A Review and Analysis of Patterns of Design Decisions in Recent Media Effects Research

W. James Potter
University of California at Santa Barbara, USA
wjpotter@comm.ucsb.edu

Abstract

This essay presents a critical analysis of patterns of research design decisions exhibited by authors of recently published empirical tests of media effects. The content of 211 articles published from 2010 to 2015 in six core communication journals was analyzed to document the design decisions made by the authors concerning their use of theory, sampling, measurement, and experiments. We also recorded the amount of variance explained by their tests and use this indicator of strength of findings to explain the patterns of methodological design decisions. The findings indicate that authors of these studies commonly select weaker design options over stronger ones. The reasons for these patterns are explored then critiqued leading to a series of recommendations calling for an evolution in thinking in the areas of method, theory, and paradigm. The methods recommendations attempt to increase (a) awareness of the advantages and disadvantages of options available for each design decision, (b) an understanding that often assumptions made to justify the selection of an option are faulty, and (c) a commitment to meeting a higher degree of challenges. The theory recommendations focus on increasing an understanding about why designers of most tests of media effects ignore the many theories available when designing their studies. Morover, the paradigm recommendations focus on examining more critically the assumptions we make about the nature of human beings, the purpose of our research as challenges evolve, and the defaults in practices we have established in an exploratory phase.

Content

Highlights

- The findings of published content analyses of the media effects literature are critically analyzed to provide a foundation for an analysis of current methodological patterns.
- The design decisions made by authors of 211 articles reporting a test of media effects that were recently published (2010 to 2015) in six core communication journals are analyzed.
- The literature continues to display an atheoretical pattern evidenced by the majority of designers of media effects studies ignoring theories when constructing a foundation for their studies.
- There is a continuing pattern of many authors of media effects studies selecting weaker options over stronger ones when designing their samples, measures, and experiments.
- A critical analysis of the likely reasons for these patterns of design decisions reveals that many of the selected design options are supported by assumptions that have been found to be faulty.
- Recommendations are presented for an evolution in thinking about methods, theory, and paradigm.

I. General Criticism of the Development of the Field of Media Effects

It is vital for research fields to examine their patterns of methodological decisions periodically to assess whether those patterns are evolving along with the challenges. Such examinations are crucial in monitoring the development of the field (Borgman, 1989).

The scholarly field of media effects is now almost a century old and has produced a large number of empirical studies (see Bryant & Oliver, 2009; Nabi & Oliver, 2009; Perse, 2001; Sparks, 2015) that was estimated to have been over 6200 published studies a decade ago (Potter & Riddle, 2007) and now is likely to be even larger. Scholars have periodically conducted content analyses of this growing literature to document various methodological features, such as the use of methods, theories, and types of samples (Kamhawi & Weaver, 2003; Lowry, 1979; Matthes, Marquart, Naderer, Arendt, Schmuck, & Adam, 2015; Moffett & Dominick, 1987; Perloff, 1976; Potter, Cooper, & Dupagne, 1993; Schramm, 1957; Trumbo, 2004; Wimmer & Haynes, 1978). The findings of these content analyses have triggered many scholars to criticize particular patterns in the way researchers have been designing their empirical tests of media effects (Fishbein & Hornick, 2008; Kamhawi & Weaver, 2003; Krcmar, 2009; Lang, 2013a, b; LaRose, 2010; Levine, 2013; Lowry, 1979; Matthes et al, 2015; Neuman, Davidson, Joo, Park, & Williams, 2008; Niederdeppe,

2014; Oliver & Krakowiak, 2009; Perloff, 2013; Potter, 2009; Slater 2004; So, 1988; Valkenburg & Peter, 2013).

This article begins with a review of the criticisms of many of these methodological patterns within media effects research. We use these criticisms as a foundation to design a content analysis of recently published media effects studies in order to determine the extent to which particular methodological patterns still exist. After reporting the results of that content analysis, I analyze those practices to uncover the assumptions that likely support methodological decisions. Then I present a system of recommendations to help our fellow media effects scholars transition more efficiently into research practices that can better meet the challenges that we currently face in explaining what media effects are and how they arise.

II. Criticisms of the Design of Empirical Tests of Media Effects

Critics have repeatedly raised concerns about a variety of methodological practices in the communication literature in general and the media effects literature in particular. In this section, I focus attention on four areas that seem to have attracted the most criticism. These four areas are the use (or non-use) of theory, sampling procedures, measurement features, and the design of experiments.

A. Theory as Foundation

Many scholars have argued that theory development and testing are essential to the overall development of the field of media effects (Kamhawi & Weaver, 2003; McQuail, 2005; Nabi & Oliver, 2009; Potter, 2009). For example, McQuail (2005) explained that because the "main purpose of theory is to make sense of an observed reality and guide the collection and evaluation of evidence" (p. 5), the use of theory helps a field grow in a more useful and efficient manner. Furthermore, Kamhawi and Weaver (2003) argued that "theoretical development is probably the main consideration in evaluating the disciplinary status of the field" and that over time, the need for theory-guided research becomes more critical because as "our field grows in scope and complexity, the pressure for theoretical integration increases" (p. 20).

Theories can be a valuable tool in growing the knowledge in a scholarly field efficiently and effectively due to their ability to integrate research findings into a system of explanation and to guide future research studies in a programmatic manner. Good theories provide empirical researchers with a map showing the most useful paths for extending knowledge as well as showing the extent of progress along those paths. They also provide researchers with a progression of knowledge about the methods, measures, and analysis strategies that have been found to demonstrate the greatest value in contrast to other design options that have been revealed to rely on faulty assumptions. Thus researchers who use a theory as a foundation for their empirical work increase their efficiency by following the theory's guidance about what to test and how to test it in the best way possible. In addition, researchers who use a theory as a foundation for their empirical work also experience the benefit of using a richer context for presenting the findings of their individual studies.

However, most of the empirical research in the field of communication in general and the sub-field of media effects in particular has been found to be atheoretical (Bryant & Miron, 2004; Kamhawi & Weaver, 2003; Potter, Cooper, & Dupagne, 1993; Shoemaker & Reese, 1990; So & Chan, 1991; Stevenson, 1992; Trumbo, 2004). For example, Kamhawi and Weaver (2003) found that only 30.5% of all articles published between 1980 and 1999 in 10 communication journals even mentioned a theory. Moreover, within the sub-field of media effects, Potter, Cooper, and Dupagne (1993) conducted an analysis of articles reporting tests of media effects published in eight competitive peer-reviewed journals from 1965 to 1989 and found that only 27.6% mentioned a theory. In an updating of the Potter, Cooper, and Dupagne study, Trumbo (2004) reported finding 42% of studies mentioned a theory. Bryant and Miron (2004) analyzed 1,806 articles dealing with mass media in three communication journals from 1956 to 2000 and found that 576 (31.9%) articles included some kind of theory. Potter and Riddle (2007) analyzed 936 articles published in 16 journals from 1993 to 2005 and found that 35.0% referred to a theory. From the findings across these five studies, it appears that only about one third of media effects research mentions a theory.

Mentioning a theory, however, is not the same as using a theory as a foundation for the empirical study. When content analysis studies of the media effects literature have gone beyond recording mentions and analyzed how the mentioned theories were used, we find an even more troubling pattern. Potter, Cooper, and Dupagne (1993) found that only 8.1% of articles were guided by a theory and provided a test of that theory while another 19.5% were tests of hypotheses but these hypotheses were not derived from a theory. Trumbo (2004) in an updating of the Potter, et al (1993) content analysis reported finding that 18% of studies published from 1990 to 2000 were guided by a theory and that an additional 24% mentioned a theory but did not use it to create either an hypothesis or a research question to guide their studies. When Bryant and Miron (2004) analyzed articles published from 2000 to 2004 in six communication journals, they found that among all articles that mentioned a theory, only 23% used the theory they mentioned as a framework for their study. To summarize these patterns, it appears that about two thirds of published studies of media effects completely ignore theory and that within the one third of the literature that acknowledges at least one theory, the majority of acknowledgements are simple mentions rather than using the theory as a framework for the empirical test of a media effect.

The persistently low proportion of empirical tests of media effects that are guided by a theory is puzzling, especially when we realize that there is a large number of such theories available for testing. For example, Potter and Riddle (2007) found more than 150 theories in use in their analysis of published research from 1993 to 2005 in 16 journals. Bryant and Miron (2004) found references to 604 different theories, paradigms, and schools of thought in their analysis of 1,806 mass media articles published in three communication journals from 1956 to 2000.

B. Sampling Procedures

While there are many different methods used for generating samples for empirical studies, all of those methods can be organized into two types: representative samples and non-representative samples. While representative samples are more difficult to construct than non-representative samples, they offer greater value to the development of the field of media effects because their findings can be generalized beyond the sample to known populations. Without representative samples, the field cannot build a base of knowledge about the scope, prevalence, and strength of media effects.

Most of the media effects literature in the past has relied on non-representative samples. A content analysis of sampling practices in media effects research found that 27.9% of media studies used probability samples and another 11.0% were population studies (Potter, Cooper, & Dupagne, 1993). A more recent study by Levine (2013) did not address the idea of representative samples directly in his content analysis of the communication empirical literature but he did point out that 50% of the samples were composed of college students, which indicates convenience sampling. He argues, "The use of expedient student data is somewhat controversial and is conventionally considered a limitation" (p. 78). Therefore, it is important to examine the samples used in recent media effects research to determine the extent to which the literature is shifting away from the expedient option of selecting non-representative samples and toward the more useful -- but challenging -- option of generating representative samples.

C. Measurement Features

This section focuses attention on four measurement issues - the use of self reports of mundane behaviors, the use of attribute variables as surrogates for active influences, the measurement of change, and the providing of evidence for the quality of measures.

Mundane behaviors.

Media effects researchers frequently need to measure mundane behaviors, such as the extent of exposure to media and messages (LaRose, 2010). Mundane behaviors are actions people habitually perform in their everyday lives. Once learned, these behaviors are governed by a process of automaticity where people no longer need to think about what they are doing as they perform those actions (e.g., typing, riding a bicycle, accessing internet sites). Because habitual behaviors are performed repetitively without much, if any, mental effort in a cognitive state of automaticity (Bargh, Chen & Burrows, 1996) and because people do not have counters in their brains that systematically record the accumulation of such behaviors, measures that ask respondents to recall how many times they performed these behaviors are fundamentally flawed (Slater, 2004; Verplanken et al., 2005). When respondents are confronted with the task of providing estimates of a mundane behavior, they cannot use recall and instead must rely on heuristics (Kahneman & Tversky, 1984). Because there is a variety of heuristics -- including anchoring, representativeness, availability, simulation, and adjustment (Fiske & Taylor 1991) -- the data generated from such questions are likely to be a conglomeration of responses generated by different heuristics and thus result in a complex of confounds. For example, self-report data of media exposure has been found to be composed of more than 20% measurement error in one early study (Bartels, 1993), and subsequent research has shown that this measurement error is likely to be much higher (Cohen & Lemish, 2003; Funch et al,. 1996; Kobayashi & Boase, 2012; Schüz & Johansen, 2007).

