Message Variability and Heterogeneity: A Core Challenge for Communication Research

Michael D. Slater,
School of Communication, The Ohio State University

Jochen Peter, and
Amsterdam School for Communication Research, University of Amsterdam

Patti Valkenberg
Amsterdam School for Communication Research, University of Amsterdam

Abstract

Messages are central to human social experience, and pose key conceptual and methodological challenges in the study of communication. In response to these challenges, we outline a systematic approach to conceptualizing, operationalizing, and analyzing messages. At the conceptual level, we distinguish between two core aspects of messages: message variability (the defined and operationalized features of messages) and message heterogeneity (the undefined and unmeasured features of messages), and suggest preferred approaches to defining message variables. At the operational level, we identify message sampling, selection, and research design strategies responsive to issues of message variability and heterogeneity in experimental and survey research. At the analytical level, we highlight effective techniques to deal with message variability and heterogeneity. We conclude with seven recommendations to increase rigor in the study of communication through appropriately addressing the challenges presented by messages.

Communication researchers share with psychologists, sociologists, and political scientists interest in how mediated and interpersonal communication informs what we believe, who we think we are individually and collectively, and the actions we take as individuals, organizations, and societies. Psychologists are concerned with the manifestations of mental activity: perception, cognition, personality, and their relationship to enacted behaviors. Sociologists study social systems, socialization processes, how they are shaped, and how they in turn shape human attitudes and behaviors. Political scientists do much the same with respect to political systems and processes. Communication researchers, as Paisley (1984) once pointed out, are interdisciplinary, exploring the role and function and impacts of communication across each of these levels of analysis. Interdisciplinary approaches have many virtues. However, in what domains of human activity do communication researchers offer a clearly distinctive expertise?

Author notes: Michael D. Slater is Social and Behavioral Sciences Distinguished Professor at The Ohio State University, 3022 Derby Hall, School of Communication, Columbus OH 43085, slater.59@osu.edu.
Patti M. Valkenburg is University Distinguished Professor at the University of Amsterdam, where Jochen Peter is Professor at the Amsterdam School for Communication Research.

We argue that messages represent one such domain. Messages, to adapt Berlo's (1960) classic formulation, are expressions in symbolic form—in verbal language, image, sound, and combinations thereof—from some individual or institutional source, via some mediated or interpersonal channel. Messages are expressions of personal and social meanings, goals, needs, and drives, characterizing humans and their social organizations. Messages have their own distinctive forms, conventions, and constraints. They are extraordinarily heterogeneous, and the task of finding meaningful patterns, distinctions, typologies, as well as methods for managing the study of this heterogeneity in any given context, is a significant intellectual enterprise.

There have been various efforts in the communication literature to address some of the conceptual and methodological challenges of studying messages (e.g., Bucy & Tao, 2007; O'Keefe, 2003; Jackson, O'Keefe, & Jacobs, 1988). However, there is no systematic, contemporary discussion across the range of these challenges of which we are aware. Our aim here is to stimulate greater awareness of the implications of the ways researchers address messages at each step of empirical research. How message variables are conceptualized, defined, and operationalized in experiments and survey research, how particular messages are selected for study, and how messages are analyzed statistically impact the value not just of individual studies but of research practice in the communication discipline as a whole. It is our hope that our discussion will lead in some cases to more research designs in which findings can be better generalized across populations of messages. In other cases, we hope to encourage careful explanation of the rationale for design decisions and increasingly thoughtful discussion of limitations and boundary conditions consequent on how message variables were employed in the study. We conclude with a series of suggestions and recommendations intended to reflect the commitment of our discipline to thoughtfulness and rigor in the study of messages, and to further progress towards increased cumulative knowledge in communication.

Defining Message Variability and Message Heterogeneity

We find it useful, in thinking about messages, to distinguish message variability from message heterogeneity. The term *message variability* refers to the explanatory potential in conceptualizing and operationally defining message characteristics so that they may serve as variables. In so doing, we can study message variables as predictors, controls, mediators, moderators, and outcomes. For example, media violence researchers might be concerned with distinguishing messages portraying justified violence from those portraying unjustified violence, and find reasonable criteria for making this distinction between violence that supports versus undermines a civil society. *Message heterogeneity* is everything else not captured by these variable definitions and operations—the undefined, unexplained, often idiosyncratic variation among messages. The violent messages may vary by the personal and physical attractiveness of the heroes, of the villains, of victims, or of anti-heroes, their gender, age, race, depth of characterization, plot predictability and complexity, popularity of actors in the drama, production quality, amount of suspense, historical epoch, pacing, length, use of music, emotional tone during the story, severity, graphicness of violence, nature of subplots, how the story ends, the outcomes for the various protagonists—all these may be considered elements of message heterogeneity. Experiments on impact of dramas featuring

unjustified violence might yield very different effects depending on whether the stimuli selected feature attractive or unattractive villains. Results of survey research assessing the effects of such exposure might be attenuated if distinctions regarding unjustified versus justified violence or attractive versus unattractive villains cannot be made in measures of exposure to violent media content. Message heterogeneity that is not captured through definition and operationalization of message variables introduces a wide range of concerns regarding generalizability of results beyond the specific messages studied, and issues regarding appropriate statistical analysis. These issues are addressed in the latter part of this paper.

Message heterogeneity captured via rigorous and replicable definition and operationalization becomes message variability, in our vocabulary. In the above example, a researcher might additionally include measurement of protagonist attractiveness and gender. As soon as such variables are explicated and operationalized, they become message variability rather than heterogeneity. We will begin by discussing some of the challenges in defining and operationalizing message variables, drawing on analyses by O'Keefe (2003) and Bucy and Tao (2007).

