MESSAGES AS REPLICATIONS:
TOWARD A MESSAGE-CENTERED DESIGN STRATEGY

SALLY JACKSON, DANIEL J. O'KEEFE, SCOTT JACOBS, AND DALE E. BRASHERS

Three significant points of controversy separate Hunter, Hamilton, and Allen's defense of single-message designs from our suggestion that messages be replicated within experiments. We respond to each of these controversies. First, we examine their claim that "controlled" single-message designs completely eliminate all confounding of manipulated variables with other possible influences on the dependent variable and show it to rest on manifestly implausible assumptions. Second, we show why researchers should plan for the possibility of nonuniform treatment effects across messages, and so use multiple-message designs; contrary to Hunter et al.'s suggestion, the available empirical evidence shows that treatment effects can and do vary across messages. Third, we discuss the advantages of having message replications both within and between studies, as opposed to Hunter et al.'s suggestion that such replication occur only between studies; multiple-message designs provide greater reliability in estimation of treatment effects, equivalent power for detection of variability in treatment effects, and easier identification of moderator variables. Other issues raised by Hunter, Hamilton, and Allen (nested vs. crossed designs, the desirability of experimental manipulation of messages, the benefits of meta-analysis) are in fact not controversial.

MUCH communication research aims at the development and justification of general propositions about messages, but certain difficult methodological problems must be confronted if these aims are to be achieved. Jackson and Jacobs (1983) formulated these problems as having to do with generalizability and considered some abstract strategies for coping with them. The principal recommendation made by Jackson and Jacobs was that multiple messages be incorporated as replications in experiments aimed at general conclusions about messages and that messages be treated as an explicit random factor in statistical analysis. These suggestions and their rationale have been the subject of ongoing discussion (e.g., Bradac, 1983; Hewes, 1983; Jackson, O'Keefe, & Jacobs, 1988; Morley, 1988a, 1988b; D. O'Keefe, Jackson, & Jacobs, 1988).

Hunter, Hamilton, and Allen (1989, henceforth HHA), in contributing to this discussion, offer a defense of single-message experiments. Their essay raises significant methodological issues, but also attacks positions no one has taken. This will surely prove distracting to those who lack familiarity with the ongoing development of issues, but we caution readers not to take at face value HHA's representations of the literature. To make clear and constructive discussion possible, the bogus disagreements need to be sorted from the genuine controversies.

One bogus issue raised by HHA concerns the merits of nesting message replications within treatments versus crossing message replications with treatments. All agree that crossing messages with treatments is generally better than nesting messages under treatments when the problem type is such that recognizably similar content can appear at each treatment level.[1] Jackson and Jacobs (1983) didn't

advocate abandoning crossed designs in favor of nested ones (as should be clear enough from their reference, on p. 173, to "matched-messages" designs); subsequent papers have dealt even more explicitly with this issue (see Jackson, O'Keefe, & Jacobs, 1988, esp. p. 129). Nor, for those cases requiring nested designs, do we dispute the need to treat messages as an explicit factor; on the contrary, we have consistently argued for treating messages as an explicit random factor, whether crossed with or nested within the levels of the message variable of interest.[2]

A second bogus issue raised by HHA concerns the benefits to be gained from experimental manipulation of the message variables of interest. Jackson and Jacobs (1983) are misrepresented as favoring sampling of naturally-occurring messages over experimental manipulation of message features. Jackson and Jacobs did suggest collection of naturally-occurring messages *as the basis for experimental manipulation*, in order to respond to concerns about generalizability. But Jackson and Jacobs clearly expected that the base messages, however *selected*, would be experimentally manipulated (to achieve experimental control) whenever the variable under study permitted that.

A third bogus issue concerns the merits of meta-analysis. HHA misrepresent us as offering "attacks" on meta-analysis, when in fact we have explicitly endorsed (and undertaken) the appropriate use of meta-analysis. From our point of view, the issue is not whether meta-analysis is useful, justified, and worthwhile (we agree that it is all of these), but whether meta-analysis substitutes for well-designed primary research (we think it does not). In the series of papers cited by HHA (Jackson, 1984; Jackson & Jacobs, 1987; Jackson, O'Keefe, & Jacobs, 1988; D. O'Keefe, Jackson, & Jacobs, 1988) we have argued that meta-analysis does not require, nor does it justify, poor design at the level of primary research—but this is no criticism of meta-analysis.

Putting aside these bogus differences, three significant controversies remain. First, HHA claim, and we deny, that "controlled" single-message designs eliminate confounding of manipulated variables with other possible influences on the dependent variable. Second, we suggest, and HHA deny, that researchers should plan for the possibility of nonuniform effects across messages for the sorts of variables researchers commonly seek to manipulate. Third, HHA suggest that replication of an experiment across messages should occur only between studies, whereas we recommend replication both within a study and between studies. We'll consider each of these issues in turn.

"CONTROLLED" SINGLE-MESSAGE DESIGNS

The first controversial question is whether rigorously-controlled single-message designs eliminate all confounding of the manipulated variable with other possible influences on the dependent variable. Our view is that single-message designs, even if very carefully controlled, are vulnerable to confounding of the manipulated variable with other unintended variations from treatment level to treatment level; the inclusion of multiple messages in a controlled design can make the design less vulnerable to such confounding. HHA argue that controlled designs "completely eliminate" all possible confounding (by matching messages so that the messages differ only with respect to the desired variable), and that therefore no replications are needed.

Note that the "controlled" single-message design is obviously an unsuitable

general solution to the problem of confounding. For many of the message phenomena studied experimentally, controlling extraneous variables by matching messages across treatments is patently impossible. For instance, messages representing Marwell and Schmitt's (1967) sixteen compliance-gaining techniques can be matched on topic, but certainly not on content; "threat" messages and "promise" messages, for instance, can't possibly be matched in the required way. Similarly, one-sided and two-sided arguments can be matched in some ways, but not for characteristics of the segments containing the refutational material. Messages on contrasting topics such as "sports" and "politics" (the topics appearing in HHA's "controlled experiment" example, where a single message represents sports and a single message represents politics) plainly cannot be thoroughly matched. The manifest inadequacy of single-message designs for variables like these means that, at best, one might argue that not *all* research problems require multiple message replications.

But even this weaker version of HHA's claim will not withstand scrutiny. The strongest case for single-message designs would seem to be in the study of apparently isolable features of a message—for example, fear appeals or language intensity. But even in these cases, an investigator cannot guarantee that everything except the treatment has been held constant. *HHA's defense of single-message designs requires such a guarantee,* because without one, the single-message design is more vulnerable to the effects of unwanted confounds than is the multiple-message design. HHA will want to claim that control over basic content provides umbrella protection against confounds, but consider the assumptions that are necessary to guarantee freedom from confounding.

First, one must assume that a message can be given a complete theoretical description as a combination of specific values on a finite number of features. If it is not possible to give an exhaustive list of all of the ways in which one message can differ importantly from another, it will not be possible to establish that two messages differ in only one isolated important respect.

This first assumption is manifestly implausible. If it is possible that a future investigator may someday devise a new message variable that will turn out to be important, then there is no guarantee that two messages matched *today* on "everything we can think of" will turn out *tomorrow* to be matched on the new variable as well. Think historically: Researchers in 1940 weren't in a position to match (equate) messages on "language intensity," because that concept wasn't yet available to them. Researchers in 1965 couldn't equate messages on "powerful vs. powerless language," because that concept wasn't available to them. And researchers in 1990 can't equate messages on variables they haven't noticed. Plainly, one cannot presume that we *now* possess every single useful way of conceptualizing message differences.

