Feshbach and Singer study (2) and the Wells study (5), while equivocal, did find that television does not have a strong effect. The strong effect indicated in the Lefkowitz et al. study (3) is tempered by the small sample size and the possible capitalization on chance. Thus, I see this study as being consistent with the literature, and not inconsistent, as others have claimed. (One worry, of course, is that a weak effect for all could be a strong effect for a few.)

Unfortunately, the whole area has become overly politicized and one wonders whether many of the actors involved really care about the data. They have already made up their minds. For those of us open to change, the NBC study is a very important piece of a puzzle that is not yet completed. Milavsky and his colleagues have told us what that piece looks like, but we need a concentrated secondary analysis of their data by other research teams if we are to determine its exact shape and just how it fits into the puzzle.

REFERENCES

1. Davis, J. A., J. L. Spaeth, and C. A. Huson. "Technique for Analyzing the Effects of Group Composition." *American Sociological Review* 26, 1961, pp. 215–225.
2. Feshbach, S. and R. D. Singer. *Television and Aggression: An Experimental Field Study*. San Francisco: Jossey-Bass, 1971.
3. Lefkowitz, M., L. Eron, L. Walder, and L. R. Huesmann. "Television Violence and Child Aggression: A Follow-up Study." *Television and Social Behavior, Reports and Papers,* Vol. 3. Washington, D.C.: National Institute of Mental Health, 1972.
4. Steiger, J. H. "Tests for Comparing Elements of a Correlation Matrix." *Psychological Bulletin* 77, 1980, pp. 245–251.
5. Wells, W. D. "Television and Aggression: Replication of an Experimental Field Study." Unpublished manuscript, Graduate School of Business, University of Chicago, 1973.

EDITORS' NOTE: Prior to publication, the reviewer's sharing of his review with the authors of the book resulted in our receiving unsolicited the comment printed below; the reviewer's response follows.

A Comment by J. Ronald Milavksy, Ronald C. Kessler, Horst H. Stipp, and William S. Rubens

David Kenny raises two kinds of issues. One has to do with substance; the other involves questions about the context surrounding our study. We will discuss each of these in turn.

Kenny starts his review by raising questions about the slow pace of the analysis and about the sponsorship of the research and resulting authors' biases.

J. Ronald Milavsky and Horst H. Stipp are in the Social Research Department and William S. Rubens is in the Research Department, all at the National Broadcasting Company, New York. Ronald C. Kessler is at the Department of Sociology and the Institute for Social Research, University of Michigan, Ann Arbor.

Pace of the analysis. When we designed the study, we had no inkling that the relationship between viewing and aggression would turn out to be as small as it did. Our efforts to grapple with that problem account for the slow pace of the analysis.

In the initial analysis we used a tabular approach. Kenny is correct when he writes that "this early work was externally criticized." In fact, it was first criticized by Kenny himself. His persuasive argument was that relationships as small as those reported could be artifactual results of measurement error and regression to the mean. This convinced us that more sensitive analysis methods than we had used initially were required to tease out the true substantive message of the data.

At that time there was no adequate way to take into account measurement error in panel studies. Jöreskog did not publish his paper offering a solution to assessing measurement error in panel studies using structural equations until 1973 (2); we first learned of it in 1975. In all, there was about a three-year period in which we consulted with a number of scholars on how best to conduct the analysis. They were in agreement that the approach we used was inadequate, but they did not agree on a preferred method. Indeed, they disagreed quite sharply on the latter.

The issues raised in these discussions were finally resolved after Ronald Kessler began his collaboration with us in 1976 and we decided that the structural equations approach was best. From then on, it was a matter of obtaining the LISREL program, getting it to run on our computers, and learning how to use and interpret it. That's when the data analysis really began, starting with redoing all the early tabular analyses.

Kenny is right when he says that we left no stone unturned. But no matter how we looked at the data, we came to the same conclusion: that there is a much smaller relationship between television violence and aggressive behavior than the experimental literature would suggest. We took a long time to come to that conclusion because we wanted to be sure that any error we might have made in underestimating the impact of television was corrected. We thought of a great many potential biasing factors—all those raised by Kenny and more—and concluded that the absence of a substantial association is real, not an artifact.

Given this persistence, coupled with the care we took to document all that we did, we are puzzled by Kenny's substantive comments. Most of his criticisms are answered in the book itself; others could have been recognized as implausible by a careful examination of the documentation.

As to Kenny's point about our possible biases, we can only say that we stated that this project was funded and carried out by NBC. We attempted to deal with the bias issue in the most direct and effective way, not by general discussion but by documenting everything we did as clearly and in as much detail as possible.

All of Kenny's substantive criticisms argue, in one way or another, that we have underestimated the impact of television violence on children: by artificially inflating cross-sectional associations or the stability of the aggression measure, by working with invalid measures of television viewing and aggression, or by choosing an inappropriate test of significance. We take up each of these criticisms in turn and show that none of them has any merit.

