

# Communication Research

<http://crx.sagepub.com/>

---

## **Distinguishing Good Speculation from Bad Theory**

Paul M. Hirsch

*Communication Research* 1981 8: 73

DOI: 10.1177/009365028100800103

The online version of this article can be found at:

<http://crx.sagepub.com/content/8/1/73>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Communication Research* can be found at:**

**Email Alerts:** <http://crx.sagepub.com/cgi/alerts>

**Subscriptions:** <http://crx.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://crx.sagepub.com/content/8/1/73.refs.html>

# DISTINGUISHING GOOD SPECULATION FROM BAD THEORY

## Rejoinder to Gerbner et al.

PAUL M. HIRSCH  
*University of Chicago*

The number of my major points which emerge undisputed and fully intact after the preceding "Comment" is impressive and noteworthy. The Annenberg team's reply combines assertions about theory, method, and findings which I shall disentangle and divide more appropriately into separate sections for this rejoinder. Within each of these, the following analytical components of their response (in descending order of importance) will be addressed: (1) *basic issues*, on which (a) they concede, (b) we are in substantial agreement, or (c) they remain silent, despite the length of the "Comment"; (2) *areas they dispute*, but for which no new supporting information, clarifications, or data are provided; (3) *questions raised*, whose answers are already found in the text and tables of my article; (4) *additional questions* on topics which have little or nothing to do with the substance of my presentation; and (5) *suggestions* that I have committed an unconscionable (scholarly) act by holding accountable to the standards of "pure science" some glaring inadequacies in publications promoting the "cumulative

---

AUTHOR'S NOTE: *This project was supported generously by a grant from the John and Mary R. Markle Foundation to facilitate dialogue between the social sciences and humanities in the field of mass communication. For critical discussion and first-rate research assistance, I am especially indebted to Tom Panelas, and gratefully acknowledge the contributions of Stuart Michaels and Sally Kilgore.*

COMMUNICATION RESEARCH, Vol. 8 No. 1, January 1981 73-95  
© 1981 Sage Publications, Inc.

results of a decade of fairly massive cooperative research and theory building."

In the pages to follow, I will respond to the more substantive portions of the comment and leave it to the reader to judge the rest.

### MAJOR FINDINGS FROM THE REANALYSIS

Let us begin by making very clear that Gerbner et al. do not dispute any of the following issues and findings:

- (1) When simultaneous multiple controls were employed to test the hypothesis that heavy television viewing "cultivates" perceptions of the world as a "mean" and "scary" place, virtually no statistical support for the claimed relationship was found in the data. Gerbner et al. restate the consequences of applying this standard statistical procedure as "simple relationships between amount of viewing and some attitudes are wiped out."
- (2) The amount of total variance in these attitude items explained by *all of the independent variables combined* never exceeded 18%, and more typically was less than 10%. Of the five independent variables, television viewing's relative contribution to this small percentage consistently ranked *last or next to last*. The entire search for, and debate over the extent to which television "cultivates" perceptions centers around data for which it can independently explain only a miniscule amount of variance, so small that elementary textbooks on research design suggest such "findings" are nonreportable as positive results and are better presented as *nonfindings*.
- (3) The application of multivariate analysis to the NORC dataset yields no evidence of any consistent patterns at all, nor evidence for linearity in the relation of amount of television viewing to responses on attitude items. Gerbner et al. here acknowledge a "convincing" demonstration of either curvilinear or negative overall relationships "in some cases." (In others, there was no relationship of any kind.) They con-

tinue: "In most cases in the GSS/NORC dataset, the aggregate associations are reduced to trivial proportions."

All of the above statements hold for *both*: the (seven) items taken from the NORC dataset by Gerbner et al. and reanalyzed in the article, as well as the (eleven) new items included in the same reanalysis. They characterize *each* set of items equally well, and continue to hold up even if all the added items (such as those tapping attitude toward suicide) are excluded from the analysis altogether.<sup>1</sup>

- (4) About 90% of the adult population views between 1 and 5 hours per day and falls within one standard deviation of the mean number of hours viewed (2.9). This is a truncated distribution from which to seek variation over a 24-hour day. Examination of respondents whose hours of viewing are furthest from the mean of the distribution found that neither nonviewers' nor extreme viewers' answers to attitude items lent support to the hypothesis that "mean" and "scary" world responses vary linearly or monotonically with amount of television viewing. While they dispute my contention that the responses of nonviewers and extreme viewers are of obvious conceptual significance for testing cultivation theory, the empirical findings are accepted by Gerbner et al., who concede that "'nonviewers' often seem [sic] to be more likely than light viewers to give 'television answers' [and that] it is not unreasonable to question whether they should be lumped together with 'light viewers.'"