A second required assumption is that contrasting message elements chosen to manipulate a variable are interchangeable with other elements that might be chosen to represent the same contrast. For example, in comparing the effectiveness of positive and negative altercasting as compliance-gaining strategies, one contrasting pair of examples would have to be assumed to be interchangeable with any other such pair for purposes of generalizing about the strategy types.

But this second assumption, too, is implausible. Consider, for instance, various pairs for contrasting positive and negative altercasting. For positive altercasting, the

options might include (1a) "A person with good qualities would comply," (2a) "Any intelligent person would do X," (3a) "I know you have enough integrity to choose X," (4a) "If you're polite you'll do X," and so on. For negative altercasting, the options include (1b) "Only a person with bad qualities would refuse," (2b) "Only stupid people refuse to do X," (3b) "Refusing X would show a lack of integrity," (4b) "You won't be very polite if you don't do X," and so on. These pair up on both substantive and formal features (and hence represent a level of control beyond what is typical in the research literature), but even so one is hardly in any position to assume that the contrast between appeals (1a) and (1b) is "the same" as the contrast between (2a) and (2b)—or between (3a) and (3b) or (4a) and (4b). Without such an assumption, however, message variables cannot be operationalized satisfactorily through individual examples of their levels.[3] Obviously, there are grounds for caution here.

Notice that our claim is that there are grounds for caution, not that one is always certain to find variation from one instantiation to another. The question of variability is, we think, an empirical question, to be answered with data. Single-message designs spurn such research data, in favor of making an unexamined assumption that all other instantiations of the experimental manipulation will have effects exactly identical to those of the selected manipulation (within sampling error).

A third assumption required for HHA's proposal is that features of messages or message elements can be individually manipulated without inducing change in other features. For example, it is necessary to assume, as HHA do, that adding two-sidedness to a message does not alter the message's values on other features (such as internal consistency), except insofar as such is inherent in two-sidedness.

But this third assumption is also implausible. A change in isolated message elements—even if carefully controlled to avoid any unwanted variation *within* the element—may well have consequences for global message attributes (such as organization, coherence, and internal consistency). Further, a change in one element may have consequences for *other* elements, as in impression-formation studies showing that the manipulation of one or two attributes in a list of traits can change the valence of other attributes held constant across lists (e.g., Delia, 1976).

In short, HHA's defense of "controlled single-message designs" rests on assumptions that are themselves indefensible. Even within controlled designs of the sort HHA envision, replication of the manipulation across multiple messages will be necessary to minimize the confounding of the treatment variable with other message differences. The single-message design requires a guarantee of rigorous, feature-by-feature matching to eliminate confounding. HHA would back this guarantee by faith in the astuteness, perceptiveness, and knowledgeability of the individual researcher to detect possible confounds. Our skepticism toward experimental control as a panacea for problems of confounding is not based on some metaphysical commitment to the ineffable mystery of meaning, but on a realistic appraisal of what is currently known about message variables and on a tough-minded preference for relying on procedural safeguards to eliminate confounding rather than trusting to the personal impressions of individual researchers. Although HHA impute to us a set of views they label "subjectivist," it is we who call for empirical evidence to address an empirical question, while their putatively "objectivist" position argues for a course of action that systematically suppresses the collection of data that might put their own presumptions to test.

MESSAGE-BY-TREATMENT INTERACTIONS

Our second disagreement with HHA concerns whether researchers should plan for the possibility that treatment effects will not be uniform across messages—or equivalently, for the possibility that the particular messages chosen for study may contribute error variance to the estimation of the treatment effect. Within the conceptual framework of a multiple-message design, nonuniform treatment effects across messages would appear as message-by-treatment interactions—as differences in the treatment effect, from one message replication to another. Our position has been that researchers should prepare for the possibility of nonuniformity, by designing research in such a way as to take account of it should it exist within the empirical domain being explored. Specifically, the nonuniformity is taken into account by treating messages as a random factor, so that, among other things, the treatment effect is tested against the message-by-treatment interaction rather than against the subjects-within-groups variance.

HHA argue against this view. Asserting that the rationale for multiple-message designs depends upon "ubiquitous" message-by-treatment interactions, they deny the need for multiple-message designs on the grounds that no message-by-treatment interactions have been documented. They acknowledge that the existence of message-by-treatment interactions would undercut the validity of conclusions drawn from single-message studies, but offer meta-analysis of collections of such studies as the optimal strategy in those cases where message-by-treatment interactions do exist. In other words, they remain skeptical about whether message-by-treatment interactions exist, but believe that even if they do exist, primary research should still be conducted using one experimental message per study. Thus there is disagreement both about whether a state of affairs obtains (whether message-by-treatment interactions exist) and about the implications of that state of affairs for alternative design strategies. We consider each of these issues in turn.

The Empirical Evidence of Message-by-Treatment Interactions

Evidence from multiple-message studies. There is one direct source of evidence for the nonuniformity of effects across messages: In studies in which multiple messages appear as replications crossed with treatments, one can search the studies for direct assessments of the size and significance of the message-by-treatment interactions. Unfortunately, such studies remain relatively rare. Jackson and Jacobs (1987) examined seven studies in which multiple messages were crossed with treatments— all that could be found with either positive or negative findings relevant to the question of whether treatments interact with messages—and noted that among these studies were a number of message-by-treatment interactions reported as significant. Jackson and Jacobs (1987) offered the findings of these studies as empirical grounds for concern about nonuniform treatment effects across messages.

HHA re-examined these seven studies (and 12 additional studies cited by Jackson, O'Keefe, and Jacobs, 1988, as illustrating ways of generating message replications) and claim to have found no interactions, "trivial" interactions, interactions that were "predicted a priori," or interactions treated as inconsequential by the researchers. Jackson and Jacobs (1987) did not represent all of their seven cited studies as showing significant message-by-treatment interactions, and of course not all studies using multiple messages offer any possibility of assessing such interactions. In all, we can identify 11 studies that have crossed replications (messages,

conversations, situations, contexts, etc.) with treatments and offered some assessment of the uniformity of treatment effects across replications, and so provide relevant data. The other eight studies mentioned by HHA incorporated multiple messages, but for various reasons permitted no assessment of the interactions between replications and treatments.[4]

The 11 relevant studies are generally available for inspection. Of these, seven plainly report significant replication-by-treatment interactions (see Bradac & Mulac, 1984, study 1, for interactions involving "message set"; Doelger, Hewes, & Graham, 1986, for interactions involving "context"; Hample & Dallinger, 1987, for interactions involving the dummy-coded "situation" variable; Housel, 1985, for interactions involving "conversation"; Rogers & Mewborn, 1976, for interactions involving "message"; Tracy, 1982, Experiment 2, for interactions involving "conversation"; and Tracy, 1983, for interactions involving "example" of conversational extension type). One study (Jackson, Jacobs, & Rossi, 1987) obtained significant interactions between treatments and replications (specific conversational acts nested within speech act type and crossed with treatment) and took them into account in testing treatment effects but did not include details in the published report. One (Planalp, 1985) used a small number of conversations as replications and reported no statistical test of the replication-by-treatment interaction, but noted that some effects were limited to one of the three conversations. Finally, only two studies reported no interactions between replications and treatments (Hewes, Graham, Doelger, & Pavitt, 1985, for "context"; Tracy, 1984, for "conversation").

That replication-by-treatment interactions have been reported is a matter of public record. What HHA mean by "trivial" is not clear, but whatever they mean must differ from commonly accepted statistical criteria. As to whether the interactions in question were "predicted a priori," (a) none of these authors did predict or could have predicted any interaction between the treatment variables and the message replications factor, whatever other interactions may have been predicted; and (b) whether predicted or unpredicted, a significant message-by-treatment interaction is a message-by-treatment interaction (that is, the fact that an interaction is predicted doesn't make it disappear).