Message Variability and Intrinsic Message Features

Communication researchers sometimes define message variability in terms of intrinsic properties of the message. Often, though, they define message variables in terms of the psychological state that the message evokes (see O'Keefe, 2003; Bucy & Tao, 2007, for their exploration of this distinction). For example, a more or less fear-inducing message will typically be defined by pretests or manipulation checks demonstrating that a given message induced more or less fear than another, typically without specifying the exact features that might give rise to greater fear. O'Keefe points out that such an approach offers us little understanding of the effects of message variables, as we do not gain any systematic understanding regarding the intrinsic message features that have led to the psychological state of interest. Consequently, O'Keefe (2003) emphasizes the importance of defining the message variable of interest in terms of intrinsic message features.

In the example of fear appeals in persuasion, one can turn to theory to identify potential intrinsic message features. Research on risk perception (e.g., Slovic & Fischhoff, 1982), for example, suggests that the extent to which messages emphasize prevalence, severity, catastrophicness, and dreadfulness of a given risk will determine responses regarding that risk. Therefore, following O'Keefe's recommendations for use of intrinsic message differences, one might manipulate inclusion of information about the prevalence of a risk in a message to increase or decrease the fear induced by the message. O'Keefe argues that in this way, we actually learn what it is about a message that induces the fear reaction. In O'Keefe's view when an intrinsic message variable such as risk prevalence is used, a question such as "how scary did you find this message" is not a manipulation check but measurement of an intervening psychological response that the researcher expects will influence the outcome. From this perspective, use of intrinsic message features as variables increases knowledge about message differences and clarifies the various elements involved in the causal process being theorized (see also Bucy & Tao, 2007, on advantages for theory-

building of incorporating mediating and moderating variables arising from differences in the processing of message content).

However, using intrinsic features to operationalize message differences is often easier said than done. A study attempting to study simultaneously each of the range of intrinsic messages features associated with fear such as prevalence, dread, catastrophicness, and severity would be extremely cumbersome. Moreover, the distinction between intrinsic features and subjective, psychological responses is not at all a clean one. What intrinsic features make for a judgment that a risk is particularly dread-inducing or severe? Clearly, a subjective element will remain. O'Keefe acknowledges this problem:

> "… the separation of intrinsic message features from recipient responses is not an unproblematic undertaking. This is a more complex matter than can be sorted out here…Leading researchers to a still more sophisticated understanding of the nature of messages is a very desirable goal. The argument here cannot be more than an initial step toward that end, however, because any easy distinction between message features and recipient responses can be no more than—to invoke Wittgenstein's (1921/1961, 6.54) image—a ladder to be climbed and thrown away." (O'Keefe, 2003; p. 270).

An likely explanation of why O'Keefe's proposals about emphasis on intrinsic message features have not been more widely adopted is that the problems and complexities to which he refers are the norm, not the exception. We provide some suggestions for rendering this complexity more tractable.

Content Analysis as a Model for Exploring Message Variability

Content analysis can play a central role in laying the groundwork for theory-development and theory-testing research employing experiment or survey methods (McLeod & Reeves, 1980; see Slater, 2013, for an extended example of this process). Content analysts propose a coding scheme based on their theoretical concerns and their observations of the messages of interest, refine the scheme empirically in the process of training coders and clarifying coding definitions, and end by identifying reliably replicable, and theoretically or substantively useful, distinctions amongst messages. In so doing, content analysis seeks to transform much of the heterogeneity of a given domain of messages into message variability.

Content analysis experts and O'Keefe's prescriptions agree about the advantages of intrinsic message features: the more objective the message variable to be coded, the more replicable and reliable the coding scheme (e.g., see Krippendorff, 2013; Riffe, Lacy, & Fico, 1998). Some message variables capture differences that are normally unambiguous, and error in coding is likely to result only from lapses of attention. Others are inherently more subjective.

For example, in a recent content analysis looking at social aggression in children's television programming (Martins & Wilson, 2012), coders assessed clearly objective content such as character's biological sex, and whether they were human, supernatural, anthropomorphized, or other. In another recent study of violence on You-Tube videos (Weaver et al., 2012),

coders coded objective features including You-Tube category, date, length in seconds, rating, number of raters, and number of comments. Other content, in contrast, requires interpretation, subjectivity, a psychological response on the part of a coder regarding the distinction proposed in the coding scheme. Coders in the Martins and Wilson (2012) study had to identify instances of social as well as physical aggression, character attractiveness, benevolence or malevolence of behavior, rewards or punishments for behavior, and humor —each of which clearly require some measure of subjective assessment. In the Weaver, et al. (2012) study, coders also had to assess whether the video was professional or amateur, and the valence of on-screen reactions to violence.

The task of content analyses is to develop definitions in the coding scheme that are sufficiently clear and objective that two coders can achieve reasonable agreement, even if as a result some nuance is lost in the process. Sometimes coders cannot come to reliable agreement on a variable that requires subjective judgment even after extensive training and rule refinement efforts. When this happens, this should suggest to the researcher that the desired variable cannot meaningfully be operationalized in the message population. Content analysis, then, provides a means to transform subjective responses (such as attractiveness, or moral justification) to operationally-defined intrinsic features through the process of creating coding rules and testing them for intercoder reliability.

