

Rebuttal Report of Dr. Arthur Lupia

October 14, 2002

I have been asked by the Brennan Center to evaluate “An Analysis of the 1998 and 2000 Buying Time Reports” by Dr. James L. Gibson. I henceforth refer to this document as “the JLG report.” Unless otherwise specified, page numbers refer to those in the JLG report.

My evaluation is organized as follows.

- An initial section entitled “Qualifications” provides relevant information about my background.
- In Section I, I evaluate claims made in the section of the JLG report entitled “Summary of Conclusions about *Buying Time 1998*.”
- In Section II, I evaluate claims made in the section of the JLG report entitled “Conclusions-*Buying Time 1998*.” There, I also draw on the JLG report’s supporting arguments (pp. 10-43).
- In Section III, I evaluate claims in the “Summary of Conclusions” and “Summary and Conclusions” sections of the JLG report that are devoted to *Buying Time 2000*. As many of the claims in this section of the JLG report repeat claims made in the analysis of *Buying Time 1998*, I focus my comments on the claims that are unique to the 2000 study.

In the process of writing this evaluation, I have read *Buying Time 1998*, *Buying Time 2000*, the Expert Report of Dr. Ken Goldstein, portions of the Expert Report of Drs. Jon Krasno and Frank Sorauf, and a series of working papers and articles by Dr. Goldstein and his co-authors. I have also consulted other materials to check claims and to provide examples for my arguments. These ancillary materials are cited below.

In summary, I find that many of the JLG report’s claims are derived from speculation rather than well-established scientific principles. In addition, I find that the JLG report, while insinuating that the entire *Buying Time*/Goldstein work lacks

credibility, in fact addresses relatively few of the work’s key findings. *Almost all of the main claims of the Buying Time reports and the Goldstein expert reports are unaffected by the JLG report.* Therefore, the JLG report’s core criticisms of the *Buying Time* reports and Goldstein work do little or nothing to undermine their key findings.

The JLG report is strongest when it presents standards for scientific inference, most notably replicability, and when it derives conclusions from methods or arguments whose logic is transparent. However, the JLG report itself fails to adhere to these standards consistently.

Indeed, to draw conclusions, the JLG report repeatedly substitutes speculation and innuendo for direct evidence and transparent, replicable methods.

This is particularly apparent with respect to two of the JLG report’s most important claims. For example, pages 31 to 35 of the JLG report ask us to discount the idea that “genuine issue ads” can actually constitute electioneering. Throughout Section II.d, I demonstrate that the JLG report’s claims are not the product of scientific research applied directly to the question at hand. Instead, they are based on false assumptions and interpretations of data that are inconsistent not only with standard practices, but also with Dr. Gibson’s previous research.

Similarly, pages 35-43 of the JLG report focus on how many current ads the proposed regulatory change would impact. The *Buying Time* and JLG estimates of this effect are determined in large part by different assumptions about appropriate denominators. In other words, they take different points-of-view on which set of advertisements should be included when assessing the regulation’s impact. The JLG

report makes strong claims about validity of its preferred denominator relative to the one used in *Buying Time 1998* -- claiming that only its preferred denominator is relevant to this case. However, I demonstrate throughout Section II.a that the JLG report's claims on this matter are indistinguishable from pure result-oriented advocacy and are of questionable scientific validity.

It is also noticeable that the Plaintiffs' expert reports no attempt to replicate the data collection procedures that the JLG report so vocally criticizes. In light of the work already performed by Goldstein, such an exercise would have been extremely easy and relatively inexpensive to conduct particularly given that Plaintiffs could have run a test on some sample of the storyboards. It would also have given the Plaintiffs' experts a more credible basis for critique than the speculations on which the JLG report repeatedly rely. The JLG report has almost no impact on my assessment of the credibility of the *Buying Time*/Goldstein endeavor. Having read two sides of the argument on this matter, I conclude that the *Buying Time* reports and the Goldstein expert report are a product of innovative scientific procedures that clarify important dynamics of political advertising.

INTRODUCTION

Overview

The JLG report has two main sections. The first section, pages 3-43 offer a critique of *Buying Time 1998*; the second, pages 44-64 offer a critique of *Buying Time 2000*. In what follows, I address the JLG report's charges against *Buying Time 1998* and *Buying Time 2000*, and show that many of the JLG report's main conclusions are speculative or erroneous.

While the JLG report is my primary focus, I feel it informative to give a brief overview of *Buying Time 1998*, *Buying Time 2000*, and the Goldstein research agenda. These efforts provide a unique and valuable source of information about political advertising. The Goldstein database, which combines information from numerous political advertisements with information about when and where they were aired, is advancing the emerging science that seeks to clarify the impact of these broadcasts. The innovative and painstaking service that Goldstein and his collaborators are providing is increasingly recognized as a valuable public good. At a minimum, it is creating a kind of database that did not exist before.

The *Buying Time* publications provide a unique and clear presentation of trends in political advertising. No other publicly available document that I know of matches these volumes in their attempts to bring this range of hard data to important questions about the correspondence between issue advocacy and electioneering and other important questions relating to political advertising. *Buying Time 1998* and *Buying Time 2000*, with their presentational clarity and dependence on data more than anecdote, provide future debates

– both scientific and legal – with important new methods, interesting results, and improved empirical and conceptual foundations from which to build.

The Goldstein research agenda is also having a broad scientific impact. For example, the methodological improvements that Goldstein has brought to studies of how political advertising impacts viewers have caused one of the most recent and important changes to the National Election Studies. The studies now ask better and more detailed questions about citizens' viewing habits – questions about when and what they watch. These new questions, which the studies designed with ample counsel from Goldstein, are providing the thousands of scientists who use the NES new opportunities to conduct more precise studies of advertising influence. Having helped to design these studies in official capacities dating back to 1998, I can tell you that the Goldstein research agenda is the direct cause of these improvements.

In sum, the Goldstein research agenda, which helped to produce the data and findings for the *Buying Time* and Goldstein expert reports, is making a unique, valuable and increasingly recognizable impact on the conduct of political science.

Providing a Context for the Critique

The JLG report primarily addresses the *Buying Time* reports and also reviews Dr. Goldstein's conclusions more generally. I begin my evaluation of the JLG report by providing a context for its actual scope. The JLG report contests very few findings in each of these publications. Figure 1, for example, sets out the principal conclusions of Dr. Goldstein's expert report.

assistants and I processed the 2000 CMAG data. I have used these data in a series of reports and articles to describe the source, content, and targeting of political advertising in the 2000 elections. In this report, I set forth nine principal conclusions that emerge from my study of the CMAG data:

1. Advertisements sponsored by parties and interest groups comprised a significant and increasing portion of political television ads broadcast in federal races over the course of the entire year and especially during the last 60 days of the 2000 election.
2. The "magic words" defined in *Buckley v. Valeo* do not provide an effective way to identify political television ads that have the purpose or effect of supporting or opposing candidates for election to a public office.
3. Interest group financed ads that depicted or mentioned candidates for federal office were concentrated in the weeks immediately preceding the election and stopped on Election Day. Interest group ads that did not mention or picture candidates for federal office were distributed more evenly throughout the year, rising and falling with the ebb and flow of the legislative calendar.
4. Interest group ads that mentioned a candidate within 60 days of the general election almost invariably targeted House or Senate candidates in hotly contested races or presidential candidates in battleground states.
5. Ads that mentioned or depicted federal candidates for office differ from ads that did not mention or picture candidates in other ways, such as their content, their distribution over time, their geographic dispersion, etc.
6. Ads that did not mention or depict a candidate are readily distinguishable from quintessential electioneering ads (such as ads run by a candidate seeking election to public office) in numerous ways such as their content, their distribution over time, their geographic dispersion, etc.
7. Virtually all (97.7 percent) political television ads aired during 2000 that would have been covered by BCRA were perceived by project coders as having an electioneering purpose.
8. BCRA requirements would have applied to 3.1 percent of ads broadcast by interest groups during 2000 which were perceived as genuine issue ads by coders.
9. Broadcasting issue ads in close proximity to an election is not necessarily an effective way to inform voters about public policy issues or generate action on such issues.

Figure 1. Goldstein principal conclusions. From Goldstein expert report, page 3.

The JLG report:

- Does not contest conclusion 1 – that advertisements sponsored by parties and interest groups comprised a significant and increasing portion of political television ads broadcast in federal races.
- Does not contest conclusion 2 – that the *Buying Time* and Goldstein reports make many interesting and important findings about the role of “magic words” in political advertisements.
- Does not contest conclusion 3 – that interest group financed ads that depicted or mentioned candidates were concentrated in the weeks immediately preceding the election.
- Does not contest conclusion 4 – that interest group ads that mentioned a candidate within 60 days of the general election tended to target hotly contested races.
- Does not contest conclusion 5 – ads mentioning or depicting a candidate differ from other political ads in ways that are meaningful to questions about their correspondence to electioneering.
- Does not contest conclusion 9 – that broadcasting issue ads in close proximity to an election is not necessarily an effective way to inform voters about, or generate action regarding, public policy issues.

On conclusions 1-5, I am persuaded by the findings and arguments of the *Buying Time* and Goldstein expert reports. In each case, the research is grounded in transparent and replicable procedures that enhance their scientific credibility. Goldstein conclusion 9, by contrast, is not based directly on his own data. Instead, it is grounded in widely accepted scientific discoveries. Given what I know of these discoveries, I am persuaded that conclusion 9 is valid.

The JLG report focuses on conclusions 6, 7, and 8 in my Figure 1 and it is on the JLG report’s critique of the matters upon which my evaluation now focuses.

SECTION I. Evaluating “Summary of Conclusions about *Buying Time 1998*.”

The JLG report begins with a list of bulleted conclusions. I begin by evaluating these claims.

1.

“*Buying Time 1998* should not be accepted as the product of scientific inquiry, but is instead policy advocacy written by people with a strong ideological commitment to a particular position on campaign finance reform.” Page 3.

“The strong policy and ideological commitments of the investigators are not compatible with the conventional canons of scientific objectivity and may have undermined the integrity of the data collection and analysis.” Page 3, fn 3.

A person’s political or ideological beliefs need not prevent them from being an effective scientist. In fact, I know of “no conventional canons of scientific objectivity” that preclude scientists from having political and ideological commitments, and Plaintiffs’ experts have not produced one.

