Political Communication
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/upcp20

All Virtue Is Relative: A Response to Prior

Seth K. Goldman , Diana C. Mutz & Susanna Dilliplane

To cite this article: Seth K. Goldman , Diana C. Mutz & Susanna Dilliplane (2013) All Virtue Is Relative: A Response to Prior, Political Communication, 30:4, 635-653, DOI: 10.1080/10584609.2013.819540

To link to this article: http://dx.doi.org/10.1080/10584609.2013.819540

Political Communication, 30:635–653, 2013

ISSN: 1058-4609 print / 1091-7675 online
DOI: 10.1080/10584609.2013.819540

Routledge
Taylor & Francis Group

The Workshop

All Virtue Is Relative: A Response to Prior

SETH K. GOLDMAN, DIANA C. MUTZ, and SUSANNA DILLIPLANE

In "The Challenge of Measuring Media Exposure: Reply to Dilliplane, Goldman, and Mutz," Markus Prior suggests that scholars should avoid using a new method of measuring exposure to political television that we evaluated in a recent article published in the American Journal of Political Science. *We respond to each of his criticisms, concluding that although no measurement approach is without its flaws, scholars should always use the best approach that is available at any given point in time.*

Keywords media exposure, measurement

> Virtue is relative to the actions and ages of each of us in all that we do.
>
> —Plato

In "Reply to Dilliplane, Goldman, and Mutz," Markus Prior suggests that a proposed new method of measuring exposure to political television, the program list technique, is ill advised on the grounds of construct validity, convergent validity, predictive validity, and reliability. We respond to each of these arguments in turn, highlighting technical as well as conceptual misunderstandings. Many of Prior's criticisms are directed at traditional survey measures of media exposure, not at the measures we propose in our article that appeared in the *American Journal of Political Science* (Dilliplane, Goldman, & Mutz, 2013). Given that few people defend traditional media exposure measures, and this problem has been much lamented elsewhere, we see little disagreement in this regard. Our goal was to improve upon the approach most widely used today, a goal we think we have achieved.

However, in addition to clarifying and correcting some of Prior's assertions about the measures we develop, we wish to highlight why adhering to a rigid conception of what scholars really want from media exposure measures may ultimately hamper the progress of research in this area. At the end of the day, assessment of any measurement technique is a

Seth K. Goldman is Honors Assistant Professor of Communication, University of Massachusetts Amherst. Diana C. Mutz is Professor of Political Science and Communication, University of Pennsylvania. Susanna Dilliplane is George Gerbner Postdoctoral Fellow, Annenberg School for Communication, University of Pennsylvania.

Address correspondence to Seth K. Goldman, Communication Department, Machmer Hall, University of Massachusetts, 240 Hicks Way, Amherst, MA 01003-9278, USA. E-mail: skgoldman@comm.umass.edu

matter of whether it is the best that one can possibly do at any given time and place. Given our available options, we suggest that the new approach is a substantial improvement.

Construct Validity

Prior's critique suggests that the construct validity of our measure is low "because it does not attempt to measure the amount of exposure, only the number of different programs a respondent watches over the course of a month." In making this assertion, he confuses traditional *operationalizations* of media exposure with the underlying theoretical construct of interest. In order to understand how best to measure any construct, it is important to consider what it is we would ideally like to measure for purposes of testing a given theory. The study of political media effects centers on the extent to which content in the media influences political attitudes/behaviors or levels of political knowledge. Even studies that focus on hypotheses such as priming or framing are ultimately interested in effects of media content on one of these two kinds of outcomes.

Prior suggests that "amount of exposure" requires tapping frequency of viewing as has been done with the traditional "how many days per week" questions that have long been in use, or possibly in terms of the minutes per day or week that respondents use a medium. As has been amply demonstrated elsewhere, asking people about their amount of news exposure is fraught with difficulties. Disagreement over what counts as "news" is only the beginning of the problems posed by such measures. Both approaches—days per week and minutes—have been tried and been found wanting (see Althaus & Tewksbury, 2007).

But more importantly, conceptualizing frequency as the only legitimate way to tap exposure represents rigid adherence to an interpretation of exposure that is outdated. There is nothing inherently appropriate or ideal about the number of minutes per week or any frequency metric for purposes of tapping media exposure. To illustrate why this is the case, it is useful to conduct a thought experiment and ask what scholars would use to test their hypotheses in an ideal world, one without practical constraints. Assuming that people were capable of reporting the exact number of minutes per week that they watched programs that a given researcher considered political, is this what a scholar would ideally use to test a hypothesis about the effects of media content on political opinions or political knowledge?

We suspect not. To test a persuasive hypothesis, for example, what one would really want to know is how many unique favorable or unfavorable arguments people had been exposed to while viewing media content. Frequency of viewing would not tell us this. If the study were about opinions toward the economy, we would want to quantify the amount of positive and negative economic coverage to which a person was exposed. Frequency of viewing would not tell us this. To predict knowledge gain, we would want to know how much novel information was conveyed in each of the programs they watched. Total amount of exposure, whether in minutes, days per week, or some other metric, is not the actual construct of interest in our theories. It is instead a crude surrogate for what we actually would like to measure: exposure to particular media content.

Is frequency of total viewing time a better measure of this than the number of programs watched? The answer depends upon how much these measures tell us about the content of what was watched. In tapping political television, sheer amount of exposure made some sense back in the days when viewers had very few news programs from which to choose, and among those limited choices, one 30-minute newscast carried pretty much the same information as another. But in the contemporary media environment, days per week watching news is an arbitrary metric that tells us next to nothing about the content that was consumed; knowing the programs that a viewer watched at least gets us closer to ascertaining the kind of content to which a person was exposed. By knowing which programs are watched, we can attempt to capture variations in the content of media exposure in a way that traditional measures of sheer frequency made impossible. This opens the door to testing a much greater range of theories and hypotheses.

Given that we cannot easily measure the actual independent variable of interest, we naturally want the best possible surrogate. Just as Webb, Campbell, Schwartz, and Sechrest (1966) found that fingerprints and nose smudges on museum exhibit glass were the best operationalizations of the popularity of an exhibit, scholars need to be open to the best way to get at the underlying construct within the limitations of their study design. No one would argue that smudges *are* popularity ratings, but if they tap how much time people spend engrossed with each exhibit, then they are tapping the right construct. What ultimately matters with any operational measure is if it taps the right construct with a strong ratio of signal to noise.

As explained in greater detail below, the media measures that are currently on offer have not demonstrated strengths in either predictive validity or true-score reliability. In contrast, our operationalization has demonstrated both. Moreover, as an empirical matter, a study cited in Prior's critique directly contradicts his claim that it is implausible for a measure of the number of programs viewed regularly to proxy for the extent of exposure. Wonneberger, Schoenbach, and van Meurs (2012) used Dutch people-meter data (a source that Prior considers credible) to create measures paralleling the usual survey questions assessing the total duration of news viewing and the number of news programs viewed. Their results indicate that these two measures tap the same construct:

> The number of news programs viewed . . . and the duration of news viewing showed very high intercorrelations . . . *with duration and the number of programs nearly perfectly related.* (Wonneberger et al., 2012, p. 9, emphasis added)

Thus, amount of exposure (what Prior claims should be of interest) and the number of programs watched (what the program list technique assesses) are virtually indistinguishable based on Prior's own standard, which is people-meter data. This evidence strongly contradicts his dismissive claim that we "end up measuring a concept that is of little theoretical relevance."