Use of content analyses to inform hypotheses and study design

While the means are methodological, the content analysis process normally leads to greater clarity concerning one's construct of interest and possible covariates, moderators, or boundary conditions to address in a given study. Consider a researcher wanting to study the impact of media messages modeling social aggression on youth attitudes and behavior. In making sense of the heterogeneity of the portrayals of social aggression through content analysis, researchers would presumably come to consider whether aggressive behaviors were rewarded or punished, intended benevolently or malevolently, the attractiveness of the perpetrator and victim, the humor or lack thereof in the context, whether the protagonists were human or cartoon characters, etc. Certain of these variables might be included as treatment levels (e.g. behavior rewarded or punished) in an experiment. Some variables might be incorporated as covariates and potential moderators (e.g. character attractiveness). Some variables might identify boundary conditions for the study (e.g. cartoons may be excluded if the focus is on human modeling of aggressive behaviors).

As researchers, we may not want, or realistically be able, to study systematically and rigorously each of these variables. We can, however, spell out what it is we are studying, and what we are not, and why. We can interpret our findings with explicit cautions regarding what we have taken into account in our approach to these message variables, and what we have not. We can suggest future research that might be most theoretically or substantively interesting related to unstudied variables or messages outside the boundaries of the present study.

Developing message difference variables and selecting stimuli from content analyses customized for the researchers' own research questions is ideal. The process of developing the content analysis requires the researcher or research team to conceptualize the aspects of

message heterogeneity that they want to explicate, define, and code. In so doing, researchers transform part of message heterogeneity in their message population of interest into message variables amenable to empirical investigation. Another major advantage of one's own content analysis is that one can select messages that have been identified in the content analysis to represent differences on variables of interest in experimentation. Still another advantage is that the process of identifying message variables (in part inductively through examination of the messages themselves), and discovering patterns among variables in the content analysis may lead to insights or hypotheses that can be addressed through further survey or experimental research (see Slater, 2013 for examples).

Such custom content analyses are time-consuming and presume a specific intellectual interest in a given domain of messages. Sometimes a research question is more general, and selection of a domain in which to test it is arbitrary and to some extent a matter of convenience. Sometimes specific messages are of interest. In such cases an extensive content analysis is hard to justify. Perhaps content analyses already available in the literature can identify reliable message variables. Typically the researcher will not have access to the messages used in the original content analyses. If so, the researcher still has to find or create his or her own message exemplars and confirm, probably through pretest, that these exemplars in fact represent different values on such variables. In the absence of such content analyses, the researcher most likely will provide a conceptual analysis of relevant stimulus variation. For example, a researcher might for theoretical reasons propose that effects of unjustified violence are contingent on attractiveness of perpetrators, and construct messages to provide a test of this hypothesis independent of a content analysis that identifies actual exemplars of these messages. Such a conceptual analysis of theoretically relevant message variables will normally be followed up with pretests or other empirical checks to assess the validity of variable distinctions made based on such analysis. In such cases, what the researcher is in fact doing is an informal content analysis based on consideration of relevant theoretical message variables assessed through personal observation and study of the literature, followed by pretest of what the researcher considers reasonable exemplars of those message variable differences. Such approaches may be justified when the domain of messages is a matter of convenience and of secondary interest, when messages are being created as a matter of experimental control, or in initial exploratory studies to assess the potential of further research into a given message domain. However, the limitations resulting from such an approach are significant and need to be clearly acknowledged, arguments for more general inference from results tempered, and the importance of research to better assess the generality of findings across actual populations of messages clearly acknowledged.

Selecting or Sampling Messages for Study and the Problem of Message Heterogeneity

Our discussion has focused on advantages that accrue from drawing upon a content-analytic mindset in conceptualizing and operationalizing message variability. Attending to lessons from content analysis also has other benefits for our thinking about message variability and heterogeneity. One such benefit is a focus on clearly defining a message population of

interest for a given study. When a researcher plans a content analysis, a key initial step is to define the population of messages of interest, and then to come up with a plan to sample or select individual messages from this population. Any time a communication researcher selects a message or set of messages to use in an experiment, the researcher has implicitly expressed research interest in a population of messages, and made a sampling decision of some kind regarding exemplars of that population.

Directly testing generality of findings across messages may come in the later stages of development of a research program and theoretical development. There is little point in investing the time and resources in more ambitious research designs that permit such testing unless one has reason to believe that there are findings likely to hold across populations of messages. However, in our view, concern with the question of generality of findings across populations of messages reflects an understanding of the challenges of communication inquiry consistent with maturation of our discipline. We therefore would like to see explicit identification of the message population of interest, the reasoning for selecting those messages used in the study, and clear acknowledgements of limitations arising from selection strategy, as hallmarks of communication research.

Below, we discuss various strategies for selecting messages for study, beginning with a random selection approach such as those often used in content analysis that may permit statistical generalization across message populations. We then discuss the various compromise approaches to message selection that are required in different experimental research contexts, including use of small numbers of messages selected for use in an experiment, creating messages to be used in an experiment, or manipulating a single message. Our focus is on the trade-offs involved and the importance of providing a clear rationale for the message selection decision and discussion of the theoretical and substantive implications and limitations associated with that decision. We also briefly address implications of message heterogeneity and variability for the conduct of survey research.

Addressing Message Variability and Heterogeneity in Experiments

The issues in selecting messages for experiments in many ways parallel those regarding selection of research participants. In survey research, for example, random selection from a population may be considered the most desirable approach in terms of potential generalization of findings. Nonetheless, surveys often use self-selected on-line panels because of the need to present information on a computer screen or for cost reasons. Sometimes they use systematic convenience samples for populations which can't be defined for sampling purposes (e.g., IV drug users or gay men). Experiments typically use relatively homogeneous convenience samples, often of undergraduate students. In each case, the compromises required are typically accepted by reviewers and editors if appropriately explained and defended, and the resultant limitations to findings discussed. In our view, however, it is crucial to keep in mind the analogous problem of message selection and the limitations arising from whatever approach is selected. We find it helpful, in thinking about these problems, to use random selection of messages as a benchmark, just as a random sample of research participants from a defined population is the benchmark in survey research against which other approaches are compared.