Scientists must make several important decisions in the process of conducting their research. One decision is what to study. A scientist’s personal interests, political or otherwise, often drive this decision. This fact is neither necessary nor sufficient to undermine a study’s scientific integrity. Therefore, when the JLG report alleges that *Buying Time* cannot be the product of scientific inquiry because its authors have an ideological commitment, it is in error.¹

¹ One counterexample consists in the case of Sir Francis Crick, who with James Watson discovered the double helix of DNA. In his autobiography, Dr. Crick writes:

“I have no doubt, as will emerge later, that this loss of faith in Christian religion and my growing attachment to science have played a dominant part in my scientific career, not so much on a day-to-day basis but in the choice of what I have considered interesting and important. I realized early on that it is detailed scientific knowledge which makes certain religious beliefs untenable. A knowledge of the true age of the earth and of the fossil record makes it impossible for any balanced intellect to believe in the literal truth of every part of the Bible in the way that fundamentalists do...It seemed to me of the first importance to identify these unexplained areas of knowledge and to work toward their scientific understanding, whether such explanations would

For example, we can agree that much cancer research is driven, at least in part, by the fact that many scientists are more interested in slowing or curing cancers than they are in promoting their spread. This does not mean that their work, if done according to established standards, is not regarded as the product of scientific inquiry. Applying the same principle to this case – any apparent or implied ideological commitments of the *Buying Time 1998* authors are not, by themselves, sufficient to show that *Buying Time 1998* is not a product of science.

The footnote’s claim “may have undermined the integrity” is pure speculation. The JLG report presents no direct evidence on this point. The JLG report does claim that there are problems with the data, but provides no direct evidence that ideological commitments are their cause.

2.

“In general, the conclusions drawn in *Buying Time 1998* should not be accepted because the data upon which the report is based are fundamentally flawed.” Page 4.

The JLG report provides no clear standard for judging data “fundamentally flawed” and in most cases, the data collection practices critiqued in the JLG report are not only widely practiced and broadly accepted in science, but also employed by Dr. Gibson in the conduct of his own research. See my point 9 in this section, my points 1, 2, 4, 6, and 7 in Section II.b, and my point 5 in Section II.d.

turn out to confirm existing religious beliefs or to refute them.” (Francis Crick. 1998. *What Mad Pursuit: A Personal View of Scientific Discovery*. New York: Basic Books. Page 3.)

Dr. Crick is not reticent to admit that a strong ideological commitment drove him to study certain topics. It is also worth noting that his emphasis on using science to find answers, whatever they may be, also highlights the critical difference between one scientific decision -- choosing what to study – and another -- choosing how to study something. The former choice is not regulated by scientific practice – scientists can and do study an incredible range of topics for many different reasons. The latter choice, how to conduct the study, is so regulated.

3.

“Indeed, to the extent that any reliable inferences can be drawn from the data, the conclusion would be that a very large percentage of the advertisements whose purpose was coded as seeking to influence an election in fact had policy issues as their primary focus.” Page 4.

The claim that “any reliable inferences” would lead to the unique conclusion described in the JLG report is speculated but not proved. The JLG report does not explore the range of other possible conclusions. I will show that other conclusions are possible and that the JLG report’s manner of deriving its preferred conclusion is itself questionable.

Two questions in the coders’ questionnaires are at issue. One question asks coders for *their opinions* of an advertisement’s “purpose.” The other asks coders to judge the ad’s “primary focus.” I accept one of the claims that the JLG report derives from these two questions, namely that “a very large percentage of the advertisements... had policy issues as their primary focus.” However, the conclusion’s implication for the question at hand is not as simple.

The JLG report bases many of its conclusions on the assumption that if a coder’s opinion of an advertisement’s “purpose” differs from his or her judgment of its “primary focus” then we must disregard the former as a credible statement of the coder’s opinion. The JLG report offers no coherent scientific justification for this assumption.

Indeed, the assumption is a “red herring” and it is wrong. There is no conflict or contradiction in viewers opining that an ad’s purpose is something akin to electioneering while judging its primary focus to be something akin to policy. Therefore, it is erroneous to assume that an ad whose purpose is perceived to be electioneering cannot also be judged to focus primarily on policy. I will also show that this error is manifest in

numerous conclusions made throughout the JLG report. The unfortunate result is that the conclusions of many of these arguments are in error as well. *See* Section II.d for a more thorough description of these errors.

4.

“This Report is not the product of any peer-review process. The Report was published by the Brennan Center, not by a commercial or academic press. Thus, the Report was not vetted in any way whatsoever prior to its publication, and consequently the normal process of explication of the project methodology, error correction, and review of substantive conclusions prior to publication did not take place. This seriously limits the confidence one can place in the Report.” Page 4.

While an academic or commercial press did not publish *Buying Time 1998*, we can agree that not all publications of scientific value are published by academic presses. A leading example in political science is *The Book of the States*, which is published by The Council of State Governments, a non-profit organization not affiliated with any university. *The Book of the States* is widely used by academics and is regarded as a credible source of information about state governance.

We can also agree that *Buying Time* does not appear to have been written with an exclusively academic audience in mind. This document, whose prose is based on scientific practices but whose words are crafted for a largely non-scientific audience, are of a kind that many leading scholars have written. Such documents can draw important lessons from scientific practices and provide valuable clarity for relevant audiences. In such cases, the fact that an academic press did not publish *Buying Time* is not determinative or informative about the credibility of its conclusions. *See* point 5 of Section I below.

The charge that *Buying Time 1998* “is not the product of any peer-review process” or “was not vetted in any way” is of doubtful significance – at best. Common peer review

processes vary greatly in the number and kind of activities they include.² Academic presses do not hold a monopoly on such activities. Moreover, the JLG report displays no apparent knowledge of whether scholars or experts had opportunities to comment on critical aspects of the *Buying Time* reports. In sum, the claim that the *Buying Time* reports were subject to no activities generally categorized as peer review is both unsubstantiated and of doubtful significance.

5.

“The Report largely consists of a series of tables filled with quantitative information. The tables are rarely explained, analyzed, or interpreted. For instance, Chapter 4 consists of one page of text (p. 87), then 21 pages of tables (pp. 88 - 108), a half page of text (p. 109), and another page with tables (p. 110). Such a chapter cannot be taken as serious scholarship. Serious scholarship would carefully explain the objectives of each of the analyses, the details of how the data were manipulated, and would carefully and clearly link any substantive conclusions with specific empirical evidence. None of this takes place in this chapter, or in this report more generally. This report was surely published “inhouse,” by the Brennan Center itself, because, based on my experience as a reviewer for publishing houses (and the sort of review I would submit were I called upon for publication advice), I doubt whether any academic publisher would publish a document like this.” Pages 4-5.

Scholars, be they new assistant professors or Nobel Prize winners, get opportunities to provide new knowledge to different audiences. Sometimes we publish in very abstract, but rigorous academic journals. In other cases, we are asked to convey lessons to audiences with varying levels of technical detail. Therefore, there are no universally accepted standards about the minimum amounts of explanation, analysis, or interpretation that qualify a piece of writing as “serious scholarship.”

Indeed, the JLG report offers no evidence that the standard of “serious scholarship” that it invokes is the product of well known and widely agreed upon

² Some publishers, for example, will publish an article on the basis of two or three reviews, while others regularly accept papers for publication that have been presented at a public conference. Peer-review processes have no single, universally accepted structure. The point of peer reviews, whatever the exact procedure of acquiring them, is to get other expert opinions on the merits of a scientific argument.

standards. Moreover, I doubt that Plaintiffs' experts can produce a widely cited document on scientific procedure that says anything about the minimum table to text ratio that qualifies a chapter or piece of scholarship as "serious." Nothing in the report reveals the first three sentences of this particular claim to be anything other than a personal opinion offered in place of widely accepted scientific standards.

In response to the claim "serious scholarship would carefully explain..." I agree that common scientific standards require transparent methods and replicable results. These standards do not, however, require that we detail our methodology every time we open our mouths or put pen to paper. There are times and places where the audience wants less detail. In such cases, the standard practice is to provide clear citations to these details. While the JLG report correctly notes that the *Buying Time* reports do not provide such citations, this fact does not imply that materials that "carefully explain the objectives of each of the analyses, the details of how the data were manipulated, and would carefully and clearly link any substantive conclusions with specific empirical evidence" do not exist. To the contrary, such details do exist in other work by Goldstein and co-authors [see, e.g., Appendices E (all), F (all), and I (pages 14-19) of "Appendix to Expert Report of Kenneth M. Goldstein."]

The relevance of the claim "I doubt whether any academic publisher would publish a document like this" is questionable – at best. Must something be written for an academic press to contain serious scholarship or to be taken seriously in a public policy debate? If so, what should we make of the status of the JLG report? It is not written for an academic press, no evidence of peer review is presented. Yet it is presented as having a scientific basis. Its credibility comes from application of the scientific method. With

respect to the correspondence between a document's credibility and its publication by an academic press, *Buying Time 1998* and the JLG report must be judged in the same fashion. Again, the fact that the *Buying Time* reports were not written for academic presses provides no direct evidence about the quality of the work.

Finally, I note that the tables in *Buying Time 1998*'s Chapter 4, which was singled out by the JLG report, are largely addressed to the temporal distribution of political ads, a conclusion that the JLG report does not question.

6.

“No data base has been (nor can be, it appears) produced that will generate the specific numbers found in this Report. This is in part because the data set is continuously being manipulated and changed. But it is also a function of the lack of transparency in the statistical analysis that underlies these tables. I will give specific illustrations of this throughout my report, but most tables do not provide sufficient information for an analyst to replicate the findings. In the social sciences, we demand that statistical analysis be replicable (that another investigator using the same data be able to reproduce exactly the same findings). This report is not replicable, and that undermines tremendously any confidence one should place in the findings produced.” Page 5.