At the outset, then, we discover that the Annenberg team has not disputed the empirical findings noted in 1-4 above. At the same time, Gerbner et al. contend that readers should also dismiss these for failing to raise any legitimate doubt as to the empirical value and theoretical basis for their assertions about cultivation effects. Their grounds for avoiding the obvious inferences to be drawn from these findings consist of three claims which are specious and lacking in credibility; first, these findings are based on a reanalysis of only one dataset, while they have analyzed numerous others and come to different conclusions. (The findings are "based on his use of a number of problematic items from virtually

one data base, which, for whatever reasons, show incongruous results.") Second, that the *National Opinion Research Center's* (NORC) General Social Survey is itself unreliable (it contains items which "exhibit some of the weakest associations we have ever found"); that is, the findings are unchallenged but, by implication, this entire dataset should be rejected or heavily discounted because it does not support the "theory." And finally, the reanalysis in Part I—which tests for *overall* effects and positive cultivation differentials—is rendered "irrelevant" for not paying attention to negative signs ("the key to the puzzle"); and also is "obliterated" for not appreciating sufficiently the "empirical evidence leading to the concept of 'mainstreaming.'"

It is clear from their acceptance of the main findings that the Annenberg team must succeed in their effort to discredit the data base from whence they came, in order to (1) avoid conceding the weakness of their "theory," and (2) restore their credibility in reporting results from all the other surveys from which they claim enough empirical support to assert the existence of cultivation effects. Gerbner et al. level a remarkable number of charges at NORC (and myself) in the process—some very serious, others more frivolous. These begin with the simple argument that "all data are flawed in some way" and the suggestion that their accordance of equal weight to surveys which vary widely in size, sampling design, and representativeness is not so much bad science as the "cumulative consistency" which actually makes their reported findings so persuasive, unarguable, and "most compelling." Having thus downplayed the greater reliability inherent in national probability samples, they move on from equating the NORC GSS cumulative N of 4536 (another 1500 questionnaires are now in the field) with convenience samples they have utilized with Ns as low as 116.

The second line of defense is to suggest that the reanalysis took items from NORC's dataset which are unreliable and untrustworthy. Among those with the "weakest associations we have ever found" are two questions concerning

fear of walking alone and approval of using physical violence. Neither is at all central to the conclusions reached in my article, but their treatment of each warrants discussion in light of the repeated claim that these NORC data are unreliable. Of particular interest here is that nowhere in previous articles, in which Gerbner et al. marshaled the same nationally acclaimed data base in support of cultivation theory, have they expressed any doubts about its reliability.

NORC's GSS does differ from most of the surveys which the Annenberg team analyzes and derives its findings from, for it has always been available at low cost to any interested researcher. In contrast, it is unclear where one can turn to gain such easy access to reanalyze responses to the proprietary surveys which Gerbner et al. claim contradict the data in the NORC file. The organizations which conducted these surveys are not even identified in their Violence Profile No. 8.<sup>2</sup> After other researchers reported that these NORC data failed to support the claims of the Annenberg team, and contained seemingly relevant items whose correlates went in the "wrong" direction, Gerbner et al. quickly announced that these data are flawed. My selection, for reanalysis purposes, of the *same* "fear" item employed earlier by the Annenberg team is attacked here on the grounds that readers should realize it is misleading. The "actual violence" item which runs so contrary to expectations from the hypothesis of "across-the-board" effects is attacked on the grounds that (a) they never reported it because the item scored poorly on a test for reliability in constructing a scale; and (b) it yielded "counterintuitive" associations with the standard background variables employed in survey research. The latter "failing" constitutes reason to report, rather than suppress the finding; and its reliability in scaling has nothing to do with its value as a single item.<sup>3</sup>

Gerbner et al. owe the research community a far more detailed and substantiated indictment of this data base before they can be taken seriously in their efforts to undercut its reputation for reliability, careful design, and high quality. NORC's GSS is probably the most widely used and

widely trusted data base employed by social scientists interested in secondary analysis. It was created by the National Science Foundation for the express purpose of permitting widespread analysis and reanalyses of items considered important by the research community. Far from being "just" one among many surveys, it (along with Michigan's CPS Survey) is the only data base utilized by the Annenberg team to which there is easy access, and which is well-known and available to all, and reanalysis of their claimed findings possible by any interested user. In light of their own documented poor performance in analyzing its contents, the burden falls on the Annenberg team to show that results they report from other, less accessible surveys have followed appropriate statistical models, assumptions, and procedures. Rather than attack NORC, they should make these other surveys available to interested researchers for reanalysis, as the GSS has been.

The third argument in the comment proposed for rejecting the damaging implications of my article is that I neglected to give proper weight or consideration to negative signs and evidence leading to the mainstreaming concept. These passages argue, in effect, that apples should be rejected for not being oranges. The "theory" tested in Part I is the "across-the-board" version of the cultivation argument, which specifies a positive sign for the cultivation differential between heavy and light viewers. When cultivation differentials either exhibit a negative sign, or fail to appear between viewing categories, *both* constitute evidence for the null hypothesis and the formulation of being tested.<sup>4</sup> That the explanation in Table 5 of Part I states this exact point does not mean, of course, that the article is unaware of or disinterested in contingent formulations or tests on specific subgroups. In fact, both possibilities were previewed and addressed in the introduction and conclusion of Part I, though not elaborated in detail until the hypotheses about subgroups and cultivation effects were formalized and more fully discussed in Part II. There, the Annenberg team's more recent "refinements" concerning subgroups, mainstreaming, and resonance are

considered and rejected as speculative, nonpredictive, unspecified, post hoc, and irrefutable. We shall see that it is only the very last of these conclusions with which their "Comment" actually takes issue.