Prior also questions construct validity on the grounds that we limit the content of interest to coverage of the presidential campaign by using a filter question asking whether respondents have heard anything about the campaign on TV. Given that this was an omnibus study of opinions toward the presidential candidates and knowledge of their issue positions, our focus on the campaign is appropriate; studies interested in political coverage beyond campaigns would naturally not use the same filter. Far from introducing systematic measurement error as he suggests, the filter question eliminates people who never watch political television, and thus shortens the survey interview for some. In this particular case, however, it made little difference because nearly everyone said they had heard something about the presidential campaign on television (between 89.1% and 91.2% of the sample in each wave). Thus, this filter did little to alter our measures. Prior also cites studies suggesting that "decomposition tends to increase reported frequencies without increasing their accuracy" and uses the Pew Media Consumption Survey to illustrate this problem. However, given that we did not ask people questions about how *frequently* they viewed each specific program, and we did not combine any frequency measures to create a summary estimate, this work seems of questionable relevance to the program list technique.

Prior also argues that the program list technique introduces a "considerable cognitive burden" for respondents, suggesting that they "can only answer the question accurately if they recognize the names of the programs they watch." Here he overlooks several key features of this approach. First, as is well known, recognition is a far easier task than recall. Relative to traditional survey questions about media use that serve as our basis for comparison, the cognitive burden of the program list measure is greatly reduced. Asking people to mentally calculate how much news they watch in an average week or how many total programs they have viewed requires lengthy recall and mathematical calculation on the spot. Recognizing regularly watched programs from a list does not require a multi-step mental calculation process.

Second, we ask specifically about programs watched *regularly*. Given the impossibility of knowing the exact content of each individual program watched by a respondent, we ask about programs viewed regularly because these are the programs with the greatest probability of affecting the viewer's knowledge and attitudes. We are not attempting to capture the kind of incidental or fleeting exposure that occurs while changing channels. Moreover, in order to watch a program regularly, a person would need to know what the program was; otherwise, how could one become a regular viewer? If one watches the same program regularly, one is very likely to be able to recognize the name of the program on a list, even if one could not recall the name of it in an open-ended format.

Finally, and perhaps most importantly when considering cognitive burden, it is worth noting that the program list technique made its debut as a means of allowing first- and third-grade children to reliably self-report their exposure to television programs (see Huesmann, Moise-Titus, Podolski, & Eron, 2003). The high true-score reliability of program-level self-reports—even at the level of individual programs (see Appendix A in our original article)—suggests a substantially lower cognitive burden than traditional measures.

Prior also suggests that scale construction is problematic because the cutoff we use for defining "regular" exposure is different from what has been used in traditional measures. As he suggests, "for individual programs, the measure does not distinguish different amounts of exposure." This is true, but the value of more fine-grained assessments of amount of exposure is debatable; more detailed estimates of minutes per day or week have been found to add no explanatory power to the traditional "days per week" measure (Althaus & Tewksbury, 2007). Given that there must be some unit of quantification, we see nothing inherently preferable about day as a unit of analysis relative to program. Few hypotheses are about particular days, whereas knowing the specific political programs a person watches and their unique content (partisan or not, talk shows versus news, and so forth) presents opportunities for scholars to test additional hypotheses. Moreover, both the Wonneberger et al. (2012) analyses already mentioned and the LaCour and Vavreck (2013) study described below in our discussion of convergent validity demonstrate that the number of programs viewed and the total amount of exposure tap the very same construct.

Prior further criticizes the program list approach because it "asks respondents to report program exposure regardless of whether they encountered political content or 'anything about the presidential campaign' while watching the program." Thus, he argues, "a respondent who watches *Oprah, Ellen*, or *The View* would receive a higher exposure score even if these programs did not cover politics." Further, he criticizes the measures for not taking into account the length of each of the programs or how often they are aired. Here Prior appears to overlook the part of the article where we experimented with measures that take into account the frequency with which a program is aired and the length of the program, as well as our coding for whether the program emphasizes presidential politics regularly, occasionally, or only rarely. As we demonstrate in the article, the optimal scale for purposes of predicting knowledge gain appears to be the one that takes the degree of political emphasis of specific programs into account.

Moreover, an additional benefit of the program list technique is that one can group the program items into categories that best fit specific hypotheses. If a scholar is interested in effects of partisan content, he or she can construct measures of left- and right-leaning programs (e.g., Dilliplane, 2011, in press). If the theory is instead about exposure to economic news, then the measures can be weighted by how much each program covers this topic. The flexibility of these items extends well beyond that of traditional exposure measures, thus offering many potential applications (see, e.g., Dilliplane, in press; Goldman, 2012).

Prior further argues that because there are no frequency of viewing estimates for each regularly viewed program, "neither would it be possible to determine the relative impact of two different programs." It is worth noting that the television exposure measures currently available in surveys do not allow this either; thus, this is not a disadvantage of our particular approach, so much as a limitation common to all available approaches. In a survey of sufficient length, it would certainly be possible to ask about the frequency of viewing for each program that a respondent reports watching regularly, and this additional information might ultimately produce more precise estimates of exposure to various kinds of content. Because of survey length constraints, the program list measures on the National Annenberg Election Survey (NAES) Panel Study did not do this. Thus, we must leave it to future researchers to explore and validate these possibilities.

Convergent Validity

Prior's claim that the program list technique lacks convergent validity relies on two main arguments: (a) that the program list measure produces overreporting relative to other exposure measures, namely Nielsen's people meters and Integrated Media Measurement Incorporated (IMMI) media use tracking technology, and (b) that the program list measure fails to show the increase in levels of exposure that would be expected to occur over the course of a presidential campaign. We find no evidence that either of these issues is problematic for the proposed purpose of studying media effects.

In order to support the argument of overreporting, Prior relies primarily on Nielsen estimates, which is problematic for several reasons. First, it is not possible to directly compare the program list estimates and Nielsen estimates, in part because of the limited availability of Nielsen data (as Prior acknowledges), but also because the two measures do not assess exposure at the same unit of analysis. The Nielsen ratings capture program or channel viewing on a given evening, whereas the program list measure taps program exposure that occurs "at least once a month." And whereas the Nielsen ratings primarily capture viewing that occurs on a conventional television set, the program list measure captures viewing that occurs on a television as well as through other media, whether computer, smart phone, tablet, broadband devices, or the like.

But more importantly, Prior's focus on Nielsen estimates makes the questionable assumption that Nielsen data are the appropriate standard for measuring levels of exposure. Nielsen's lack of transparency makes it unsuitable for use as a validation criterion in scholarly research. There has been no scholarly assessment of the reliability and validity of Nielsen's people-meter data, and studies that have attempted to assess the quality of Nielsen's people-meter system find "severe faults" (Milavsky, 1992, p. 114) and "high levels of uncertainty regarding the degree to which measured audiences provide an accurate and reliable reflection of actual audiences" (Napoli, 2003, p. 81). For example, questions have been raised about the Nielsen sample's representativeness (Milavsky, 1992) and the reliability of Nielsen data for programs with small audiences, which are increasingly common in the highly fragmented high-choice media environment (Napoli, 2003).