It is important that a number of *statistically significant* message-by-treatment interactions appear within these studies. Achieving statistical significance is not intrinsically impressive, but these significant interactions mean, necessarily, that the interaction variance was larger than the variance computed for subjects-within-groups (or other error term), since an effect can never be significant unless the ratio between effect variance and error variance is greater than 1. Even if such an effect looks small, it will obviously have some consequences for the estimation of the treatment effect. Specifically, if there is any message-by-treatment interaction, the estimate of the treatment effect will contain sampling error of the ordinary sort, and more error due to the "message sample." These sources of error afflict the estimate of the treatment effect regardless of whether the experimenter uses single or multiple messages, but only the multiple-message design allows for calculation of the degree to which the treatment variance must be discounted for the two sources of error. To show that message-by-treatment interactions may safely be ignored, HHA would have to show that it makes no difference to the testing of the treatment effect whether message-by-treatment variance is taken into account in the test of the treatment effect and estimation of true treatment variance. But this they cannot do: The appearance of significant message-by-treatment interactions is sufficient evidence that it *does*

make a difference. (Readers may consult the detailed literature on these matters stimulated by Clark, 1973).

Thus the direct evidence for message-by-treatment interactions, derived from multiple-message studies, suggests that the possibility of variability in treatment effects across messages is indeed something to be taken seriously.

Evidence from meta-analyses. To bolster their conclusion that message-by-treatment interactions may safely be ignored, HHA look to meta-analyses of the persuasive message effects literature, under the presumption that if treatments had nonuniform effects across messages this fact would or should become apparent in meta-analysis. HHA assert that "a general finding" of such meta-analyses is that treatment effects *are* uniform from study to study (and hence, presumably, from message to message), once variance due to artifacts, moderators, and sampling error is removed. This is a defective argument, for two reasons: The relevance of the meta-analytic results to the point at issue is uncertain; and even if the results were relevant, they wouldn't underwrite what HHA propose.

First, the relevance of the meta-analytic results (to the issue of message-to-message variability) is uncertain. Within collections of single-message controlled designs, message-by-treatment interactions are always undetectable within individual studies but *may* appear as nonuniformity in treatment effects across studies. Of course, the existence of nonuniform effects across studies does not establish that there are message-by-treatment interactions, since many factors other than the specific messages used vary from study to study. But according to HHA's reasoning, the *absence* of nonuniform effects would be sufficient to establish a presumption of uniformity across messages (or indeed across any feature that might contribute to the between-studies variability).

In principle, this is incorrect: Uniformity across *studies* may or may not indicate uniformity across *messages*. The reason is that messages used in one study may have been repeated in later studies, so that effect size estimates from several different studies may all be based on the same message. Obviously, in such a circumstance, finding little study-to-study variability wouldn't necessarily mean that there was little message-to-message variability.

And in practice, such message repetition is not uncommon, particularly in the persuasive effects literature cited by HHA. For example, in the fear appeals research, the same base message appears in the series of studies reported by Dabbs and Leventhal (1966), Leventhal, Jones, and Trembly (1966), Leventhal, Singer, and Jones (1965), and Rosen, Terry, and Leventhal (1982); in this research area, several other messages make repeat appearances in pairs or triplets of studies. Likewise, in the language intensity literature, the six studies reported in Burgoon and Chase (1973), Burgoon, Jones, and Stewart (1975, experiments 2 and 3), Burgoon and King (1974), and Chase and Kelly (1976, studies 1 and 2) appear to use a total of two base messages.

Given this research practice, the proper meta-analytic evidence concerning message-to-message variability would be obtained by aggregating effect sizes based on an individual message and allowing each message to contribute only one estimate to the analysis. If K is the number of studies in the entire collection and M is the number of different messages used in the studies, then both the variance of the effect sizes across messages and the estimate of sampling error should be based on M, not K, separate effect sizes.

Unfortunately, the usual meta-analytic procedures (including those used in the published meta-analyses cited by HHA) treat each of the K effect sizes separately, instead of aggregating effect sizes within messages and calculating variance among these aggregate estimates. For purposes of calculating the average effect size weighted by study N, it makes no difference whether M or K effect sizes are tallied, since the combined sample sizes for a given message remain the same either way. However, both the variance among the effect sizes and the estimate of sampling error are affected by whether the calculation is presumed to represent K separate estimates or only M separate estimates. Exactly how much the effect size variance is affected will depend upon the particular circumstances, but the effect on the estimate of sampling error can be precisely specified. If each single-message study is allowed to contribute an effect size, the estimate of sampling error will be K/M times as great as the estimate that would result if each message were allowed to contribute only one effect size:

$$\frac{4K(1 + \bar{d}^2/8)}{N} \quad \text{vs.} \quad \frac{4M(1 + \bar{d}^2/8)}{N}$$

This suggests the strong possibility that the standard meta-analysis procedures, applied to assessment of study-to-study variability, will yield biased assessments of message-to-message variability. Notice that although message-to-message variability remains in principle just one component of study-to-study variation, the non-independence among the study effect size estimates makes it impossible to interpret corrected study-to-study variability as "containing" message-to-message variability. Even substantial message-to-message variability could disappear under a swollen estimate of sampling error based on an inflated count of the number of independent observations. The larger the ratio of K to M, the greater the swelling. In a small subset of studies, a single repetition could lead to a large overestimate of sampling error and a consequent underestimate of message-to-message variability.

In short, meta-analytic procedures for testing of homogeneity of effect sizes across studies are relevant to the issue of homogeneity across messages only if each study uses a different unique message, and that condition is not met in all of the lines of research cited by HHA. If meta-analytic findings are to be brought to bear on the issue of message-to-message uniformity (message-by-treatment interactions), the meta-analyses need to be conducted in such a way as to examine variability among messages, not studies.

We make this point for reasons that transcend HHA's invocation of meta-analytic findings. Our claim is not, for example, that all of the lines of research figuring in HHA's argument are afflicted by non-independence (though obviously some of them are, and that's relevant to the immediate discussion). The larger matter to which we wish to draw attention is the distinction—necessitated by the research practice of re-using messages—between study-to-study and message-to-message variability. Correspondingly, we wish to stress both the need for caution in interpreting existing meta-analyses and the need for greater care in the future application of meta-analysis to the message effects literature. But in any case, the meta-analytic findings cited by HHA are at best uncertainly relevant to the question of the existence of message-to-message variability in treatment effects.

Second: Even if the meta-analytic results were relevant, they would not provide evidence that treatment effects are uniform across messages. That is, even in the

absence of potential problems connected with message repetition, these meta-analytic findings do not underwrite HHA's proposed use of single-message designs.

As background, it is important to understand the meta-analytic procedures involved. In the procedures advocated by Hunter, Schmidt, and Jackson (1982) and HHA, if a group of studies shows significant variability from study to study (after sampling error and artifacts are taken into account), the studies are searched for some feature that differentiates those with effects at one extreme of the range from those with effects at the other extreme. Once such a moderator is identified, its contribution to the study-to-study variability is extracted. This search for moderators continues as long as there is significant unexplained variability from study to study, with the aim of eventually sorting all studies into homogeneous subsets and thereby eliminating all unexplained variability in effect sizes.

Taken at face value, several such meta-analyses would seem to provide evidence of significant variability in treatment effects across messages. In a number of meta-analyses of the persuasive effects literature—including several cited by HHA—the moderator variables used to account for between-study or between-message variability have themselves been message features or message-specific features. For instance, Stiff (1986) concluded that the effect of evidence on persuasive effectiveness varies as a function of how involving the message topic is (for other examples, see Buller, 1986; Dillard, Hunter, & Burgoon, 1984; Jackson & Allen, 1987; D. O'Keefe, 1987).