### SUBGROUPS AND REFORMULATIONS: CULTIVATION THEORY AND THE "WORLD OF TELEVISION"

In their comment, Gerbner et al. offer two "refinements" of cultivation theory, which they state are intended to "deal with the exceptions" in those (rare?) instances in which the evidence fails to support their continued assertion that "cultivation is often a virtually across-the-board phenomenon." Apparently these concepts, "mainstreaming" and "resonance," need to be invoked only infrequently. In any event, the Annenberg team suggests that if variations in viewers' responses to attitude items fail to fit the patterns called for in one of the "theory's" formulations, it is likely to be compatible with either of the remaining two. As one example, they cite viewers' responses to the NORC question concerning approval of the use of physical violence, notwithstanding their reservations about its possible unreliability. Here, the finding that increased television viewing is associated with *disapproval* of physical violence is not only acknowledged as possible, albeit dubious, evidence *against* the "across-the-board" version of the theory which predicts the opposite; it also is presented as evidence *for* "mainstreaming"—the newer refinement which embraces findings in which response patterns can contradict those anticipated by the more "global" formulation yet to be cited as evidence supporting cultivation theory. This point, stressed in Part II of my article, is only affirmed once again in their response.<sup>5</sup>

Gerbner et al. correctly note that I question whether a "theory" which exhibits such a remarkable facility for

taking as its own confirmation virtually any percentage movements in any direction by survey respondents can be treated seriously as anything more than a tautology. They see nothing peculiar about advancing "the proposition that many, if not most, of the specifications which emerge when overall relationships disappear . . . can be explained by one of these two concepts." After their "decade of theory building," they freely acknowledge that the theory of cultivation cannot anticipate in advance of the data analysis what it expects to find or which version, reformulation, or refinement of the argument will best "explain" whatever surprises or findings emerge from the data. More specifically, the authors provide no information or guidelines concerning the selection of appropriate subgroups, no predictions about whether responses by members of any subgroups to particular items will be consistent or inconsistent, or will vary positively or negatively with amount of television viewed. When explanations are offered for any findings obtained, they are ad hoc or post hoc, except for the recurring insistence that whatever emerged from the data is related somehow to the "cultivation" of respondents' perceptions by the television medium. Indeed, while my article is criticized for failing to distinguish between their own "clearly presented speculations" and "firm conclusions," it seems to me that only Gerbner et al. are able to distinguish between the latter and the former in their own work.

The only exceptions Gerbner et al. actually take to my article's critique of the concepts of "mainstreaming" and "resonance" concern (1) whether such specifications in advance are necessary or desirable, and (2) if conditions exist under which their "theory" could ever be disconfirmed or refuted. Far from being "overstated," as Gerbner et al. characterize the issue of specification, it is precisely this failure to specify the conditions under which mainstreaming and resonance will occur that makes them unacceptable in their present form as explanations for response patterns. Specification is necessary if only to avoid the appearance that these concepts represent little more than selective

reporting and, for mainstreaming, the artifact of regression effects. The absence of operational procedures, for determining the point where the "mainstream" resides for the population and for the comparison of subgroups, or for predicting which items and which subgroups should exhibit mainstreaming patterns as opposed to resonance, and vice versa, demonstrates that, at best, these concepts remain at a primitive stage of development. Ideally, as pointed out in the article, the "mainstream" for an item should be defined independently of amount of television viewing. Gerbner et al. thus consistently ignore the distinction between the testing or confirmation of a hypothesis versus its generation.

They also claim to have demonstrated "clearly" that mainstreaming and resonance are indeed falsifiable. Of the six hypothetical graphs offered as examples, however, at least three also could be taken as support for one or another version of the cultivation hypothesis.<sup>6</sup> And, of the eleven possible outcomes, we can identify only three which could not easily be claimed as confirmation for some version of the cultivation hypothesis. Gerbner et al. have thus provided enough formulations ("specifications") of their theory to allow 73% of the patterns located to also provide support for the "theory." As stated in Part II of my article (n. 9), this "puts forth a hypothesis with an unusually high probability of being confirmed by chance." It is difficult to believe that Gerbner et al. are seriously concerned with establishing criteria for the falsification of their concepts when their own text consistently betrays a strategy of groping relentlessly to find ever new post hoc and conditional formulations for any datum that cannot be explained by any existing formulation. Indeed, nothing in their comment persuades one that where no relationship is found they would *not* propose that television's influence was "canceled out" by the interaction of variables with different subsets of the same group.<sup>7</sup>

A third, alternative hypothesis concerning subgroups was offered and tested in my article. This is the one-tailed extrapolation from the Violence Profiles linking the findings

of "message system analysis" directly to the cultivation hypothesis. Unlike the "concepts" of mainstreaming and resonance, it formalizes a theoretical position, states predictions in advance of the data analysis, specifies the direction in which responses should "move" in order to confirm or disconfirm the hypothesis, and performs a test on items employed previously by the Annenberg team to see if there is supporting evidence for the relationships which are thus specified:

If viewers extrapolate estimates of their own life chances from those befalling their demographic counterparts on television, as Gerbner et al. propose, heavy viewers from subgroups whose members are most often victimized or killed should appear more afraid and suspicious than light and medium viewers from the same groups. Conversely, viewers from the subgroups whose members are least often victimized or killed should reflect television's representation of their relative invulnerability; i.e., their levels of fear and suspicion should decline with increased viewing . . . *Given* the absence of the claimed relationship across the population, the most logical inferences from the Annenberg team's own content analysis is that (a) heavy viewers among those subgroups portrayed as the most victimized would reflect television's "cultivation effect" by appearing more alienated, anomic, or "scared" than light viewers in the same subgroups; and (b) heavy viewers among those subgroups portrayed as more dominant and less often victimized would also reflect television's "cultivation effect" by becoming *less* alienated, fearful, or "scared"—or else showing no attitude change—when compared to light viewers in their own subgroups. Note that this formulation, derived from the texts of Violence Reports 7-10 (1976-1979), specifies *in advance of the data analysis* both a theory and predictions about (1) *which* groups will be affected differentially, (2) *why* one should expect these patterns to emerge, and (3) *in which direction* the attitudes of heavy viewers in particular subgroups are expected to "move." . . . These hypotheses are easily tested, since the Violence Profiles code television characters' attributes in terms of the same standard demographic variables into which survey re-

search sorts respondents. . . . Following the content analysis reported by Gerbner et al., we examined the attitudes of . . . population subgroups reported to exhibit the highest frequency or probability of seeing their demographic counterparts victimized or killed on television. . . . Across 18 items, including those analyzed by Gerbner et al. in the same dataset, we found no empirical support for these contingent propositions derived from the Violence Profiles. None of the "victimized" or "dominant" subgroups exhibited any consistency in appearing more or less "cultivated" in the directions which follow from cultivation theory in its original version [from Part II of Hirsch article].

Gerbner et al. assert that the failure of the audience data to support this clearly formulated hypothesis casts no doubt on the logic of cultivation theory. It is described as a "far from uninteresting contribution," which is tested for the first time in my article and has been selected by them for testing as well; but it also is characterized simultaneously as "probably oversimplistic," and "nowhere to be found in our theory." It is, however, acknowledged to appear in their own published "speculations," which is where I said it derived from in the first place—i.e., the "text" of the Violence Profiles. The absence of empirical support for this formulation of the "theory" provides a suggestive (and speculative) insight about why the Annenberg team ventured forth with "concepts" which resist formalization, fail to specify in advance the groups to which they apply, and fail to predict the direction of responses by the same subgroups to different items. That is, quite simply, because when this *is* done, based on the results I reported, one finds out *there simply are not consistent patterns in any direction among the critical subgroups to which "mainstreaming" and "resonance" are presented as most applicable.*<sup>8</sup> Quite clearly, there is no advantage to be gained for asserting cultivation effects by making the formulation testable enough to become theoretically meaningful but also, at the same time, falsifiable.

## THE "WORLD OF TELEVISION"

An even more fundamental issue is whether cultivation analysis actually has any relation at all to the "world of television" portrayed in the "message system analysis" portions of the Violence Profiles. The original version of the cultivation hypothesis, which posited across-the-board monotonic effects, also was based on the premise that any survey item used to test the hypothesis had to have a specifiable "television answer" which could be contrasted with one that more accurately described life in the "real world." The "television answer" was derived from the content analyses of what "messages" heavy viewers have been exposed to, which are reported annually by Gerbner et al. in their Violence Profiles. This version of the hypothesis predicted that heavy viewers are more likely to give the "television answer" than light viewers, i.e., the sign of cultivation differentials should be positive. If, as we are now told, *any* difference in the responses of heavy and light viewers, whether in a positive or negative direction, is attributable to television's "cultivation" of that discrepancy, then (a) it makes no difference how members of particular subgroups are portrayed in the "world of television," since the directionality of a subgroup's response pattern is no longer specified by the knowledge of whether that group is portrayed on the screen as dominant or vulnerable. In other words, it is no longer a problem for the "theory" if a subgroup like black women, who are most likely to be seen as victims on television, exhibit a positive association between viewing and the "television answer" on one item, and a negative association on another. Also, (b) this makes even more problematic efforts to relate their acceptance of response patterns which move in either direction, to the idea of each item having a *single* television answer.

Post hoc speculations about the relative salience of *each* item to members of each subgroup, based on their responses to every item (taken singly or grouped arbitrarily) does not constitute a theory. Rather, it provides an ad hoc

hypothesis, applied on a case by case basis, in place of an appropriate explanation. This is why the table presented in the article relating amount of television viewed and respondents' attitude to astrological sign is every bit as plausible (or valueless) as Gerbner et al.'s illustrations of their new "concepts" in Violence Profile No. 11.

Indeed, as the article makes clear in more detail, the "concepts" of mainstreaming and resonance, by attributing any percentage movements (including regression effects) to hours of television viewed, not only continues the practice of imposing abstract formulations about "message systems" on the interpretive mind of the viewer. They also indeed "cut loose" cultivation analysis from the content analysis portion of the Violence Profiles by adding post hoc assertions about the salience of each item for each subgroup, and by refusing to predict how respondents' attitudes will correspond to the "messages" transmitted during all the hours watched by heavy viewers.<sup>9</sup>

The Annenberg team's contention that the content analyses they report were *never intended* as predictive of what heavy viewers believe also is simply belied by their long-standing division of survey items of all kinds into "television" and "nontelevision" answers on the explicit assumption that larger proportions of heavy viewers would provide the latter response. Regardless of who "cut loose" each from the other, the denial that television content and viewer attitudes need be related only weakens further the logic and credibility of the entire enterprise. Without a theoretical link between them, it is a mystery how a theory of cultivation effects can even be logically possible.