When making this claim, Dr. Gibson violates his own clearly articulated standards for scientific argument. He offers no description of his own replication procedures. As a result, there is no transparency. Someone attempting to replicate the JLG report's replication attempts would have far less guidance than is offered in *Buying Time 1998*. By Dr. Gibson's standards, we should question the credibility of the JLG report's replication claims. We cannot tell whether Dr. Gibson tried a number of ways to replicate the data or if he made one careless or half-hearted attempt and then gave up. We cannot tell whether he was assisted in his replication attempt by a team of leading experts or if he derived his answer by less rigorous or unscientific means. He offers no evidence about whether any of a large number of possible and scientifically valid changes to his own

procedures (from very minor to quite major), would result in successful replication. For more on the JLG report's failed replication attempts, *see* my points 8 in this section, my point 10 in Section II.b, and my point 2 in Section II.c.

More generally, it is important to consider the difference between *replicated* and *replicable*. Replication is a single act by a scientist. Replicability is a general attribute of a study. That a particular scientist fails in her attempt to *replicate* a study does not show that the study is not *replicable* any more than your defeating a child in a single chess match means that you are “unbeatable” at chess. The claim that “the report is not replicable” is not proved in the JLG report.

Indeed, the JLG report establishes that Dr. Gibson replicated some of the *Buying Time* results, but it focuses on a small number of instances in which the replication attempt did not succeed. The JLG report does not prove that replication of these few items is impossible.

It is also worth noting that the Plaintiffs and their experts passed up the opportunity to resolve their concerns by replicating the data collection procedure itself. The lack of such an attempt is surprising given the JLG report's emphasis on replicability and the ease with which this activity could have been undertaken – particularly in light of the groundwork already laid by Dr. Goldstein's work. My understanding is that Plaintiffs had most of the storyboards from the 1998 study and all of the storyboards from the 2000 study. Therefore, it would have cost relatively little and taken relatively little time to recruit coders, show them the story boards, administer the questionnaires and then record and report their results. This reluctance to replicate the Goldstein/*Buying Time* data

collection procedures has repeated relevance in my evaluation. *See* my points 2, 4, and 5 of Section II.b, my point 5 of Section II.c and my point 3 of Section II.d.³

The JLG report also claims that “the data set is continuously being manipulated and changed.” Large data sets regularly change. In many cases, new information or discovery of minor coding errors prompts the changes. It is common for scientists to accept that database management entails such changes if the changes are done in a transparent and replicable manner. Therefore, the existence of changes in a database as large as *Buying Time*’s is neither uncommon nor is it directly relevant to the validity of *Buying Time*’s results. The relevant question is about the manner in which the data was changed and the extent to which these changes are consistent with standard scientific practices.

7.

“The Report is filled with questionable statistical techniques and applications.”
Page 5.

The JLG report presents no evidence that Goldstein or the authors of *Buying Time* used any improper statistical techniques or followed improper statistical principles. Indeed, it identifies no mistakes in technique (such as confusing a mean for a mode, or correcting for stratification when the real problem is unobserved latent variables), as many political scientists would read the comment to imply. Instead, many of the JLG report’s questions about “statistics” hail from complaints about the data – he finds limited differences in various versions of the database and that his attempt to replicate certain

³ It is also worth noting that Dr. Gibson faced a handicap when analyzing this data that most scientists do not. Best practices in managing a scientific database include giving all variables clear labels, giving each value of each variable a clear label, and making all of these labels available to other scholars either in the data itself or, as more commonly occurs in political science, in a codebook. Actual practices vary somewhat, especially with respect to value labels – they are not always complete. Had Dr. Gibson encountered trouble with the data under more normal scientific conditions, he would have asked Dr. Goldstein directly for additional information.

analyses failed. In other words, the JLG report identifies no mistakes in the application of statistical procedures. Indeed, beyond the data complaints (addressed individually in my point 6 above) the JLG report's critiques tend to reflect a difference in point-of-view on how to categorize certain events that has nothing to do with statistical techniques *per se*.⁴

8.

“Finally, I must note that the authors of the Report apparently were not involved in collecting either the data about the airings (from CMAG) or the data describing the ads (from Professor Goldstein). As with any secondary analysis of data collected by others, this most likely limits the authors' understanding of the nuances and peculiarities of the data base. This is particularly so since this is a demanding data base involving large numbers of cases, multiple units of analysis, various data infirmities (e.g., missing data), and variables based on highly subjective judgments made by undergraduate students. Moreover, I note that one of the authors (Mr. Seltz) seems to have little if any training in statistical analysis, apparently learning whatever skills of data analysis he possesses on-the-job, in the course of preparing this report. Given that this data base is large, complicated, and difficult to analyze, it is extremely worrisome that the results are so heavily dependent upon the limited skills of an author who is a novice analyst.” Pages 5-6.

To the claim that “any secondary analysis of data collected by others ... most likely limits the authors' understanding of the nuances and peculiarities of the data base” it should be noted that Political Scientists commonly conduct “secondary analyses of data collected by others” for reasons of efficiency and effectiveness. Examples are numerous, particularly for scholars who engage in the kind of work that Dr. Gibson does. *See*, for example, Laura L. Vertz, John P. Frendreis, and James L. Gibson, “Nationalization of the Electorate of the United States” *American Political Science Review* 81(3): 961-966, which uses “county-level electoral data for the period 1962-1982 from the historical archive of the Inter-University Consortium for Political and Social Research” (Vertz et al,

⁴ For example, the JLG report discusses at length (pp. 38-43) an important mathematical denominator. This part of the JLG report does not identify a mistakenly applied statistical technique. Instead, Dr. Gibson merely advocates a unique way of counting events that privileges a conclusion favored by the Plaintiffs.

p. 965, fn 3); and James L. Gibson “Political Intolerance and Political Repression During the McCarthy Red Scare” *American Political Science Review* 82(2): 511-530, in which “The source of the opinion data is the Stouffer survey, conducted in 1954” (Gibson, page 517). The Red Scare article won the American Political Science Association’s award for the best paper in that journal that year.

To the claim that variables are “based on highly subjective judgments made by undergraduate students,” I offer two points. First, the data used to defend most of the conclusions in both versions of *Buying Time* and the Goldstein expert report *do not* depend on the data in question – that is coders’ responses to two questions about an advertisement’s “purpose” and its “primary focus.” It does not question the validity of much of the other data. Second, with respect to those two questions, I will demonstrate that the JLG report’s representation of coders’ judgments as “highly subjective” contradicts standard scientific interpretations, including those that Dr. Gibson adopts in his own research. *See* my points 6, 7, and 12 in Section II.b below.

To the claim about Mr. Seltz’s statistical training, it is worth emphasizing that no advanced statistics are used in *Buying Time 1998* or in the JLG report. Almost all of the claims depend on simple multiplication and division. Others are based on simple descriptive statistics and sorting by attributes. Such skills are taught in advanced high school and introductory college courses.

SECTION II. Evaluating “Conclusions-Buying Time 1998.”

In this section, I first address the JLG report’s four conclusions relating to *Buying Time 1998* and then take up in detail the key arguments offered to support those conclusions.

The conclusions in this section are numbered from 1 to 4. I address them in order:

“1. Careful examination of the Brennan Center/Professor Goldstein 1998 data provides little confidence that the data were collected and assembled accurately and reliably. For instance, the data provided for this analysis cannot reproduce the findings reported in the 1998 Report, and several obvious errors exist in the data base.” Page 43.

Most scientists would agree that “careful examination of the data” is not the correct procedure for judging whether the data were “collected and assembled accurately and reliably.” Such judgments are credible only to the extent that they depend on evidence on how the data was collected, which is often housed separate from the data itself. The JLG report describes only a very limited analysis of this separate evidence, much of which is available in Sections E, F, and I of the appendix to Dr. Goldstein’s report.

The claim that “the data provided for this analysis cannot reproduce the findings reported in the 1998 Report” is speculative and should be treated as unreliable. *See* my point 6 in section I of this evaluation regarding the unproven character of the claim “cannot reproduce.”

“2. The methodology of coding the advertisement is deeply flawed.” Page 43.

This claim is too vague to allow a precise response. Doing a search on the word or word-part “flaw” in the parts of the JLG report pertaining to *Buying Time 1998*, I find two non-redundant references to flaws in coding the ads. I will focus on these.

- “If I am correct that the student coders were not trained, then this is a flaw of considerable proportion.” Page 10.
- “...no empirical evidence exists to indicate that the coders’ subjective assessments of these ads were accurate. This is a very serious flaw in the methodology of the study, especially since the crucial variables under consideration require highly subjective judgments.” Page 18.

In both cases, the purported flaws are, in fact, procedures that are entirely consistent with standard scientific practices, including those of Dr. Gibson. *See* my points 6 and 7 of Section II.b.

“3. The practice of engaging in idiosyncratic, standardless, and highly subjective post-hoc alteration of the data base by partisans severely undermines the credibility of the data set and the conclusions drawn therefrom.” Page 43.

The JLG report finds that several versions of the 1998 database are not identical. Much of this aspect of the JLG report, however, itself fails to articulate clear and widely accepted standards for management of a scientific database. Moreover, these sections are littered with speculation and innuendo about other scholars’ motives; the JLG report continually alleges that other scholars’ alleged personal ideological commitments (about which he has very limited information) provide the explanation for their database management decisions.

It is very important to recognize that the changes in question are easy to see – they are transparent. Given the information I have, it appears that Dr. Goldstein has not attempted to hide anything. It is common, when scholars request each other’s databases for the purpose of replicating an analysis, to receive only one version. I seriously doubt whether most scholars could present every version of every data set they has ever used or

even more than one version of many such sets. Dr. Goldstein is not acting like someone with something to hide.

Large academic databases change for legitimate reasons, so the mere existence of the relatively small changes cited in the JLG report provide no basis to negate the project's credibility (*see* my point 6 in Section I). What matters is why and how the changes were made. On this matter, the questions raised about the *Buying Time*/Goldstein practices are legitimate, but the JLG report's answers to them are of varying and questionable credibility -- as I have argued in point 1 of Section I and as I demonstrate throughout Section II.

“4. The 1998 data provided by the Brennan Center/Professor Goldstein not only do not support the conclusions announced in the Report, but they in fact lead to exactly the opposite conclusion: The vast majority of ads broadcast within 60 days of the 1998 election, which depicted one candidate or the other, had as their primary focus issue advocacy.”

I revisit this claim, addressing it directly and thoroughly at the conclusion of Section II.d.