When researchers compare self-reported data to electronically recorded data of mundane behaviors, they find little correspondence, leading them to conclude that self-reported data has serious validity problems in many fields of study including psychology (Nisbett & Wilson, 1977; Schneider, Eschman, & Zuccolotto, 2002), sociology (Lauritsen, 1998), anthropology (Bernard, Killworth, Kronenfel, & Sailer, 1984), political science (Sigelman, 1982), criminology (Hindelang, Hirschi, & Weis, 1981), and behavioral medicine (Stone & Shiffman, 2002).

While the field of psychology has been demonstrating a movement away from the use of self reports of mundane behaviors (Haeffel & Howard, 2010), it appears that the same cannot be said about the field of communication. Lowry (1979) analyzed 546 empirical articles published in seven communication journals from 1970 to 1976 and found that about three quarters relied on participant self reports. Levine (2013) in his analysis of the general communication literature of quantitative tests found that about 80% of all published studies used self reports and that this trend seemed to be growing; he found that 71% of the studies he analyzed from 1988-1991 used self reports while 88% of studies published

2008-2011 did. These findings led him to argue that "an over-reliance on self-report survey" items is "especially responsible for slowing intellectual progress" (p. 72).

Attribute variables as surrogates.

Although attribute variables (such as sex and age) are typically easy to measure, they are often imprecise surrogates for active variables (such as gender socialization and cognitive development). When researchers intend to measure biological differences (such as hormones or body changes throughout adolescence), then biological sex is a valid measure, of course. However, often a measure of biological sex is used as a surrogate to represent an active influence, such as a pattern of gender socialization, and this substitution has been found to raise serious problems with validity. Also, a child's chronological age has often been used as a surrogate for a child's developmental maturity. However, research over the years has shown that this is a poor surrogate for cognitive development (King, 1986) and moral development (Van der Voort, 1986).

The use of attribute variables as surrogates for active influences generates a higher degree of measurement error than the use of more valid measures. For example, gender role socialization is a continuum of degrees of maleness and femaleness; collapsing everyone into two categories loses much of that richness of variation and runs the risk of misclassifying many people who are a blend of gender-related characteristics.

Measuring change.

Almost all media effects research assumes change, that is, media researchers perceive an effect as something that is altered in an individual (knowledge, attitude, belief, emotion, behavior) that can be attributed to some kind of media exposure. However, measuring such change requires an assessment of research participants at least two points in time. Typically one measure of the effect variable is taken before exposure to some media message and a second measure is taken after the exposure. When this minimum of two measurement points is *not* met, researchers have no foundation for claiming whether or not a change occurred; instead change must be assumed.

The assumption of change is widespread within media effects research. For example, experimenters will typically design studies that focus on group differences but then use those differences across group means to claim that their treatments are responsible for the differences in means across treatment groups. Such a claim implies change, that is, experimenters assume that the group of participants in each treatment condition are equivalent before experiencing their assigned treatments. So if non-equivalence is observed following the treatments (i.e., difference in means across treatment groups), then this difference is assumed to represent change, but this is an assumption, not evidence of change. For example, let's say that the treatment group displays a higher mean on the outcome variable compared to the mean of the control group. This difference could mean that the control group did not change while the treatment group did (increase on the outcome variable). Alternatively, perhaps both groups changed but the treatment group experienced a bit more change than the control group. Or perhaps the control group changed (scores were reduced for some reason) while the treatment group did not, so that the treatment group ended up with a higher mean on the outcome variable. With only one measure taken on the outcome variable, it is impossible to tell which pattern occurred and hence where the change was. When researchers shift away from using assumptions to support their findings and shift towards using more direct tests of their claims that generate valid evidence, the value of the findings from research studies increases.

It appears that in media effects research, there has been a dominant practice of depending on assumptions about change rather than measuring actual change. In his content analysis of 546 empirical articles published in communication journals, Lowry (1979) found that 87% of studies relied on purely cross-sectional data. Levine (2013) criticized the general communication empirical literature for not treating communication as a process, but instead conducting cross-sectional studies that focus on differences and relationships at one point in time.

Evidence for quality of measures.

There is growing criticism about the lack of attention to the validity of measures used in media effects research, especially measures of media exposure (Fishbein & Hornick, 2008; Gentile & Bushman, 2012; Niederdeppe, 2014; Slater 2004). This criticism points out that researchers frequently describe their measures without providing a supporting argument for the validity of those measures, which is a troubling oversight.

D. Design of Experiments

The experiment is a useful method that allows media effects researchers to focus on a particular factor of influence then test whether that factor makes a difference in bringing about a media effect. The experiment has been a popular method in testing for possible media effects, although not as popular as the survey. In his analysis of empirical articles published from 1970 through 1976 in seven journals, Lowry (1979) found that 30% used survey, 19% used experiments, and 13% used content analysis. In their analysis of articles published from 1980 to 1999 in 10 communication journals, Kamhawi and Weaver (2003) found that 33.3% used survey method; 30.0% used content analysis; 13.3% were experiments; 4.7% used historical method; 10.3% used other qualitative methods; and the remaining 8.4% used a combination of methods. Potter and Riddle (2007) found that survey method accounted for 32.0% of their analyzed studies, experiments for 28.8% (with 261 out of 277 being laboratory-based experiments), qualitative methods for 15.4, secondary analysis of an existing database for 8.4% and content analysis for 2.5%.

In order for an experiment to generate results that are useful, designers of experiments need to demonstrate (1) that their treatment groups are equivalent and (2) that the treatments delivered to each group have conveyed the meaning that researchers expected those treatments to deliver. As for the first task, researchers typically use random assignment and checking their experimental groupings for balance. As for the second task, researchers conduct manipulation checks.

Random assignment.

The primary advantage of using random assignment is that it "protects significance testing: Without random assignment, a test of statistical significance of between-group differences is not conventionally interpretable" (Krause & Howard, 2003, p. 753) because "random assignment protects statistical significance by preserving the applicability of the logical model upon which significance testing is based. Treatment and control groups being compared for significance testing are assumed by the model to be alike in dependent variable expected values with respect to everything but possible differences due to treatment effects" (p. 761).

Balance check.

Random assignment of individuals to treatment conditions still requires researchers to assume they have achieved equivalency. In order to test for equivalency, researchers conduct a balance check. Such a check is necessary to determine whether the assignment of participants to conditions has in fact resulted in a match of groups on the critical factors that are being tested as influencers of the outcome variable. Testing for group equivalency is especially important when researchers must use intact groups and cannot randomly assign participants to conditions; in this situation there is no basis for assuming equivalency, so a balance check is essential.

Manipulation check.

When experimenters do not conduct a manipulation check, they default to two assumptions, and one or both of these assumptions is likely to be faulty. One of these assumptions is that the research participants in each treatment group are accepting the meaning in the treatment that the experimenters intended. However, people are interpretive beings who are used to making their own judgments about the meaning of media messages (Barthes, 1967; Fiske, 1987; Krcmar, 2009). People have been shown to exhibit a range of interpretations about whether a message is entertaining (Bartsch, 2012), humorous (Boxman-Shabtai & Shifman, 2014), and whether characters are good or bad (Krakowiac & Oliver, 2012). Therefore, the default should be skepticism about what meaning is being received by participants, and until researchers can demonstrate that participants in a treatment group did in fact interpret the meaning in the way that researchers intended, it is typically faulty to assume that they did.

The second assumption experimenters often make is that if there is variation among interpretations across participants in the same treatment group, then that variation is unimportant. This assumption is built into the ANOVA statistical procedure where within-group variation is regarded as error that is used as the denominator in computing an F ratio with the numerator being the between treatment group variation. Thus ignoring the differences across individuals in their interpretations of any media message creates an artificially low ceiling on the amount of variance that can be explained (Oliver & Krakowiak, 2009; Valkenburg & Peter, 2013). Conducting a manipulation check gives researchers the ability to determine whether their participants received the same message and therefore the same stimulus. Because if there is a large variation within a treatment group on the interpretation of meaning, then researchers must realize that while

there was one stimulus administered to a particular treatment group, there was likely a range of stimuli received across participants within that group. To illustrate, we designed an experiment to test whether exposure to different levels of violence in a video resulted in a differential effect on an outcome variable (Potter, Pashupati, Pekurny, Hoffman, & Davis, 2002). With three treatment groups (exposure to a video with a low amount of violence, medium amount of violence, and high amount of violence). Our manipulation check revealed that the mean rating of violence in the stimulus video was highest in our high group and lowest in our low group. But we also found considerable within-group variance such that there were many participants in our "low violence exposure group" who believed they saw a moderate amount of violence. Also, there were some participants in our "moderate violence group" who believed they saw a low amount of violence and some who believed they saw a high amount of violence. If our purpose was to determine whether the group means were different in the expected direction, we designed a successful study. However, if our purpose was to extend knowledge about how people process stimuli, make judgments about media stories, and how those judgments influence other types of outcomes, then our design and the simple comparison of means offered limited value.

E. Reporting Effect Sizes

The reporting of effect sizes is something that more scholarly journals are requiring of authors (Matthes et al., 2015; Sun & Fan, 2010). In their content analysis of experimental communication research published in four flagship journals from 1980 to 2103, Matthes and colleagues (2015) found that 57.3% of experimental studies reported effect sizes. The finding that more than 40% of studies did not report effect sizes led Matthes et al. (2015) to say, "this finding is still alarming" given that these articles were published "in the field's flagship journals" (p. 202). In a similar study, Sun and Fan (2010) analyzed all articles published in four communication journals over four years (2003 - 2006) and found that the effects sizes were reported 79% of the time when the statistical tests were significant but only 55% of the time when statistical tests were non-significant.

The increases in the proportion of the empirical literature that reports effect sizes is a positive trend because it serves to shift the focus away from using statistical significance as the main indicator of the importance of findings and placing the focus instead on the proportion of variation the study is able to explain. Gigerenzer and Marewski (2015) criticized the use of statistical significance saying that it has become a surrogate for good research. They say it is practiced in a compulsive, mechanical way — without judging whether it makes sense or not. Abelson (1985) argued that in his field of psychology, researchers "sometimes tend to rely too much on statistical significance tests as the basis for making substantive claims, thereby often disguising low levels of variance explanation" (p. 129).

