Survey, June 2010. http://people-press.org/reports/pdf/627.pdf

people to assess than whether or not they approve of the job the President is doing. Pew thus introduced the question with, "as you may know, by law most states must balance their budgets, but many are now unable to do so." The question then proceeds, "which is the better way to address this problem?"[63]

As another example, a regular Washington Poll question seeks to measure how people perceive the initiative and referendum process. Again, if we assume that this is not something people think about often nor understand well, a neutral introductory prompt is needed. In this case, the introduction reads, "Not all states use the initiative or referendum process, where people get to vote on some important issues directly. Washington state does, however, and people here often vote on policy issues." After this statement is read, respondents are asked their opinion of "statewide initiative elections."[64] Rather than assuming that people understand how state budgets work, or what initiative elections are, these introductions allow the survey instrument to produce more accurate measures of how people respond to the actual concept the researcher is interested in.

B) Measuring the Concept of "Official Party Nominee"

Dr. Manweller acknowledges the difficulty that voters might have understanding the "specific legal and political meaning" of the concept "nominee" that he is attempting to measure.[65] He notes, "Not all voters may be cognizant of the term's narrow definition."[66] The discussion in his paper clearly states that the key concept here is

[63] Pew Research Center Poll. June, 2010. http://people-press.org/reports/questionnaires/628.pdf
[64] http://www.washingtonpoll.org/archives.html. See Question 5, 2007.
[65] Manweller, p. 16.
[66] Manweller, p. 16.

whether voters perceive that a candidate has the nomination or endorsement of some official political party organization.

Dr. Manweller states one of his key research questions is whether or not voters mistakenly believe that candidates are the "*official* nominee" of a particular party (emphasis added).[67] He notes further, "the most important question" is whether voters believe the candidates "are perceived as the *official* nominees of political parties." (emphasis added).[68] He also notes his survey was "measuring how often respondents incorrectly answered questions about the hypothetical candidates' status as the *official* nominee of a political party...and that "from the Court's perspective, respondents are "confused" if they identify the candidates on the Top Two ballot as *official* party nominees" (emphasis added).[69] In his conclusion, Dr. Manweller again states that he is presenting evidence that "voters perceive an *official* relationship" between candidates and political parties, and that the courts will need to determine if "identifying candidates as *official* nominees of a party meets Justice Thomas' standard of widespread voter confusion" (emphasis added).[70]

Dr. Manweller thus regularly refers to the concept of interest as voter perception of candidates as "official party nominee." The key concept that needed to be measured is whether voters perceive that candidates listed on the ballot are somehow officially approved by a legal party organization. This definition of the concept closely mirrors the

[67] Manweller, p. 12.
[68] Manweller, p. 12-13.
[69] Manweller, p. 13.
[70] Manweller, p. 20

concept of an official party "endorsement" that Dr. Manweller references while quoting from *Washington State Grange v. Washington State Republican Party.*[71]

C) Failure to Measure the Concept of Official Party Nominee

However, Dr. Manweller's survey instrument was not designed to measure perceptions of whether candidates were the official nominee of a party. Nor was it designed to measure perceptions of whether a candidate was endorsed by a party. Nor was it designed to measure if voters perceived that a party approved of or associated with candidates. The instrument (Appendix B in the Manweller paper) asks people, "Based on the information presented in the sample ballot above, would you consider Candidate John Smith the nominee of the Democratic Party." In other questions, the words "nominee of" are replaced by "affiliated with," "representative of" and "associated with."

The wording of Dr. Manweller's survey questions fail to describe accurately the potential for an active relationship between a party organization and a candidate. Rather, Dr. Manweller's questions measure the potential for a passive relationship between the candidate and a party. As such, the questions are not designed to measure how voters might perceive the true legal and political relationship between parties and candidates. At no point does the survey ask if voters consider the candidate to be an *official* nominee of a party. It does not ask if voters perceive candidates to be endorsed by a political party. The words "endorsed" and "endorsement" are not included in the survey. The word "approved" and the words "approved by the party" are not included on the survey.

[71] Manweller, p. 9, citing *Grange* p. 12.

Rather than asking if "the *party* associates with"[72] a candidate, Dr. Manweller asks people if they think a candidate is associated with a party. Not only does the use of passive rather than active wording fail to accurately reflect the legal and political relationship between parties and candidates, but such wording differences are known to produce vastly different responses in survey experiments. Passive wording can produce dramatically higher rates of agreement in responses.[73] Given Dr. Manweller's coding of "voter errors" (where agreement with his questions about the Top Two ballots are "errors" and agreement with the same questions about the Partisan general are correct) this wording inflates the rate of "voter error" that Dr. Manweller attributes to the Top Two ballot while deflating "error" associated with the partisan ballot.