II.a. Evaluation of “The 7% Figure in *Buying Time 1998*-Deconstructing Figure 4.22 and Page 109.”

“For the purposes of my analysis, perhaps the most relevant section of *Buying Time 1998* is Figure 4.22 (page 110) and the text purporting to interpret this figure (page 109).”⁵ [Pp. 35-36]

Pages 35-43 of the JLG report focuses on how the proposed regulatory change will affect ad sponsoring interest groups. Some of the critique in this section repeats claims registered elsewhere about the quality of the data, and I will not repeat them here.

⁵ It is interesting to note that there is no attempt to qualify the word “purposes” in this statement. If we follow the standard articulated on page 34 of the JLG report, which chides Goldstein for not using the word “primary” to qualify the word “purpose” in Question 6, we cannot know whether the purposes in question are primary, secondary, or otherwise ranked. Can we agree that this unconditional use of the term in such contexts is conventionally understood to mean “primary?”

A focal statistic is 7%. *Buying Time 1998* claims that this is the percentage of “genuine issue” advertisements that would be covered by the regulatory change. The JLG report argues that the figure should be higher, perhaps as high as 64%. In evaluating this section, I first address the debate over which cases (airings) should be included in the denominator of the fractions that produce the debated figures. I then analyze both positions in this debate in a way that clarifies their implications for questions about the likely future impact of the proposed regulations.

The *Buying Time* and JLG estimates are determined in large part by different assumptions about appropriate denominators. In other words, they take different points-of-view on which airings should be included when assessing the regulation’s impact. The *Buying Time 1998* denominator counts all genuine issue ads (that is, all group sponsored ads found by the coders to have purpose of providing information). Of this denominator the JLG report says on page 38: “Of course, using a denominator of all issue ads broadcast in 1998 for these calculations is arbitrary and makes little sense.” The question I pose in what follows is: By what scientific standard is a denominator not arbitrary, in Gibson’s sense?

The JLG report then suggests other denominators, favoring a particular one in the end. Instead of counting all issue ads aired in 1998, Gibson proposes counting only ads that feature a candidate and are aired within 60 days of an election. He justifies this choice by positing that the overriding question before us is: What percentage of the time would the assumption that such ads constitute electioneering be in error? At the same time he claims that

“The denominator chosen by the authors of *Buying Time 1998* relates to an entirely different, and virtually meaningless question: What percentage of the

total ads run throughout the year that mentioned a candidate by name and were coded as providing information or urging action appeared *within* 60 days of the election, rather than *earlier than* 60 days before the election?” Page 39.

We can agree that the two questions are different. But the question prompting the *Buying Time 1998* denominator is reasonable.

To see why, note first that the scientific standard by which the *Buying Time* question is judged to be “virtually meaningless” is neither apparent nor described. Is the question meaningless or isn’t it? If so, what is the concrete and widely accepted scientific standard upon which we should base our judgment of a question’s “meaning?” The JLG report offers no answers to these questions. Indeed, the report contains no evidence that the “virtually meaningless” standard it wishes to impose is the product of scientific reasoning rather than just result-oriented advocacy.

Second, I will not pretend to know how Congress or the litigants in this case define the “meaning” of questions. But I do feel qualified to raise the following question: Is it not possible that a reasonable person would agree that both questions have merit if we want to understand the extent of the new regulation’s restrictiveness? In my opinion, the answer is yes. If I were asked to assess the proposed regulation’s restrictiveness, the JLG report’s fraction could provide information about the impact during a particular time period, while *Buying Time 1998*’s fraction could provide a better measure of the regulation’s impact on issue advocacy more generally. The groups who fund these ads may be able to exercise discretion in choosing *when* they air advertisements. So, if I observed that all such groups aired all of their ads within 60 days, I might draw a different conclusion about the regulation’s restrictiveness than if I observed a more constant year-round flow. Page 41 of the JLG report states “I can see no justification for

making the denominator equal to all issue ads aired in 1998.” I have just presented one with little effort, and there are many others.

Gibson calls the *Buying Time* denominator “arbitrary” because it focuses on a period of time other than “60 days before an election.” That Gibson wants us to focus exclusively on this period, as the basis for drawing conclusions about the likely impact of the proposed regulations, is not a scientific argument – it does not follow logically from a transparent and widely accepted set of first principles. Therefore, I do not agree that *Buying Time 1998*’s denominator is meaningless or arbitrary (as Gibson uses the term), but if it is, then Gibson’s preferred denominators are also meaningless and arbitrary.

I now turn to a different topic. The Gibson and *Buying Time* denominators are reasonable conceptualizations of the question about how the proposed regulations will affect groups in the present and future if groups act exactly as they did in the past. If, however, we want to evaluate the regulations’ likely future impact we should consider the possibility that groups will adapt to the new regulations in different ways.

If, by analogy, we want to understand the effect of a proposed tax on red cars, we must account for the possibility that some people who drive or would have driven red cars without the tax may adapt to the new incentives by switching to cars of other colors. Continuing the analogy, both positions in the denominator debate are based on an assumption equivalent to “Everyone who drove red cars in the past will continue to do so in the future.”

Now consider that groups may adapt in ways that include the following:

- Running an ad within 60 days of a general election that mentions or pictures a candidate and having to submit to more rigorous disclosure requirements and restrictions on funding sources,

- Running the same ad minus the candidate reference within 60 days of a general election and not having to submit to more rigorous disclosure requirements and restrictions on funding sources, and
- Running the same ad either more than 61 days before and election or more than one day after the election and not having to submit to more rigorous disclosure requirements and restrictions on funding sources

It is speculative, at best, to assume that all affected groups will choose option 1.

But literal interpretations of the claims of both sides in the “denominator debate” *as predictions about the regulations’ likely future impact* are undermined if not all groups make this choice. To the extent that affected groups are able to choose the other strategies listed above, *both estimates in the denomination debate may exaggerate the extent to which this aspect of the new regulation will restrict the groups’ abilities to express themselves in the future.* The extent of these exaggerations depends on when, and how many, groups would choose strategies such as the second or third item listed above. To the extent that we agree that such groups will adapt in various ways, the credibility of high-percentage estimates of the likely future impact of the proposed regulations on interest groups is severely undermined.

II.b. Evaluation of “The Sources of the Data for *Buying Time 1998.*” (Pp 6-19)

This section of the JLG report focuses on data quality. It begins with a series of bulleted claims. In what follows, I draw attention to items that are demonstrably false or based on speculation.

1.

“CMAG does not monitor all broadcasts; instead, its coverage is limited to only 75 media markets (out of 210 such markets in the country). Given this limitation, one must be particularly careful about generalizing the findings of this study to all political communications.” Page 7.

To the claim that “one must careful about generalizing,” I note that Dr. Goldstein’s expert report (page 5) adds the detail that the 75 markets in question are the “top 75 markets

(compromising approximately 80 percent of the nation's population)". Scientists, be they working in fields ranging from medicine to politics, draw general inferences from sampled populations. Sampling is a general scientific practice undertaken to increase research efficiency. It occurs regularly in the medical and natural sciences as well as the social sciences. Consider, for example, some of Dr. Gibson's own work. His article "Social Networks, Civil Society, and the Prospects for Consolidating Russia's Democratic Transition" *American Journal of Political Science* 45 (1) 51-69, uses a sample of 2059 respondents⁶ to draw the conclusion that "Russians are not atomized and socially isolated, and that aspect of Russian political culture has important consequences" (p. 51). In other words, a sample of 2059 individuals, or about less than 1 percent of the entire Russian population, is used to draw general conclusions about the whole.⁷

It is regular scientific practice to draw general inferences from limited data. The credibility of general claims depends on the procedures by which the cases were selected – the procedures should be transparent and replicable. The cited claim, however, does not criticize the procedure, only the fact that not all markets are monitored. The JLG report presents no evidence or reason to believe that that including advertisements from the markets not covered would change *Buying Time's* results.

⁶ Gibson page 55: "Interviews were attempted with 2,442 respondents, with a resulting response rate of 84 percent." Gibson page 55 FN 10 then explains, "Of the 383 interviews not completed..." 2442-383=2059.

⁷ "According to preliminary forecasts, the average number of citizens permanently living in Russia was 144.5m in 2001 and 143.6m in 2002." Source, Johnson's Russia List, Report 6404 (www.cdi.org/russia/johnson/6404-5.cfm.)

2.

“Complete visuals are not provided, meaning that some information depicted in the ad may be excluded and not captured, thus compromising the coders’ ability to analyze the complete content of the ads.” Pages 7-8.

As in my point 1 of this section, the suggestion here is that sampling procedures are inadequate. Here, the CMAG program obtains a sample of all available images from ads. As stated above, sampling is standard practice in science, including Dr. Gibson’s research. A critical question, therefore, is whether the sampling procedure is transparent and replicable. The detail in which many of the Goldstein documents and the JLG report describe CMAG procedures is evidence that its collection procedure is well documented, and easy to replicate with the right equipment.

To justify with more than speculation the conclusion that coders’ abilities were compromised, the JLG report must provide evidence that CMAG’s sampling procedure impeded coders. The JLG report contains no evidence to this effect. Therefore, the JLG report’s claims about the extent of coder’s compromised abilities are speculative.

3.

“Crucial information was missing in the CMAG data, as for example, in more than a quarter of the broadcast data, the identity of the sponsor was not included in the data (*Buying Time 1998*, 8). Without accurate information about the sponsor of any given ad, the fundamental distinction drawn in *Buying Time 1998* among ads sponsored by candidates, parties, and groups becomes suspect. It appears that in some instances the ad sponsor was identified through methods independent of the CMAG data collection and analysis, although these methods are not thoroughly documented in the Report (e.g., *Buying Time 1998*, p. 8).” Page 8.

The claim that “crucial information was missing in the CMAG data” does not imply that the information was missing from the data analyzed or that any data beyond that explicitly mentioned is missing. Moreover, of the data the JLG report claims to be missing, it also reports that some of the data– it does not say how much – was filled in. It

is common practice for scholars to add new information to datasets. It happens in every large-scale database with which I am familiar. What matters in such practices is whether the changes are conducted with integrity (e.g., in a transparent and replicable manner). Therefore, the most relevant question before us is: Is the added information accurate?