Researchers know that they can always achieve statistical significance by continuing to increase the size of their samples. Thus using statistical significance as a threshold for determining the value of findings is a weak criterion, especially when we consider the work of Meehl (1990) who found that almost all constructs of interest to social scientists (attitudes, stereotypes, beliefs, impressions, and expectations) were weakly related to one another. He referred to this widespread pattern of low-level association throughout the social sciences as "the crud factor" to warn social scientists that it is specious to present their findings as important when they find only weak correlations, whether they are statistically significant or not. Schneider (2007) further developed this point by arguing that "the challenge for social scientists resides in their ability to conduct studies that will go beyond identifying which variables are related (as this is likely to be almost all of them)" and instead to determine "which variables are the most strongly related . . . in theoretically or practically important ways" (p. 182).

We know from meta-analyses of various topics within the media effects literature that the effect sizes found in empirical research studies are modest (Valkenburg & Peter, 2013). Effect sizes typically fall in the range of only 2% to 10% of variance explained for even the most highly researched topics such as violence, sex, and advertising (see reviews in Preiss, Gayle, Burrell, Allen, & Bryant, 2007). While these meta-analyses of parts of the media effects literature are valuable as indicators of how powerful various systems of explanation are, we lack a more general picture of the level of explanation across the entire field of mass media effects. But even more important is the need to move beyond an exploratory perspective on research where any findings are perceived to make a contribution as long as they are statistically significant and focus more on designing studies that can continually "move the needle" upward in the amount of variance explained. To do this, researchers need

to move variance from the denominator to the numerator when computing F ratios, that is, to reduce our tolerance for unexplained -- or error -- variance by thinking more extensively about possible patterns of systematic variance. The progress of a scholarly field would seem to be attributable more to increases in explanation rather than how many analyses result in statistically significant findings.

III. Current Patterns in Effects Research

In order to assess current patterns of methodological decisions, we conducted a content analysis. This section describes the design of that content analysis then reports results.

A. Generating Data to Document Recent Patterns

Sample.

Riffe and Freitag (1997) characterized a field's mainstream journals as "the barometer of the substantive focus of scholarship and research methods most important to the discipline" (p. 873). Scholars who have looked at bibliographic citation patterns (Dominick, 1996; Reeves & Borgman, 1983; Rice, Borgman, & Reeves, 1988; Rice et al, 1996; So, 1988) have concluded that media research is largely clustered within a sub-set of four communication journals – *Communication Research*, *Journal of Broadcasting & Electronic Media*, *Journal of Communication*, and *Journalism & Mass Communication Quarterly*. Since those early bibliographic studies, two new journals have been found to publish a good deal of media effects research (*Journal of Children and Media* and *Media Psychology*) so these were added to our sample. For each of these six journals, we randomly selected two years from the period 2010 to 2015. Thus our sample was composed of 12 journal/year units. We read through all issues throughout those 12 journal/year units and coded all articles that presented a test of a media effect. We did not code editorials, book reviews, introductions to symposia, or editors' reports. We did include articles labeled as "Research in Brief" or similar designations.

In order to be coded for this study, an article's authors needed to make some claim or provide some evidence that a medium exerted some kind of influence leading to a recognizable effect. Those effects could be on individuals or at a more macro level (such as an institution, the public, society, the economy). Using these criteria, we identified a total of 211 articles as providing tests of mass media effects.

Coding variables.

The study measured 16 characteristics of each published article that was identified as dealing with media effects. Four of these were bookkeeping variables (journal, year, pages, and authors' names), 10 variables focused on design features, and the remaining two variables measured whether authors reported the amount of variance explained and the extent of that variance.

The 10 research design variables were developed from previous research studies then modified in a series of pilot tests in order to develop a list of options that reflected positions on a quality continuum for each of the 10 variables.

Use of theory.

This variable had five values as follows: (a) *a priori* theory, where the authors used an existing theory to deduce hypotheses and test them; (b) theories are mentioned as a background rationale for the study but the authors did not use those theories to deduce hypotheses; (c) the authors developed their own speculative model and tested the elements in that model; (d) the authors presented and tested hypotheses derived from their review of empirical findings but not from a theory; and (d) the authors did not mention any theory as a foundation for their study. Because there is a range of definitions for what a theory is in the social science literature, we did not start with an a priori definition; instead, we let authors tell us whether they used a theory or not. If they referred to something as a theory, we counted it as a theory mention. Also, if they referred to something that is generally regarded as a theory (e.g., cultivation hypothesis, uses & gratifications) without using the word "theory," we counted it as a theory mention.

Sampling.

This variable had two values: Representative sample and non-representative sample. In order to be coded as a representative sample, the authors needed to claim that their sample was randomly selected from a particular population or sampling frame.

Measures.

We collected data on five measurement characteristics

– self-reported mundane behaviors, attribute variables, change scores, and arguments for quality of measures both reliability and validity.

As for self-reported mundane behaviors, we first examined whether researchers asked their respondents to self report on their own behavior. Then we determined whether the behaviors were mundane, which were defined as habitual behaviors performed in a state of automaticity. Examples of a question eliciting a self report of a mundane behavior are: How many hours of TV did you watch last week? How many tweets do you send each day? Examples of a question eliciting a self report of a non-mundane behavior are: Did you watch a movie in a commercial theater last week? Did you pay to subscribe to a new website in the last month?

As for attribute variables, we recorded the use of characteristics of participants that were used as factors of influence on a media effect. These characteristics were typically the attributes of biological sex in place of gender socialization or the attribute of chronological age in place of level of development (cognitive, emotional, etc.).

As for change scores, we recorded whether researchers gathered data on their effect variable at more than one point in time. In experiments, we looked for evidence that authors administered a pre-test before a treatment followed by a post-test. In surveys, we looked for evidence of authors administering a questionnaire or interview at more than one point in time.

As for making a case for the quality of measures, we looked for the reporting of indicators of reliability and validity for the measures used in the study. The coding of reliability was based on whether researchers provided indicators of internal consistency for items on their scales (i.e., yes, no). As for validity, coders assigned one of four values to each article depending on how the authors treated the issue of validity: (a) argument for validity provided; (b) authors presented citations of other published studies using the same measures; (c) authors presented citations of published studies using measures that the authors adapted for their own use; and (d) authors did not address validity, that is, they simply described the measures they used.

Experiments.

If the study was an experiment, coders looked for three features. First, coders recorded whether the authors said that their research participants were randomly assigned to experimental conditions (i.e., yes or no). Second, coders re-

corded whether treatment groups were tested for balance (i.e., yes or no). Third, coders recorded whether authors reported a manipulation check (i.e., yes or no).

Strength of findings.

Finally we recorded whether authors reported figures for proportion of variance explained in their statistical tests (yes, no). If such figures were reported, we recorded what those figures were. Typically authors who reported these figures did so for tests run on each of their hypotheses, and their reportings were usually R-squares for correlational tests and eta-squares for tests of differences. We did not manufacture our own figures (such as squaring simple correlation coefficients) but confined ourselves to recording only indicators of proportion of variance explained as presented by the authors.

Testing reliability.

Approximately 20% of the sample was coded by two coders to create an overlap that could be used to test reliability. First the unitizing was tested and it was found that there was agreement 99% of the time with the yes-no decision of whether to include the article in our study. Essentially , this is a decision about whether the article dealt with a media effect or not. Second, we computed the percentage of agreement between coders on each variable, then corrected these percentages of agreement for chance agreement using Scott's pi taking into consideration all valid values and a "can't tell" option on all variables. Scott's pi figures are as follows: sample, .92; use of self reports of mundane behavior; .83; attributes as surrogates, .81; presentation of validity information, .78; use of change scores, .90; use of theory, .79, and reporting of proportion of variance explained, .88. In addition, if the article was coded as an experiment, it was also coded for random assignment (.90), testing for balance (.84), and checking for manipulation (.81). These reliability figures are relatively high because almost all coding decisions were based on manifest indicators and few required judgments from latent content.

B. Findings

Reporting the use of theory.

Of the total of 211 articles analyzed, 59 (28.0%) were theory driven. That is, the authors referenced an existing

theory, deduced hypotheses from that theory, and conducted their study to test those hypotheses. About half of the coded articles mentioned at least one theory but did not use any of those theories to generate hypotheses; instead these studies either developed their own hypotheses (33.6%) and tested them or developed their own model (16.6%) and tested that new model. In the remaining 45 articles (21.3%), the authors mentioned no theory; these authors presented a study largely driven by an exploratory type question.

The findings of this current study show a continuing trend toward the greater use of theory as a foundation for a media effects study. The content analysis from 1965 to 1989 found 8.1% of published tests of media effects was guided by a theory (Potter, Cooper, & Dupagne, 1993); the replication of this content analysis found that 18% of published tests in 1990 to 2000 were guided by a theory (Trumbo, 2004); and now this figure has increased to 28.0% in published tests from 2010 to 2015.

Sampling decisions.

Four out of five (80.6%) studies were found to use non-representative samples. Thus it appears that the use of non-representative samples is increasing when we compare our current results with those of Potter, Cooper, and Dupagne (1993) who found that 67.1% of empirical studies published in 8 major communication journals from 1965 to 1990 used non-representative samples. A likely reason for this trend is the increase in the proportion of experiments that almost never use probability samples. Within the group of surveys, one third used representative samples but with experiments, only 5.3% used representative samples.

Measurement decisions.

Although there are no previous studies to use as benchmarks of comparison with our results about measures, we can still see that there is room for improvement in three areas. First, 64.8% of coded articles measured mundane behaviors with self reports. Of these 136 studies, 65.4% used self reports of exposure habits to media, 10.3% used self reports of behavioral intentions, and the remaining 24.3% used self reports of other mundane behaviors, typically estimations of habitual behaviors performed automatically in respondents' everyday lives.

Second, 43.6% of the coded studies were found to use attribute variables as surrogates for active influences, typically age and sex. Of course, many studies used age and sex

as control variables . But these were not counted as surrogates unless the authors appeared to be using these variables as an indicator of something like level of cognitive development or gender role socialization. Third, only 26 (12.7%) measured the effect variable at more than one point in time.

As for making a case for the quality of measures, authors were found to be much more focused on reliability than validity. As for reliability, authors of 166 studies (78.7%) reported assembling individual measures into scales, and 141 of these studies (85.1%) displayed tests of the reliability of their scales. Across these 141 studies that used scales, there were 559 scales reported and reliability coefficients ranged from .27 to .99 with a median of .83.