It is difficult to see how HHA believe such findings show that treatment effects are uniform across messages. On the face of it, such results suggest that treatment effects vary from study to study depending on such message-specific features as message strength, audience favorability to the advocated position, substance of appeal, and beneficiary of the requested action. (It may well be that when the studies are divided into subclasses based on the message-variable moderator, nonuniformity doesn't appear *within* the subclasses. But this is irrelevant; the appearance of nonuniform effects across message subclasses is what suggests the more general conclusion that treatment effects vary across messages.)

Now one should not assume that these results (indicating an apparent role for message-related moderator variables) are necessarily good evidence of message-by-treatment interactions, since these meta-analyses generally assess study-to-study, not message-to-message, variability. If meta-analyses tailored to evaluation of message-to-message variability were to show results similar to the results of the available meta-analyses, such results would obviously be evidence for the need to replicate messages within studies. But presumably the available results will, if anything, underestimate message-to-message variability; at a minimum, thus, the extant meta-analytic findings intimate the possibility of variability in treatment effects across messages—and certainly do not support HHA's claim of uniformity.

On the other hand, one should be careful about taking certain kinds of "negative" meta-analytic results to mean that nonuniform treatment effects across messages do not exist. HHA appear to suggest that if there is no remaining significant variability in effect sizes after taking into account sampling error, artifact, and non-message-variable moderators, then treatment effects must be uniform across messages. But this is wrong. Although the procedures used to evaluate and decompose study-to-study variability can serve useful purposes, they beg the question in dispute here. The reason is that the application of these procedures to any closed set of studies virtually *guarantees* that the creative meta-analyst can succeed in accounting for all

the variance, and can do so without risk of discovering message-variable moderators that might prove inconvenient for arguments such as HHA's.

Consider, for example, Dillard et al.'s (1984) two meta-analyses (of foot-in-the-door and door-in-the-face techniques of persuasion). Each showed substantial initial study-to-study variability in effect size (here represented by correlations between strategy and compliance). In each case, the individual studies were therefore coded for some number of possible moderators (four for the FITD studies, two for the DITF studies), and the correlations of these moderators with effect size were calculated. The moderators *most highly correlated with effect size* were used to "subset" the studies—into four groups for the FITD studies and two for the DITF studies. After subsetting, the study-to-study variability was again computed within subsets, and again compared to sampling error. Whenever a subset showed variability beyond that expected from sampling error, the subset was further subdivided (resulting in three "final" subsets for the DITF studies). Note that this process of subdividing can continue until all of the variability from study to study is exhausted; if no systematic basis for subsetting can be found, an ad hoc basis can. In Dillard et al.'s meta-analysis, the final result (100% of the variability accounted for) was obtained only by *discarding* deviant studies from within subsets, justified in each case by a particularistic, speculative rationale for the study's "oddity." (These analytic procedures are carefully spelled out on pp. 474–479 of their report; apart from the treatment of the "oddities," these procedures are also fairly typical of other meta-analyses cited by HHA.)

Obviously, variance accounted for in this way cannot constitute a compelling case for uniform effects of treatment across messages. The procedure is post hoc and provides no independent test of the ability of the moderators to define homogeneous subsets of new messages. The meta-analyst's position is analogous to the position of a primary researcher, who, after looking at statistical results of an experiment, fools around with the respondents until some blocking variables are discovered that can be used to reduce the within-groups variance. Even if a given moderator appears to account for all the study-to-study variability, one can't be sure that the moderator would exhaust the message-by-treatment variance in an independent test. At a minimum, a justified claim for the role of such moderators will need a new set of independent studies confirming the existence and size of such moderator effects.

Moreover, there is no way to establish that some proposed moderator (or set of moderators) uniquely accounts for the variability. Indeed, different meta-analyses of a line of research sometimes use different moderators to explain variability (e.g., for FITD and DITF techniques, the results of Dillard et al. might be compared with the results of Beaman, Cole, Preston, Klentz, & Steblay, 1983, or Fern, Monroe, & Avila, 1986). Hence the fact that a given moderator set "explains all the variance" is no reason to think that no other moderator set could do so. Consequently, even if sampling error and non-message moderators leave no residual "unexplained" variability, this gives no reason to suppose that no message-related moderator could have significant effects. That is, such results constitute no reason to think that treatment effects are uniform across messages.

Finally, it may be worth remembering that meta-analytic reviews of the persuasion literature consistently find that, even after removing the variability attributable to sampling error and artifact, there still remains enough study-to-study variability to justify the search for moderators (see, e.g., Boster & Mongeau, 1984; Dillard et al., 1984; Jackson & Allen, 1987; Stiff, 1986). Our claim is not that this

positively demonstrates the existence of variability in treatment effects across messages (it doesn't), but only that the existence of such residual study-to-study variability is consistent with the existence of nonuniform treatment effects across messages. And, as just noted, the ability of a creative meta-analyst to "explain" this study-to-study variability post hoc with non-message moderators is not evidence that treatment effects are uniform across messages.

Summary. We have considered two sorts of evidence that might bear on the question of whether treatment effects ever vary from message to message. Studies in which multiple messages are crossed with treatments supply direct evidence as estimates of message-by-treatment interactions and tests of their significance; only a few such studies are available, but these report enough interactions involving message replications to preclude any presumption that treatment effects will not vary from message to message. Indirect evidence could be gathered through meta-analysis, but published analyses commonly assess study-to-study, not message-to-message, variability; even so, there is no reason to think (and good reason to disbelieve) that these meta-analytic results indicate uniformity in treatment effects across messages.

The Import of Message-by-Treatment Variability

Recall that the main question in this section is whether researchers should plan for the possibility of nonuniform treatment effects in designing experiments, by employing multiple-message designs. As we have seen, the empirical grounds for HHA's defense of single-message designs, even taken at face value, support no premise stronger than this: Some portion of the nonuniformity found in certain lines of research can be accounted for in terms of moderators identified post hoc, some of which are message-specific features. But this premise is not remotely adequate to the task of justifying single-message designs as a general practice.

Now HHA apparently believe that if one doesn't know whether treatment effects (for the variable under study) are uniform across messages, one should assume that they are, and so use a single-message design. But a general presumption of treatment invariance is hardly reasonable in view of the evidence. There are too many reported cases of significant message-by-treatment interactions (and too many suspected message-variable moderators) to allow one to assert confidently that treatment effects will not vary across messages. The inability to justify such a general presumption of uniformity, however, underscores the importance of multiple message replications: *A multiple-message design is preferable whenever one does not know in advance that the treatment variable of interest will have uniform effects across the domain of messages to which it might be applied.*

Given that a particular single-message design cannot be defended on the basis of some general presumed treatment invariance, the only other possible basis would be good evidence that the *specific* treatment effect being investigated will be invariant from message to message. But obviously this evidence could only come from empirical examination of the treatment effect across multiple messages—which again suggests that multiple-message designs are to be preferred.

Now it might be thought that the choice (between single- and multiple-message designs) comes down to a matter of timing: Even if single-message designs are used, any nonuniformity in treatment effects across messages will eventually be detected and explained through subsequent meta-analytic reviews—and so (it might be reasoned) one needn't worry about having multiple-message designs. In the next

section we discuss more thoroughly the issue of having message replications only between (single-message) studies as opposed to within (multiple-message) studies, but for the moment we offer a reminder that the point of the primary investigation is to estimate the treatment effect, not to produce grist for the meta-analyst's mill. And whenever there is (or might be) nonuniformity in treatment effects from message to message, and the sources of that uniformity are not known in advance, a single-message study is inadequate as a basis for estimating or testing the treatment effect.