Sampling approaches to message selection in communication experiments—
When conducting an experiment intended to shed light on how messages are experienced, it is wise to explicitly define at least conceptually, and ideally operationally, the population of messages of interest. One sound approach is to define a message population of interest as a content analyst would—e.g., "M" rated video games available in Dutch stores in 2013, speeches made by U.S. senatorial candidates in the 2012 election as available from certain archives, episodes of the three highest-rated police procedural shows on U.S. television from the past 3 seasons, etc. In some ambitious studies, it is often possible to randomly select a large enough number of such messages to make possible generalization of findings to the larger population of messages. (What "large enough" is depends on the size of the expected effect and how variable the messages are, a complex power problem that is beyond the scope of the present paper, e.g., see Snijders, 2005.)

There are two possible approaches to using randomly sampled messages. One approach is to simply take a random sample of messages of two or more different types and compare them. For example, the social aggression researcher might employ as experimental stimuli several dozen examples of cartoons from the content analysis sample in which social aggression is rewarded and contrast them with several dozen in which social aggression is punished. An attractive part of this sampling approach to operationalizing a variable is the potential for generalizability. It may be that the kinds of cartoons in which social aggression is rewarded are also quite different in other ways than cartoons in which social aggression is punished. Perhaps the former cartoons use more anthropomorphic characters, for example, or have less complex plots. Nonetheless, if sampling is random and sample size is adequate to reasonably represent variability across the message types of interest, the nature of this confounding represents that confounding as it exists in the actual population of messages under study. In other words, in the real world, cartoons with rewarded social aggression also (in this hypothetical example) have more anthropomorphic characters and simpler plots. The researcher therefore can draw conclusions about the effects of real-world message populations of cartoons using this design in ways analogous to the survey researcher drawing conclusions about human populations (see citation withheld for an example of such a study).

The most obvious reason to employ such a design is to be able to draw the kind of generalizable conclusions about the impact of real-world messages that can influence social policy—in our view an important role for communication researchers. Another possible reason for such a design might be to test theoretical claims from more tightly-controlled experiments against real-world message populations. For example, consider the well-known elaboration likelihood model (Petty & Cacioppo, 1986), which has demonstrated that people respond to argument quality differences much more in messages that are personally relevant than in messages that are not. One might randomly select letters to the editor or on-line comments regarding proposed drinking age enforcement policies (high relevance to undergraduate research participants) or proposed public school closures (low relevance) in their community, code and sort them into high, medium, or low argument quality based on a content analysis, and thereby test the generalizability of this aspect of the model against these message populations. The weakness of selecting messages of a given type is that it does not permit unambiguous attribution of effect to a given message variable apart from the

various other message characteristics (heterogeneity) with which it might be associated. If the combination of traditional experiments, which are internally valid, and the approach sketched here, which is externally valid, across multiple studies elicits homogeneous results, we have obtained casually rigorous and generalizable knowledge.

A second, alternative approach is to manipulate some random sample of messages obtained from a defined population. Slater, Hayes, Goodall, and Ewoldsen (2012) did this by incorporating or removing alcohol mentions from a random sample of 60 news stories. The exact wording involved in creating the manipulation is unique to each story, producing heterogeneity in the manipulation that reflects real-world news story wordings that should increase validity of the manipulation. Effects are unlikely to be due to some idiosyncratically effective manipulation in a story or two. Still, the manipulations are created by the researcher and therefore half of the stories (the ones manipulated into the condition in which they did not originally appear) approximate rather than directly represent real-world differences. Each respondent reads just one story, and each story is read by several respondents, so the effects of message heterogeneity—all the effects associated with the various story differences besides the manipulation—can be statistically estimated and incorporated appropriately in the analysis. The advantage here, of course, is that one can make relatively confident assertions about the influence of the manipulation, independent of other executional elements which may tend to correspond with the presence of that element (e.g., if stories that actually reported alcohol as a factor also tended to more often involve youthful perpetrators or victims). Even more important, the effect of the manipulation can be generalized with reasonable confidence to the population from which messages were sampled. This is a resource-intensive approach, and can be best justified when generalizability is important in terms of implications for social policy.

This approach also might be used to assess generalizability of previous theoretical claims based on findings obtained from study of only a few messages. Using the previous example about the elaboration likelihood model, one might take a randomly sampled set of on-line comments or letters about proposed drinking age enforcement changes, and manipulate them, per standard practice in elaboration likelihood research, to be about the students' own community or one far away. It seems to us that assessing the generalizability of theoretically interesting findings from experiments that used a few carefully-selected messages to populations of real-world messages is a valuable and distinctive approach for communication scientists to explore.

<u>Random effects/multilevel analysis of sampled messages:</u> Development of multi-level models (MLM), also known as hierarchical linear models (HLM; Raudenbush & Bryk, 2002), provides an efficient and practical way to analyze studies with large numbers of (hopefully randomly-sampled) messages. MLM is comparable to random effects analyses of variance previously recommended in multi-message research (e.g., Jackson, O'Keefe & Jacobs, 1988), but is more flexible and is generally to be preferred for such analyses. Multi-level modeling is applicable to any research context in which observations are nested within a larger unit (for introductory texts, see e.g., Hayes, 2006; Henry & Slater, 2008; Park, Eveland, & Cudeck, 2008; Slater, Snyder, & Hayes, 2006). For example, it is possible to set up a model in which one has multiple observations on an individual over time nested within

the individual, who is in turn nested within a community. In the present case, one may analyze the responses of different individuals who are responding to (i.e., are nested within) the same message.