As for making a case for the validity of any of their measures, 54.0% of authors ignored this task. Of the 97 studies that did offer some kind of support for the validity of at least one of their measures, 11.3% presented an argument for validity, 42.3% simply cited other published studies that used the measures the cited authors had used, and the remaining 46.4% cited measures from other published studies that these authors then adapted for their own purposes.

Design decisions in experiments.

Of the total 211 articles coded, 95 (45.0%) reported experiments, 108 (51.2%) reported surveys, and the remaining 8 used a qualitative method. This shows a continuation of a trend reported by Matthes et al. (2015) where 43.4% of articles in four mainstream communication journals from 2010 to 2013 were experiments.

Of the 95 articles that were experiments, 52 (54.7%) indicated random assignment of their participants to treatments; 26 (27.4%) reported they conducted a manipulation check of their treatments; and 13 (13.7%) said they conducted a balance check on the assignment of participants to treatments.

Reporting variance explained.

Of the 211 articles examined, 131 (62.1%) reported the proportion of variance explained of at least one of their statistical tests for a total of 681 reportings of a proportion of variance explained in a statistical test. The range of this distribution of proportions went from one proportion reported to 15 reported. The mean of this distribution is 5.3 and the median is 5 reportings.

As for the strength of those figures of proportion of variance explained by each of those 681 tests, the range went from a low of zero (reported by 42 tests) to a high of 84%

(reported in one test). The median of this distribution was 8% of variance explained, with one quarter of those reported tests explaining 3% or less of the variance. On the high end, one quarter of those tests reported more than 17% of the variance.

IV. Big Picture Patterns

The reporting of the findings on the individual variables above shows some indications that patterns are shifting toward using stronger design options, although there is still considerable use of weaker options. The terms "weaker" and "stronger" refer to the comparative ability of design options to generate knowledge about media effects that is even more useful. For example, when designing a sample, researchers have two options -- representative samples and non-representative samples. Both types of samples are useful in generating data that can be used to describe patterns in the sample, but with a representative sample, researchers can also use inferential statistics to estimate the confidence level of their descriptions as reflecting patterns in the populations they represent. Thus representative samples are stronger than non-representative samples. Weaker design options themselves are not necessarily faulty because they still have value in generating knowledge about media effects, although researchers can use a design decision in a faulty manner, such as when they use a non-representative sampling procedure but then generalize their findings beyond their samples.

It is time to consider some explanations for why these patterns persist. In this section, we will first analyze the patterns of weaker design options in the context of theory use and strength of findings. Then we analyze the patterns of methodological decisions in order to determine the assumptions that underlie the selection of the most prevalently used design options.

A. Theory-Methods-Findings Nexus

The three main findings of this study are: (1) the majority of published studies continue to avoid using a theory as a foundation, (2) there are patterns of design decisions that indicate the selection of weaker options over stronger ones, and (3) the proportion of variance explained by most studies is fairly small. Perhaps these three patterns are related to one another such that authors who use a theory as a foundation for their studies are more likely to select stronger design options which then lead to stronger findings as represented by explaining a higher proportion of variance.

It is reasonable to expect a nexus among these three characteristics. The use of theory as a foundation for empirical studies should be expected to guide study designers away from selecting options that previous tests of the theory have been found to be less useful and toward the use of methods, samples, and measures that have been found to be more useful and valid. Because theories provide a structure for programmatic research, designers of theory-based studies should be more likely to focus on the most promising concepts and propositions thus increasing the explanatory value of each subsequent study, so theory-guided research studies should be expected to explain a greater proportion of variance in their findings. That is, when studies are constructed with stronger design options, those studies should generate findings composed of smaller proportions of random error and therefore explain greater proportions of variance.

We found patterns to suggest some support for this theory-design-findings nexus. Table 1 displays an analysis of methodological patterns by the role of theory in the design of the published studies. When we compare the percentages of the middle two columns, we can see that authors who used a theory to deduce their hypotheses also selected better design options compared to those authors who did not use a theory. That is, authors who were guided by a theory were more likely to use a representative sample compared to those who did not use a theory (28.0% to 20.5%); were more likely to provide support for the validity of their measures (57.6% to 39.3%); and were more likely to compute change scores (16.9% to 11.8%). Also, authors of theory-based studies were more likely to design experiments (49.2% to 41.9%), to randomly assign participants to conditions (64.3% to 51.0%), and to conduct a manipulation check (28.6% to 24.5%). They were also more likely to avoid making design decisions based on faulty assumptions as reflected in being less likely to use attribute variables as surrogates for active influences (37.3% to 48.7%) and being less likely to use self reports of mundane behaviors (62.7% to 65.5%). While these comparisons indicate a relatively consistent pattern of better design options compared to studies that did not use a theory as a foundation for their studies, the differences themselves are small and none are large enough to be statistically significant in our tests. Although it is likely that those differences might still

exist if we were to increase the power of our test to a point where those differences would become statistically significant, we argue that statistical significance is a secondary concern to the more primary concern of consistency, which is an argument also emphasized by other scholars (Abelson, 1985; Gigerenzer & Marewski, 2015).

When we move on to the nexus' third component -- proportion of variance explained -- the expected inter-relationships do not appear (see Table 2). It is puzzling to see that while authors of theory foundation studies are more likely to report figures indicating the proportion of variance that their tests explained (5.77 to 4.96), those median proportions were lower (9.8% to 14.5%) in studies using theory compared to those that did not. Also, when we analyze the differences in proportions of variance explained by individual design decisions, we see that the proportion of variance explained does not differ significantly on any of these design decisions. For example, the 21 studies that used a representative sample

had a mean (.131) virtually the same as the group of 109 studies that did not use a representative sample (.139). However, when interpreting this finding, we must also consider that almost 38% of the analyzed studies did not report any proportions of variance explained and that many of these were non-theoretical studies; if these authors had reported their proportions, perhaps the means of the non-theoretical group would have been much lower.

Thus the nexus argument appears to have some value. That is, scholars who use a theory as the foundation for designing their studies show pattern of being slightly more likely to select stronger design options over weaker ones. However, while these theory-foundation authors are more likely to report slightly greater proportions of variance explained, the mean of those proportions is lower than the mean of the distribution of proportions reported by non-theory foundation authors.

Table 1. Uniformity of Patterns Across Theory Usage

Data Gathering	All	No Theory	Theory Deduced Hs	Tested Model			
Sample Represent	19.4%	20.5%	28.0%	16.6%	X2 = 0.07	df = 2	p = .701
Attribute Surrogate	43.6%	48.7%	37.3%	37.1%	X2 = 2.80	df = 2	p = .247
Behavior Self Report	64.8%	65.5%	62.7%	65.7%	X2 = 0.15	df = 2	p = .927
Validity Support	46.0%	39.3%	57.6%	48.6%	X2 = 5.41	df = 2	p = .067
Change Scores	12.7%	11.8%	16.9%	8.6%	X2 = 1.57	df = 2	p = .456
Experiment Method	45.0%	41.9%	49.2%	48.6%	X2 = 4.14	df = 4	p = .387
Random Assignment	55.3%	51.0%	64.3%	52.9%	X2 = 1.32	df = 2	p = .518
Balance Check	13.8%	14.3%	10.7%	17.6%	X2 = 0.44	df = 2	p = .801
Manipulation Check	27.7%	24.5%	28.6%	35.3%	X2 = 0.75	df = 2	p = .686
Variance Expl - Num	5.24	4.96	5.77	5.25	F = 1.094	df = 2	p = .338
Variance Expl - Med	13.8%	14.5%	9.8%	18.2%	F = 2.604	df = 2	p = .078

(back to text)

B. Analyzing Practices for Supporting Assumptions

We now shift away from testing for a nexus of patterns across the findings on our individual coding variables and attempt to dig deeper into the findings by examining assumptions in three areas: Assumptions traceable to foundational beliefs about research, assumptions that have been found to be faulty, and assumptions about the continuing value of exploratory research (see Table 3). The assumptions in the first category cannot be considered right or wrong by any objective standard because they are traceable to scholars' beliefs about ontology and epistemology. However, the assumptions in the other two categories do have standards that can be used in a relatively objective manner to determine whether they are faulty.

1. Assumptions attributable to research perspective.

There are some research practices that will trigger a

Table 2. Strength of Findings by Design Decisions

		Mean	S.D.	n		
Representative Sample:	No	.139	.144	109		
	Yes	.131	.130	21	$F = .053$	$p = .818$
Attributes as Surrogate:	No	.142	.150	67		
	Yes	.133	.133	61	$F = .123$	$p = .726$
Mundane Behavior Self Report:	No	.123	.108	47		
	Yes	.147	.158	81	$F = .017$	$p = .358$
Validity Argument:	No	.125	.120	62		
	Yes	.149	.149	66	$F = .931$	$p = .336$
Change Scores:	No	.130	.132	112		
	Yes	.189	.194	16	$F = 2.420$	$p = .122$
Exp - Random Assignment:	No	.145	.177	27		
	Yes	.115	.109	41	$F = .749$	$p = .390$
Exp - Balance Check:	No	.127	.142	60		
	Yes	.129	.127	8	$F = .002$	$p = .965$
Exp - Manipulation Check:	No	.134	.153	50		
	Yes	.110	.096	18	$F = .372$	$p = .544$

Table 3. Analyzing Assumptions Underlying Selections of Methodological Options

Assumptions Attributable to Research Perspective

- 45.3% of experiments indicated no random assignment of their participants to treatments

  Assumption: Participants are fungible

- 86.3% of experiments reported no balance check on the assignment of Ss to treatments

  Assumption: Treatment groups are matched on key characteristics

- 72.6% of experiments reported no manipulation check of their treatments

  Assumption: All Ss within each treatment group perceive same stimulus

  Assumption: All Ss within each treatment group perceive stimulus as intended

- 87.3% measured the effect variable at only one point in time

  Assumption: In surveys, cross sectional data is adequate to suggest change.

  Assumption: In experiments, post exposure data is adequate to suggest change, even when those single point measures from individuals are averaged to form group means.

Faulty Assumptions

- 56.7% reported inferential statistics without providing a basis to warrant their use.

  Assumptions: Inferential statistics do not require the use of representative samples in surveys or random assignment of participants to treatment conditions in experiments.