Rather than using relatively clear, less ambiguous wording about a candidate's potential legal relationship with a party that he uses in the text of the paper (e.g. "endorsed by" and "official nominee"), Dr. Manweller uses more ambiguous phrases in the survey instrument (e.g., "associated with" and "nominee.") As noted above, since people may lack a clear understanding of what concept is being measured, a survey should include a clarifying introductory statement. In this case, something as simple as stating, "*State and local party officials often nominate candidates they approve of.* Based on the information presented in the sample ballot above would you consider Candidate John Smith the *official* nominee of the Democratic Party." Likewise, another more valid

---

[72] This is the language used on the Top Two ballot disclaimer statement.

[73] Moore, David W. 1992. *The Superpollsters*. Four Walls Eight Windows Press, p. 334; Shuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. San Diego: Academic Press. Also, on wording effects, see Smith, Tom. 1987. "That Which we Call Welfare by Any Other Name Would Smell Sweeter: An Analysis of the Impact of Question Wording on Response Patterns." *Public Opinion Quarterly*. 51:75-83.

measure would be produced by using the introduction, *State and local party officials often endorse candidates they approve of.* Based on the information presented in the sample ballot above would you consider Candidate John Smith the *official* endorsed candidate of the Democratic Party."

D) Effects of the Flawed Survey Wording on Measures of Voter Confusion

These omissions in the survey design are highly problematic because the use of the ambiguous phrases like "associates with" and "nominee" likely inflate affirmative responses to questions about the candidates. The use of less ambiguous phrases, such as "official nominee" or "endorsed by party officials" are more directly related to the concept that Dr. Manweller claims to be measuring. If these precise definitions had actually been used in the survey experiments, it is likely that fewer people would agree that the candidates were nominated *by the parties.* Furthermore, had the survey been designed with a standard clarifying statement informing voters about what it means to be an official nominee of a party organization (or officially endorsed by a party organization) we would likely observe lower rates of voter agreement when people were asked if candidates are nominees of a party.

The differences in results obtained from surveys that specifically define the meaning of a concept can be dramatic. Consider attempts to use surveys to measure public attitudes about "abortion." Surveys produce dramatically different results depending on how the potentially ambiguous term "abortion" is qualified and specified. For example, surveys record 81% to 90% support for policies that allow "abortion to save the health or life of the woman" but only 7% to 13% support for policies that allow

"abortions in the third trimester."[74] This illustrates why survey researchers use specific language in their questions. Questions that attempt to measure public attitudes by simply using the ambiguous phrase "abortion" without clarification would produce highly ambiguous results. The same can be said about questions that simply use the ambiguous phrase "nominee."

The issue is important here, because voters likely have several ways they might understand of the meaning of the simple word, "nominee." To voters, this word could be viewed as synonymous with a candidate having simply received more votes than other candidates in a primary election. Such perceptions likely exist among voters independent of the ballot type they interact with. If voters have this conception of "nominee" Dr. Manweller's experiments will code their responses as being "errors." Given that the concept "nominee" was not explained in the survey, it is not clear what the survey instrument is measuring when it asks about "nominee."

IX. Mistaken Interpretation of Voter "Errors"

In the real world, candidates who appear on Washington's Top Two primary ballots and general election ballots are regularly nominated and endorsed by official party agents. Dr. Manweller's definition of a "voter error"[75] would require that if a voter responds, correctly, to a survey question asking if a candidate was nominated or endorsed by a party, that the voter would have been confused and have made an "error."

---

[74] For example, surveys record 81% to 90% support for policies that allow "abortion to save the health or life of the woman" but only 7% to 13% support for policies that allow "abortions in the third trimester." http://www.gallup.com/poll/9904/public-opinion-about-abortion-indepth-review.aspx#2

[75] "respondents are confused if they identify candidates on the Top-Two ballot as the official nominee of a political party." Manweller, p. 13

A) Many Candidates on Top Two Ballots are Endorsed by Political Parties

As I write this report (prior to the August 2010 Washington primary) the official Whatcom County Democratic party web page lists Patty Murray as their endorsed candidate for the US Senate.[76] The party also list endorsements for three state legislative candidates. These candidates are listed on the Top Two primary election ballot. Those who win will be listed on the General Election ballot in November 2010. The Democratic Party's web page also lists the Party's endorsements from 2008. The party endorsed Christine Gregoire for Governor, Rick Larsen for the US Representative, and several other candidates for statewide office and legislative office. These candidates all appeared on the Top Two general election ballot.[77]

The Whatcom County Republican party's official web page also lists endorsements prior to the primary 2010 primary. The party endorsed several legislative candidates who will appear on the 2010 Top Two general election ballot, and two candidates who will appear on the 2010 Top Two Primary ballot.[78] The publication of party endorsements of candidates listed on the Top Two primary and Top Two general ballots appears common. Prior to the 2010 primary, King County Democrats also listed endorsements in the US Senate contest, and state and local races.[79]