Thus even under a strong presumption that any nonuniformity of treatment effects across messages will eventually be detected and explained, the single-message design is an unattractive choice for the primary researcher. Under any weaker presumption, the single-message design becomes not merely unattractive, but wholly untenable. And a weaker presumption is what seems warranted, if one considers the situation currently faced in studies of message effects: There is reason to suppose that treatment effects may vary from message to message; one cannot know in advance what will explain any such observed variation; and one cannot know how long it will be before explanations are in hand. In such a circumstance, the prudent researcher will prefer the multiple-message design, since unlike the single-message design, the multiple-message design provides empirical data on the existence, nature, and extent of any nonuniformity and allows for this nonuniformity in evaluation of treatment effects.

REPLICATION BETWEEN OR WITHIN STUDIES

Despite their defense of the single-message design, HHA appear to be keenly aware of its shortcomings, as is attested by the centrality of meta-analysis to their proposal. HHA explicitly acknowledge that without multiple studies done to their specifications, no generic claim about messages can be defended unless it is already known that the treatment variable of interest has uniform effects across all messages to which it may have relevance. HHA's complaint that single-message studies are unjustly being denied publication is inconsistent with their tacit acknowledgment that such studies are not *individually* a sound basis for generalization.

So for HHA and us alike, credible generic claims require assessment of multiple messages. In this section we consider just how such multiple-message assessment is best undertaken. It can be done straightforwardly, by incorporating multiple messages within the design of any individual study (as Jackson & Jacobs, 1983, suggested), or it can be done through meta-analysis of many individually produced single-message studies (as HHA suggest).

It is important to be clear about what separates these views. As noted earlier, the choice is not between using and not using meta-analysis; any advantages of meta-analysis can be enjoyed by either single-message or multiple-message designs. The relevant question is the locus of the replicated messages: having such replications only between studies (HHA's single-message proposal) as opposed to having such replications both within and between studies (the multiple-message proposal we defend).

We have elsewhere discussed in general terms the possibility of using meta-analysis of single-message designs as an alternative to multiple-message designs (Jackson, O'Keefe, & Jacobs, 1988; D. O'Keefe, Jackson, & Jacobs, 1988), noting that proposals such as HHA's are afflicted by various practical problems. Many of these problems follow from an inability to guarantee that there will be enough

studies done on any given topic to permit meta-analysis at the time the conclusion is wanted. Others follow from the fact that individual researchers do not (and cannot) plan their experiments with the meta-analyst's interests in mind. These practical problems remain unresolved, but here we turn to a series of statistical issues raised implicitly or explicitly by HHA: reliability in estimation of treatment effects; power for detection of variability in treatment effects; and moderator identification.

Reliability in Estimating Treatment Effects

Given that one's objective is some generic conclusion about the effect of a certain treatment (e.g., message variable), one would like the estimate of the effect to be as reliable as possible. An individual researcher with a fixed number of respondents can either use a single base message, dividing the available respondents among treatment levels, or use multiple message replications, dividing respondents within treatment levels among the various messages. In the latter case, the number of respondents per message will be considerably lower than in the former case, although the total N remains the same. The estimates of treatment effects for each individual message will (obviously) grow less reliable as the number of respondents per message decreases, but the key question concerns the overall estimate of the treatment effect.

Obtaining the overall estimate of the treatment effect is crucial to the objective (a generalization about the effect of the treatment variable). From the standpoint of generalization, the reliability with which one estimates the effect of the variable on any specific message is entirely secondary to the reliability with which one estimates the average effect across messages. When messages are chosen arbitrarily as replications, the estimates of treatment effects for each individual message are of no intrinsic importance (any more than individual human respondents are in such designs); the main problem is to estimate treatment effects across messages.

Now HHA suggest that having reliable estimates of treatment effects is important only when treatment effects are uniform across messages. In HHA's view, if there are significant message-by-treatment interactions, estimation of the treatment effect should no longer be of interest. This mis-application of a common practical principle arises from a failure to see the implications of treating messages as *replications*. In fact, the existence of variability in treatment effects across replications—whether subjects or messages or other sorts of replications—does not diminish one's interest in the main effects of the treatment. (This is one of the many differences between factors whose levels are in themselves important—"fixed" factors—and factors whose levels are mere representatives of a larger class of possible levels—"random" factors.) The fact that persuasive effects of, say, sidedness are not entirely uniform from message to message will not mean there is no value to establishing a general or average effect. Suppose, for instance, that it turns out that two-sidedness has a general advantage, but that this advantage reverses—even reverses unpredictably—for some messages. Both the researcher and the persuader still gain something from knowing the general advantage, since consistent use of the two-sided strategy will produce greater benefits than consistent use of the one-sided strategy.

Applying HHA's alleged "rule" that main effects should not be interpreted in the presence of significant interactions would mean that any time replications showed any degree of case-to-case variability, one would give up generalizations about the treatment variable in favor of "simple main effects" of the treatment for each arbitrary replication studied to date. But the "simple main effects" for each

replication are of no general worth or interest, being restricted to this particular message or that particular situation.

Consider the parallel case of respondent-by-treatment interactions. Such interactions do exist (not only in repeated-measures designs, but also—confounded with other things—in independent groups designs), and once in a while they turn up as explicitly tested effects in research reports (as in Tracy, 1982, 1983, 1984). But the presence of a large respondent-by-treatment interaction in a repeated measures design would of course not nullify our interest in the average treatment effect—yet by HHA's reasoning it should divert that interest into a search for moderators (in this case, moderators of the treatment effect for the individual respondents, that is, "blocking" variables). Certainly if it is possible to break down a respondent-by-treatment interaction, one might want to do so, and of course something is generally gained (both substantively and statistically) when conclusions about the treatment effect are drawn separately for each level of the blocking variable. But HHA go much further than this, suggesting that treatment effects (even those representing simple main effects of treatment at levels of a moderator) are meaningless whenever there is any sort of undepleted treatment-by-replication (e.g., treatment-by-respondent) interaction.

Commonplace (and well-founded) social-scientific practice rejects this notion that a respondent-by-treatment interaction somehow nullifies the validity, desirability, or utility of generalizations about the treatment effect—and parallel notions about message-by-treatment interactions should similarly be rejected. Messages occurring as replications crossed with treatments have the same status vis-a-vis generalization as do respondents in a repeated-measures design. The fact that message replications interact with treatments may invite a search for moderators (or "blocking" variables), but whether that search succeeds or fails, the main effect (generalized across replications) will still be of primary interest. To return, then, to the general point: Contrary to what HHA suggest, securing reliable estimates of treatment effects is important, even given significant interactions between treatments and replications.

So consider, first, the question of the relative reliability of an individual single-message study, as opposed to an individual multiple-message study, for estimating the treatment effect. Expressed in concrete design terms, the choice is between having 1000 respondents respond to one message-pair, and having them respond to 10 message-pairs (with 100 respondents each). As perhaps is apparent, the choice is between getting a very good estimate for one arbitrary message, or getting pretty good estimates for 10 different messages—and insofar as one's objective is to estimate treatment effects across messages, plainly the multiple-message design is to be preferred. Notice that if effect size is not variable from message to message (as would be expected if HHA were correct), dividing the total N among multiple messages would make no statistical difference as compared to spending all the respondents on one message; and if effect size *is* variable from message to message, increasing the number of message replications will naturally increase the reliability with which we estimate the general underlying effect size. In other words, the multiple-message design is never less reliable than the single-message design, and whenever there is any message-by-treatment interaction—even a small one—the multiple-message design is more reliable than the single-message design. HHA do not dispute this, of course.