In the case described, the answer is discoverable. The missing data refers to publicly available information about who sponsored an ad. It is, therefore a straightforward – though admittedly tedious – exercise to systematically compare the added data in the *Buying Time*/Goldstein database – against available records. Without such evidence, Dr. Gibson’s criticism is pure speculation.

To the claim that CMAG’s methods “are not thoroughly documented in the Report,” I ask, “What is the standard for “thorough” documentation?” *Buying Time 1998* describes the methods as follows “We were able to group them to our own satisfaction using CMAG’s original coding (which accurately provides the sponsor of the ad in well over 95 percent of cases), examining the content of the ad, and, in a few cases, by phoning television stations” (p. 8). I concur that the *Buying Time* authors could have given a more detailed description of how they acquired this supplementary data, but such brief descriptions of supplementary data are not uncommon in social science.

Consider, for example, Laura L. Vertz, John P. Frendreis, and James L. Gibson, “Nationalization of the Electorate of the United States” *American Political Science Review* 81(3): 961-966. Its main source of data is “county-level electoral data for the period 1962-1982 from the historical archive of the Inter-University Consortium for Political and Social Research.” (page 965, fn 3). However, the authors of that paper also wanted to draw a conclusion about 1984, but data on this election was not included in

their ICPSR data. In other words, *in some instances, relevant information was identified through methods independent of the ICPSR data collection and analysis*. Moreover, the article's entire attempt to document the methods of incorporating this additional data set is as follows: "The 1984 election returns were collected directly from each state (generally from the office of secretary of state)." (page 965, fn 3). By what standard is this one sentence explanation more thorough documentation than that offered on page 8 of *Buying Time 1998*? The JLG report offers no scientific basis for questioning one of these practices while accepting the other without question.

Moreover, this claim neglects the fact that a script of the text read during the advertisement accompanied each storyboard, which can provide additional information. The JLG report does not question the reliability of the scripts.

4.

"Finally, no evidence has ever been adduced documenting the accuracy of the CMAG data. Indeed, insofar as the 1998 data were collected and assembled through methods similar to those apparently used in 2000, the lack of accuracy of the data has been documented." Pages 8-9.

Footnote 6, following this sentence:

"No detailed description of the 1998 CMAG data has been published or otherwise made available to me. However, in Appendix E to Professor Goldstein's Report for this litigation, he discusses various problems in 2000 with the CMAG data and technology. The data difficulties seem to be numerous and formidable. Given the general trend for technology to improve over time, it is difficult for me to imagine that the 1998 CMAG data are any less problematic than the 2000 data."

On the point that "no evidence has ever been adduced documenting the accuracy of the CMAG data," I make two arguments.

First, the accuracy of many scientific data sets is not directly tested. Dr. Gibson, in his many articles using his – or other – data, does not insist that others adduce its accuracy before he bases scholarly claims on it. And he is not expected to. Instead, he is

expected to describe the method by which he collected the data in a way that allows others to do adduce its accuracy – perhaps by attempting to replicate his procedures. In many cases, he is also expected to make the data publicly available to be checked by others.

Second, and perhaps more importantly, such a test is easy to do (e.g., on a limited random sample of ads). Why speculate about the extent of accuracy – as the JLG report does – when it would have been very easy to replicate this aspect of the studies on a random sample of ads?

The point that “the data difficulties seem to be numerous and formidable” is simply vague. If the JLG report is trying to imply that the data have attributes that prevent any scholar in any circumstances from making scientific contributions, the JLG report does not come close to proving this point.

5.

“it is unclear how the students were recruited, what expertise they had prior to being employed for the project, whether the students had been exposed to Professor Goldstein’s classes, whether the students had ideological and/or policy commitments to a particular outcome in the project, etc...The absence of answers to these questions raises questions about the overall accuracy of the data collection process.” Page 9.

My evaluation of this claim has two parts:

First, to the claim that the absence of information about the coders “raises questions about the overall accuracy of the data collection process” it should be noted that the JLG report does not challenge most of the data collected. It focuses on only two of the many questions coders were asked. Even if the JLG report’s contentions about these two questions were valid, and they are not as I show below, to say that these

contentions affect the “overall accuracy of the data collection process” is simply not supported by the evidence that the report introduces.

Second, the JLG report cites no evidence about the means by which the students were recruited, their prior exposure to Goldstein’s ideas, or their own policy commitments. If we did have such evidence, we would want to pair it with information about the procedures by which they made their coding decisions because what they do is much more relevant than who they are. In other words, the coders’ background information alone is not sufficient to make their participation grounds for labeling the study non-scientific (*see* my point 1 of Section I). The JLG report is not based on such information. Moreover, and as I demonstrate in my point 6 of Section II.c below, the JLG report ultimately bases its explanation of coders’ choices on very strange speculation that has no apparent scientific reference.

To support the claim that these students acted in a way that damages the data’s credibility, one can speculate or one can attempt to replicate the procedure with coders who are less likely to have such attributes. In this case, such a replication would have been relatively simple to conduct (for instance, with a sample of the storyboards), and would have allowed the JLG report to rely less on speculation when alleging that measurable attributes of Goldstein’s coders affected the data collection or analysis.

6.

“If I am correct that the student coders were not trained, then this is a flaw of considerable proportion. Not only are undergraduates at Arizona State University (or, in the case of the 2000 study, the University of Wisconsin) indisputably not a representative sample of the “average viewer,” in the absence of training, the students were apparently free to exercise unstructured discretion in coding the ads. Without instruction and guidelines for what constitutes the difference between “providing information or urging action” and “generate support/opposition for candidate” — without training, practice coding, and discussion of coding rules based on the results of the practice coding — I do not

believe that undergraduate students coders can make accurate assessments on highly subjective characteristics of these ads. Page 10.

The students are explicitly asked for opinions. The three words at the beginning of question 6, “In your opinion” could not make this fact any clearer.

Are opinions objective or subjective? In thinking about the answer to this question, it is worth noting that thousands of scientists in several disciplines use interview techniques to measure opinions (some refer to these objects as attitudes). There is even a field called “public opinion” with which Dr. Gibson is often associated. A fundamental premise of such work is that opinions exist and can be measured, aggregated and compared. If responses to opinion questions were generally regarded as entirely subjective, not only would operations such as aggregation and comparison be difficult to conduct or defend, but large, diverse scientific communities would have no common basis for developing best practices for analyzing opinions statistically. But aggregation and statistical comparison is a regular phenomenon and broad communities agree on best practices concerning the measurement of opinions.

It is, in fact, standard practice in Political Science, social psychology, and in leading economics surveys to treat self-reports of opinions as the basis for making scientific claims about generalized reactions to social phenomena. Many of Dr. Gibson’s publications, and almost all of his most famous ones, rely on such reports being treated as credible. In both instances, these scholars would treat highly subjective data very differently.

The claim also focuses on the students, raising the question about their competence to complete the task presented to them. On this matter as well, the JLG report makes claims that are inconsistent with standard scientific practices, including

those of Dr. Gibson. Indeed, Dr. Gibson bases many important scientific claims on surveys of random Russians and Africans, most of whom have far less formal education than the undergraduates used in Goldstein's procedures. Are we to deem Dr. Gibson's work credible only if the Russians and Africans are offered instructions and guidance for what constitutes the difference between the categories used in his opinion questions? The scientific community has not chosen to view Dr. Gibson's work in this light, nor have I. If we apply the same standard to Dr. Goldstein's work, we must also judge his coders as credible to answer questions about their own opinions.

The claim that Dr. Goldstein's undergraduates do not form a representative sample of the "average viewer" may or may not be true. However, when the JLG report claims that these students are "indisputably" not representative, it speculates. The JLG report presents no data comparing the students' attributes with those of Gibson's "average viewer" (a label that is otherwise undefined in the JLG report). Moreover, other parts of the JLG report offer an explanation of the coders' psychology that, if we were to accept it, requires that they not be all that different than the rest of the population. *See* my point 6 in section II.c.

But suppose, for a moment, that we elevate this speculation to the status of acceptable evidence. Would this threaten the quality of the data? It would, but only if we had evidence that the way in which the undergraduates were unrepresentative caused *Buying Time's* claims to differ from what a representative population would have produced. The JLG report presents no such evidence.

In sum, the claim that the coders' undergraduate status and lack of training constitute a "flaw of considerable proportion" is based on speculation about the

undergraduates' skills and an inconsistent application of scientific interpretations of opinion questions. Here too, I find it surprising that the JLG report describes no attempts by Plaintiffs or their experts to replicate this aspect of the procedure – complete with the instructions that they might want coders to receive. An experiment, for example, where one set of coders is treated as Dr. Goldstein did and where another set receives the instructions that the JLG report desires would provide direct evidence about the extent to which varying instructions yield different responses. They did not have to speculate on this matter.

Moreover, it should be noted that in the field of psychology, and in the subfields of experimental economics and political psychology, important discoveries about mental states such as attitudes are often generated from studies that ask undergraduates to answer opinion questions after viewing paper-based stimuli. This practice has wide acceptance in social science and is the source of many important and socially valuable discoveries.

7.

“other crucial attributes are far from being objective characteristics of the ads; instead, they are highly subjective and judgmental. For instance, consider Question 6 (from page 193, *Buying Time 1998*, emphasis in the original):
6. In your opinion, is the purpose of this ad to provide information about or urge action on a bill or issue, or is it to generate support or opposition for a **particular candidate**?

1. Provide information or urge action
(If so, skip to Question #19)
2. Generate support/opposition for candidate
3. Unsure/unclear

To answer this question requires a large number of subjective assessments. First, some judgment must be made about *whose* purpose is under consideration here. For some ads, the sponsor is readily apparent, but for others it is not, so the coders were often faced with the difficult and ambiguous task of considering to whom or what to attribute “purpose.”” (page 12).

While the claims in the *Buying Time* reports rely on data from many different sources and questions, Question 6 is the prime target of the JLG report.

As demonstrated in my point 6 of this section, the idea that this question requires subjective assessments is inconsistent with standard practice in interview-based scientific procedures and in Dr. Gibson's own research. Again, it is worth noting that this question begins with the words "In your opinion." The question is asking about the coder's opinion – a mental state. Numerous political scientists, including Dr. Gibson, regularly base their scholarly reputations on the assumption that answers to these questions can be used as the basis for characterizing these mental states.