- 64.8% measured mundane behaviors with self reports

  Assumptions: Ss either provide accurate recall of details in their responses, or Ss are all using the same heuristic to construct their responses

- 43.6% reported using attribute variables as surrogates for active influences

  Assumption: Surface characteristics are valid indicators of active factors

- 54.0% made no attempt to establish a case for validity of measures

  Assumption: All items are measuring what authors intend them to measure

difference of opinion across scholars with some scholars believing that those practices exhibit the selection of adequate -- or even strong -- design options while other scholars believe that those same practices are not adequate. For example, let's examine the design of experiments and focus on decisions about random assignment of participants to treatment groups, checking for balance, and checking the manipulation. Within the set of 95 experiments we examined, 54.7% reported randomly assigning participants to their treatment conditions; 13.5% conducted a balance check; and 27.4% reported using a manipulation check. Taken together, these three practices can be regarded by some scholars as indicators of flawed designs. In contrast, other scholars might not be bothered by these practices and regard the techniques of random assignment, balance check, and manipulation check as not needed in order to design strong experiments. What could account for such a difference in judgment? The answer is likely to be scholars' fundamental beliefs concerning ontology (i.e., the nature of the phenomenon being studied) and epistemology (i.e., humans' ability to generate knowledge about the phenomenon). To illustrate this difference in beliefs, I will present three perspectives on media effects research that demonstrate how a legitimate difference in foundational beliefs can lead to justifying, or criticizing, design choices.

Perspective One is characterized by a mechanistic type ontological belief where scholars regard humans as primarily physical entities, much like any other physical entity such as ears of corn. For example, all ears of corn are the same in terms of needing the same nutrients, they all grow the same way and on the same schedule, and they all produce kernels that are the same. While there may be some ears of corn that have a few more kernels than other ears, that variation is trivial. Some scholars regard humans as primarily biological and chemical systems that all require the same nutrients, follow regular patterns of growth and maturity, have the same organ systems and chemical make-up, have a brain that is hard-wired to process information in a standard manner, etc. These scholars acknowledge that while humans may vary a bit in some characteristics, those variations are trivial; therefore all humans are regarded as essentially fungible.

Experimenters who hold this set of Perspective One beliefs would perceive no need to randomly assign participants to treatment conditions nor to check for balance, because they believe that all humans are interchangeable. Researchers who believe that all humans are fungible would auto-

matically conclude that all groups would be equivalent regardless of how they were assembled. Therefore conducting a balance check or randomly assigning participants to treatment groups are regarded as tasks that have costs without benefits. And once experimenters assume equivalent treatment groups, the only explanation possible for a finding of differences in group means is that the participants in the control group did not change but that the participants in the treatment groups did change. Therefore there is no reason to measure participants' values on the outcome variable before they experienced the treatment.

Perspective One researchers also perceive no need for manipulation checks because they believe that the characteristics they build into the messages in their treatments will all be perceived uniformly by all participants due to the hard-wired nature of human perceptions and the fairly standard ways humans attribute learned meaning to symbols. Furthermore, these researchers believe that because they too are human and using the same perceptual and cognitive processes as their participants, the characteristics they perceive in their designed treatments will trigger the same perceptions in their participants.

Perspective Two is characterized by a belief that while humans are alike in many ways (e.g., as organic physical systems), humans are also different in other ways (e.g., inborn trait differences and broad scale socializing influences). However, these differences are less individual and more category-based such that people differ across categories but within a category, they are very similar. Stage theories of human development are evidence of this thinking where all children at age 3, for example, are regarded as having the same cognitive and emotional abilities explained by that stage of development and that all 3 year olds are very different than all 6 year olds because those older children all exist in a category of humans with a higher level of cognitive and emotional development. Because all humans within a category are believed to be similar, there is no need for random assignment, a balance check, or a manipulation check *as long as all participants are from the same category*. This belief would also explain why so much of the media effects research focuses on categories such as demographics (sex, age cohort, SES, educational level, income level, etc.) and levels from stage theories.

A third perspective is characterized by a belief that in some ways humans are all alike, in other ways humans differ by categories, and in other ways humans differ in impor-

tant ways individually. These scholars search for evidence that humans are fungible for a certain kind of media effect and when they find it, they are comfortable assigning participants to conditions without going to the trouble of random assignment and assume with confidence that there will be equivalency across groups. However, when researchers are dealing with characteristics of humans that exhibit a great deal of variation across individuals, study designers realize they can have little confidence in group equivalency unless they randomly assign participants to groups. Furthermore, these researchers are likely to be skeptical about their ability to generate equivalent groups *even when using random assignment*. To illustrate, let's take the example of how people react to violent messages in the media. Past research (for a review see Potter, 1999) has shown that people's interpretation of the violence is influenced by dozens of factors about the message (degree of gore, degree of justification for the act, type of perpetrator, type of victim, portrayal of degree of harm, etc.), factors about the viewer (emotional maturity, socialization to aggress, history of bullying, moral training, range of intensity of physiological reactions, etc.), and factors about the exposure situation (peer pressures, authority demands, etc.) These scholars also acknowledge that sets of these factors work together to exert their influence rather than acting alone and independently. Thus, even if we consider only half a dozen of these facts and simplify each of these six factors to three values each, that would compute to 729 factor combinations. Let us consider that we want to design a two group experiment (control and treatment) with 30 participants each. The probability of finding one person's configuration on these 729 factor combinations in the control group to be an exact match to a person in the treatment group would be very small, and the probability of matching all 30 people in each group would be infinitesimal. Thus, while using random assignment would increase the probability of achieving matched groups, that increase would not be large enough to remove all skepticism about non-equivalency. Is this a hopeless situation? Not necessarily -- if we removed the *need* for equivalent groups. The way to do this would be to reconceptualize the value of experiments by focusing on two beliefs: (a) the belief that humans are individually different in their sensitivity to media effects because of their past experiences and reinforcement patterns, and (b) the belief that the process of influence with any effect involves multiple factors in complex interactions. Given this conceptualization of humans and the effects process, experimenters

would need to insure three design features: (a) a measurement of participants' key characteristics in order to plot their sensitivity to different effects (b) building in a recognition of the constellation of factors likely active in the process of influence with the particular effect being studied, and (c) a pre-treatment as well as post-treatment measure of the outcome variable. Experimenters can still separate participants into treatment groups but they need not assume equivalency because the group is no longer the unit of comparison, so group equivalency is no longer important; instead, the unit of comparison is the individual. Researchers can still design differential treatments to focus on a particular message element to see if varying that element across treatments contributes to an effect. However, they would not be limited to comparing group means and would have many other more valuable options for examining how the featured message element interacted with other factors to explain the degree of effect, rather than simply reporting whether there was a group difference or not.

Perspective Three requires higher costs in the design and execution of research studies. Designers need to be more careful in analyzing the findings in the published literature to identify larger sets of active factors in the process leading to the effect they are studying. They need to design a battery of measures to be able to plot their participants throughout a process of influence leading to the manifestation of the effect being studied. While all this additional design work requires higher costs to research designers, it also delivers a much higher payoff in the form of a much richer context for understanding the findings of any experiment. In this way designers could deliver much more conceptual leverage to the field. Even more importantly, this would move us away from a reliance on relatively simple designs based on faulty assumptions to study complex phenomena.

If our purpose is to grow our knowledge about media effects, then we need to grow our research designs beyond the limitations that keep the ceiling on our understanding lower than it needs to be. Compared to a design that measures the effect variable only once, designs that measure the effect variable at multiple times are superior because they can document -- rather than assume – change. Moreover, they can plot the shape of that change over time to determine if the change is a short blip or a longer term alteration that can build in strength or decay over time.

2. Faulty assumptions.

Four of the research practices we found seem to be traceable not to a legitimate difference in beliefs but to an acceptance of faulty assumptions. The first of these is the widespread use of inferential statistics regardless of whether they are warranted or not. Except for the few studies in our sample that used a qualitative methodology, all the studies used inferential statistics. However, only 19.4% relied on representative samples. When researchers conduct surveys with nonrepresentative samples, they typically demonstrate a belief that their use of inferential statistics can produce meaningful results, but this belief is faulty (e.g., Babbie, 1992; Gigerenzer & Marewski, 2015).

Although few experiments use representative samples, scholars have argued convincingly that inferential statistics can be used in experiments as long as participants are randomly assigned to their treatment groups (Courtright, 1996; Kruglanski, 1975, Lang, 1996; Sparks, 2015). Lang (1996) points out that because experimental researchers are not trying to generalize their findings to larger populations but instead are focused on "attempting to determine if some variable (often called the treatment variable) is the cause of some effect" (p. 425), they can use inferential statistics to determine if the means on outcome variables across treatment groups are large enough to be statistically significant *as long as the participants were randomly assigned to those treatment groups*. "The procedure of random assignment of subjects to treatment conditions not only results in equal groups but also forms the conceptual basis for the calculation of the theoretical distribution of all possible random assignments on which experimental inferential statistics are based" (p. 425). Given the requirement that researchers must randomly assign participants to treatment groups in order to establish a basis for using inferential statistics, it is a concern that we found that all the experiments we examined used inferential statistics although only 54.7% of those authors stated that they randomly assigned their participants to treatment groups.

It is puzzling that so much of the use of inferential statistics is unwarranted. Among experiments, we found that only 52 out of 95 studies reported random assignment and that among surveys only 36 out of 108 studies used a representative sample, which means that only 88 out of 203 studies (43.3%) presented an adequate basis for the use of inferential statistics. Why are reviewers and editors of top scholarly journals so consistently willing to overlook this problem? The posing of this question, is not meant to imply that the research that displays this fault -- or any of the other short-comings -- is worthless. Instead, there is potential value offered by all authors who present clear descriptions of valid patterns they found in their samples. The problem lies in the practice of using inferential statistics when there is no basis for doing so and for then using those specious p-values as thresholds for claiming whether the descriptive statistics are significant or not. This practice has led some scholars (Abelson, 1985; Gigerenzer & Marewski, 2015) to criticize researchers for their automatic, mechanical, and even compulsive use of inferential statistics without considering whether their use is warranted or not.

We need to pay less attention to p-values, which have no meaning in most media effects research, and much more attention to the strength of findings. And we need to move past the simple grouping of statistical results into significant findings of non-significant findings solely on p-values and think more about the significance of findings in terms of their degree of meaningful contribution to our knowledge about media effects.

Another widespread practice based on a faulty belief is the use of attribute variables as surrogates for active influences. Over time as the field has identified an increasing number of active influences in the media effects process, the marginal utility of using an attribute variable as a surrogate has diminished to a point where this practice no longer has value.

Perhaps the most puzzling practice is that in almost half of the published studies of media effects, the authors have made no attempt to convince readers of the validity of their measures. Scholars have a large literature showing a variety of measures on many variables. This literature demonstrates that with any given variable, there is likely to be a range of measures that vary in terms of how much error variance they generate, how strongly they are related to other constructs that they should be related to, and that the scaling of some measures are much more reliable than the scaling of other measures. There seems to be no justification for authors not showcasing a critical analysis of the relevant measurement literature to support their decision to use particular measures from other studies or to support their decision to design their own measures.