But now consider the relative reliability when there are multiple single-message studies and multiple multiple-message studies—specifically, when there are enough such studies to warrant meta-analysis. The question at issue is whether meta-analytic combination of single-message studies will yield as reliable an estimate of overall treatment effects as will meta-analytic combination of multiple-message studies.

A relevant small-scale Monte Carlo study, focused on the stability of meta-analytic assessment of the treatment effect, was reported by Jackson and Jacobs (1987). For purposes of the simulation it was assumed that the treatment effect in question was not uniform across individual base messages, but varied around a central value. Within each cell of the study design, 500 simulated meta-analyses were generated, and means and standard deviations computed for the meta-analytic estimate of effect size. Variables manipulated were size of the meta-analysis (15 or 30 studies), size of the individual studies ($N = 200$ or 1000), distribution of effect size (d) within the message population (four distributions: means of .2 or .4, standard deviations of .1 or .2), and number of message replications per study (1, 10, or 20). Since in all cases the N-per-study was fixed (at either 200 or 1000), for studies with multiple messages the number of respondents per message replication was the relevant fraction (1/10 or 1/20) of the total study N. Meta-analyses of single-message studies using $N = 200$ were compared with meta-analyses of multiple-message studies using $N = 200$, and meta-analyses of single-message studies using $N = 1000$ were compared with meta-analyses of multiple-message studies using $N = 1000$.

The estimate of the underlying effect size averaged over many simulations (500 "runs" by either 15 or 30 studies per run) were very similar for single-message and multiple-message studies. But the variability of the estimates depended on whether the studies used single or multiple messages. Greater reliability in the estimate of the treatment effect would be indicated by smaller variability from one simulated meta-analysis to another. For all 16 cells in the design, the standard deviation of the meta-analysis results was considerably smaller for meta-analyses of multiple-message studies than for meta-analyses of single-message studies (although the advantage of 20 messages over 10 messages was neither consistent nor pronounced).

HHA discuss this study, but apparently misunderstood the study's point, procedures, and results. They argue that this increased reliability in estimating the treatment effect is disadvantageous, because (they claim) it means that one is less likely to see that the treatment effect is variable from message to message. This conclusion is false; it appears to derive from confusing (a) the stability of results across simulated meta-analyses and (b) the ability of an individual meta-analysis to evaluate the stability of results across individual studies or messages. (We discuss this latter issue shortly, in considering the power of alternative research strategies for detecting message-to-message variability in effects.)

Thus a set of multiple-message experiments yields a more reliable estimate of treatment effects than does a set of single-message experiments, even with the same N-per-study (that is, even when a given study N is spread across multiple messages as opposed to being spent on just one). So whether one compares the individual single-message study against its multiple-message counterpart, or a set of single-message studies against a corresponding set of multiple-message studies, the more reliable estimate of treatment effects is obtained with multiple-message designs.

Power in Detection of Nonuniform Treatment Effects Across Messages

HHA argue that if treatment effects do vary across messages, such variability will be more difficult to detect within multiple-message studies than through meta-analysis of single-message studies, and hence single-message designs are to be preferred. This argument (1) confuses message-to-message and study-to-study variability, (2) overlooks the possibility of applying meta-analytic procedures to multiple-message designs, and (3) misrepresents the relative power of the relevant procedures.

First, only multiple-message designs provide direct assessment of the size and significance of treatment variability across messages, in the message-by-treatment interaction. Meta-analysis of single-message studies provides only an indirect indication of the possibility of message-by-treatment interaction, through the existence of study-to-study variability (in which, as noted earlier, between-study differences in messages are confounded with other between-study differences). If message-by-treatment interactions exist, an individual multiple-message study allows for estimating and testing such effects, and more important, for taking those effects into account in evaluating the treatment effect. By contrast, an individual single-message study renders the interaction invisible, and even many single-message studies on a common topic do not provide any focused estimation of the message-by-treatment interaction (as opposed to estimation of study-to-study variation).

Second, if meta-analytic statistics were more powerful than analysis of variance statistics (or had some other advantage), a multiple-message design could be analyzed by treating each of the message replications in the experiment as though it were a separate study. Indeed, this was precisely the form of analysis used by Allen, Agee, Dillon, Ray, Shanahan, and Stafford (1989) for a multiple-message study of argument sidedness.

But, third, as it turns out, meta-analysis procedures do not have any particular logical or statistical advantage over analysis of variance statistics. The power of the F-test to detect message-by-treatment interactions within a multiple-message study is neither better nor worse than the power of meta-analysis to detect study-to-study variability within the same set of replications.

We base this conclusion on a second Monte Carlo study (modelled after Spector and Levine, 1987).[5] To understand the logic of the comparison, imagine 1000 respondents divided among 10 messages subjected to a common treatment, and analyzed in two different ways. One way is to consider the 10 messages as constituting 10 separate studies (as Allen et al. did) and to assess message-by-treatment interaction by comparing the variance in the effect sizes for each message with the variance expected from sampling error; this is the Hunter-Schmidt procedure. The second way is to consider the 10 messages as levels of a replication factor crossed with the treatment factor, and to test the message-by-treatment interaction directly, using the mean-squares for subjects-within-groups as the error term. (When the total available N is divided evenly among the replications and the within-cells variance can be assumed homogeneous, F has a simple algebraic relationship to the variance in the effect sizes, so these two proposals are not without common ground.)

Briefly, the Monte Carlo study compared the power of the Hunter-Schmidt procedure and the power of the interaction F-test under a variety of conditions

similar to those examined by Spector and Levine (1987). Where Spector and Levine (1987) generated Monte Carlo estimates of the power of Hunter et al.'s (1982) procedures to detect differences in underlying correlations, we substituted effect sizes (*d*) for correlations and also obtained estimates of the power of *F*. The important results for the present discussion are these: As in Spector and Levine's study, some common ways of applying the Hunter et al. procedure involve de facto Type I error rates ranging between 15% and 40%, but whenever the Type I error rate for the Hunter et al. procedure is brought under control, its power to detect variability of treatment effects across messages is identical to a corresponding *F*-test with the same Type I error rate (a result that is consistent across study sizes and across varying amounts of variability). Meta-analytic procedures in fact enjoy no advantage over ordinary *F*-tests so far as detection of nonuniform treatment effects is concerned.

In sum, the (unevidenced) suggestion that multiple-message studies lack power to detect variability in treatment effects across messages, as compared with equivalent meta-analyses of single-message studies, is false. And it is worth recalling that an individual single-message study has *no* power to detect such variability in effects, so the design advocated by HHA is impotent unless profusely replicated.

Identification of Moderators

A final defense for the "controlled" single-message study is that it facilitates the identification of moderators—factors that can be offered as an account of the study-to-study variability in effect sizes. According to HHA, single-message designs are advantageous for this process because they make estimates of the treatment effect for single messages as stable as possible and thereby minimize the occurrence of "misclassification."[6]

This argument is an extension of earlier arguments (concerning the comparisons of single- and multiple-message designs with respect to reliability and the power to detect nonuniformity), and our initial response to it derives from our response to those arguments. The use of multiple-message designs would multiply the number of cases available for subsetting, thereby increasing the reliability of the average effect size measure within each subset. Just as the estimate of the effect of a treatment is more reliable if based on more messages, the estimate of the effect of the treatment at some particular level of a moderator is more reliable if based on more messages. In fact, for any moderator other than message-specific characteristics (e.g., study features such as design, measurement instruments, respondent characteristics, etc.), the moderator search doesn't even require that the experimental report give a separate treatment effect measure for each message studied (as opposed to one overall measure).