Moreover, contrary to what is widely assumed, Nielsen is not a completely passive indicator of exposure, but rather requires substantial input from household members in the form of button pushing. Fatigue leads to steadily declining compliance with even this minimal input; as time wears on, people increasingly fail to push the buttons on the device used to capture viewing behavior (Milavsky, 1992). Prior suggests that the compliance problem is resolved in single-member households. But even in these homes, after watching television for 70 minutes, the viewer must hit a button on the set-top box indicating that he or she is still viewing or else the system no longer considers that person to be a viewer, and thus potentially underestimates actual viewing in the home. People must also indicate when they leave the room in order to avoid overestimates.

Even more problematic in this particular case, Nielsen's shortcomings have been suggested to lead to consistent underreporting of exposure, especially among lower-income and less educated segments of the population. Nielsen people meters do not capture out-of-home viewing (Napoli, 2003), therefore missing the many instances of TV exposure that occur at friends' or relatives' homes, at work, or in public places. Nor do Nielsen estimates currently capture the growing amount of viewing that occurs online or on a mobile device such as a smart phone or tablet.[1] By incorporating the multiple modes of viewing currently possible, the program list measure purposely captures additional instances of exposure that Nielsen ratings do not.

Within the television industry, Nielsen is no longer perceived as a reliable source of viewership estimates because their techniques are old-fashioned. As a recent article in *The Economist* ("Counting Couch Potatoes," 2013) echoed, "Consumers' media viewing habits have changed too fast for Nielsen to keep up." A network executive concurred, "Everyone is unhappy." Media firms "say privately they would welcome new competitors to wake Nielsen up. But Nielsen retains a near-monopoly." The review of Nielsen's problems concludes by noting, "For TV firms, which earn a living by promising specific numbers of eyeballs to advertisers, such wild uncertainty is alarming. A happy ending—in the form of a reliable measurement standard—seems years away" ("Counting Couch Potatoes," 2013). If a solution is years away for reliable aggregate estimates that can be used by media industries, it is much farther away for academic researchers who want individual-level estimates of viewing behavior that can be combined with attitudinal measures, all at an affordable cost.

Finally, even if one accepts Nielsen as the appropriate benchmark against which to compare the program list technique, the two aggregate estimates differ very little. For example, according to Prior's analysis, the percentage who checked off the "Fox News" item on the program list was 34 in Wave 2, 35 in Wave 4, and 35 in Wave 5 of the NAES panel, while the Nielsen estimates of Fox News Channel viewers for the corresponding time periods were 26%, 30%, and 29%, respectively.[2] Slightly higher estimates would be expected given the broader range of viewing contexts captured by the program list measure, as well as the fact that respondents may have interpreted the "Fox News" item as including local news, not just national news. Moreover, because the NAES panel data and the Nielsen data are both based on samples rather than a census, these estimates should be interpreted as bounded by confidence intervals. We know what those intervals are for NAES data, but the lack of information available from Nielsen about their sample and confidence intervals makes it impossible to know if there is, in fact, any significant difference between the estimates.

Our own analyses, reported in our original article in the *American Journal of Political Science*, provide additional evidence that the Nielsen estimates generally correspond with the program list estimates. By rank-ordering the programs by their popularity, we found that the program list estimates highly correlate with Nielsen estimates.[3] Thus, regardless of any overall inflation, the rank order of program viewing yielded by the program list technique is consistent with Nielsen's estimates.

Beyond Nielsen comparisons, we find additional evidence for the convergent validity of the program list technique based on a comparison with estimates using a new technology that passively tracks the audio portion of media via cell phones (collected by IMMI). In his critique, Prior argues that studies (Jackman, LaCour, Lewis, & Vavreck, 2012; LaCour, 2012) comparing the IMMI measures to traditional self-report measures (which did not appear on the NAES panel survey) provide evidence of overreporting in the program list measure. However, given that the studies he cites compare the new technology to traditional self-report measures of exposure—not the program list technique that was designed to improve upon these measures—these comparisons cannot tell us about the extent of overreporting by the program list measure.[4]

More importantly, a recent study directly comparing the number of programs watched to the amount of time watched using the passive IMMI tracking technology that Prior advocates provides strong evidence of convergent validity (LaCour & Vavreck, 2013). Time spent watching news programs (measured with the IMMI tracking technology) and the number of programs those same people watched were found to be very strongly correlated. This evidence corroborates findings from the Netherlands with people-meter data (Wonneberger et al., 2012) while using a U.S. sample and an entirely different passive measurement technology. Thus, it appears that the total number of programs watched is indeed a very good surrogate for the total number of minutes watched.

To be clear, we would not be surprised to find that people overreport some kinds of media exposure. Past analyses of traditional self-reports of media exposure have suggested that people tend to overreport exposure to network news (e.g., Price & Zaller, 1990) but to underreport viewing of entertainment programs. Dutch people-meter studies suggest that people overreport their frequency of news viewing while underreporting the duration of their viewing (Wonneberger et al., 2012). But whatever the category of programming that is of interest, neither systematic underreporting nor overreporting are particularly problematic for purposes of testing most media hypotheses, so long as a measure rank orders individuals properly. In other words, if a measure accurately differentiates lighter viewers from heavier viewers while also capturing changes over time as the exposure levels of these viewers increase and decrease, our statistical tests should still give us the correct answers.

Prior's second main argument for why the program list technique lacks convergent validity is that it "barely picks up" the increase in exposure levels that one would expect to occur during a presidential campaign. His evidence does not warrant this conclusion. For example, Prior states that "the average primetime audience for FNC nearly doubled between the first quarter and October 2008, according to Nielsen"—an increase that he argues the program list measure failed to capture. However, we do not see any evidence of this increase in the Nielsen data he provided.

In fact, based on the Nielsen data that Prior provided to us, the two measures actually provide similarly modest estimates of change over time. Using our program list data, the monthly program list estimates of the Fox News audience are 31% (January), 34% (February), 36% (March), 31% (September), 33% (October), and 33% (November) of the sample. When Nielsen's monthly estimates for FNC are translated into the same metric, percentage of the voting-age population, the changes over time are relatively modest, just as they are in the program list estimates: 24% (January), 25% (February), 25% (March), 30% (September), 27% (October), and 33% (November).[5] According to

Nielsen, there is only a 3-percentage-point difference between viewing estimates during the early primary period of January and at the height of the general election campaign in October.

In addition, the Nielsen data from Prior show that exposure to FNC actually *decreased* between September and October, contrary to Prior's suggestion that exposure should increase as the election drew closer. The Nielsen estimates for FNC then increased to their highest average level in November. Given that election day was November 4, it seems odd that exposure should jump to its very highest average levels during the month *after* the election ended. A few days of high viewership in early November should not outweigh a whole month of high viewership in October.