[76] http://www.whatcomdemocrats.com/content/candidates-2010
[77] http://www.whatcomdemocrats.com/content/past-elections
[78] http://www.whatcomgop.com/Issues/WCRP-2010-Primary-Endorsements.aspx
[79] http://wa-demchairs.org/kcdems/2010/endorsements.php

B) Many Candidates on Top Two Ballots are Nominated by Parties

Official agents of the Washington State Democratic Party also nominate candidates who are listed on the Top Two primary and general election ballots. Indeed, Democratic State Chair Dwight Peltz stated that "The Top Two Primary will only serve to winnow down the field to two candidates; and will no longer nominate a candidate for one party or the other." As a result, Mr. Peltz notes that, "It is the policy of the WSDCC to have a nominee in contested partisan races."[80] Democratic Party officials (including Precinct Committee Officers, and Legislative District Committees) thus nominate candidates prior to the primary election. This would suggest that a reasonable person would conclude that most Democratic candidates listed on the Top Two general election ballot are in fact the official nominee of the Party.

C) So-called Errors are not Errors.

Party endorsements and nominations of candidates are publicized by the political parties, and by the news media.[81] Candidates publicize the fact that they are party nominees. This means that had the Manweller experiment been conducted in a real world setting, many people would respond, correctly, that a candidate listed on the Top Two general or primary ballot was nominated by or endorsed by a party. As noted above, Dr. Manweller would consider these responses to be errors.

The fact that parties regularly nominate and endorse candidates listed on Top Two ballots contaminates Dr. Manweller's experimental results. It is reasonable to expect that well-informed voters may be aware of the fact that certain candidates are nominated by or

[80] http://www.stevenscountydemocrats.org/Links_and_faqs.htm
[81] insert URLs

endorsed by a party. If such a voter were a subject in Dr. Manweller's experiments, they would be asked to "pretend that you will be participating in an election."[82] If well-informed voters were in Dr. Manweller's samples (Section III this report demonstrates that the samples are biased against including an accurate proportion of well informed voters), and they were aware of the WSDCC's policy of endorsing candidates or any party's practice of endorsing candidates, the may assume that a mock Top Two general election ballot listed candidates nominated or endorsed by a party. Such responses are counted as "errors" in the results reported by Dr. Manweller, and thus further inflate measured rates of "error" he associates with the Top Two ballot.

Furthermore, as discussed in Section VIII, Part D of this report, it is not clear why responses that Dr. Manweller considers to be "wrong" are in fact "errors." Independent of any ballot format, many confused voters responding to conventional language may assume that a candidate who claims an affinity for particular party is casually affiliated with or casually associated with that party. They may assume that a person who has ever voted for a particular party's candidate is "affiliated with" or "associated with" that party. Survey instruments attempting to measure how ballot structure affects voter perceptions of candidates' relations with parties must account for this. As noted in Section VIII, Part D, Dr. Manweller's passive, vague survey language does not provide the ability to measure accurately perceptions of the party's association with the candidate nor the specific legal and political relationships between candidates and parties. Apart from the use of the word "nominee," Dr. Manweller's survey items simply measure perceptions of casual, passive affinity a candidate might have for a party.

[82] Manweller, Appendix A p. 22 - 24

Dr. Manweller's misrepresentation and misinterpretation of these concepts inflates his reported measurements of "voter error." As another example, in the text of his report he claims that "Forty-two percent felt candidates were *the party's* representative"[83] (emphasis added). Again, the survey instrument failed to ask subjects if they considered a candidate to be "the party's representative" or "the party's official representative." This language was not used in the survey. Rather, the survey instrument used vague, passive language. It asked, if people if they considered "Candidate John Smith a representative of the Democratic Party." Conceptually, these differences in language are not trivial. Consider how people might evaluate a person they assume to be Canadian. We would get different responses to the following questions: 1) Do you consider John Smith to be representative of Canadians? 2): Do you consider John Smith to be Canada's representative? 3) Do you consider John Smith to be Canada's official representative? Dr. Manweller's questions more likely measure the concept embedded in Question 1 here, rather than the concepts embedded in Question 2 or Question 3.

X. Conclusion

This report demonstrates that the Manweller paper is deeply flawed on several levels. Dr. Manweller claims to measure voter confusion,[84] but his paper provides no evidence that most of the subjects in his experiments were people who voted. I demonstrate that major flaws in Dr. Manweller's research, including sampling issues, measurement issues, and failures in experimental design, act individually and cumulatively to inflate the measured rates of "voter confusion." My report also

[83] Manweller, P. 17
[84] Manweller, p. 2; p. 12; p. 13; p. 14;

demonstrates that none of the conclusions in the Manweller paper about a potential relationship between the Top Two ballot and voter confusion are defensible, and that there is no way to generalize from Dr. Manweller's biased samples to the population of actual voters in the state of Washington.