## CONCLUSION

In their "Comment," Gerbner et al. present a variety of tables to buttress their attack on my article. Several seek to provide evidence for mainstreaming or resonance, missing my point that it is the illogic of these concepts which the

article addresses, rather than their capacity to generate mounds of printout whose relations are uninterpretable, explain practically no variance, and are subject to regression effects. Two other tables provide "tests" for linearity in order to demonstrate that for uncontrolled bivariate relationships there are patterns in the NORC dataset which show a linear trend from the lowest to the highest viewing categories. This pattern, however, which disappears when multiple controls are placed on these items, has no bearing on the conclusions reported in my article.<sup>10</sup>

A second methodological issue is the objection by Gerbner et al. to the inclusion of new batteries of items in Part I's reanalysis of NORC/GSS data. As noted earlier, their removal does not affect any of the major findings reported. In Part I of the article, I provided the wording of each of these items, the basis on which "television answers" were constructed, and a rationale for including them.<sup>11</sup>

The Annenberg team's contention that some of the items added have nothing to do with the content of television raises a larger, more interesting issue: Do the items we *both* used, and those to which they did *not* object, have anything to do with television either? While the procedure of devising "television answers" is supposedly tied to a comparison of knowledge about the relative frequencies of events in the world with their portrayal on television, there is nothing reported in their message system (content) analysis which provides a basis for designating any of the following responses as "television answers":

- "The lot of the average man is getting worse, not better."
- "It's hardly fair to bring a child into the world with the way things look for the future."
- "Most public officials are not really interested in the problems of the average man."

Indeed, little guidance is available from the variables reported in their message system analysis and overall presentation which allow for any unambiguous selection of items and "television answers" from surveys to conform to the criteria that these tie up directly to their own content analyses. As it appears throughout their formulations in general, there is a minimum of formalization, and difficulties arise for any effort seeking to follow and operationalize their own descriptions.<sup>12</sup>

Finally, of the three remaining questions posed about my use or abuse of appropriate methods (and criteria for evaluating their use by others), the answer to one is apparent in the tables and text of the article; a second should also be obvious to Gerbner et al., and the third is a matter of judgment.<sup>13</sup>

All this leads me to restate several conclusions drawn here and in my article. The first emphasizes that the theory of cultivation effects is an interesting speculation, but lacks supporting evidence to allow one to argue it is any more than that. The "theory" fails to predict in advance how television will affect audiences, claims that it has explanatory value for interpreting virtually all differences found in the attitudes of heavy versus light viewers, regardless of their direction and notwithstanding the reduction of television's contribution as an independent variable to "trivial proportions" when it is entered at the same time as other statistical controls. Second, the Annenberg team's "Comment" accepts most of the substantive findings in my reanalysis of their own work, but tries to avoid their serious implications by attacking the NORC dataset from which the items and responses were taken. This is an understandable effort to maintain credibility under duress, but it is also insulting to the national research community of social scientists who supervise the construction and administration of this prestigious annual survey, and to its many users. Significantly, if one discounts Gerbner et al.'s argument

that the cumulative years of the GSS constitute merely one data base among many, and that it is less worthy than the nonpublic and unreanalyzed surveys they claim show "stronger" results, then their defense of the theory's across-the-board version—especially when based on evidence from their *own* questionable analyses—collapses.<sup>14</sup> Prior to the reanalyses of GSS data by Hughes and myself, empirical questions about the Annenberg team's "findings" came only when others encountered difficulty in replicating them on different samples. To find that they fail to replicate on reanalysis of the very same data from which they were reported originally poses more serious problems of confidence for Gerbner et al.'s presentation of cultivation "findings."<sup>15</sup>

Third, beyond questions of whether this or that item is appropriate for testing the "theory" lie two far more serious questions: Does the corpus of the Annenberg team's publication actually contain a clear, operationalizable, testable, formal theory which meaningfully explains or predicts statistical associations found in data, and which is subject to disconfirmation? If, as I propose, it does not, then many of the problems addressed in the Annenberg team's response deserve less careful attention than they have been accorded here and by communication researchers in general. A related question is whether it makes much sense at all to base our tests of television's "effects" on locating correlates of a single survey item on number of hours viewed. The assumptions that it is of no consequence which programs are seen, that they all transmit the same world view, that viewers interpret stories solely in terms of the characters' gross demographic characteristics, that single incidents are seen by viewers separate from the narrative, context, and genre in which they occur, that viewers do not distinguish what is fantasy from reality in popular entertainment, and that it is not the lifestyles and prior attitudes of individuals which account for the amount of television they view (rather than the opposite) are all

untested assumptions built into the framework and rhetoric of the Annenberg team's Violence Profiles as well as the research strategies of others in this field (see Hirsch with Robinson; 1980). Research into the validity of these assumptions, and the conditions under which each may operate, is needed as much as or more than continued iterations of atheoretical efforts to locate (weak) statistical associations with hours viewed.