3. Continuing value of exploratory research.

What can account for the selection of the weaker options and the use of faulty assumptions to justify them? The reason

is not likely to be ignorance because social scientists are taught early in their training about basic requirements for sampling, measurement, and data analysis. For some design choices, cost may be an explanation. For example, it is much more costly to generate a representative sample than to use a convenience sample. Therefore, cost is a likely explanation for why the literature has so many more studies that use non-representative samples compared to the number that use representative samples. But cost does not also explain why survey researchers who use a non-representative sample then also use inferential statistics. Also, while greater costs may explain why designers cut corners, it does not explain why editors of scholarly journals as well as reviewers remain comfortable publishing studies designed with weaker options based on faulty assumptions.

Perhaps there is a widespread belief that our literature only needs to be exploratory and does not need to grow beyond those limitations. This belief would explain why researchers, editors, and reviewers are continually satisfied with the patterns of weaker design features because findings only need to be able *to suggest what might be* media effects and factors of influence rather than to be able to make more definitive statements about *what is*. A research literature that provides possible suggestions would not be concerned as much about patterns of weak design decisions compared to a literature that is more oriented toward building defensible elements of knowledge.

When the field of media effects was new, exploratory research was the only option because the field had no body of knowledge to reveal that some assumptions were faulty. It had no history of design and measurement to inform researchers about which practices were stronger or weaker. It had no clear picture of what a media effect was, how many there were, or the possible factors that might be influencing any effect. So any research study was able to contribute findings of value about what might be a media effect or a factor of influence. Now that we have generated so much suggestive evidence for many different effects and many different possible factors of influence, the challenge has shifted to providing answers to the following kinds of questions: How many effects are there and how are they organized to work together? What are the most powerful effects? What are the most widespread effects? What are the effects that the greatest number of people experience? What are the most powerful factors of influence from the media? Although these are extremely important questions, we have yet to focus on them

directly by designing the kinds of research studies that would be required to generate useful answers to them. To continue to evolve as a scholarly field, we need to shift our focus as we recognize how our challenge is evolving.

V. Recommendations for Moving Forward

The findings of this study form the foundation for three sets of recommendations. First, we present a set of recommendations addressing the need to evolve away from the continued use of design options that are based on faulty assumptions. However, these recommendations have little chance of being enacted unless we also evolve in our thinking about theory as well as critically examine our paradigmatic beliefs.

A. Methods Recommendations

It is tempting to recommend that scholars move towards selecting stronger design options in place of weaker ones across the board. However, it would be foolhardy to expect a sudden revolution where each of these recommendations would be adopted and where methodological weaknesses would suddenly be eliminated. Instead, it is essential to recognize that the existing practices are entrenched, so it will take a gradual evolution for researchers to wean themselves from the habits that keep these practices continuing. Thus we present our methods recommendations in stages so that the patterns that appear to be easier to change can be altered first. Those changes that are likely to require relatively little effort to make while delivering relatively large increases in quality are labeled "higher leverage changes." Other methods recommendations will require much more effort to change, and although these changes will also likely deliver a large difference in quality, they will encounter greater resistance. We label these "lower leverage changes."

Recommendations with higher change leverage.

There appear to be four changes that can be made with relatively little effort in comparison to the increases in quality they will contribute to the research literature (see Table 4). First, we recommend that scholars continue the trend toward reporting the proportion of variance explained. There is almost no cost to doing so, because it is easy to request these proportions when running statistical packages. The benefits

Table 4. Recommendations by Costs and Benefits

Recommendations	Costs	Benefits
Report proportion of variance explained	Low; statistical packages easily provide this	High; places focus on power of explanation instead of p-values
Present arguments for validity of measures	Low; large literature of measures already exists	High; reduces use of measures of low validity
Measure influences with active variables instead of attribute variables as surrogates	Moderate; measuring active variables is more challenging	High; greatly increases validity of data
Use electronic recording of mundane behaviors instead of self reporting	Was high but has been shrinking with newer technologies	High; greatly increases validity of data
Measure effect variable at more than one point in time to document change	High; considerable challenge in avoiding threats to validity	High; provides much more direct measure of magnitude of effects
Increase usage of probability samples	High; difficult to generate acceptable response rate	High; essential for answering questions about prevalence, etc.
Increase use of theory as foundation for studies	High; to break entrenched exploratory perspective	High; gain efficiencies in designing studies and integrating findings

<div align="center">In experiments:</div>

Recommendations	Costs	Benefits
Randomly assign Ss to treatment conditions	Low unless dealing with intact groups	High; provides requirement for using inferential statistics
Check for Ss balance across treatment groups on active variables	Moderate; requires more measurement	High; documents validity of assuming equivalent groups
Conduct manipulation check to document what Ss perceive as stimulus	Moderate; requires more measurement	High; documents validity of assuming uniformity of treatment

(back to text)

will be substantial, especially if this practice helps to shift researchers' basis for claiming the significance of their findings away from using arbitrary thresholds of p-values and toward showing an increase in the proportion of variance explained compared to previous studies on the same topic. Moreover, a change in focus on significance is likely to lead to stronger research designs that would reduce the so-called error variance, which is likely to be generated by weaker design decisions. The most important indicator of the health of our field as a social science is the continual increase in our ability to explain the variance we generate in our empirical studies.

Second, we recommend that more designers of tests of media effects present arguments for the validity of their measures. There appears to be no justification for ignoring this task. The cost of completing this task is relatively low because there are many published studies on virtually every variable of interest to media effects scholars, and these studies can be easily identified through electronic searches. Furthermore, the advantages are substantial, because this practice would encourage designers to become more aware of the relative quality of existing measures when selecting measures. Further, it would discourage designers from adding to the proliferation of measures when quality ones already exist.

Third, we recommend continuing the trend of measuring influences on media effects with active variables and eliminate the use of attribute variables as surrogates. The cost of making this change has been decreasing as the literature increases in size with a corresponding increase in information about measures of active influences, especially with measures of gender role socialization (in place of biological sex) and human development in many areas (cognitive, emotional, moral, etc.) in place of chronological age.

Fourth, we recommend that more experimenters randomly assign their participants to conditions. With laboratory experiments, there is typically no cost to doing this, and the benefits are substantial. With field experiments that require the use of intact groups, we recommend that researchers use the groups instead of individuals as the units of analysis and randomly assign groups to conditions. This procedure may increase costs if new intact groups must be found in order to create a sample with enough units of analysis, but the increase in units will substantially increase the value of the research study's findings.

Recommendations with lower change leverage.

We present six additional recommendations for changes that will take more effort, so the leverage between costs and benefits is not as great as with the four recommendations in the previous section. However, each of these would deliver a substantial gain in quality of the research design.

Three of these recommendations encourage the use of additional measures over what are typically used. First, we recommend multiple measures of the media effect variable spaced out over different points in time so that change can be documented instead of assumed. Second, we recommend that designers of experiments include measures to check for balance on important participant characteristics. Third, we recommend that experimenters check their manipulations to insure that participants perceived the stimulus as the designers intended.

Building these additional measures into research designs increases the threat of reactivity of measures. However, using multiple measures should be viewed less as a barrier and more as a challenge, because there are ways to avoid those potential threats, although their costs are higher (Podsakoff, MacKenzie, Jeong-Yeon, & Podsakoff, 2003). Campbell and Stanley (1963) have provided detailed suggestions to minimize eight different threats to internal validity and four threats to external validity in the design of experiments. For example, when designers of experiments decide to measure the outcome variable not just after the treatment but also before the treatment so that they can document change, they run the risk of the pre-treatment measurement sensitizing participants to the treatment itself and this is a potential threat to the validity of their findings. However, if researchers construct a Solomon Four-Group Design, as Campbell and Stanley (1963) suggest, they can determine whether this potential threat actually appears, and if so, they have the means to remove its influence in the analysis of data. This design increases the cost because researchers would need to include four groups for each treatment in their experiment.

Design choices force researchers to consider whether keeping costs low is more valuable than increasing the precision of their planned study and the potential validity of their results. Such a choice has implications beyond the design of any one study and has a cumulative impact on the overall field. Given what we know about the field's modest ability to explain variance and its tolerance for weaker design decisions, it would appear that our pressing challenge is to design studies that can explain a greater proportion of variance.

Valkenburg, Peter, and Walther (2016) remind us that because media effect sizes are typically so small, it is essential that researchers carefully consider all threats to validity and make design decisions that would minimize those threats as much as possible. When we let costs be the primary driver of design decisions, we will likely continue with patterns of weaker designs, and this will serve to institutionalize a lower ceiling on the validity, power, and usefulness of our findings.

Our fourth lower leverage recommendation is to encourage designers to shift away from relying so heavily on non-representative samples and be more willing to accept the challenge of generating representative samples. This challenge comes with the higher costs involved with constructing an adequate sampling frame of their population, randomly selecting units from that sampling frame, and insuring an adequate response rate. These can be very challenging tasks, especially insuring an adequate response rate. Until we can do a better job of meeting the challenges of designing studies that will allow for generalization and inferential statistics, we keep a narrow scope on the kinds of questions our research studies can answer.

Fifth, we recommend the substantial reduction and eventual elimination of the use of self reports as measures of mundane behaviors. Prior to the development of technologies that define the new media environment, there were few alternatives to relying on self reports of mundane behaviors. The main alternative was using observers to follow people around in their everyday lives, and the cost for this was typically high. However, now with most people equipped with mobile devices they carry everywhere, there is a range of alternatives that vary in costs that are lower to substantially lower than using observers. Mobile devices (smartphones, tablets, laptops) as well as products that people use everyday in mundane ways (cars, household products, etc.) consistently keep electronic records of the behaviors of people who use them. Also, a wide variety of companies now routinely gather information about media use (Google, telephone companies, cable TV provides, websites, etc.).

The challenge no longer lies in designing ways to gather information on mundane behaviors but in (a) getting permission to use some of the massive amount of data that is already being generated every day and (b) analyzing those data. While it is costly to get access to some of these databases (such as from professional media usage monitoring firms like A. C. Nielsen), access to other databases is free (such as Google Analytics). And in situations were no database exists

to meet researchers' particular needs, those researchers now have the low-cost capability of texting participants at random times to ask them to report their specific behavior at that time. Researchers could ask people who agree to participate in studies to download keystroke recorders on laptops or to grant permission to collect data from the existing surveillance software (such as built-in cameras) on phones and laptops. There may be public resistance to allowing researchers to collect such data, however, given the public's demonstrated acceptance of the willingness to pay the price of an invasion of privacy as payment for the use of these devices, the population may have a much higher tolerance for others to observe and record their behaviors.