More importantly for our purposes, there is clear evidence that the program list measure registers sensible changes in exposure levels over time. Given that the date of interview within waves was randomized in our sample, the best test of the measure's ability to capture change over time is to plot weekly estimates, as shown in Figure 1. As the figure illustrates, the total number of political programs watched increased over the course of the general election campaign, just as one would expect. The peak levels in mid-November make sense given that respondents were asked about their exposure during the past *month*. In addition, Figure 1 shows the kind of post-election changes that one would reasonably expect. Exposure levels started to drop during the post-election lull, then ramped back up prior to Obama's inauguration.

Further, in follow-up interviews conducted in the fall of 2010, we find additional evidence that the program list measure captures sensible patterns in exposure over time. The mean number of political programs viewed declined significantly after coverage of the campaign and Obama's inauguration died down.[6] Indeed, as shown in Figure 2, this decline in program exposure occurring between the high-profile campaign and inauguration periods (Waves 4 and 5) and the 2010 follow-up (Wave 6) was observed across different genres of programming, from political talk shows and newscasts to TV newsmagazines. The only type of programming that did not show a significant drop in mean exposure levels was political satire. In short, the program list measure reflected the over-time trends in

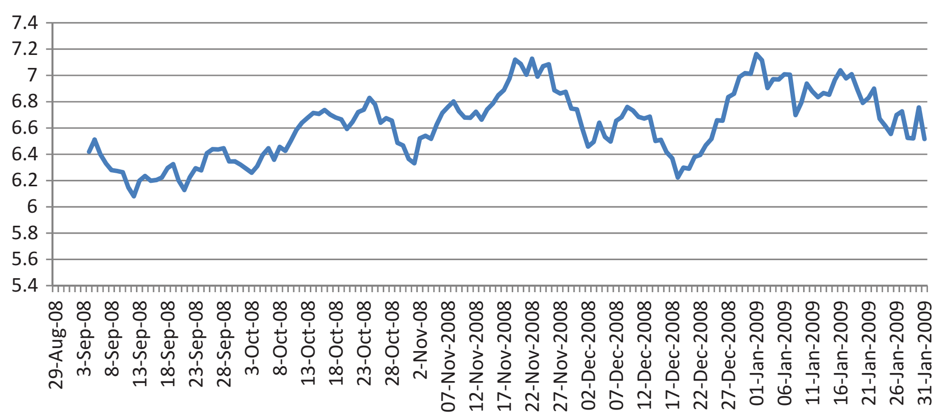

Figure 1. Mean number of political TV programs watched during the 2008 general election campaign and pre-inaugural period. The figure shows 7-day moving averages. Total number of political TV programs is the same indicator used in the reliability and validity analyses reported in our original article in the *American Journal of Political Science* (color figure available online). *Source:* 2008 NAES Internet Panel.

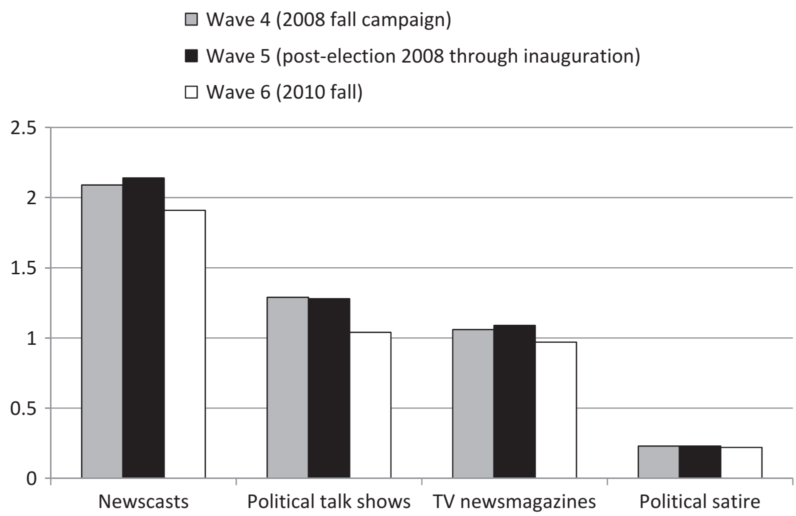

Figure 2. Mean number of political TV programs watched during pre- and post-election periods, by program genre. The figure presents the mean number of programs watched in each wave. Because the 2010 follow-up panel only included non-Hispanic Whites, the figure is limited to this subset of the population ($N = 3{,}263$). Wave 4 interviews were conducted from August 29 to November 4, 2008; Wave 5 interviews were conducted from November 5, 2008 to January 31, 2009; and Wave 6 interviews were conducted from September 21 to October 6, 2010. With the exception of political satire, the mean number of each type of program watched was significantly lower in Wave 6 compared to Wave 5.

exposure one would expect: Once the suspense of the election and the excitement of a historic inauguration was over, Americans watched less political TV. Contrary to Prior's argument, the available evidence *supports* the program list technique's convergent validity rather than calling it into question.

Predictive Validity

Prior critiques our evidence of predictive validity on several grounds. First, he argues that political knowledge gain is not an appropriate validation criterion. In fact, he questions whether there is any relationship between political knowledge and political television exposure, suggesting that because "the true individual-level relationship between news exposure and knowledge is . . . fairly uncertain, it cannot serve as a reliable benchmark for predictive validation." Prior admits, however, that "experimental studies do show that exposure to television news has a positive average treatment effect on political knowledge (e.g., Neuman, Just, & Crigler, 1992)." Many other experimental studies also demonstrate that news exposure produces political learning (e.g., Norris & Sanders, 2003; Tewksbury & Althaus, 2000).

Nonetheless, Prior suggests that any knowledge gained from exposure is likely to be quickly forgotten, and thus there is no empirical basis for expecting exposure to increase levels of knowledge over successive panel waves. Yet observational data show accumulating knowledge of candidate issue positions over time at the aggregate level (e.g., Johnston, Hagen, & Jamieson, 2004), and one experimental study found no decay from exposure over a 2.5-week period (Norris & Sanders, 2003). Even if the effects of a single exposure do fade, our measure taps regular viewing, allowing us to assess the effects of repeated exposure.

Field experiments and observational studies that do not rely on survey self-reports or laboratory manipulations point to the same conclusion: Political media exposure produces gains in political knowledge (Barabas & Jerit, 2009; Jerit, Barabas, & Bolsen, 2006; Milner, 2002). When so many different research designs converge on a similar finding, it should add to our collective confidence about any causal relationship. Thus, knowledge gain remains the best criterion for predictive validity, the "gold standard" designated by many scholars before us (e.g., Norris, 2000).

In contrast, Prior argues that "if there is a 'gold standard' for validation, it is convergent validity, not predictive validity" and suggests discarding "a measure with high predictive validity but low convergent validity." We address the convergent validity of the program list technique in a prior section of this response, but we would add that assessing the utility of a new measure should include multiple tests of validity whenever possible. Relying on convergent validity as the one and only test of a measure's utility has little theoretical or empirical basis, and is particularly problematic for assessing a measure of media exposure given the lack of solid data on media use.

In his final argument against using political knowledge gain as a validation criterion, Prior suggests that "the predictive validity test incorrectly assumes that television news exposure is necessary and sufficient for political learning. . . . Someone can learn about the candidates without watching television news and watch television news without learning about the candidates." We make no such assumptions about the relationship between exposure to television news and political knowledge; the idea that television news exposure causes increased knowledge does not require that exposure be either necessary or sufficient.