While a detailed discussion of MLM is beyond the scope of this paper, it is important to briefly highlight some of the advantages of MLM in the present context. Whenever observations on individual study participants are nested within one of several messages, the variability and heterogeneity associated with that message will influence the overall analysis and should be accounted for in the statistical analysis. One approach is to treat each message used as a level on a fixed effect instead of using MLM, an approach that is discussed later. Another particularly powerful approach, which is clearly most appropriate when a relatively large number of messages are used as stimuli, is to treat the message as an upper-level clustering or random effect in MLM.

Statistically assessing message heterogeneity with MLM: In the case of research using multiple messages, and any given participant is exposed to one of those messages, MLM can simultaneously model message variable effects while adjusting participant effects for effects of message heterogeneity—the clustering effects of being nested within a message. MLM can incorporate message variables, if they have been coded and identified, in the model, as well, providing statistically appropriate tests of the direct effects of those message variables as well as possible interactions with research participant characteristics or experimental manipulations.

There are two types of clustering effects (or effects associated with message heterogeneity) that MLM can account for: random slopes and random intercepts (see Raudenbush & Bryk, 2002, Hayes, 2006). Random slopes represent the way an independent variable's effects vary between messages. The random intercept allows the average response between messages to vary. MLM permits the researcher to assess whether message heterogeneity has sufficient impact on the outcome to require testing treatment effects against this heterogeneity as represented by the random slope. If the impact of such heterogeneity is small enough to be ignorable, the model will typically not properly converge due to lack of variance associated with message heterogeneity, which also would usually be reflected in a trivially small intra-class correlation coefficient. In such cases, it may be more appropriate to use MLM to adjust treatment effect tests for the clustering of participants within each message without testing effects against the random variability attributable to the individual messages, i.e. incorporating random intercepts but not random slopes into the model (see Raudenbush & Bryk, 2002).

When messages have been selected randomly from some population, such tests can be used to make possible generalization of effects to that population of messages (see Slater et al., 2012; Goodall, Slater, & Myers, in press). For at least some research questions, as noted above, being able to generalize findings to populations of messages should be as attractive to communication researchers as being able to generalize from a sample to a human population is for sociologists and political scientists.

Studies purposively selecting a small number of messages as study stimuli—
Use of a large number of randomly-sampled experimental stimuli is a resource-intensive
strategy. As such, it is hard to justify early in the process of theoretical development. For
some research questions, the requirements of tight experimental manipulation preclude use
of randomly-sampled stimuli, as only certain examples may lend themselves to such
manipulation or messages must be created to permit tight manipulation. In such cases, the
number of messages that may realistically be employed in an experiment are too small to
permit random effect, multi-level tests as a viable option. Often, instead, researchers select
several messages, to decrease the likelihood that effects found are unique to the particular
message selected for use as the experimental stimulus, and then manipulate each of those
messages. This small-N message approach requires use of experimental manipulations, and
is typically used to facilitate use of tight experimental manipulation that cannot readily be
applied to other than a carefully selected or constructed message or set of messages.

Unlike the studies mentioned earlier, with their use of random effects to statistically assess
the impact of heterogeneity across sampled messages, small-N multiple message studies in a
sense seek to replicate the effects of a manipulation across several messages in a single
study (see Jackson, O'Keefe, Jacobs, & Brashers, 1989; Slater, 1991). Psychologists, to
accomplish a similar end, might run an experiment multiple times with changes in the
experimental procedure and stimuli to demonstrate that effects are not due to a single
stimulus or manipulation. However, when several different messages are manipulated as
experimental stimuli such replication in effect takes place within a single study. Participants,
after all, are randomly assigned to each message, and each message is separately
manipulated. This is one advantage of message research—in many contexts (e.g., using text
rather than video-type stimuli), it may be relatively easy to manipulate several different
message stimuli within the same study, so findings aren't dependent on a single stimulus and
manipulation, and without having to rerun the same study multiple times with slightly
different message stimuli.

The persuasiveness of such replication across messages, however, depends on the extent to
which the stimuli selected or created by the researcher are varied representations of some
type, category, or population under study. If the messages selected are very similar, and in
some ways apparently atypical of the larger population of messages of interest, the use of
multiple messages does little to increase confidence in the robustness of findings. Therefore,
clear articulation of the message domain under study, qualitative or quantitative
consideration of message diversity and representativeness of the exemplar stimuli used, and
clear acknowledgements of limitations and likely boundary conditions consequent on
message selection decisions, are desirable indeed.

Multiple-message studies—do effects depend upon the message used?—
When looking at a series of related replication studies, using different stimuli with different
instantiations of manipulations, one looks for consistency in results across these studies.
Similarly, in a single study using several messages, the reader would want the ability to
assess consistency of findings across messages. While the use of random effects model for
studies even with only a few messages has been proposed since the 1980s (e.g., Jackson et

al., 1988, 1989), it has been contested (e.g., Hunter, Hamilton, & Allen, 1989; Slater, 1991) and has not been widely adopted.

Another possibility is to attempt to assess effects of message heterogeneity to the extent possible within the traditional fixed effects framework. When only a few variant base messages are used as the basis for creating experimental manipulations, the base message used is likely to contribute to variation in results. The researcher and the reader would both want to know how great this contribution is. The simplest way to examine this question within an ANOVA or regression model is to look at interactions between the experimental manipulation and the different messages used. One might find an interaction pattern indicating that some messages seem to show no effect but some have strong enough effects to generate significance overall. Such a pattern suggests that there are message variables at work that were not identified at the outset responsible for these contingent effects.