To clarify my objection to the JLG report's characterization of Question 6, I introduce a simple analogy:

You are standing in front of a tree. There are many questions I could ask about your situation, including the following:

Question type 1: What is the height of the tree?

Question type 2: In your opinion, how tall is the tree?

In the first case, I am asking a question about a physical attribute of the tree. In the second case, I am asking about an impression, a mental state. The two questions need not seek the same information.

It may be important to us to know what people think about the tree. If so, actually measuring the tree with a ruler does not provide the information in which we are interested (i.e., if the tree is 50 feet tall, but it looks 70 feet tall to you – the latter is what we care about). Much of the JLG report's discussion is analogous to insisting that if we want to know how the tree is perceived, we must get out the ruler and measure the tree. The error in this argument is that we can be interested in opinions. And the question, very explicitly, asks for an "opinion."

The practice of treating answers to opinion questions as objective phenomena is common in science. Consider, for example, the article “Social Identities and Political Intolerance: Linkages within the South African Mass Public” by James L. Gibson and Amanda Gouws, which appears in the April 2000 edition of the *American Journal of Political Science* on pages 278 to 292. The article’s main conclusion, on page 278, is that “strong group identities pose a difficult challenge for the consolidation of democracy in South Africa.” Evidence for this claim emanates from a survey run by the authors in the fall of 1996. The critical variable in the study on group identity (itself a mental state) is drawn from responses to the following questions (p. 281, fn 6):

“The respondents were asked: “People see themselves in many different ways. Using this list, which of these best describes you? Please take a moment to look at all of the terms on this list.” The respondents were then asked: “Still looking at the card, do you think of yourself in any of the other terms as well?” The next question was “Still looking at the card which would you say most strongly does NOT describe you.”

Gibson and Gouws use answers to these questions to measure respondents’ group identities. Of course, others could chide them for not seeking data on the respondents’ actual identities (after all one can “identify with” Europeans while not actually being one or being only partially European). However, I am not aware of other scientists who have chided them on this point, which is not surprising given that so many understand that what the authors really want to know is the respondents’ impressions of their group identities.

The JLG report (p.12) calls Goldstein’s coding process “unreliable” because answers to a question of type 2 (in the tree analogy) do not provide the same information as answers to a question of type 1. Using this standard, there is no basis for differentiating the Gibson and Gouws study from Goldstein’s. However, I disagree that either is

unreliable because in both cases, relevant audiences are interested in the mental state (opinion) left by the objects of interest. In other words, it is wrong to judge answers to questions of type 2 unreliable because they provide different information than responses to questions of type 1.

As a consequence, the claim that coders' views of "whom or what to attribute purpose" are not directly relevant. What matters here are the coders' opinions of the ad.

8.

"This ad is especially interesting for several reasons. First, it seems obvious that the central focus of the ad is on the policy issue of whether to ban partial birth abortions." Page 13.

This quote represents a judgment call by Dr. Gibson. Indeed, the scientific basis for the claim's "seems obvious" component is not apparent. We can treat this claim as an opinion of Dr. Gibson or of a statement of fact about the ad. As an opinion, the claim belongs to Dr. Gibson and *we need not share it*. As a statement of fact – about a physical attribute of the advertisement – *it may be a complete conjecture*. Either way, the JLG report provides no scientific reason for believing the claim.

9.

"If one had to speculate about the motives behind the ad, one might reasonably judge that the ad sought to capitalize on the widely publicized incident in Delaware as a means of generating support for a congressional ban on partial birth abortions. One might also reasonably conclude that one purpose of the ad was to elicit support for the National Pro-Life Alliance. The most reasonable overall assessment of this ad is that it is an example of issue advocacy by an interest group." Pages 13-14.

What is the basis of an "overall assessment?" What makes the assessment offered in the JLG report the "most reasonable?" Is there a "reasonability" test whose results can, using a transparent process, be applied to an "overall assessment" scale?

To these questions, the JLG report offers no clear answers, no tests, and no transparent processes. Confusing matters further is the fact that in the two sentences before this, Gibson himself first provides two very different assessments, but then – without explanation or clear basis for judgment – proclaims that his final conjecture is “most reasonable.” The JLG report provides no standard for ranking these and other assessments; therefore, its ultimate recommendation is, at best, a matter of opinion.

10.

The next section of the paper contains the most serious charge that the JLG report levels against Dr. Goldstein. The charge regards Dr. Goldstein’s changes to the data set. Throughout this passage, questions are raised about Dr. Goldstein’s motives. For example, on page 10, we are told that:

“the motives for making such changes are important. To the extent that one only examines codings that undermine the preferred conclusions, and one does not examine codings supporting the preferred conclusion, asymmetrical bias is introduced in the data set. Under such conditions, confidence in the ability of the data set to produce useful results and conclusions plummets.”

Throughout this section, numerous allusions are made about the extent of these changes. The JLG report does identify differences in different versions of the data set, but the report makes no showing that any change was made for non-scientific reasons.

In one passage, however, Dr. Gibson replicates Dr. Goldstein’s procedure. The conclusion he draws is that

“the changes in the data base are entirely asymmetrical: In not a single instance in these storyboards was a change made on an ad originally coded as having candidate support or opposition as its “purpose.” Page 15-16.

How should we judge the credibility of this claim? Transparency and replicability are a good start, but here Dr. Gibson does not fare well. He bases the claim in question on

an examination of only 25 storyboards. We are not told why 25 storyboards and not more or less were examined. We are not told how these 25 cases were selected. Were they selected at random, were they given to Dr. Gibson by counsel (as is true in an analogous replication attempt reported on page 62 of the report and addressed in my point 3 of Section III), or were they chosen by some other procedure? The JLG report provides no information. Indeed, the JLG report provides no basis for rejecting the hypothesis that the “asymmetry” claim is an artifact of the cases being selected in a way that is biased towards this generating this particular result. Without such evidence the claim in question is not credible.

11.

“One common meaning of reliability is the ability to replicate or reproduce results. Consequently, a proper methodology for assessing inter-coder reliability would involve the following procedures....” Page 16.

This passage argues that a “proper methodology” has three enumerated necessary conditions. While I concur that the listed procedures could be used, the JLG report provides no evidence that the particular sequence of actions named is widely accepted or exclusive in the rating assigned to it – “proper.” My specific contention is with the idea that “proper” procedure precludes expert or experienced coders for the subsequent coding in a reliability assessment. At a minimum, we should be furnished with a quotation from a widely cited text on research design. Moreover, it is sometimes the case that scientists want to know how people with different backgrounds perceive a given stimulus. If experts and novices constitute two such groups of people, then “the use of “expert” or highly experienced coders for the subsequent coding” (page 16) would be of interest. The notion that a “proper methodology” “precludes” such practices is incorrect.

12.

“Social scientists care about more than reliability; they also care dearly about validity. ... It is possible, for instance, that coders could reliably but invalidly code an attribute of the ads. For instance, assume that the coders are asked to judge whether the “purpose” of an ad is to “generate support/opposition for candidate” [sic — from the 1998 coding form]. Assume further that this is a highly subjective judgment ... unlike academic research based on subjective coding, no empirical evidence exists to indicate that the coders’ subjective assessments of these ads were accurate. This is a very serious flaw in the methodology of the study, especially since the crucial variables under consideration require highly subjective judgments.” (pages 17-18).

Again, the question to which this passage refers begins “In your opinion.” It seeks to measure an opinion. The argument here is that the subjective nature of the question threatens the data’s reliability. See my point 7 of this section to see why this claim is inconsistent with standard scientific practices.

The final paragraph of this section of the JLG report, which continues through page 19, treats Question 6 as if the passage “In your opinion” was not part of it. This fact undermines all of the claims in the section’s final paragraph. The paragraph argues that coders are unlikely to have known enough about an advertisement’s sponsor or featured candidate to answer Question 6 in a reliable way. Again, this is a red herring.

Regarding the information in question, we can agree that for other findings in the *Buying Time* report, knowing whether or not a person mentioned or depicted in an ad is a candidate for office is relevant (e.g., In findings that answer the question “What percentage of issue ads featured or mentioned a candidate.) However, *when the object we are studying is an opinion, as is the case in Question 6, then information about the politicians is not directly relevant.* A coder can reasonably opine that an ad’s purpose is electioneering without being certain of the status of a politician named in the ad.

II.c. Evaluation of claims about the analysis in *Buying Time 1998*. (Pp 6-19)

1.

The following section, entitled “Analysis of the 1998 Data” (pp. 19-23) characterizes the data set as a whole, as opposed to any particular part. On the whole, it is not critical of *Buying Time* or Dr. Goldstein. While this section of the JLG report notes that the merging of the database’s components “was not executed flawlessly” (p. 20), footnote 22 is the only place in it that describes an actual flaw. In this footnote, the JLG report finds some confusion in the way that two variables are labeled, but concludes, “In most instances, the errors seem to be harmless...” In addition, the footnote describes one “more serious” mistake – two different ads have the same identification number. The JLG report, however, provides with no evidence about the extent to which this error impacts the results.

2.

The next section, entitled “General Comments on the Accuracy of the 1998 Data” documents instances where Dr. Gibson failed to replicate figures in *Buying Time*.

The first concrete demonstrations are offered on page 24. The demonstrated discrepancies here are small – the difference between the figures of the JLG report and *Buying Time* are in the 1-1.5% percent range. The JLG report provides no evidence that such changes affect any of *Buying Time*’s major claims.

3.

“Many variables in the data set are not properly documented; indeed, many variables have no documentation whatsoever.” Page 25.

Improper documentation is claimed but the JLG report’s criteria for “proper documentation” are not explained. In my point 6 of Section I, I convey a set of best practices. They involve variable labels, value labels, and, often, codebooks. The JLG

report cites relatively few instances of documentation that it judges confusing. Moreover, we do not know whether and how Dr. Gibson attempted to obtain such documentation. Lacking any evidence of a comprehensive search for documentation, we must treat the claim “many variables have no documentation whatsoever” as premature and speculative.

4.