Overall, this discussion belies the usual social scientific understanding of probabilistic causation. Smoking is neither necessary nor sufficient for getting lung cancer; many smokers never get lung cancer, and some people get lung cancer who have never smoked. Nevertheless, smoking is unquestionably an important cause of lung cancer. It increases people's probability of getting lung cancer just as watching political television increases their probability of knowing candidates' issue positions. Any measure of how much people smoke that is unrelated to their long-term probability of lung cancer would be suspect and rightly discarded, just as any measure of political television exposure that is unrelated to gains in political knowledge would also be suspect. In short, we need make no assumption that television news exposure is either necessary or sufficient for political learning in order to use knowledge gain for purposes of assessing predictive validity.

To support his argument, Prior attempts to draw a parallel between Boudreau and Lupia (2011), who argue that political interest is not a suitable *surrogate measure* of political knowledge because it is neither necessary nor sufficient for knowledge, and our analysis of predictive validity. Importantly, we do not advocate using the program list measure as a surrogate for political knowledge. Instead, we merely suggest that exposure to political programs is a source of knowledge gain. As a result, this comparison does not make a good case for requiring that exposure be a necessary or sufficient condition for learning.

Oddly, even as Prior argues that political knowledge gain is not an acceptable standard for assessing predictive validity, he simultaneously argues that "many of the traditional news exposure measures that Dilliplane, Goldman, and Mutz dismiss do in fact exhibit high predictive validity." Prior then cites several cross-sectional studies that found positive associations between self-reported news exposure and political knowledge. But as we point out in our original article, cross-sectional analyses cannot capture over-time *changes* in knowledge. Even when cross-sectional analyses include a variety of control variables, they remain vulnerable to spuriousness from unmeasured and/or unobservable confounders, not to mention reverse causality. Simply put, evaluation of how well exposure measures predict knowledge *gain* requires panel data.

Prior points to only one study that employs panel data to address the relationship between self-reported exposure and knowledge. And although Chaffee and Schleuder (1986) had panel data from a small sample of Wisconsin parents and adolescents, they collapsed their measures of media exposure over all waves into a single index, and therefore did not analyze the effects of change. Other studies have also used panel data without examining the effects of change (e.g., Eveland, Hayes, Shah, & Kwak, 2005; McLeod et al., 1996). To date we are not aware of an analysis that tests the ability of change in self-reported exposure to predict change in knowledge.

In our study, we demonstrate that change in exposure causes change in knowledge by employing three waves of panel data and fixed effects models of within-person change. Prior suggests that fixed effects regression has only limited utility—that we only "guard against the simplest version of omitted variable bias, the spurious effect of stable predictors of knowledge." In fact, fixed effects regression represents a huge improvement over other observational designs, including other panel models (Allison, 2009). With fixed effects, each individual is compared to himself or herself at an earlier point in time, so *all* differences *between* individuals drop out of the equation. By contrast, "in individual-level cross-sectional studies, differences in opinions between those exposed to the media and those who remain unexposed may simply reflect preexisting differences between the two groups in political attitudes or characteristics" (Bartels, 1993, p. 267). Fixed effects regression discards all between-person variance, and instead uses only within-person variance that occurs over time. Stable factors such as education, income, age, ongoing political interest, and party identification drop out of the model, as do *all* other variables (whether observable or unobservable) that are constant over time.[7] Scholars using cross-sectional measures try to measure and control for the most likely spurious confounders, but what we should really be worried about is all of the *other* variables that we are unaware of or cannot measure (i.e., "unobserved heterogeneity bias"). This is the single greatest threat to interpreting relationships as causal in observational research, and fixed effects regression solves it (Allison, 2009; Hallaby, 2004).

Importantly, fixed effects regression does a "superior job" of controlling for stable confounders compared to traditional panel analyses using lagged dependent variable models or, by extension, cross-lagged models (England, Allison, & Wu, 2007, p. 1245; Hallaby, 2004). Although scholars often refer to lagged dependent variable models as if they assess change at the individual level, this is misleading because these models still rely on between-person variance and only assess change in the limited sense of differences over time in the rank order of individuals. Further, due to measurement error alone, the measure of the lagged dependent variable provides imperfect control for preexisting differences (Allison, 1990).[8] Fixed effects regression, on the other hand, perfectly controls for preexisting differences by only comparing each individual to him- or herself at an earlier point in time.

In fixed effects regression, only variables that change over time can produce spurious associations. Fortunately, spuriousness arising from time-varying factors in fixed effects models is far less likely than spuriousness arising from individual differences in between-person models. With fixed effects, a confounding variable would have to (a) change over time, (b) explain change over time in the independent variable, and (c) explain change over time in the dependent variable. Especially in the relatively short time span of a presidential campaign, very few potential confounders meet these requirements. Moreover, as we explain in our original article, by including a dummy variable for survey wave in each equation, we efficiently capture the average total effects of all other time-varying influences (Hallaby, 2004). For example, to the extent that political interest rises across the board during the campaign, the effects of the wave variable capture this change.

Nonetheless, as our extensive appendices to the original article demonstrate, we also controlled for a variety of time-varying factors. Consistent with the idea that the wave variable already captures these influences, including numerous other controls had no impact on the estimates of media influence. Prior suggests that we do not appropriately control for several factors, including change in exposure to other media. This should be understood not as a threat to predictive validity—given that these variables merely represent other forms of media influence—but rather as relevant to discriminant validity, that is, how well the effects of television exposure can be distinguished from effects of other closely related political media.

Second, Prior claims that "some of these controls have little variance (e.g., dummy variables measuring exposure to the presidential campaign in newsmagazines and on the Internet)," implying that all of the controls are dummy variables. On the contrary, only two variables are dummies, while the number of hours of TV watched the previous night has six values, exposure to newspapers has three values, and exposure to political radio has 15 values. Yet, including all five controls simultaneously barely reduces the size of the political TV coefficient (from .54 to .47).

Prior also argues that we "omit several other time-varying predictors of knowledge, including exposure to political advertising, the party conventions, and the presidential debates." Notably, these are not sources of potentially spurious relationships. The extent of overall political television viewing *should* capture exposure to advertising, the conventions, and the debates. However, the program list technique was not designed to allow for fine-grained distinctions between the effects of particular campaign events and other campaign content. Any influence not captured by the program list measure—such as through special one-time programs that were not asked about—should be captured by the number of hours of television watched the previous night.

Finally, Prior claims that "heavily covered issues may lead to better performance on knowledge questions not because more people watch the coverage, but because they generate more interpersonal discussion, command greater attention, or are more easily remembered." It is not at all clear to us how media coverage of political issues can "command greater attention" or be "more easily remembered" without exposure to coverage. Similarly, interpersonal discussion is not a spurious influence, but rather a mediator of mass media influence. For example, the two-step flow of communication posits that some people are exposed to political media and that these "opinion leaders" spread the message to many other people, leading to substantial indirect effects of media exposure (Katz & Lazarsfeld, 1955). With respect to political media, however, the two-step flow has received no empirical support to date (Bennett & Manheim, 2006; Chaffee & Hochheimer, 1985; Gitlin, 1978).[9] Nonetheless, we replicated our analyses controlling for the time-varying impact of interpersonal discussion and changes over time in political interest and found no change in our original findings. Given that exposure to political television is a *cause* of increased discussion and interest during a campaign, these are extremely stringent tests.