The finding that there is a significant interaction due to the messages used does not mean the results of the study are not robust and are therefore unpublishable. It does mean there are some boundary conditions evident within the study that demand attention. In the event of a message-by-treatment interaction, a qualitative examination of message content would likely generate ideas regarding message content differences that might explain the pattern of results. In other words, some message variables were lurking among the stimuli selected that had not been anticipated in the initial theorizing and concept explication. If the researcher is fortunate and has been careful conceptualizing causal processes and measurement, it is possible that *post hoc* analyses can be used to test such *post hoc* explanations. The resulting insight should add to, rather than detract from, the scientific value of findings.

Similarly, the lack of an interaction does not fully moot concerns regarding evidence for possible boundary condition effects. If power is relatively low, but just enough to show main effects, non-trivial message by treatment interactions may still not be statistically significant. Descriptive discussion of possible differences in findings by message is still desirable. Unfortunately, however, it appears to be common practice in communication research with multiple messages to simply average across the impact of the messages used without also assessing message by treatment interactions, or descriptively summarizing the presence or absence of possible differences in effects across messages.

Of course, the lack of an interaction, or the presence of descriptively similar findings across messages, does not in itself provide statistical evidence for generality of effects across messages in the real world. After all, messages in such studies have not been randomly selected and treated as a random effect. Such findings simply demonstrate that effects were robust across the messages used in the experiment. Issues of possible boundary conditions due to message differences not captured in the messages studied still require careful thought and discussion.

Messages created for or within a study—wrestling with issues of validity—A variant of multiple-message studies involves message stimuli that are created for the study or as a function of the study, rather than being sampled or selected from the social environment. Such approaches can be very attractive in terms of making possible rigorous

manipulations and tests of theory. At the same time, they raise issues of validity; the researcher must grapple with the problem of the extent to which findings might be extended to real-world phenomena, and address these questions to the satisfaction of the reviewer and reader.

For example, in a study designed to explore relational factors influencing experience of hurtful messages, relationship partners were recruited, and one partner was trained to be supportive or unsupportive in conversation with their boyfriend or girlfriend (McLaren et al., 2012). The study created a "real-time" set of interactions. In other words, the researchers created a situation in which actual messages were generated, in manipulated conditions. So, the question becomes how such messages might not fully represent real-world interactions. The researchers addressed this challenge in two ways. First, they articulated possible limitations, discussing boundary conditions associated with the modest hurtfulness of the conversations that were possible given ethical concerns. Second, they empirically addressed the larger concern, concerning how representative these conversations were of actual conversations by these romantic partners, by including measures of the typicality and realism of the conversations to assess the validity of the stimuli created.

Single-message studies—Using a single message as a base stimulus for manipulation inherently leaves a greater room for the possibility that study results are consequent on idiosyncrasies of the message studied and how the manipulation is carried out. Nonetheless, there are circumstances where such research is readily justified. The example of studies of narrative film and television programs comes to mind in this regard. Finding a film or program that permits manipulation of a theoretically significant message variable with reasonable plausibility and rigor is often quite difficult. The necessity of using different dependent measures tailored to the story content also complicates attempts to do multiple message studies in this context. Results in such cases might not be readily combined into a single analysis. Moreover, single instantiations of messages such as films and television programs may inherently be of substantive interest given their reach, visibility, and potential impact. In such cases, it is necessary to make the case for why a single message instantiation is used, and address in discussion consequent limitations and plausible boundary conditions for findings that might be addressed in future research.

Addressing Message Variability and Heterogeneity in Survey Research

Our discussion to this point has focused on addressing message variability and heterogeneity in communication experiments. Similar issues are faced in survey research, though in different ways than in experimentation. Strategies are available for contending with questions of message variability and heterogeneity in surveys.

Specificity of exposure/attention measures—Typically, survey research involving communication asks respondents about exposure and often attention to particular types of media content or interpersonal discussion (see Fishbein & Hornik, 2008; Slater, 2004). Questions asking about exposure and attention to some type of communication (let us say news) will usually account for differences in channel—television, newspaper, internet, magazines, interpersonal discussion of news. Perhaps this will be broken down further. The

researcher may be interested in differences in ideological slant of news used, and ask about Fox, MSNBC, the particular magazines read, and the types of internet news sites viewed, and the ideology of discussion partners. Perhaps use of breaking news versus analysis, opinion, and panel discussion will be distinguished. In our view, the appropriate level of specificity of such questions is critical. News, like many categories of message content, is heterogeneous. Identifying and measuring relevant specifics turns some of this heterogeneity into meaningful variability. The heterogeneity of message content within each category (e.g. variation within Fox news broadcasts), however, does not pose the kind of problem in such survey research as it does in experimentation. The researcher in this case is essentially averaging across the variation in the Fox news broadcasts seen, by using amount of exposure to Fox news as the variable. In a sense, this is equivalent to the experimenter who randomly samples Fox news broadcasts and does not attempt to manipulate them, allowing the natural heterogeneity to represent the message type.

The ability to get such fine-grained data on media use in surveys is often limited, especially when conducting secondary analyses of survey data sets not primarily concerned with communication questions. We grant that some data on media use is better than none, and that interesting findings are possible even when there are only a few media use items available. However, we strongly discourage communication researchers from using general exposure measures when they have the opportunity to advocate for or to create more specific exposure measures. The greater the specificity regarding content used, the more meaningful analyses can become. In fact, the greatest specificity is possible when surveys are used to assess exposure to specific messages in the social environment, or are combined with content analysis data sets.