The conclusion of this section is as follows: “The most important implication is simple: The data provided cannot be used to replicate the findings of *Buying Time 1998*.” (page 26). On the credibility of this claim, see my point 4 of Section I. In sum, evidence of a failed replication attempt of questionable quality is not sufficient to conclude that the data in Dr. Gibson’s possession “cannot be used to replicate the findings of *Buying Time 1998*.”

5.

On page 30, the JLG report claims

“the confusion in the instructions regarding Questions 7 through 18 may have introduced a degree of bias into how the students coded Question 6 by suggesting that any advertisement that included the name of a candidate should be coded as having a purpose of promoting or opposing a candidate.”

I know of no established theory of responses to opinion questions, nor does the JLG report cite one, that yields this unique solution to the question of what voters were thinking. It is a guess, the basis of which is not apparent to me despite my experience in this area. Moreover, this is a claim that would have been easy to evaluate through replication. Using the storyboards in Plaintiff’s possession, one could have shown one set of coders with the instructions regarding questions 7 through 18 visible while showing another set of coders Question 6 without the instructions or subsequent questions. Holding all else constant, and replicating the other Goldstein procedures, the experiment

would yield direct evidence about whether these instructions had the alleged impact. But the JLG report offers no direct evidence of such an effect. It only speculates as to what might have happened if the questionnaire was differently organized and its speculation, in my view, is highly questionable.

6.

“It seems highly likely to me that the student coders coded these three questions (6, 7, and 8) virtually simultaneously: A candidate (or what the coder thought was a candidate) was observed in the ad, and then Question 6 was coded as electioneering (in part because the coders knew that the presence of a candidate was not coded if Question 6 was coded as providing information), and then the student made the determination of whether the candidate was “the favored candidate” (Question 7) or the “favored candidate’s opponent” (Question 8).” (P. 30).

This claim is a wild guess. It has no apparent scientific basis, which matters because the claim in question includes a very detailed statement about an exact sequence in coders’ cognitive processes. Validating such a claim requires non-trivial amounts of psychological theory and experimentation – of which the JLG report presents neither. Moreover, I am quite familiar with the current scientific literature on the psychology of responses to opinion questions and this claim follows nowhere from it.

Moreover, the conclusion that follows immediately afterwards

“Thus, the entire relationship — empirical and logical — between Questions 6 and Questions 7, and 8 renders the data set of little utility for answering important questions about these ads and airings”

is unsupported by logic or empirical evidence.

Moreover, the fact that Dr. Gibson believes that he can assert this generalization, without specialized knowledge of the unique cognitive attributes of the undergraduate coders, implies *a dispute* between this characterization of coder psychology and his page 10 claim that these coders are “indisputably not a representative sample.” If the coders

are as different as the page 10 quote implies, what is the basis for Dr. Gibson's confidence in his page 30 claim? If it is not specialized knowledge of coder psychology, then it must be the assumption that the way that coders approach these questions *is not too different from the general population*.

II. d. Evaluation of "The Focus of the Ads"

This section supports three conclusions, which are stated on page 35. They are:

- "1) The coding in Question 6 is deeply flawed.
- 2) When Question 6 and Question 22 clash (i.e., the coded attributes differ), the coding of Question 22 should be considered more valid and reliable.
- 3) According to the coding, the vast and overwhelming majority of ads said to be examples of illegitimate electioneering (by virtue of promoting candidates) in fact were judged by their own coders to have "policy matters" as their "primary focus."

In what follows, I will show that the first two claims do not follow from the evidence offered. The third conclusion is true, but it does not support Dr. Gibson's assertion that ads that have policy matters as their principal focus cannot also be regarded as having been broadcast to serve electioneering purposes.

In this section, two questions and coders' responses to them are compared. One question is Question 6, a focus of discussion in my point 7 of section II.b.

6. In your opinion, is the purpose of this ad to provide information about or urge action on a bill or issue, or is it to generate support or opposition for a **particular candidate**?

1. Provide information or urge action
(If so, skip to Question #19)
2. Generate support/opposition for candidate
3. Unsure/unclear

The second is Question 22, which reads as follows:

22. In your judgment, is the primary focus of this ad on the personal characteristics of either candidate or on policy matters?

1. Personal characteristics
2. Policy matters
3. Both
4. Neither

The JLG report notes that 98.1% of the 6896 airings that appeared within 60 days of an election and mentioned or depicted a candidate were coded as having a “primary focus” on “policy matters.” The JLG report then questions how coders could have opined an ad’s purpose to be electioneering at roughly the same time that they judged its primary focus to be policy. It concludes that this pair of responses is due to flaws in Question 6 and the superiority of Question 22.

1.

I will turn to the comparative attributes of these questions in a moment. But first, I ask you to consider whether flaws in question 6 are a necessary condition for someone perceiving an ad’s purpose to be electioneering while judging its primary focus to be policy. In so doing, I offer an analogy.

Watch a professional football game on television this weekend. These programs contain many commercials. Quite a few are sponsored by brewing companies.

If you watch these advertisements, you will notice the following: Many of them devote almost all of their time to depicting men (usually in their early to mid twenties) engaged in a range of activities that we can call “wild nights out.” What is the purpose of this ad? Perhaps we could agree that the purpose of the ad is to sell a particular brand of beer. But the ads tend not to focus on beer. Gone, for the most part, are the days when images of brewers are matched with voiceovers about “choice hops” and “master brewers.” I concede that these ads exist, but again we are watching a football game. Therefore, the ad combines some form of bathroom humor or rock music with dancing,

driving fast, laughing with friends, and stereotypes of women that are, for lack of a better term, colorful. Is it unreasonable for an audience member, even an undergraduate, to perceive that the purpose of the ad is to get them to buy Coors, but to judge its primary focus as wild times and, for lack of a better term, girls? A viewer's perception of an ad's purpose and their judgment of its primary focus need not be the same.

This point is important because the JLG report asks us to assume that when a coder's opinion of an ad's purpose (Question 6) and their judgment of its primary focus (Question 22) "clash (i.e., the coded attributes differ), the coding of Question 22 should be considered more valid and reliable" (page 35). But this assumption implies that people cannot distinguish what they think an ad is trying to do to them from the means by which it is trying to do it. The JLG report offers no justification for this assumption and I doubt that one can be produced.

Having demonstrated that the second of the three listed conclusions does not follow from the evidence presented, I turn to claims about flaws in Question 6.

2.

The JLG report raises the possibility that the difference in responses to Questions 6 and 22 is due to the fact that question 6 does not include a response such as "both" or "neither." In particular, it claims that

"In the event of mixed content, the undergraduate coders were forced to make a dichotomous judgment about the ad's purpose." (pp. 33-34).

This claim is false. As is readily apparent, Question 6 also offered the third⁸ option of "unsure/unclear." For more on why the JLG report fails to show that response options

⁸ Hence negating the "dichotomy."

such as “both” or “neither” affect the *Buying Time*/Goldstein results, see my point 4 of this section.

3. The JLG report also alleges that Question 6

“provides no guidance whatsoever as to how to code mixed-content ads.”

This claim is true. A relevant question, however, is “Does this fact affect the results?” About this matter, the JLG report only speculates. Indeed, the JLG report provides no direct evidence that any other way of asking this question would generate different or “better” results. In addition, if instructions were given to the coders, it could be argued that they would bias the coders towards particular response options, which would make the questionnaire less effective in recording viewer impressions of the ads.

Finally, as I have noted above, it would not have been difficult to replicate this study – in light of the groundwork done by Dr. Goldstein. The experiment would give one group the questions as written and another group the questions rewritten as the JLG report suggests. If everything else about the presentation is held constant and adheres as closely as possible to Dr. Goldstein’s procedures, then the experiment would provide direct evidence about how changing the questions changes the answers. The choice to speculate rather than replicate in this circumstance is surprising.

4.

“In contrast, the construction and coding of Question 22 is an improvement over Question 6 in a number of respects. First, the question allows the coders options of “3. Both” and “4. Neither”. Thus, the problem of forcing a choice between different parts of the manifest content of the ad is resolved by allowing a coding of “mixed” content.”

That Question 22 “resolves” such a problem, if it exists, is speculative and where it is not speculative it is incorrect. Suppose, for example, that instead of thinking that an

ad's primary focus is both policy and electioneering or "neither of the two," the coder is unsure or unclear about their judgment. Question 22 forces the coder into a category that describes them less well than the third category, "unsure/unclear," offered in Question 6. The JLG report therefore mischaracterizes Question 22's alleged resolving powers.

Moreover, the JLG report offers no direct evidence on how answers to the questions would have changed had we allowed the responses "both" and "neither" in Question 6 or the response "unsure/unclear" in Question 22. Again, the Plaintiff's experts could run a simple and very standard question wording experiment of the type described in my point 3 of this section. Therefore, the JLG report can only speculate about whether Question 22 or Question 6 is an improvement on the other.

And to the extent that the notion of "improvement," depends on the assumption that both questions must be measuring the same phenomena, I refer to my point 1 of this section – if you want a person's opinion of an ad's purpose, a question asking for a judgment of its primary focus is not an improvement.

5.

"Second, the question provides at least some guidance for how to make the judgment required, telling the coder to consider the "primary focus" of the ad."

Here, the emphasis is on the word primary. The claim is that since Question 6 did not explicitly ask about a *primary* purpose it is less reliable than Question 22. I would first note that, as established above, these questions are written to seek different kinds of information. Therefore, the notion that the questions' reliability can be measured on a common scale is based on faulty premise. Henceforth, I focus on the role of the word "primary" in questions such as Question 6.

Adding the qualifier “primary” to Question 6 may or may not clarify matters for coders. The JLG report presents no evidence to this effect. Alternatively, we may ask whether treating the coder’s opinion of the ads “purpose” as a measure of the coder’s opinion of its “primary purpose” is standard practice in Political Science. To this end, I reintroduce “Social Identities and Political Intolerance: Linkages within the South African Mass Public” by James L. Gibson and Amanda Gouws, which appears in the April 2000 edition of the *American Journal of Political Science* on pages 278 to 292.

The critical variable in the study, on group identity, itself a mental state, is drawn from responses to the following questions (p. 281, fn 6):

“The respondents were asked: “People see themselves in many different ways. Using this list, which of these best describes you? Please take a moment to look at all of the terms on this list.” The respondents were then asked: “Still looking at the card, do you think of yourself in any of the other terms as well?” The next question was “Still looking at the card which would you say most strongly does NOT describe you.”