My article also commended the Annenberg team for its pioneer efforts to empirically investigate linkages between the content of television and its relation to viewers' perceptions and attitudes, and for assembling a vast archive of content analyses within the limited framework of its assumptions about which aspects of content should make a difference. My criticism of their work is directed not at these shared goals, but at the methodological errors and unwarranted claims about theories and findings which characterize their effort to implement and reach these common objectives.

## NOTES

1. The Annenberg team has objected to items on suicide because they bear no relation to the "world of television." While this is arguable, simply on the grounds that they bear no less relation than items which they did choose (e.g., "It's hardly fair to bring a child into the world with the way things look for the future"), the important point here is that when *all* of the items which they did not take from the NORC file are excluded from the analysis, the negative findings remain. When multiple controls were placed on just the items they used, the article showed there is no stronger relationship with amount of television viewing; the amount of variance explained is no higher; and the absence of linearity remains the same. For the seven items they reported, the responses of nonviewers and extreme viewers ran counter to the direction predicted by the hypothesis in 9 of the 14 tables (64%) described in my article, while for the 11 added items, the number of tables running counter to the direction predicted by the hypothesis is 15 out of 22 (68%).

2. Gerbner et al. begin their critique by asserting that I failed to note their reliance on national probability samples other than those of NORC and the Michigan Survey Research Center. They imply that this "omission" was deliberate and exemplifies my "misrepresentation" of their procedures. A rereading of

Violence Profiles 7-10, to which Part I of my reanalysis is directed, confirms that prior to Summer 1980 no other national surveys are cited or named by the Annenberg team. Further, only in Violence Profile No. 8 does the possible utilization of alternative national surveys arise. This possibility, however, is contained only in an obscure reference (pp. 177-178) to a "quota sample" and to one "national probability sample." I mention these datasets in note 5 of Part I (p. 452) but was not able to identify their sources because they were not given. This manner of citation is hardly conducive to making information accessible to researchers wishing to reanalyze their reported results. To read that I failed to take into account their use of Harris, Starch, and ORC data is news to me, for only the Opinion Research Corporation's Survey has ever been cited by name in the Violence Profiles, and that only appears in No. 11 (1980).

3. Even if an item's reliability were questionable, the appearance of theoretically meaningful items on a dataset from which other questions are analyzed and reported naturally leads one to suspect the worst (i.e., selective reporting) when no mention is made of findings which contradict the argument made for a theory. In addition to the item on approval of "actual violence," the NORC/GSS codebook contains additional questions about crime and ownership of firearms whose response patterns run counter to predictions from the cultivation hypothesis (Hughes, 1980), are also unreported by Gerbner et al., and whose reliability they have not questioned.

4. To assert that Part I's test of the "across-the-board" variant of the hypothesis is rendered "irrelevant" by a more recent formulation which instead argues for subgroup effects (while still claiming support for the more global version) has no bearing on whether or not the first formulation is supported by data rather than mere assertion. Conceptually, because the first formulation predicts positive cultivation differentials, the one-tailed test is appropriate. The introduction of "refinements" which fail to predict directionality require that the criteria for empirical support be relaxed and that two-tailed tests be administered. These two formulations are analytically distinct and require separate tests and treatments, as provided in my article. In criticizing the first for not being the second, Gerbner et al. confound and obscure elementary canons of hypothesis testing and research design.

5. Throughout Part II of the article, the use of the term "cultivation analysis" by Gerbner et al. to embrace, take credit for, and account for contradictory patterns in the data is both highlighted and criticized. One example provided, which also speaks to the way in which Gerbner et al. used the "physical violence" item in their comment, appears early in Part II's comparison of subgroups. When the responses of black men and black women reversed from one item to the next, we interpreted this as evidence for no pattern, but continued by noting: "Gerbner et al.'s contradictory formulations 2 and 3 ('mainstreaming' and 'resonance') assemble and interpret very similar statistics as *support* for the cultivation hypothesis; i.e., regardless of whether the cultivation differentials are larger or smaller. This will be discussed further in the following section."

Along these same lines, it is relevant to note the Annenberg team's utilization of Morris Rosenberg's (1968) standard text. Their invocation of his injunction that "survey analysis involves the hot pursuit of an idea down paths and byways which have little to do with one's original hypothesis"; and "a reluctance to follow the

lead of the findings may stultify and abort a good deal of promising research" is curious in light of my article's findings and critique. Indeed, I could not agree more with Rosenberg's observations. If high-quality social research entails going where the data take us, then we must be prepared to *relinquish* hypotheses and formulations which fail to find confirmation. This is clearly the case for the "across-the-board" version of the cultivation hypothesis. However, Gerbner et al. continue to cling to it tenaciously, even as they offer "refinements" which contradict it. Their reluctance to drop this discredited formulation, or to finally specify the conditions under which it obtains, violates Rosenberg's wise counsel more than it follows it.