Perhaps the most significant challenge in using electronically recorded documentation of mundane behaviors is designing ways of looking for meaningful patterns in those huge data bases. In order to engage this challenge, researchers will have to think more deeply about why they want to collect these data. Also, researchers will have to conceptualize media exposure more precisely, for example, in deciding whether a click on a website counts as exposure, whether there are thresholds in time (does a one-second scan of a website news headline count as exposure), and when a sequence in web page surfing is meaningful. These are essential conceptual issues that have been finessed in the past, so confronting a challenge that requires more scholarly work would seem to be a positive development.

Sixth, we recommend that designers use theory as a foundation for empirical tests of media effects. This recommendation is likely to require a relatively high cost from scholars who are unaware of the many theories available and therefore must sift through the hundreds of available media effects theories to find those of most relevant to their interests. However, once scholars make this investment, they then are able to achieve the efficiencies offered by the selected theory. Those efficiencies include guidance in deducing hypotheses from the theory's propositions, selecting the best measures, using appropriate analyses of data, and contextualizing the findings.

Even when atheoretical researchers engage in a line of programmatic research, which appears to be rare (Lowery & DeFleur, 1988) and can learn from their mistakes in design, they are still at a disadvantage compared to researchers who design their tests with a theoretical foundation. This is because non-theory driven empiricism provides no basis for falsification, as Karl Popper (1959) has pointed out. A test

that falsifies a proposition in a theory offers a more useful finding than does a test that supports a proposition, because weak findings allow theoreticians to carve away the parts of their systems of explanation that have been found not to work and instead concentrate more on those parts that offer greater potential for explanation. Thus over time, theories can offer a greater degree of guidance that maps out where the more promising avenues of research are and thereby directs researchers to design studies with more potential to explain much higher proportions of variance. This opportunity is especially underutilized in media effects research where so many studies fail to explain more than a tiny proportion of variance.

B. Theory Recommendations

In the above section, we recommended the use of theory as a foundation for empirical tests of media effects, but that recommendation has little utility if there are no theories or if the existing theories provide researchers with little guidance. We know the first is not the case, that is, there are hundreds of media effects theories available, but there may be a serious question about the second -- that the existing theories are useful as guides in designing empirical tests. Given the persistently small proportion of the media effects literature that is guided by a theory, it appears that most research designers do not believe that the available theories are useful.

We need a rigorous analysis of the existing theories of media effects in order to determine if this low use of theory is due to a misperception by study designers or whether the theories themselves are indeed weak in their ability to guide research. Therefore we recommend that active researchers be interviewed to find out what their beliefs are about the usefulness of theories. We also recommend that the existing theories be critically analyzed to determine their potential usefulness by engaging in the challenge to answer questions in two areas.

The first area includes questions about the current value of media effects theories as guides that are keyed to clarity and completeness. Do the available theories provide clear enough definitions of their concepts so that researchers can easily operationalize measures and research procedures? Do theories provide a set of propositions that form a system of explanation or do the propositions instead appear as single isolated ideas or as a random list with gaps?

The second area of questions focuses on the openness and integrative nature of theories. Do the theoreticians keep up with the growing literature -- both of direct tests of their theories as well as contiguous literatures -- in order to provide critical analyses rather than simple descriptive inventories of assorted findings? To what extent have media effects theories altered their systems of explanation to respond to tests of falsification? That is, have theoreticians altered their initial propositions to make them less general and more contingent? As research tests accumulate, theoreticians learn that their initial propositions might not apply to all people, in all situations, with all media content, so these propositions need to be scaled back from their general claims in order to eliminate those types of people, situations, and media messages that do not apply. Theories that do this well offer more guidance to researchers by directing them to focus on more promising avenues of explanation, and this strategic re-direction should result in a progression in the proportion of variation they can explain. Furthermore, do theoreticians continually update their calculus, that is, the recommended measures and designs? To what extent have media effects theoreticians been interested in examining the epistemological assumptions that underlie their theories? Are the theoreticians aware of these assumptions, and if so, are they willing to use what they learn from the literature to acknowledge when an assumption has been found to be faulty? And ultimately, are theoreticians willing to alter their theory to direct researchers away from initial parts of theory that were supported by assumptions found to be faulty and re-orient those researchers towards other parts of theory that have a stronger foundation?

Theoreticians who regard their theories as fixed and spend their careers defending their initial conceptualizations can be admired for their initial creativity and continued tenacity if the empirical literature of tests substantially support those initial claims. However, rarely is a literature so uniformly supportive; instead, empirical literatures typically display equivocal and contradictory findings that stimulate the need for constant re-examination in order to ferret out weaker conceptualizations and operationalizations in favor of stronger ones that can progressively increase the theory's explanatory value. Scholars who regard their theories as a tentative step in a progression towards more insightful and powerful explanations will make more valuable contributions to their fields. And theoreticians who view their role as less of an fixed authority and more as a guide to researchers

will make more useful contributions to other scholars who struggle to design better empirical tests that will produce more valuable results to the field. This kind of theoretician is needed for our field to evolve out of an initial exploratory phase.

C. Paradigm Recommendations

Our most fundamental recommendation is that scholars in the field of media effects continue moving from an exploratory perspective into a more explanatory perspective. This is most fundamental, because until this movement reaches a critical point, exploratory studies will continue as the dominant form of research, and the field will continue to labor under a low ceiling because the exploratory perspective provides an easy justification for researchers to select weaker design options.

The exploratory phase is an essential first step in the development of a scholarly field. When a scholarly field is new, researchers have little guidance from the literature to help them figure out how to assess their focal phenomenon. Researchers are limited to designing exploratory studies to start building an inventory of rudimentary ideas. They must focus on identifying what is possible, so almost every study regardless of topic examined, methods used, or strength of findings is likely to make a contribution. When designing these initial exploratory studies, researchers must rely on many untested assumptions as support for their decisions. Over time as a research field grows in size and as knowledge accumulates, many of those assumptions are found to be faulty. Recognizing a faulty assumption forces researchers into a dilemma as they are confronted with the decision of (a) staying on the same familiar path with its diminishing ability to generate findings that would increase knowledge about the field's focal phenomenon or (b) taking the risk of making substantial changes in one's thinking and practices for the opportunity to generate much more meaningful knowledge. Kuhn (1970) has referred to this dilemma as the essential tension in the development of any scholarly field. Scholars are confronted with a choice between two paths forward essentially must choose between comfort of efficiency and the challenge of becoming more effective.

With media effects research, the efficient path is a continuation of exploratory studies where it is much easier to remain using familiar options even though the assumptions used to support those options have been found to be either faulty or to be based on beliefs that have long since lost their usefulness. Given the size of the literature on media effects, the marginal utility of designing another study that is limited to suggesting another media effect that has not already been suggested has dwindled to a trivial point.

In contrast, progress down the explanatory path entails higher costs but it also offers much higher benefits in terms of generating useful knowledge in under-addressed or non-addressed areas as indicated by the following questions. Which of the suggested media effects are the strongest, the most prevalent (i.e., occur most often), and have the greatest scope (i.e., likely to effect the greatest number of people)? How can all these suggested effects be organized in a meaningful way so that we do not call the same thing by different names or use the same name for very different kinds of effects? Which of these effects work together in terms of existing simultaneously or in an unfolding progression over time? Which factors of influence are the strongest (most influential) and the most prevalent (leading to many different kinds of effects)? What is the nature of the relationships among these factors (simple relationships, non-linear, asymmetric, thresholds, ceilings, etc.)? And how do these factors work together in bring about media effects (straight line, recursive, indirect, etc.)?

The field of media effects will continue to grow in terms of attracting scholars and generating a greater number of studies, but in order for the field to also grow in terms of increasing useful knowledge about the nature of media effects, then scholars will have to evolve in their thinking about the nature of the field and in their practices in designing research studies. Such an evolution begins with individual scholars examining the beliefs they take for granted, identifying which are faulty, then displaying the courage to reject those beliefs that have been found to be faulty. In his book *Basic dilemmas in the social sciences*, Blalock (1984) says "the more information that is missing, the more untested assumptions we have to make in order to compensate." He continues, "whenever one is in doubt about an assumption, the temptation is to hide it from view possibly by using vague language or simply playing it down by embedding it in a number of innocuous assumptions or a technical discussion that most readers are unlikely to follow" (p. 135). The findings of the current study show that many of the assumptions that had value in allowing us to make progress in an initial exploratory phase of building the field of media effects have lost their value and now form barriers that slow continued

progress.

When we reject the assumptions we relied on in the exploratory phase of research, we can orient more toward greater precision in measurement and design, which will reduce the opportunities for error to enter our data. By reducing measurement error and increasing our understanding of the nature of relationships and differences, we can systematically improve our ability to explain media effects. Some of these patterns should be relatively easy to change, but others will require greater costs. However, more important than costs are the benefits. When we focus more on the purpose of our field, the more we will be concerned with eliminating weaker elements in the design of our individual studies so that we can achieve greater satisfaction in our contributions.

References

Abelson, R. P. (1985). A variance explanation paradox: When a little is a lot. *Psychological Bulletin*, *97*, 129-133.

Babbie, E. (1992). *The practice of social research*, 6th edition. Belmont, CA: Wadsworth.

Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230–244. doi:10.1037/0022-3514.71.2.230.

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

Barthes, R. (1967). *Death of the author.* Aspen, 5, 6.

Bartsch, A., (2012). As time goes by: What changes and what remains the same in entertainment experience over the life span? *Journal of Communication*, *62*, 588-608.

Bernard, H. R., Killworth, P., Kronenfeld, D., & Sailer, L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology*, *13*, 495-517.

Blalock, H. M, Jr. (1984). *Basic dilemmas in the social sciences.* Beverly Hills, CA: Sage.

Borgman, C. L. (1989). Bibliometrics and scholarly communication. *Communication Research*, *16*, 583-599.

Boxman-Shabtail, L., & Shifman, L. (2014). Evasive targets: Deciphering polysemy in mediated humor. *Journal of Communication*, *64*, 977–998.

Bryant, J., & Miron, D. (2004). Theory and research in mass communication. *Journal of Communication*, *54*, 662-704.

Bryant, J., & Oliver, M. B. (Eds.). (2009). *Media effects: Advances in theory and research* (3rd ed.). New York: Routledge.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cohen, A. A., & Lemish, D. (2003). Real time and recall measures of mobile phone use: Some methodological concerns and empirical applications. *New Media & Society*, *5*, 167–183.

Courtright, J. A. (1996). Rationally thinking about nonprobability. *Journal of Broadcasting & Electronic Media*, *40*, 414-421.