The footnote continues:

““Primary” social identities are the initial responses given.”

But the word “primary” appears nowhere in the question. The authors of this article do not ask readers to be wary of their results as a consequence of their choosing not to include the word “primary” in the question. And nowhere in the article do they attempt to defend this choice of wording. Indeed, they regard their choice to be so innocuous and inconsequential that they talk about it only in a brief footnote.

In fact, Gibson and Gouws proceed as if subjects understood that their first answer would be treated as their “primary” allegiance. They report no attempt to run an experiment testing whether inserting the word “primary” in the question or explaining beforehand to respondents that the authors “will treat your first answer as primary and

most important to you, your second response as secondary and less important than the first, and so on” would change their results.

What Gibson and Gouws are doing is standard practice. With Question 6, Goldstein follows this same practice. The JLG report provides no tangible evidence or scholarly reference that Question 6 is inconsistent with standard scientific practice.

8.

Recall from the beginning of Section II that I left one of the four items from “Conclusions-Buying Time 1998” to be answered at a later time. The item was as follows.

“4. The 1998 data provided by the Brennan Center/Professor Goldstein not only do not support the conclusions announced in the Report, but they in fact lead to exactly the opposite conclusion: The vast majority of ads broadcast within 60 days of the 1998 election, which depicted one candidate or the other, had as their primary focus issue advocacy.”

My evaluation of this statement has two parts. First, consider the claim that the data leads to the conclusion “The vast majority of ads broadcast within 60 days of the 1998 election, which depicted one candidate or the other, had as their primary focus *issue advocacy*.” Is the claim *true* or *false*?

The claim is drawn from coders’ responses to Question 22, which reads:

22. In your judgment, is the primary focus of this ad on the personal characteristics of either candidate or on policy matters?

1. Personal characteristics
2. Policy matters
3. Both
4. Neither

In this question, the coders’ response options are “personal characteristics” and “policy matters.” “Issue advocacy” is nowhere mentioned in the question. The *claim is false*.

Perhaps, however, this claim in the JLG report was supposed to end “had as their primary focus policy matters.” I introduce this possibility, as it is more consistent with the argument the JLG report actually contains. Given this interpretation, *the claim is still false*.

To see why, note that the claim that the 1998 data lead to “exactly the opposite conclusion” of *Buying Time* refers to the conclusion that some coders perceived issue ads as having electioneering as their purpose. The JLG report’s conclusion follows from the false assumption that we must ignore a coder’s opinion of an advertisement’s purpose (culled from a completely different question, as I describe in Section II.d above) if it conflicts with their Question 22 judgment of its primary focus (see my point 3 of section I). This assumption is a red herring. While the data does show that “[t]he vast majority of ads broadcast within 60 days of the 1998 election, which depicted one candidate or the other, had as their primary focus *policy matters*,” this fact does not contradict coders’ reports of their opinions about an ad’s “purpose.” *Buying Time* finds that coders are able to make this distinction with respect to ads that mention or depict a candidate, and JLG report provides no evidence to the contrary.⁹ Therefore, the “exactly the opposite conclusion” claim follows not from Goldstein’s data but – as Section II.d has shown – from a flawed assumption in the JLG report’s attempt to characterize this data.

⁹ It also finds that they can make this distinction with respect to candidate ads.

SECTION III. Evaluating “Conclusions-Buying Time 2000.”

The main emphasis of the section of the JLG report devoted to *Buying Time 2000* is that it shares many *Buying Time 1998*'s faults. As Sections I and II of my evaluation show, many of the JLG report's claims on these matters are false or speculative. In this section, I will show that claims pertaining to the 2000 *Buying Time* study suffer a similar fate. Like the part of the JLG report that focuses on *Buying Time 1998*, these claims are often based on confounding the concepts of “replicable” and “replicated”, a failure to replicate particular claims in an unstated but perhaps limited number of attempts, complaints about outcomes and procedures that Dr. Gibson could have attempted to replicate but did not, claims about Goldstein's intent regarding the management of his database whose basis in fact is questionable, and the JLG report's insistence that an opinion question about an advertisement's “purpose” must be treated as an inferior attempt to measure its “primary focus” despite presenting no logically coherent justification for such treatment.

Indeed, a great deal of attention is again paid to the “purpose” and “primary focus” questions, with the report drawing the conclusion that only the latter is credible. It remains true, however, that coders were asked for their “opinion” of the ad's “purpose.” The question is quite explicit in not asking for an exclusive physical characteristic of the ad, rather it seeks a mental impression – a verbal report of an internal physical characteristic of the person. It asks about a person's thoughts after seeing an ad, it does not ask about the ad itself. It is standard scientific practice to ask people questions of this nature and to treat their responses as sufficient measures of their mental impressions.

Henceforth, I will not revisit these issues. Instead, I will focus on claims in the JLG report that are unique to the 2000 study.

1.

“I adduce evidence that the *Buying Time 2000* data base produces such wildly divergent estimates of the number of airings with certain characteristics (such as issue ads that are aired within 60 days of the election and which depict a candidate), that the data cannot be used to provide useful conclusions about the characteristics of such ads and airings.” Pp. 44-45.

“many specific findings from *Buying Time 2000* cannot be replicated.” Pp. 47-48.

This claim “cannot be used to provide useful conclusions” is analogous to claims seen earlier in the JLG report. It is based on demonstrations such as the one at pages 47 and 48. There, it is reported, “many specific findings from *Buying Time 2000* cannot be replicated.” The JLG report’s only evidence on this matter, however, is a brief and vague description of a failed attempt to reconcile a difference in the number of airings between two versions of the dataset. Like the JLG report’s claim about the irreproducibility of the 1998 study (see my point 6 of Section I), the JLG report does not divulge aspects of the replication attempt that would render it transparent and replicable (e.g., on procedures used or number of replication attempts). The JLG report also provides no evidence as to whether and the few differences in data that it identifies affects the substance of *Buying Time 2000*’s major claims. Therefore, the claim that such data “cannot be used to provide useful conclusions” is speculative and the claim that “findings cannot be replicated” is not even close to proven.

2.

“I have adduced in this report indirect logical and empirical evidence indicating that the student coders over-reacted to the appearance of a politician in these ads, often falsely coding issue advocacy ads as electioneering ads.” Pp. 55-56.

“Indeed, the results I produce are entirely logical given the artificial nature of the coding decision required by Question 11. Question 11 produces the results it produces because (a) the coders failed to code mixed content ads, since Professor Goldstein provided them no coding category with which they could record mixed-content ads, and (b) they also over-reacted to the presence of politicians in the ads when they coded Question 11.” Pp. 57-58.

“Surely the coders judged the “purpose” of this ad to be candidate promotion because they over-reacted to the mentioning of a candidate and because they were not allowed to code such ads as having mixed content.” Page 60.

I include these quotes because they are stated differently than in the earlier sections of the report and may be interpreted as new claims. As I showed in my points 5 and 6 of Section II.c, the “indirect logical and empirical evidence” is actually several empirically unsupported speculations about the impact of question wording joined by a wild supposition about the coders’ psychology for which the JLG report offers no tangible evidence or scientific reference. Moreover, given the possible interest of participants in this case about what people think of political advertisements, Question 11 in the 2000 study (which is parallel to Question 6 in the 1998 study) is entirely appropriate. The coding decision it requires – it asks coders to choose their opinion using a small number of discrete categories – is not only consistent with standard practices, it is also common.

3.

“Just as important as the accuracy or inaccuracy of the data cited in the 2000 Report, however, is the extent to which changes in a relatively small number of the highly subjective codings can affect the results reported and the conclusions reached. To make this point, let us consider 30 specific ads from the 2000 data base. These ads are: 676, 704, 1270, 1367, 1381, 1422, 1552, 1648, 1757, 1761,

1844 2065, 2088, 2089, 2107, 2158, 2301, 2344, 2502, 2558, 2588, 2695, 2712, 2810, 2900, 2935, 3158, 3170, 3220, 4001.” Page 62.

A footnote follows this sentence:

“This list of 30 ads was provided to me by counsel. The storyboards for these 30 ads are attached as Exhibit 15 to this report.”

How should we judge the credibility of the claim that “changes in a relatively small number of highly subjective codings” can affect ... the conclusions reached?”

Having countered the “highly subjective” status of the codings in my point 6 of Section II.b, I turn to the rest of the claim.

As was true in my point 11 of Section II, I begin with an assessment of the transparency and replicability of the reported procedure. As was the case earlier, the JLG report does not fare well on these grounds. Its claim is based on an examination of 30 storyboards. We are not told why 30 storyboards and not more or less were examined.

In this case, however, we are told that these cases were *provided by counsel*. How did counsel select these cases? The JLG report does not reveal whether counsel selected cases at random or whether they simply chose cases that are biased towards a conclusion that they favor. If the storyboards are not sampled in an unbiased way, then we must have less confidence that the inaccuracies identified in the JLG report represent the quality of the database as a whole. Indeed, the JLG report provides no basis for rejecting the hypothesis that the alleged inaccuracy is an artifact of the cases being selected in a way that is biased towards this generating this particular result.

4.

“The conclusion one draws from a review of this e-mail is that the investigators were committed to drawing a particular set of substantive conclusions from the data. When the conclusions were not forthcoming, the data were scrutinized further and alterations were made in the data base. ... Such strong *apriori*

commitments to drawing certain substantive conclusions from the data base seriously undermines the credibility of the reports.” Page 64.

The JLG report presents quite limited and indirect evidence on this matter. While the JLG report claims that data was changed in order to achieve a particular kind of outcome, it repeatedly fails to connect database updates to changes in the main themes of *Buying Time*. In some cases, the JLG report suggests that the changes could make such difference without demonstrating such effects directly. In other cases, the JLG report bases these conclusions on interpreting responses to standard opinion questions in ways that contradict common scientific standards or by asking us to accept a denominator that is one of many that are justifiable (see my points in Sections II.b and II.d). Therefore, the statement “*a priori* commitments to drawing certain substantive conclusions seriously undermines the credibility of the reports” is, at best, premature and, with certainty, not proven in the JLG report.

Arthur Lupia
October 14, 2002