Reliability

As we point out in our original article, the true-score reliability of the program list technique is quite high relative to traditional measures of political television exposure (.84).

However, Prior's reply indicates some confusion regarding terminology. As explained by Zaller (2002, p. 315):

> Use of the term "reliability" needs to be clarified, as it is deployed in two different ways in the research literature. The most common usage is that of Carmines and Zeller (1979) (p. 11–13): "reliability concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials." They distinguish reliability from validity, which is the extent to which a measure "measures what it purports to measure." They add: "just because a measure is quite reliable, this does not mean that it is also relatively valid." Thus, over-reporting of media exposure could, in their senses of the terms, make a media exposure variable more reliable but less valid. Lord and Novick's (Lord and Novick, 1968) classic treatment offers a different conception of reliability: "The reliability of a test is defined as the squared correlation, $\rho_{XT}{}^{2}$, between observed score and true score" (p. 61). In this usage, the distinction between reliability and validity does not exist.

Thus, Zaller distinguishes between test-retest reliability and true-score reliability. To clarify further, he uses the term "validity" here for the extent of alignment of an imperfect operational measure of a variable with the true value of the variable. This is to be distinguished from "construct validity," which is alignment of the operational measure with the theoretical concept it purports to proxy. For a measure to adequately capture a concept, both notions of validity are required; that is, the measure must adequately represent the true value of the variable, and the variable must fit the theoretical concept. Because we have already discussed construct validity, we focus the discussion here on Zaller's discussion of reliability.

Cronbach's alpha is a measure of test-retest reliability that relies on correlations among multiple measures all purporting to assess the same concept. Measuring true-score reliability is preferable to Cronbach's alpha because it is a higher standard. But there is a hitch: The definition would seem to require the true score, which is precisely what we do not know. In the case of multi-wave panel data, under the assumption of uncorrelated error, Heise (1969) discovered a way to infer the squared correlation between observed and true score without directly observing the true score. Thus, with three or more waves of panel data, and only in this case, can one indeed measure true-score reliability.

Prior dismisses the high true-score reliability of the program list measure because he says high reliability does not matter when a measure is not valid. Like most scholars, we think reliability and validity are *both* relevant criteria for good measurement. Moreover, we use a far more exacting standard for reliability than previous studies because we have the advantage of multi-wave panel data. Reliability, combined with predictive validity, convergent validity, and construct validity, would seem to produce a powerful combination.

The potentially exaggerated reliabilities that Prior cites from Zaller (2002) are examples where Zaller expresses doubt about the measures because they correlate weakly with political knowledge. As Zaller puts it, "I find it difficult to take seriously a measure of news exposure that has so weak an association with what ought to be the effect of news exposure" (Zaller, 2002, p. 310). Zaller suggests that the high reliabilities despite low predictive validity could be due to overreporting. For the program list measures, by contrast, we find evidence of high reliability *and* high predictive validity.

Importantly, the reliabilities Zaller is referring to are based on Cronbach's alpha. While useful for some purposes, Cronbach's alphas merely reflect the average intercorrelations among survey items. This is not the same conception of reliability as a true-score reliability; the two are simply not comparable. Moreover, even the one panel study that Prior cites (Chaffee & Schleuder, 1986) does not report a true-score reliability for television exposure. Instead, they collapsed measures from across their panel waves into a single indicator and report a Cronbach's alpha. As we point out in our original article, we have found only one other reported true-score reliability for a political television measure administered to the U.S. population (Bartels, 1993), and ours is substantially higher. There simply is no literature documenting true-score reliabilities from previous attempts to measure Americans' political television exposure.

Prior correctly notes that all reliability estimates can be affected by correlated errors; to the extent that respondents consistently overreport (that is, the same person by similar factors in repeated waves), correlated errors could result in inflated reliability estimates. He therefore suggests that high reliabilities are really not a good sign. It is certainly correct that even with true-score reliabilities, correlated errors can inflate the estimates. Nonetheless, they are still the best possible measures of reliability that are independent of stability. A high true-score reliability could mean either a very reliable measure or that individuals systematically overreport. Interestingly, neither case is particularly problematic for our purposes.

For the sake of argument, let us take the worst case scenario and imagine that we have evidence that the latter is most definitely the case: That is, individuals are systematically overreporting. The estimation technique used in our panel analyses makes any consistent overreporting of exposure completely unproblematic to our results. The fixed effects panel model that we use already *assumes* correlated errors over time. To the extent that a person consistently overreports, an analysis of individual-level change over time will be unaffected. Whatever inflation factor exists for a given person drops out of the model altogether. This makes consistent overreporting a non-issue when analyzing panel data in this fashion. It renders individual differences in overreporting unimportant because each person is only being compared to him- or herself, not to others who might inflate by different factors. Previous analyses have not had this advantage because none of them—including the Chaffee and Schleuder (1986) panel study cited by Prior—used fixed effects models of within-person change in their analyses.[10] In short, when using differences over time as predictors, a within-person constant displacement is irrelevant.

Another problem might be if different people overreported by *different* amounts at *different* times. How can we be certain that this is not occurring? What if a person inflates his or her estimate by one factor at Time 1 and a different factor at Time 2? If this were the case, it would have devastating effects on the true-score reliabilities of our measure. Given the high true-score reliabilities that were obtained, we have convincing evidence that this is not occurring.

A third potentially problematic scenario is if, at different points in time, everyone in a study inflates their viewing more or less simultaneously. Fortunately, the analysis approach used in our study also accounts for this potential problem. All of the fixed effects analyses presented in our article include variables representing wave of interview in the models, that is, the component of change that is common across all respondents from one wave to the next (Hallaby, 2004). Thus, any simultaneous inflation effects of this kind (due, for instance, to overall increases in political interest at a given point in time) would be removed from the analyses by the wave variables.

Could these same measures be problematic if a researcher attempts to make a case for media effects using cross-sectional data? Absolutely, but the legions of problems involved in using cross-sectional data to establish media effects go far beyond problems with the reliability of exposure measures. In cross-sectional data, high levels of political knowledge may go along with high levels of political news exposure for many spurious reasons that have nothing to do with television's ability to increase levels of political knowledge. Convincing observational accounts of media effects inevitably must involve observation of change in the dependent variable at the very least; otherwise, stable individual differences are hopelessly conflated with the outcome of interest.

Conclusion: What Do Scholars Really Want?

Prior provides a litany of criticisms of our alternative approach to measuring political television exposure. He admonishes scholars against making use of these measures in their future research using the American National Election Studies or NAES panel surveys. Unfortunately, he does not offer any better alternatives. Virtue is, indeed, relative; we would certainly not claim our approach to be without its faults, but we think it is a substantial improvement over what scholars have used for many decades. And while the approach we offer is not perfect for all purposes, it has clear virtues that current approaches lack.