For message-specific factors (message variables or variables that pertain to individual messages rather than to all messages within a study) suspected as moderators, the situation is a bit different. For such moderators, the fact that a study might report only an overall treatment effect rather than message-by-message treatment effects could represent a comparative disadvantage for the multiple-message design, in that studies could not be cleanly divided into subsets based on the suspected moderator unless all messages used in a study just happened to share the same value on the moderator.

But an approach based on multiple-message designs offers an alternative route to the same goal: The search for "moderators" can be done at the level of the individual multiple-message study. That is, where large variability in treatment effects across

messages (whether significant or not) occurs, the messages can be searched for blocking variables the same way a meta-analyst can search a group of studies for moderator variables. (Indeed, once again one sees the possibility—ignored by HHA—of using meta-analysis-like procedures for analyzing data from multiple-message designs.)

This search, whether done by the primary researcher or the meta-analyst, is always a post hoc procedure, and one that works better as a basis for hypothesizing than as a basis for testing. But if it is aimed at discovering message variables that interact with the treatment, searching within a (multiple-message) study has numerous advantages over searching between (single-message) studies. Among these are direct access to the messages (rather than to brief descriptions of the messages in research reports) and control over other potential between-studies influences (population, measurement, etc.). Indeed, a side benefit of this strategy is that it clarifies the analysis of non-message-based moderators at the between-studies level, since the study-to-study variability due to message samples will be reduced to a degree determined by the size of the message samples within each study.

In short, rather than being disadvantageous as compared with searching for moderators in meta-analysis of single-message studies, searching for moderators within and between multiple-message studies would be quite advantageous. Besides, the idea that one will be able to search for moderating message variables between studies is hardly more than a nice fantasy, since individual experimental messages are rarely available for inspection.

CONCLUSION

The "controlled single-message design" is manifestly a poor basis for generalization; the case against such designs is surely closed. Even the defenders of such designs see that they depend on a (still-problematic) meta-analytic salvage operation to be undertaken at some indefinite future time. This is insupportable if any alternative can be found. Multiple-message designs are such an alternative.

Naturally, the incorporation of multiple messages into experiments presents various sorts of practical difficulties, but these are not intractable. How to generate messages, and how many messages to generate, are serious questions needing attention. But progress toward resolving these practical difficulties requires a recognition that we cannot retreat into single-message designs and a corresponding determination to create designs in which message replications fit as naturally as human respondents.

NOTES

[1]Although this essay will focus exclusively on crossed (message-by-treatment) designs, it should be noted that nested designs have legitimate uses in communication research, namely in comparison of message categories that define essentially unrelated subcategories from some more abstract category. HHA acknowledge this point readily enough, but nevertheless assert that nested designs are weak because of their inability to deal with interaction—a non sequitur, since for the sort of message categorization that calls for nesting, interaction isn't possible. It is not possible, for example, for television program content to interact with a classification of program type (say, situation comedy vs. dramatic series vs. variety show), because the content "belongs" to one type or another, and cannot "appear" at multiple levels of type. Variables that might require nesting of replications within levels of the message variable include classifications of speech act types, speech genres, speech situations, and speech topics. HHA give as an example of a moderator just such a classification: sports vs. politics. One cannot control message content across levels of the variable whose levels are sports and politics, hence efforts to generalize about the variable require examination of multiple cases nested within each level. (It will be noted that despite their acknowledgement that variables of this kind require nesting of multiple messages under each of the types to be contrasted, HHA nevertheless propose a

study in which a single sports message and a single political message are each crossed with a language intensity treatment in order to uncover the intensity × topic interaction.)

[2]HHA incorrectly describe Jackson (1984) as recommending that data from messages nested within levels be "pooled." HHA apparently misinterpreted a passage concerning whether a research report should include cell means for each separate message included in a multiple message design. But how one treats messages in the statistical analysis and what level of detail one includes in a report are two distinct issues, and Jackson's (1984) discussion concerned the latter.

[3]In case it is thought that this example depends on having chosen a variable that uniquely defies control, consider the parallel task of creating a high fear appeal for an anti-smoking communication (for a study of high versus low fear appeals). For constructing the high fear appeal, there is a wide range of alternative appeals: the threat of death from cancer, the threat of slow death from emphysema, the threat of disfiguring surgery on the mouth or throat, the threat of chemotherapy and its scary side effects, the threat of lingering pain, and so on. It might be the case that it doesn't matter which of these is picked—but then again it might not be. Even if death from cancer and disfiguring oral surgery evoke equal levels of fear in the audience (and so are equivalent operationalizations with respect to evoked fear), this doesn't mean that these two appeals are identical in every other way. A single-message design is committed to the belief that it doesn't much matter which one is picked: A "rigorously controlled" single-message design putatively permits one to generalize from the effects of one fear appeal contrast to the effects of all other fear appeal contrasts, from the effects of one explicit-versus-implicit-conclusion contrast to all others, from the effect of one argument strength manipulation to all others. But the implausibility of such a belief is already suggested by what little empirical evidence is available; for example, Tracy (1983) found that identically classified instantiations of types of conversational extension showed significantly different patterns of subject selection, and Jackson and Backus (1982) found that identically classified instantiations of Marwell and Schmitt (1967) compliance-gaining techniques showed significantly different patterns of selection.

[4]The following studies incorporate multiple messages, but in ways that do not permit any assessment of replication-by-treatment effects: Cody, Greene, Marston, O'Hair, Baaske, and Schneider (1986); Jackson and Backus (1982); Jackson, Jacobs, Burrell, and Allen (1986); Mulac, Bradac, and Mann (1985); Mulac, Lundell, and Bradac (1986); B. O'Keefe and McCornack (1987); Planalp, Graham, & Paulson (1987); and Tracy, Craig, Smith, and Spisak (1984). Four involved nesting of replications within categories rather than crossing of replications with treatments.

[5]Details of the procedure and results are available in a separate report from Sally Jackson, Department of Communication, 780 Van Vleet Oval #331, University of Oklahoma, Norman, OK 73019.

[6]This argument, by invoking the idea that the effect size for a specific message is crucial to its "classification," presumes a particular image of the moderator search (an image that portrays the searcher as brazenly capitalizing on chance). The reliability of the estimate for a specific message is important only if one means to begin by dividing the studies into subsets based on effect sizes and end by searching for features to differentiate one set from another—rather than first dividing studies into subsets based on a feature of interest and then evaluating the reduction in variability within sets. Note that if one partitions the studies into subsets with large and small (or positive and negative) effect sizes, the ability to find a "moderator" is limited only by one's tolerance for the incoherent, the implausible, and the bizarre.

REFERENCES

Allen, M., Agee, P., Dillon, K., Ray, C., Shanahan, W., & Stafford, S. (1989, May). *Testing a model of message sidedness and message generalizability*. Paper presented at the annual meeting of the International Communication Association, San Francisco.

Beaman, A.L., Cole, C.M., Preston, M., Klentz, B., & Steblay, N.M. (1983). Fifteen years of foot-in-the-door research: A meta-analysis. *Personality and Social Psychology Bulletin, 9*, 181–196.

Boster, F.J., & Mongeau, P. (1984). Fear-arousing persuasive messages. In R.N. Bostrom (Ed.), *Communication yearbook 8* (pp. 330–375). Beverly Hills, CA: Sage.

Bradac, J.J. (1983). On generalizing cabbages, messages, kings, and several other things: The virtues of multiplicity. *Human Communication Research, 9*, 181–187.