Assessing exposure to campaign messages—Some survey research is concerned with the possible effects of specific messages present in the social environment. The simplest approach, often used in advertising and public health evaluations of campaigns with relatively small numbers of messages, assesses recognition of the messages via description or sample images. Recognition memory is generally quite good (Shapiro, 1994), and a tendency to falsely report recognition, though commonplace, can be controlled for using recognition of foils or pseudo-messages (Slater & Kelly 2002; Southwell, et al., 2002).

As the number of messages in an advertising or advocacy campaign is relatively small, it is easy to analyze the content of these messages. However, the small number of messages also creates challenges. Message differences of interest are likely confounded with idiosyncratic executional differences. The effects of public health messages that, for example, emphasize social normative concerns compared to personal risk are confounded with how the particular campaign addresses social normative concerns versus personal risk. In such cases, the rationale for such comparison must be made on the basis that the execution of these messages is a substantively important example of how such execution takes place in the social world. In a presidential election the way character attack ads are constructed may be a function of the personalities involved and the advertising agencies employed; however, how they are executed in that campaign is in fact, substantively, what matters. The confounding of message variable with execution represents the natural confounding present in the social world at the time of the study. It seems to us important for the researcher to discuss the

distinctive approach to the message type taken in the campaign under study when interpreting results in the discussion section, and to discuss how that approach might have contributed to findings and possible implications of alternative approaches.

Combining surveys and content analyses—Linking survey data to content analyses provides another means to examine the impact of exposure to specific message content. For example, researchers interested in effects of popular movies on adolescent smoking and alcohol use content analyzed hundreds of such movies, and by simply asking teens what movies they have seen, are able to assess the effects of exposure to various message elements represented in these movies, and to control for other elements also present in these movies (see Sargent, Worth, Beach, Garrard, & Heatherton, 2008). A similar approach has been taken to the impact on adolescents of sexual content in media (Brown, et al. 2006). Media diaries can be used to assess respondent exposure, and the actual content seen by the respondent can be content-analyzed.

Using geographic differences to model message variability—Another approach to studying effects of message variables in surveys is based on data regarding the geographic distribution of messages in the social environment. In these studies, differences in content (e.g., as a function of differences in media advertising buying or cable penetration by market, for example), are assessed regionally and the influences of message exposure to a given type of message assessed based on place of residence, using multi-level modeling, with residents of a media market nested within the media-market level data. In this case, content differences are conceptualized as environmental differences, and studied as such. These methods can be applied to overcome problems of self-report in campaign evaluations and other effect studies (e.g., Snyder, Milici, Slater, Sun, & Strizhavkova, 2006). They can also be employed to examine the effect of message variability identified in content analyses that vary by region, such as differences in news coverage associated with the news practices or ideological slant of regional news outlets (e.g., Hoffman, 2012).

Summary of Recommendations and Conclusion

We conclude this paper by summarizing our recommendations concerning ways to more thoughtfully and consistently address message variability and heterogeneity in communication research. As noted earlier, we do not advocate a single analytic or design strategy as "the" solution, but prefer flexibility and adaptation to the message domain and research question.

1. Start with Daniel O'Keefe's recommendation (2003) as an aspirational goal: to define and operationalize message variables based on intrinsic message features instead of defining message differences based on people's responses to messages.

2. When distinguishing message variables requires some subjective judgment, such operationalizations can be accomplished through formal content analyses or by using prior content analyses conducted by others. At the least, researchers can challenge their own thinking about message features through careful conceptualization and definition of message differences as they would if creating a content analysis coding scheme, as well as using validation through pretest.

3. Explicitly identify the message population of interest. In the Methods section, be clear about the approach and rationale used for sampling or selecting messages from this population, and why the approach was reasonable given the research question and context.

4. When feasible, consider defining a message population and randomly sampling a large N of messages to use in the study, analyzed with multi-level models. Obviously, this is a priority when the purpose is to make policy-relevant observations or critiques, or when trying to demonstrate the robustness of a more mature theory across a range of real-world messages. It seems to us that the ability to generalize to populations of messages has the potential to be as significant to communication scholars as generalizing to human populations for sociologists and political scientists.

5. The use of large numbers of randomly selected messages is typically not feasible, or may not be a sensible use of resources when conducting initial tests of a theoretical proposition. In such cases, explicit discussion is needed regarding ways in which the messages selected may not represent actual variation in the message population of interest and possible boundary conditions to findings associated with unstudied message differences.

6. In survey research, endeavor to maximize specificity of exposure and attention measures; if possible, link survey responses to analyses of message content.

By no means do we suggest that communication researchers currently ignore these challenges associated with message variability and heterogeneity. Indeed, as we look at our major journals, typically there is a serious effort to address at least some of these issues in each article. However, it is our general impression that in many articles, at least some of the research problems associated with message variability and heterogeneity are overlooked or passed over quickly. This then leads to another recommendation:

7 We would like to see reviewers and editors encourage more explicit discussion of message variability, message selection, and boundary conditions, viewing such discussion as indicators of intellectual rigor rather than of methodological weakness or limitation (unless, of course, the choices made cannot be reasonably justified).

The problems of message variability and heterogeneity, and the resultant limitations and uncertain boundary conditions for findings, are not an embarrassing family secret that we should want to sweep under the rug. Our understanding of message variability and heterogeneity, our attention to these challenges in our conceptualizing and theorizing, our thoughtful choices in our research design, our willingness when appropriate to take on more ambitious and complex message stimuli designs, and our careful interpretation of findings in the light of these issues, can increasingly become trademarks of our field. To the extent we do so, the distinctive contributions of the communication discipline to the social sciences are likely to become increasingly evident.