The appropriate standards for measurement are reliability and validity. On these two dimensions, the virtues offered by our approach are many. They include (a) confirmed construct validity based on the strong relationship between number of programs viewed and total extent of viewing as documented by multiple studies (see LaCour & Vavreck, 2013; Wonneberger et al., 2012), (b) high true-score reliability based on panel data, (c) convergent validity with Nielsen estimates and with the logical points in time when exposure should increase and decrease, and (d) strong predictive validity in models predicting change over time in knowledge with change over time in exposure. If, as Prior suggests, we are "reliably measuring something other than the thing we want to measure," then it is incumbent upon him to explain why change in this measure predicts change over time in candidate knowledge so consistently. Indeed, ours is the first evidence from a representative panel sample demonstrating that changes in television exposure correspond to changes in political knowledge over time. Many experiments have shown this, and some previous panel studies have suggested it, but without actually measuring change over time or using a statistical model that eliminates the impact of stable individual characteristics that could confound the results.

While we, like many others, find that television viewing is mostly habitual (e.g., Adams, 2000; Wood, Quinn, & Kashy, 2002), the measures of exposure to all kinds of political programs—including standard newscasts, talk shows, and newsmagazines—all showed declines after the election, just as one would predict. Prior suggests that the program list measure registers "barely any increase in news exposure as the 2008 presidential election approached," and yet the increase in Figure 1 seems quite apparent—and quite similar to the modest increases recorded by Nielsen. Moreover, if exposure were completely stable in our estimates as Prior suggests, there would be no conceivable way for change in exposure to predict change in knowledge.

Finally, with respect to construct validity, we think our measure offers scholars more of what they really want from exposure measures than what is offered by our current frequency of exposure measures. What scholars really want is to know the kind of political television content to which people are exposed, including content on programs that have not traditionally been included in exposure measures, and including content that is not viewed in real time or even necessarily on television sets. Because our measures allow scholars to match the content of programs to the people who watch them regularly, this opens scholarship up to far more interesting hypotheses than previous measures of television exposure.

If there are other feasible approaches to measurement that have already surpassed the program list technique when using these same stringent tests of validity and reliability, we are unaware of them.

The research that Prior cites on passive measurement techniques offers aggregate comparisons to some traditional media exposure self-reports, but *not* to the program list technique, so a direct comparison of these two approaches is impossible for the moment. Using a small sample from a single Designated Market Area collected by IMMI, LaCour (2012) looks at exposure to television programs, and Jackman and colleagues (2012) examine advertising exposure. While both studies suggest that their results are promising, their validation analyses do not include evaluation of the extent to which exposure was underestimated (i.e., individuals were exposed to content, but the audio was not registered by their cell phones and/or the matching technology) or overestimated (individuals were not actually exposed, but their cell phones and the associated technology registered exposure). LaCour (2012) compares his aggregated results to Nielsen's aggregated results, but is not able to provide individual-level validation.

Aggregate comparisons of people exposed to political ads or to whole networks of programs cannot speak to issues of reliability and validity for individual-level measurements of program exposure. One would assume that passive measures would include less error, but technologies, as well as humans, are prone to errors, especially when humans must cooperate to activate these technologies (see Jackman et al., 2012, for a discussion). The early studies suggest that these approaches are not completely without their own problems and inconsistencies. For example, Jackman and colleagues (2012) suggest that these estimates underreport viewing relative to Nielsen estimates, that time-shifted viewing can be problematic to capture, and that compliance in using the technology (keeping the battery charged and the cell phone on the body at all times) is less than 100%, even among a nonprobability sample. We assume that passive approaches will still experience fewer problems than have been encountered with self-reports, but it is simply too early to tell when such technology will be available.

Although Prior is strongly opposed to the use of the measures we have developed, his only suggestion for the moment is to use "technologies developed by audience measurement companies" that collect data passively. What he does not mention is that, at least within the United States, there is no such company at present. Even if a researcher had unlimited resources, as most of us do not, as of this writing no such opportunity exists. The company responsible for the small-scale study of passive measurement that he mentions (IMMI) is no longer in business. Curiously, in addition to the studies using IMMI data, Prior also cites approvingly the data used in Gentzkow and Shapiro (2011). However, for the extent of people's exposure to television, newspapers, and newsmagazines, these authors use survey-based self-reports of precisely the kind that have been roundly denounced.

All measurement consists of the best one can do at any given point in history; we must make do with what is on offer. Audience measurement companies currently provide purely aggregate data on samples of unknown and unknowable quality. At least for now, these data cannot facilitate academic research on media effects that combines survey reports on attitudes and behaviors with media exposure measures. Prior's central concern about self-reports appears to be systematic overreporting, and yet panel data combined with fixed effects regression models make such concerns irrelevant.

Because virtue is indeed relative to a given time and place, we look forward to a future when demonstrably better measures will be both available and affordable in the United States. For now, however, the only publicly available observational data on individual media exposure—particularly measures that allow scholars to marry exposure to measures of opinion and knowledge—are from surveys. Given that survey-based exposure measures remain essential to the field at this point in time, it is incumbent upon political communication scholars to do our best to improve them as much as possible. Granted, our existing exposure measures make an easier target to improve upon than some as yet unknown and unvalidated future technique. But the alternative—to cease everything except experimental political communication research for the time being—would be unfortunate indeed.

Notes

1. Nielsen has recently announced plans to start incorporating into its ratings Internet streaming to TV sets as well as viewing occurring on broadband and mobile devices, but a number of these changes are still only in the early stages of planning (Bauder, 2013).

2. In addition to listing specific Fox News Channel programs, the program list on the NAES survey instrument included an item labeled "Fox News." This is the item Prior uses in his comparison of Nielsen estimates and the program list estimates.

3. Using program as the unit of analysis, we created a variable representing the percentage of the NAES sample viewing each program, and another variable representing the Nielsen rating (from September 2007, when the panel began) for each program that could be matched to it. The Pearson correlation between these two variables was .86 ($p < .001$). We also ranked the Nielsen ratings of programs from high to low and ranked the programs based on survey percentage viewing. Spearman's rho was .76 ($p < .001$) for this rank-ordered association.

4. Prior also cites his analysis of rolling cross-sectional survey data, which yielded evidence that self-reports of exposure to a single presidential debate lacked validity. As that analysis did not actually provide a test of the convergent validity of the program list measure, it seems at best only tangentially related.

5. These percentages are based on Nielsen's monthly estimates of the number of people watching Fox News Channel for 6 minutes or more: 56 million (January), 58 million (February), 58 million (March), 69 million (September), 63 million (October), and 75 million (November)

6. The 2010 interviews were limited to a random subsample of non-Hispanic White respondents from the 2008 NAES panel, as the primary purpose of the follow-up panel was to assess post-election changes in racial attitudes among Whites.

7. More specifically, fixed effects regression controls for the *constant* effects of individual characteristics, so we also controlled for the time-varying impact of individual characteristics by including interactions between the wave variable and education, age, gender, income, race, strength of ideology, strength of party identification, and political interest.

8. Even *after* adjusting for measurement error, "the use of Y_1 as a regressor variable seems to underadjust for prior differences" (Allison, 1990, p. 99).

9. Recent extensions of this idea under the heading of the communication mediation model also posit that interpersonal influence mediates the effects of media exposure on political knowledge and participation, though the hypothesized process differs from the two-step flow (e.g., Cho et al., 2009; Shah et al., 2007). Notably, however, these studies show significant *direct* effects of media exposure on knowledge even after controlling for interpersonal discussion. At the same time, these studies rely on cross-sectional analyses and so provide weak causal evidence with regard to both media and interpersonal influence.