6. Graph f is clearly an example of an across-the-board effect, thus conforming to the original formulation of the cultivation hypothesis and having no need for the introduction of either mainstreaming or resonance. Graph i meets the conditions for "mainstreaming," since the distance which separates heavy viewers is less than that between light viewers. Graph j is interpretable as a case of resonance, since one of the subgroups which starts out high on the dependent variable gives the television answer with increased frequency as its consumption of television increases. For graphs of this sort, taking only two subgroups, each of which can either increase, decrease, or remain the same across viewing categories, a total of nine discrete patterns can be generated, plus two additional cases where they would cross.

7. Taking two examples from their comment alone, we offer the following illustrations: (1) "further examination of differences within [a] group—controlling for additional variables—might reveal a theoretically intelligible pattern underlying the apparent lack of relationship with television viewing . . . the implementation of additional controls might uncover their presence in distinct subsets of the groups. . . . This kind of elaboration is a fundamental task of scientific analysis—to delve ever more deeply into phenomena, to examine layers of association, with an ever sharper focus." (2) "But an overall, aggregate relationship is simply the product of subrelationships which may pull at each other in different directions and with varying intensity."

The problem these statements suggest is that in light of their past performance, the Annenberg team is simply too committed to finding television effects to stop "delving" and further dividing up subgroups long enough to consider their hypothesis disconfirmed.

8. While Gerbner et al. credit my article for presenting the first test of this proposition, they nevertheless challenge the validity of its test "because of the small number of groups examined and the comparison of inappropriately matched groups." Although the Annenberg team's content analysis could perhaps generate additional subgroups for which this formulation could be tested, race, sex, and age are clearly three of the principal axes of differential victimization in the "world of television." Failure to find response patterns which correspond to those on television for racial and sexual subgroups is severely damaging to the hypothesis, even if the test is not exhaustive.

The second objection, that we examined "inappropriately matched groups," is also of dubious merit. Gerbner et al. charge, for example, that we "equate" nonwhite characters with black respondents, thus clouding the value of the test provided. But, as they are well aware, the percentage of nonwhites other than

blacks in the NORC dataset (and in other probability samples) is less than 1%, not enough to make a difference in any multivariate analysis. Relatedly, we note that to confine our test to black respondents is not to "equate" them with all nonwhite characters on television. If the violence profiles indicate, as Gerbner et al. say, that to be "other than clearly white" makes one more vulnerable, then blacks (who are other than clearly white) should exhibit the expected response patterns when compared to whites in any event, but they did not.

9. Gerbner et al.'s assertion that I "imply" viewers be "conscious" of how television's role in allegedly "cultivating" misperceptions about the "real world" is unfounded and does not follow from the text (or notes) in the article.

10. Gerbner et al. readily concede that as soon as multiple controls are placed on the relation of television viewing to attitude items, the linear "pattern" found in the less appropriate, uncontrolled bivariate context is reduced to "trivial proportions." The 22 tests they report in Table 1 of the comment bear no relation to this central point for, once again, their tests are for zero-order relationships, prior to any application of even a single control. No information or citation explaining the source of either test is provided, but based on the table alone, the following points are relevant: The logic of any such test weights the proportions observed by the number of cases in each cell. Here, over 90% of the sample falls in the three middle categories; hence, because light, medium, and heavy viewers' responses to the attitude items follow a pattern of upward movement before multiple controls are introduced, the results for the entire distribution are indeed "significantly linear."

The Annenberg team's use of this "finding" is misleading on several counts. It is *precisely the problem raised in my article*, which noted that the marked deviation of responses by nonviewers and extreme viewers is masked and obscured when they are collapsed with those of viewers in other categories (light and heavy, respectively), whose cell sizes are much larger. What is most damaging to the cultivation hypothesis conceptually—e.g., that nonviewers provide the "television answer" in larger proportions than light viewers and others—is the very issue which the Annenberg team uses these "tests" to avoid. Both tests, driven by sample size, serve to restate the obvious point that the viewing distribution is highly concentrated, with most adults reporting 1-5 hours per day. If any ambiguity surrounds the meaning of my suggestion early in Part I, that the formulation of the hypothesis as linear is challenged when respondents at both tails of the distribution fail to provide answers in the expected direction, it should be obvious from all of the text which follows that the assertion of linearity is undercut totally—for the entire distribution—as soon as multiple controls are placed on the zero-order relationships.

11. To be sure, I also questioned the theoretical logic of this whole procedure as the right way to go about linking television content to audience impact and interpretation: "There is, of course, some question about whether examining a single survey item and its correlates is adequate to this task. . . . However, accepting this presumption for the sake of argument, this article presents a reanalysis of a significant body of data on which much in the cultivation analysis reports are based."

The article noted that of all the items selected for analysis by the Annenberg team, only the question about fear of going outdoors enabled a comparison of the

likelihood of victimization in the "real world," by subgroup, with the television portrayal of each subgroup's probability of being victimized. We explicitly excluded the following two items, on the grounds that "Gerbner et al.'s basis for deriving a 'television answer' for them is too obscure for us to feel comfortable in adopting it." (p. 454):

—"Do you expect the United States to fight in another war within the next ten years?"

—"Do you think it will be best for the future of this country if we take an active part in world affairs, or if we stay out of world affairs?"