Bradac, J.J., & Mulac, A. (1984). A molecular view of powerful and powerless speech styles: Attributional consequences of specific language features and communicator intentions. *Communication Monographs, 51*, 307–319.

Buller, D.B. (1986). Distraction during persuasive communication: A meta-analytic review. *Communication Monographs, 53*, 91–114.

Burgoon, M., & Chase, L.J. (1973). The effects of differential linguistic patterns in messages attempting to induce resistance to persuasion. *Speech Monographs, 40*, 1–7.

Burgoon, M., Jones, S.B., & Stewart, D. (1975). Toward a message-centered theory of persuasion: Three empirical investigations of language intensity. *Human Communication Research, 1*, 240–256.

Burgoon, M., & King, L.B. (1974). The mediation of resistance to persuasion strategies by language variables and active-passive participation. *Human Communication Research, 1*, 30–41.

Chase, L.J., & Kelly, C.W. (1976). Language intensity and resistance to persuasion: A research note. *Human Communication Research, 3*, 82–85.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Cody, M.J., Greene, J.O., Marston, P.J., O'Hair, H.D., Baaske, K.T., & Schneider, M.J. (1986). Situation perception and message strategy selection. In M.L. McLaughlin (Ed.), *Communication yearbook 9* (pp. 390–420). Beverly Hills, CA: Sage.

Dabbs, J.M., & Leventhal, H. (1966). Effects of varying the recommendations in a fear-arousing communication. *Journal of Personality and Social Psychology, 4*, 525–531.

Delia, J.G. (1976). Change of meaning processes in impression formation. *Communication Monographs, 43*, 142–157.

Dillard, J.P., Hunter, J.E., & Burgoon, M. (1984). Sequential-request persuasive strategies: Meta-analysis of foot-in-the-door and door-in-the-face. *Human Communication Research, 10*, 461–488.

Doelger, J.A., Hewes, D.E., & Graham, M. (1986). Knowing when to "second-guess": The mindful analysis of messages. *Human Communication Research, 12*, 301–338.

Fern, E.F., Monroe, K.B., & Avila, R.A. (1986). Effectiveness of multiple request strategies: A synthesis of research results. *Journal of Marketing Research, 23*, 144–152.

Hample, D., & Dallinger, J.M. (1987). Individual differences in cognitive editing standards. *Human Communication Research, 14*, 123–144.

Hewes, D.E. (1983). Confessions of a methodological puritan: A response to Jackson and Jacobs. *Human Communication Research, 9*, 187–191.

Hewes, D.E., Graham, M.L., Doelger, J., & Pavitt, C. (1985). "Second-guessing": Message interpretation in social networks. *Human Communication Research, 11*, 299–334.

Housel, T.J. (1985). Conversational themes and attention focusing strategies: Predicting comprehension and recall. *Communication Quarterly, 33*, 236–253.

Hunter, J.E., Hamilton, M., & Allen, M. (1989). The design and analysis of language experiments in communication. *Communication Monographs, 56*, 341–363.

Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Jackson, S. (1984, April). *Meta-analysis and generalization in communication research.* Paper presented at the annual meeting of the Central States Speech Association, Chicago.

Jackson, S., & Allen, M. (1987, May). *Meta-analysis of the effectiveness of one-sided and two-sided argumentation.* Paper presented at the annual meeting of the International Communication Association, Montreal.

Jackson, S., & Backus, D. (1982). Are compliance-gaining strategies dependent on situational variables? *Central States Speech Journal, 33* (1982), 469–479.

Jackson, S., & Jacobs, S. (1983). Generalizing about messages: Suggestions for design and analysis of experiments. *Human Communication Research, 9*, 169–181.

Jackson, S., & Jacobs, S. (1987, May). *The search for systematic message effects: Contributions of meta-analysis and better design.* Paper presented at the meeting of the International Communication Association, Montreal.

Jackson, S., Jacobs, S., Burrell, N., & Allen, M. (1986). Characterizing ordinary argument: Substantive and methodological issues. *Journal of the American Forensic Association, 23*, 42–57.

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

Jackson, S., Jacobs, S., & Rossi, A. (1987). Conversational relevance: Three experiments on pragmatic connectedness in conversation. In M.L. McLaughlin (Ed.), *Communication yearbook 10* (pp. 323–347). Beverly Hills, CA: Sage.

Leventhal, H., Jones, S., & Trembly, G. (1966). Sex differences in attitude and behavior change under conditions of fear and specific instructions. *Journal of Experimental Social Psychology, 2*, 387–399.

Leventhal, H., Singer, R.P., & Jones, S. (1965). Effects of fear and specificity of recommendation upon attitudes and behavior. *Journal of Personality and Social Psychology, 2*, 20–29.

Marwell, G., & Schmitt, D.R. (1967). Dimensions of compliance-gaining behavior: An empirical analysis. *Sociometry, 30*, 350–364.

Morley, D.D. (1988a). Meta-analytic techniques: When generalizing to message populations is not possible. *Human Communication Research, 15*, 112–126.

Morley, D.D. (1988b). Reply to Jackson, O'Keefe, and Jacobs. *Human Communication Research, 15*, 143–147.

Mulac, A., Bradac, J.J., & Mann, S.K. (1985). Male/female language differences and attributional consequences in children's television. *Human Communication Research, 11*, 481–506.

Mulac, A., Lundell, T.L., & Bradac, J.J. (1986). Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the gender-linked language effect. *Communication Monographs, 53*, 115–129.

O'Keefe, B.J., & McCornack, S.A. (1987). Message design logic and message goal structure: Effects on perceptions of message quality in regulative communication situations. *Human Communication Research, 14*, 68–92.

O'Keefe, D.J., Jackson, S., & Jacobs, S. (1988). Reply to Morley. *Human Communication Research, 15*, 148–151.

O'Keefe, D.J. (1987). The persuasive effects of delaying identification of high- and low-credibility communicators: A meta-analytic review. *Central States Speech Journal, 38*, 63–72.

Planalp, S. (1985). Relational schemata: A test of alternative forms of relational knowledge as guides to communication. *Human Communication Research, 12,* 3–29.

Planalp, S., Graham, M., & Paulson, L. (1987). Cohesive devices in conversation. *Communication Monographs, 54,* 325–343.

Rogers, R.W., & Mewborn, C.R. (1976). Fear appeals and attitude change: Effects of a threat's noxiousness, probability of occurrence, and the efficacy of coping responses. *Journal of Personality and Social Psychology, 34,* 54–61.

Rosen, T.J., Terry, N.S., & Leventhal, H. (1982). The role of esteem and coping in response to a threat communication. *Journal of Research in Personality, 16,* 90–107.

Spector, P.E., & Levine, E.L. (1987). Meta-analysis for integrating study outcomes: A Monte Carlo study of its susceptibility to type I and type II errors. *Journal of Applied Psychology, 72,* 3–9.

Stiff, J.B. (1986). Cognitive processing of persuasive message cues: A meta-analytic review of the effects of supporting information on attitudes. *Communication Monographs, 53,* 75–89.

Tracy, K. (1982). On getting the point: Distinguishing "issues" from "events," an aspect of conversational coherence. In M. Burgoon (Ed.), *Communication yearbook 5* (pp. 279–301). Beverly Hills, CA: Sage.

Tracy, K. (1983). The issue-event distinction: A rule of conversation and its scope condition. *Human Communication Research, 9,* 320–334.

Tracy, K. (1984). The effect of multiple goals on conversational relevance and topic shift. *Communication Monographs, 51,* 274–287.

Tracy, K., Craig, R.T., Smith, M., & Spisak, F. (1984). The discourse of requests: Assessments of a compliance-gaining approach. *Human Communication Research, 10,* 513–538.