10. Chaffee and Schleuder (1986) instead use a lagged dependent variable approach with constant measures of exposure. This approach is not the same because lagged dependent variable models do not directly address individual-level change over time, and are known to produce biased and inconsistent estimates (see, e.g., Allison, 1990, 2009).

References

Adams, W. J. (2000). How people watch television as investigated using focus group techniques. *Journal of Broadcasting & Electronic Media*, *44*, 78–93.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. *Sociological Methodology*, *20*, 93–114.

Allison, P. D. (2009). *Fixed effects regression models*. Beverly Hills, CA: Sage.

Althaus, S. L., & Tewksbury, D. (2007). *Toward a new generation of media use measures for the ANES*. Retrieved from http://www.electionstudies.org/resources/papers/Pilot2006/nes011903.pdf

Barabas, J., & Jerit, J. (2009). Estimating the causal effects of media coverage on policy-specific knowledge. *American Journal of Political Science*, *53*, 73–89.

Bartels, L. M. (1993). Messages received: The political impact of media exposure. *American Political Science Review*, *87*, 267–285.

Bauder, D. (2013). *Numbers game: Nielsen finally taking account of time-shifted viewing*. Retrieved from http://www.theledger.com/article/20130226/news/130229506

Bennett, L. W., & Manheim, J. (2006). The one-step flow of communication. *Annals of the American Academy of Political and Social Science*, *608*, 213–232.

Boudreau, C., & Lupia, A. (2011). Political knowledge. In J. N. Druckman, D. P. Green, J. H. Kuklinski, & A. Lupia (Eds.), *The Cambridge handbook of experimental political science* (pp. 171–186). New York: Cambridge University Press.

Chaffee, S. H., & Hochheimer, J. (1985). The beginnings of political communication research in the United States: Origins of the "limited effects" model. In E. M. Rogers & F. Balle (Eds.), *The media revolution in America and Western Europe* (pp. 267–296). Norwood, NJ: Ablex.

Chaffee, S. H., & Schleuder, J. (1986). Measurement and effects of attention to media news. *Human Communication Research*, *13*, 76–107.

Cho, J., Shah, D. V., McLeod, J. M., McLeod, D. M., Scholl, R. M., & Gotlieb, M. R. (2009). Campaigns, reflection, and deliberation: Advancing an O-S-R-O-R model of communication effects. *Communication Theory*, 19, 66–88.

Counting couch potatoes. (2013). Retrieved from http://www.economist.com/news/business/21571471-measuring-tv-audiences-keeps-getting-harder-counting-couch-potatoes

Dilliplane, S. (2011). All the news you want to hear: The impact of partisan news exposure on political participation. *Public Opinion Quarterly*, *75*, 287–316.

Dilliplane, S. (in press). Activation, conversion, or reinforcement? The impact of partisan news exposure on vote choice. *American Journal of Political Science*.

Dilliplane, S., Goldman, S., & Mutz, D. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, *57*, 236–248.

England, P., Allison, P., & Wu, Y. (2007). Does bad pay cause occupations to feminize, does feminization reduce pay, and how can we tell with longitudinal data? *Social Science Research*, *36*, 1237–1256.

Eveland, W. P., Hayes, A. F., Shah, D. V., & Kwak, N. (2005). Understanding the relationship between communication and political knowledge: A model comparison approach using panel data. *Political Communication*, *22*, 423–446.

Gentzkow, M., & Shapiro, J. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, *126*, 1799–1839.

Gitlin, T. (1978). Media sociology: The dominant paradigm. *Theory and Society*, *6*, 205–253.

Goldman, S. K. (2012). Effects of the 2008 Obama presidential campaign on White racial prejudice. *Public Opinion Quarterly*, *76*, 663–687.

Hallaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, *30*, 507–544.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, *34*, 93–101.

Huesmann, L., Moise-Titus, J., Podolski, C., & Eron, L. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology*, *39*, 201–221.

Jackman, S. D., LaCour, M., Lewis, J., & Vavreck, L. (2012). *Digital fingerprints: A new method for measuring political advertising*. Unpublished manuscript.

Jerit, J., Barabas, J., & Bolsen, T. (2006). Citizens, knowledge, and the information environment. *American Journal of Political Science*, *50*, 266–282.

Johnston, R., Hagen, M., & Jamieson, K. H. (2004). *The 2000 presidential election and the foundations of party politics*. Cambridge, England: Cambridge University Press.

Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. New York, NY: Free Press.

LaCour, M. J. (2012). *A balanced news diet, not selective exposure: Evidence from a direct measure of media exposure*. Unpublished manuscript.

LaCour, M. J., & Vavreck, L. (2013). *Improving media measurement: Evidence from the field*. Unpublished manuscript.

McLeod, J. M., Guo, Z., Daily, K., Steele, C. A., Huang, H., Horowitz, E., & Chen, H. (1996). The impact of traditional and nontraditional media forms in the 1992 presidential election. *Journalism & Mass Communication Quarterly*, *73*, 401–416.

Milavsky, J. R. (1992). How good is the A. C. Nielsen people-meter system? A review of the report by the Committee on Nationwide Television Audience Measurement. *Public Opinion Quarterly*, *56*, 102–115.

Milner, H. (2002). *Civic literacy: How informed citizens make democracy work*. Hanover, NH: University Press of New England.

Napoli, P. M. (2003). *Audience economics: Media institutions and the audience marketplace*. New York, NY: Columbia University Press.

Neuman, W., Just, M., & Crigler, A. (1992). *Common knowledge: News and the construction of political meaning*. Chicago, IL: University of Chicago Press.

Norris, P. (2000). *A virtuous circle: Political communications in post-industrial societies*. New York, NY: Cambridge University Press.

Norris, P., & Sanders, D. (2003). Message or medium? Campaign learning during the 2001 British general election. *Political Communication*, *20*, 233–262.

Price, V., & Zaller, J. (1990). Evaluation of media exposure items in the 1989 NES pilot study. Report to the Board of Overseers, American National Election Studies.

Shah, D. V., Cho, J., Nah, S., Gotlieb, M. R., Hwang, H., Lee, N. J., . . . McLeod, D. M. (2007). Campaign ads, online messaging, and participation: Extending the communication mediation model. *Journal of Communication*, *57*, 676–703.

Tewksbury, D., & Althaus, S. L. (2000). Differences in knowledge acquisition among readers of the paper and online versions of a national newspaper. *Journalism & Mass Communication Quarterly*, *77*, 45–79.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago, IL: Rand McNally.

Wonneberger, A., Schoenbach, K., & van Meurs, L. (2012). Dimensionality of TV news exposure: Mapping news viewing behavior with people-meter data. *International Journal of Public Opinion Research*, *25*, 87–107.

Wood, W., Quinn, J. M., & Kashy, D. (2002). Habits in everyday life: Thought, emotion and action. *Journal of Personality and Social Psychology*, *83*, 1281–1297.

Zaller, J. (2002). The statistical power of election studies to detect media exposure effects in political campaigns. *Electoral Studies*, *21*, 297–329.