The responses taken as "television answers" for items unreported by the Annenberg team were chosen on the basis of my own inferences from the text of their Violence Profiles. Except for contending that suicide items bear no relation to the "world of television," they do not take issue with these decisions on the appropriate "television answers" for the seven additional items taken from the NORC/GSS file.

12. The procedure of devising "television answers" is supposedly tied to the results of message system analysis, as in the use of questions about crime rates and the percentage of the workforce employed in law enforcement by Gerbner et al. in Violence Profile No. 8. With the questions on law enforcement officers and crime rates, the prevalence of a given phenomenon in the real world may be compared with the frequency of its appearance on television because there are operational procedures for determining both (i.e., statistical measurement). From the discrepancy between the prevalence of an event in the real world and its frequency of appearance on television, it is possible to determine which of an item's response categories is closest to the "television answer." This is only possible, however, when a phenomenon's rate of occurrence in *both* the real world and the world of television may be determined. (For example, content analysis reveals the percentage of television characters employed in law enforcement, while official statistics supply this percentage for the real world.) The researcher's inability to make *either* of these determinations for a given survey item means that that item cannot be said to have a television answer and should therefore be disqualified from use in cultivation analysis. This indeterminacy is clearly present in the anomia items above and, in our judgment, is also evident in the "mean world" items used by Gerbner et al. and us. Our acceptance of these items has been for argument's sake, since we continue to find their use for this purpose puzzling, at best.

In the reanalysis, my rationale for examining items tapping respondents' attitude toward suicide was included in the following summary: "Whereas a fair inference from the Violence Profiles is that television viewing 'cultivates' doubts about life's sanctity . . . we find that both television nonviewers and extreme viewers are *least likely* to see any of these circumstances as justifying the taking of one's own life" (pp. 425-426, italics in original).

In the absence of any more logical explanation for their inclusion of items which have no clearer bearing on the "world of television" than the suicide battery, it is difficult to see why Gerbner et al. so peremptorily dismiss my use of these items in their comment. Since the percentage of the total sample responding

"yes" to items like "Public officials are not really interested in the problems of the average man" is as high as 62%, there is a stronger basis for asking whether perhaps *all* these items warrant exclusion than for seeking to determine only which of them is *most* inappropriate.

13. The criticism that the years and Ns for items used are unclear from my presentation is unwarranted. The article states clearly that all questions come from the surveys of 1975, 1977, and 1978 (the three years for which the item on viewing habits is available). The N provided in each table makes it easy to determine whether an item is from one, two, or three years, and I note each of the years for which every item is available when they are introduced in the text (even if each corresponding table does not restate this).

Equally uncalled for is the complaint about no explicit statement about whether the variables in the Multiple Classification Analysis (MCA) were entered as factors or covariates. Since the MCA tables provide breakdowns for all the independent variables used, it is clear that they were entered as factors, since covariates are continuous variables for which MCA yields no such discrete breakdowns. Since this is explained in all writeups of the program, and since Gerbner et al. report they have used it themselves, it seems odd that they failed to understand this.

The issue of whether it is advisable to dichotomize television viewing by breaking the sample at the median, as Gerbner et al. began doing in 1979's Violence Profile (No. 10), is a matter of judgment. In my view, this practice has two principal drawbacks which recommend against it: (1) It eliminates the category of medium viewers and makes it far more likely that any curvilinearity they represent or contribute will escape detection; and (2) it leads to a loss of comparability across samples since the categories of light and heavy viewing will be defined differently for each one.

Finally, Gerbner et al. and I agree that their statement—that each new violence profile takes precedence over those of previous years and renders the earliest reports inoperative—is also open to interpretation. I have taken it to mean, among other things, that the data analysis procedures used in the more recent profiles are considered by the Annenberg team to be improvements over those utilized in the earlier profiles, and that they view their current practice of dichotomizing television viewing as an improvement over earlier, more differentiated codings.

14. Part II of my article presented three alternative formulations to interpret findings from the GSS data. First, that heavy viewing is simply a function of more free time and availability. The demographic characteristics of heavy and extreme viewers, for example, show a disproportionate percentage of housewives and unemployed and retired people. If viewing and alienation are (weakly) associated, it could also be that heavy viewers watch more because they are alienated to begin with, rather than that television viewing makes them that way. Second, that the concept of status inconsistency offers a plausible explanation for some of the more interesting findings. It would account for higher mean scores on anomia, for example, among highly educated viewers who watch a great deal of television. Finally, since only weak relationships, where there are any, emerge from the data, the possibility that they are caused by random variation cannot be excluded or discounted. Gerbner et al. address none of these alternative formulations in their comment.

15. I am not saying here that the same findings did not appear when identical tests were conducted following the same procedures; rather, the analysis strategies were inappropriate. The former charge would be more serious, but is clearly unwarranted.

## REFERENCES

- HIRSCH, P. with J. ROBINSON (1980) "A research agenda for approaching the study of television." Commissioned for the Aspen Institute Conference on Proposals for a Center for the Study of Television, October. (unpublished)
- HUGHES, M. (1980) "The fruits of cultivation analysis: a reexamination of some effects of television watching." *Public Opinion Q.* 44 (Fall): 287-302.
- ROSENBERG, M. (1968) *The Logic of Survey Analysis*. New York: Basic Books.