Dominick, J. R. (1996). Citation analysis of the Journal of Broadcasting & Electronic Media. *Journal of Broadcasting & Electronic Media*, *40*, 427-438.

Fishbein, M., & Hornik, R. (2008). Measuring media exposure: An introduction to the special issue. *Communication Methods and Measures*, *2*(1–2), 1–5.

Fiske, J. (1987). *Television culture.* London, England: Routledge.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.

Funch, D. P., Rothman, K. J., Loughlin, J. E., & Dreyer, N. A. (1996). Utility of telephone company records for epidemiologic studies of cellular telephones. *Epidemiology*, *7*, 299–302.

Gentile, D. A. & Bushman, B. J. (2012). Reassessing media violence effects using a risk and resilience approach to understanding aggression. *Psychology of Popular Media Culture*, *1*, 238-151.

Gigerenzer, G., & Marewski, J. N. (2015). Surrogate science: The idol of a universal method for scientific inference. *Journal of Management*, *41*(2), 421-440.

Haeffel, G. J., & Howard, G. S. (2010). Self report: Psychology's four-letter word. *American Journal of Psychology*, *123*(2), 181–188.

Hindelang, M. J., Hirschi, T., & Weis, J. G., (1981). *Measuring delinquency*. Thousand Oaks, CA: Sage.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341-350.

Kamhawi, R., & Weaver, D. (2003). Mass communication research trends from 1980 to 1999. *Journalism & Mass Communication Quarterly*, *80*(1), 7-27.

King, P. M. (1986). Formal reasoning in adults: A review and critique. In R. A. Mines & K. S. Kitchenor (Eds.). *Adult cognitive development: Methods and models* (pp. 1-21). New York: Praeger.

Kobayashi, T., & Boase, J. (2012). No such effect? The implications of measurement error in self-report measures of mobile communication use, *Communication Methods and Measures*, *6*(2), 126-143, DOI: 10.1080/19312458.2012.679243

Krakowiac, K. M., & Oliver, M. B. (2012). When good characters do bad things: Examining the effect of moral ambiguity on enjoyment. *Journal of Communication*, *62*, 117–135.

Krause, M. S., & Howard, K. I. (2003). What random assignment does and does not do. *Journal of Clinical Psychology*, *59*(7), 751-766.

Krcmar, M. (2009). Individual differences in media effects. In R. L. Nabi & M. B. Oliver (Eds.). *Media processes and effects* (pp. 237 - 250). Los Angeles, CA: Sage.

Kruglanski, A. W. (1975). The human subject in the psychology experiment: Fact and artifact. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 101 -147). New York: Academic Press.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd edition). Chicago: University of Chicago Press.

Lang, A. (2013a). Discipline in crisis? The shifting paradigm of mass communication research. *Communication Theory*, *23*, 10–24.

Lang, A. (2013b). Discipline in waiting. *Communication Theory*, *23*, 334–335.

LaRose, R. (2010). The problem of media habits. *Communication Theory*, *20*, 194–222.

Lauritsen, J. L. (1998). The age-crime debate: Assessing the limits of longitudinal self-report data. *Social Forces*, *77*(1), 127-154.

Levine, T. R. (2013). Quantitative communication research: Review, trends, and critique. *Review of Communication Research*, 1(1), 69-84. doi: 10.12840/issn.2255-4165_2013_01.01_003

Levine, T. R., Weber, R., Hullett, C., Park, H. S., & Lindsey, L. L. (2008a). A critical assessment of null hypothesis significance testing in quantitative communication research. *Human Communication Research*, *34*, 171–187.

Lowry, D.T. (1979). An evaluation of empirical studies reported in seven journals in the '70s. *Journalism Quarterly*, *56*(2), 262-282.

Lowery, S. A., & DeFleur, M. L. (1988). *Milestones in mass communication research* (2nd ed.). White Plains, NY: Longman.

Matthes, J., Marquart, F., Naderer, B., Arendt, F., Schmuck, D., & Adam, K. (2015). Questionable research practices in experimental communication research: A systematic analysis from 1980 to 2013, *Communication Methods and Measures*, *9*(4), 193-207. DOI: 10.1080/19312458.2015.1096334

McQuail, D. (2005). *McQuail's mass communication theory* (5th ed.). London: Sage.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often interpretable. *Psychological Reports*, *66*(1), 195-244.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.

Moffett, E. A., & Dominick, J. R. (1987). Statistical analysis in the JOB 1970-85: An update. *Feedback*, *28*, 13-16.

Nabi, R. L., & Oliver, M. B. (Eds.) (2009). *The Sage handbook of media processes and effects* (pp. 1 - 5). Los Angeles, CA: Sage.

Neuman, W. R., Davidson, R., Joo, S-H., Park, Y. J., & Williams, A. E. (2008). The seven deadly sins of communication research. *Journal of Communication*, *58*, 220–237.

Niederdeppe, J. (2014). Conceptual, empirical, and practical issues in developing valid measures of public communication campaign exposure. *Communication Methods and Measures*, *8*(2), 138-161. DOI: 10.1080/19312458.2014.903391

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Oliver, M. B. & Krakowiak, K. M. (2009). Individual differences in media effects. In J. Bryant & M. B. Oliver (Eds.), *Media Effects: Advances in Theory and Research* (3rd edition) (pp. 517-531). New York: Routledge.

Perloff, R. (1976). Journalism research: A twenty-year perspective. *Journalism Quarterly*, *53*, 123-26.

Perloff, R. M. (2013). Progress, paradigms, and a discipline engaged: A response to Lang and reflections on media effects research, *Communication Theory*, *23*, 317–333.

Perse, E. M. (2001). *Media effects and society*. Mahwah, NJ: Erlbaum.

Podsakoff, P. M., MacKenzie, S. B., Jeong-Yeon, L., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal Of Applied Psychology*, *88*(5), 879.

Potter, W. J. (1999). *On media violence*. Thousand Oaks, CA: Sage.

Potter, W. J. (2009). *Arguing for a general framework for mass media scholarship*. Thousand Oaks, CA: Sage.

Potter, W. J. (2012). *Media effects*. Los Angeles, CA: Sage.

Potter, W. J., Cooper, R., & Dupagne, M. (1993). The three paradigms of mass media research in mainstream journals. *Communication Theory*, *3*, 317-335.

Potter, W. J., Pashupati, K., Pekurny, R. B., Hoffman, E., & Davis, K. (2002). Perceptions of television: A schema explanation. *Media Psychology*, *4*, 27–50.

Potter, W. J., & Riddle K. (2007). Profile of mass media effects research in scholarly journals. J*ournalism & Mass of Communication Quarterly*, *84*, 90-104.

Preiss, R. W., Gayle, B. M., Burrell, N., Allen, M., & Bryant, J. (2007). *Mass media effects research: Advances through meta-analysis*. New York, NY: Erlbaum.

Reeves, B., & Borgman, C. L. (1983). A bibliometric evaluation of core journals in communication research. *Human Communication Research*, *10*, 119-136.

Rice, R. E., Borgman, C. L., & Reeves, B. (1988). Citation networks of communication journals, 1977-1985: Cliques and positions, citations made and citations received. *Human Communication Research*, *15,* 256-283.

Rice, R. E., Chapin, J., Pressman, R., Park, S., & Funkhouser, E. (1996). What's in a name: Bibliometric analysis of 40 years of the *Journal of Broadcasting* (*& Electronic Media*). *Journal of Broadcasting & Electronic Media*, *40*, 511-539.

Riffe, D., & Freitag, A. (1997). A content analysis of content analyses: Twenty-five years of Journalism *Quarterly. Journalism & Mass Communication Quarterly*, *74*, 873-882.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *EPrime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Schramm, W. (1957). Twenty years of journalism research. *Public Opinion Quarterly*, *53*, 91-107.

Schüz, J., & Johansen, C. (2007). A comparison of self-reported cellular telephone use with subscriber data: Agreement between the two methods and implications for risk estimation. *Bioelectromagnetics*, *28*(2), 130–136.

Schneider, S. L. (2007). Experimental and quasi-experimental designs in behavioral research: On context, crud, and convergence. In W. Outhwaite & S. P. Turner (Eds.) T*he Sage handbook of social science methodology* (pp. 172-189), Los Angeles: Sage.

Shoemaker, P. J., & Reese, S. D. (1990). Exposure to what? Integrating media content and effects studies. J*ournalism Quarterly*, *67*, 649-52.

Sigelman, L. (1982). The nonvoting voter in voting research. *American Journal of Political Science*, *26*(1), 47-56.

Slater, M. D. (2004). Operationalizing and analyzing exposure: The foundation of media effects research. *Journalism and Mass Communication Quarterly*, *81*(1), 168–183.

So, C. Y. K. (1988). Citation patterns of core communication journals: An assessment of the developmental status of communication. *Human Communication Research*, *15*, 236-255.

So, C., & Chan, J. (1991). *Evaluating and conceptualizing the field of mass communication: A survey of the core scholars*. Paper presented at the Annual Meeting of the AEJMC, Boston, MA.

Sparks, G. G. (2015). *Media effects research: A basic overview*, 5th edition. Boston, MA: Cengage Learning.

Stevenson, R. L. (1992). Defining international communication as a field. *Journalism Quarterly*, *69*, 543-53.

Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, *24* (3), 236-243. DOI 10.1207/S15324796ABM2403_09

Sun, S., & Fan, X. (2010) Effect size reporting practices in communication research, *Communication Methods and Measures*, *4*(4), 331-340, DOI: 10.1080/19312458.2010.527875

Trumbo, C. W. (2004). Research methods in mass communication research: A census of eight journals 1990 to 2000. *Journalism & Mass Communication Quarterly*, *80*(2), 417-436.

Valkenburg, P. M. & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. *Journal of Adolescent Health*, *48*, 121–127. doi: 0.1016/j.jadohealth.2010.08.020

Valkenburg, P. M., Peter, J. (2013). The differential susceptibility to media effects model. *Journal of Communication*, *63*, 221-243.

van der Voort, T. H. A. (1986). *Television violence: A child's-eye view*. Amsterdam: North-Holland.

Verplanken, B., Myrbakk, V., & Rudi, E. (2005). The measurement of habit. In T. Betsch & S. Haberstroh (Eds.), *The routines of decision making* (pp. 231–247). Mahwah, NJ: Erlbaum.

Wimmer, R., & Haynes, R. (1978). Statistical analyses in the Journal of Broadcasting, 1970-1976. *Journal of Broadcasting*, *22*, 241-48.

Copyrights and Repositories

Corresponding author

W. James Potter
Department of Communication
4005 Social Sciences & Media Studies
University of California at Santa Barbara 93106-4020
wjpotter@comm.ucsb.edu