References

Berlo, DK. Process of communication: An introduction to theory and practice. New York: Holt Rinehart and Winston; 1960.

Brown JD, L'Engle KL, Pardun CJ, Guo G, Kenneavy K, Jackson C. Sexy media matter: Exposure to sexual content in music, movies, television, and magazines predicts black and white adolescents' sexual behavior. Pediatrics. 2006; 117:1018–1027. [PubMed: 16585295]

Bucy EP, Tao CC. The mediated moderation model of interactivity. Media Psychology. 2007; 9(3): 647–672.

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

Goodall CE, Slater MD, Myers TA. Fear and anger responses to local news coverage of alcohol-related crimes, accidents, and injuries: Explaining news effects on policy support using a representative sample of messages and people. Journal of Communication. (In press).

Hayes AF. A primer on multilevel modeling. Human Communication Research. 2006; 32(4):385–410.

Henry, KL.; Slater, MD. Assessing change and intraindividual variation: Longitudinal multilevel and structural equation modeling. In: Hayes, AF.; Slater, MD.; Snyder, LB., editors. The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research. Thousand Oaks, CA: Sage Publications; 2008. p. 55-88.

Hoffman LH. When the world outside gets inside your head: The effects of media context on perceptions of public opinion. Communication Research. 2012

Hunter JE, Hamilton MA, Allen M. The design and analysis of language experiments in communication. Communications Monographs. 1989; 56(4):341–363.

Jackson S, O'Keefe DJ, Jacobs S. The search for reliable generalizations about messages: A comparison of research strategies. Human Communication Research. 1988; 15(1):127–142.

Jackson S, O'Keefe DJ, Jacobs S, Brashers DE. Messages as replications: Toward a message-centered design strategy. Communications Monographs. 1989; 56(4):364–384.

Krippendorff, K. Content analysis: An introduction to its methodology. Thousand Oaks, CA: Sage; 2013.

Martins N, Wilson BJ. Mean on the screen: Social aggression in programs popular with children. Journal of Communication. 2012; 62(6):991–1009.

McLaren RM, Solomon DH, Priem JS. The effect of relationship characteristics and relational communication on experiences of hurt from romantic partners. Journal of Communication. 2012; 62(6):950–971.

McLeod, J.; Reeves, B. On the nature of mass media effecs. In: Withey, SB.; Abeles, RP., editors. Television and social behavior: Beyond violence and children. Mahwah, NJ: Erlbaum; 1980. p. 17-54.

O'Keefe DJ. Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. Communication Theory. 2003; 13(3):251–274.

Paisley, WJ. Communication in the communication sciences. In: Dervin, B.; Voigt, M., editors. Progress in communication sciences. Vol. 5. Norwood, N.J.: Ablex; 1984.

Park, HS.; Eveland, WP.; Cudeck, R. Multi-level modeling: Studying people in contexts. In: Hayes, AF.; Slater, MD.; Snyder, LB., editors. The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research. Thousand Oaks, CA: Sage Publications; 2008. p. 219-246.

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods (Vol. 1). Thousand Oaks, CA: SAGE Publications, Inc.; 2002.

Riffe, D.; Lacy, S.; Fico, F. Analyzing media messages: Quantitative content analysis. Mahwah, NJ: Lawrence Erlbaum; 1998.

Sargent JD, Worth KA, Beach M, Gerrard M, Heatherton TF. Population-based assessment of exposure to risk behaviors in motion pictures. Communication Methods and Measures. 2008; 2(1–2):134–151. [PubMed: 19122801]

Shapiro, MA. Signal detection measures of recognition memory. In: Lang, A., editor. Measuring psychological responses to the media. Mahwah, NJ: Erlbaum; 1994.

Slater MD. Use of message stimuli in mass communication experiments: A methodological assessment and discussion. Journalism & Mass Communication Quarterly. 1991; 68(3):412–421.

Slater MD. Operationalizing and analyzing exposure: The foundation of media effects research. Journalism & Mass Communication Quarterly. 2004; 81(1):168–183.

Slater MD, Hayes AF, Goodall CE, Ewoldsen DR. Increasing support for alcohol-control enforcement through news coverage of alcohol's role in injuries and crime. Journal of Studies on Alcohol and Drugs. 2012; 73(2):311–315. [PubMed: 22333339]

Slater MD, Kelly KJ. Testing alternative explanations for exposure effects in media campaigns: The case of a community-based, in-school media drug prevention project. Communication Research. 2002; 29:367–389.

Slovic, P.; B. L. S. Fishhoff, BLS. Judgement under uncertainty: Heuristics and biases. New York: Plenum Press; 1982. Facts versus fears: Understanding perceived risk; p. 463-489.

Southwell BG, Barmada CH, Hornik RC, Maklan DM. Can we measure encoded exposure? Validation evidence from a national campaign. Journal of Health Communication. 2002; 7(5):445–453. [PubMed: 12455763]

Snijders; Tom, AB. Power and sample size in multilevel linear models. In: Everitt, BS.; Howell, DC., editors. Encyclopedia of Statistics in Behavioral Science. Vol. 3. Chicester (etc.): Wiley; 2005. p. 1570-1573.

Snyder LB, Milici FF, Slater M, Sun H, Strizhakova Y. Effects of alcohol advertising exposure on drinking among youth. Archives of Pediatrics & Adolescent Medicine. 2006; 160(1):18. [PubMed: 16389206]

Weaver AJ, Zelenkauskaite A, Samson L. The (non)violent world of youtube: content trends in web video. Journal of Communication. 2012; 62(6):1065–1083.