Scaling procedures and wave 1 correlation. Kenny suggests that we artificially inflated the wave 1 correlation between television and aggression in our scaling procedures for the two measures. The size of correlations did not enter in any way into decisions about how to scale the television and aggression measures. Indeed, it is difficult for us to understand the genesis of Kenny's idea, given that we devoted two full chapters to discussions of scaling procedures. On p. 59 we stated that the aggression measure was created "by adding together the percentage scores on each of the separate items," and on pp. 75–80 we described the construction of the television measures, where reported viewing for each program was converted to half-hour units, a violence weight specific to individual programs was multiplied by the units, and these weighted products were summed to yield the violence exposure score.

In any case, we cannot understand Kenny's placing such emphasis on a pattern of wave 1 correlations being larger than others, as this pattern is not consistent. It does not exist among boys who are valid reporters of their television viewing—among whom the key analyses were done—nor does it exist for three of the four aggression measures among teenagers.

Finally, Kenny suggests that the wave 1 correlations should be smaller than others because they contain high proportions of second and third graders, "for whom past research has shown weaker correlations between television violence and aggression." We do not know what the basis for this assertion is, but major surveys conducted by Eron and Huesmann (1) and by J. and D. Singer (3) do not show pronounced or consistent differences in exposure and aggression correlations between children in these grades and older children. In short, there is nothing that suggests that there is anything unusual about our wave 1 correlations.

Stability of the aggression measure. Kenny also raises the possibility that nomination procedures on which the aggression measure was based inflated its stability by introducing a "nomination bias"—a between-classroom difference in the probability of children with the same behavior being nominated as aggressive because of differences in the nomination behavior of those who rate them. He notices that the split-half reliability of the nominations of boy and girl raters averages a good deal below that of the inter-item reliability, which is based on all raters. Kenny suggests that this might have come about due to systematic differences in the nominating behavior of boys and girls and that this may have caused a "modest" underestimation of the coefficients for the effect of television. We did find a bias of this sort: girls are more likely than boys to nominate their classmates as aggressive. However, as reported in our book on p. 65, this difference introduces only a trivial amount of bias into the aggression score, because there is very little variability in the sex-ratio of classrooms. Indeed, in that same discussion, we reported the cumulative effects of five classroom context characteristics that we considered as potential biasing factors. As shown there, the total variance in aggression accounted for by all these factors was so small that we decided to ignore them in the substantive analyses.

A direct test to answer the point raised by Kenny confirms our decision to ignore these biasing factors. We repeated all the basic regression analyses—that is, fifteen wave-pair regressions estimated separately for boys, girls, and for boys and girls combined—with a revised aggression score that takes out between-classroom variance by subtracting the class mean from each individual's score. This score is the individual aggression score minus the class mean. If Kenny's assumptions were correct, we would expect the regression coefficients of television predicting later aggression to be larger in these equations than in those based on the original aggression measure. Instead, they were slightly smaller than those obtained in our original analyses.

Thus, there is no evidence for the existence of nomination bias that deflates estimates of television's effects in our data.

Validity of the aggression measure. Kenny uses the "striking result" of a lagged association of .3 between average classroom aggression and the individual's aggression to cast doubt on the validity of the aggression measure. As reported on p. 187, the cross-sectional correlation between classroom aggression and individual aggression contributes a large portion of this association, which makes the result far from remarkable. To find that there is a correlation between average classroom aggression and the aggression of the individual is to do nothing more than show that we were successful in realizing one aim of our sampling strategy: to select schools that differed considerably in aggregate socioeconomic context as a way of maximizing variance on our dependent variable. There is nothing in this that calls the validity of the aggression measure into doubt.

Validity of the exposure measure. As described in the book (pp. 69–75), our television exposure measure is based on viewing information collected from each child in each wave on about fifty programs and televised movies, covering all parts of the day. It is perhaps the most elaborate exposure measure ever used in studies on this issue. As Kenny points out, we devised a number of checks on the validity of this measure.[1] Kenny raises three additional issues about its validity.

First, he notes that we present only aggregate validity data—the correspondence between viewing reports obtained by children in the study and commercial ratings data for children of their ages. These parallels are, of course, not definitive in establishing validity and we never claimed that they were. Rather, we present these data, among several others, as *indications* of the validity of our data. Kenny, though, chooses to focus on the limitations, noting that these results say nothing about the individual data which could, in his view, "be invalid." True enough; they could be, but it strikes us as unwarranted to begin the next paragraph with the introduction, "*Further* doubts about the validity of the viewing measure," as if the observation that an aggregate test is not definitive somehow casts doubt on our measure's validity.

The second issue deals with Kenny's "further doubts" based on patterns in the data that he finds puzzling. One is that the correlation between violence viewing and overall viewing is .92 for boys, a result Kenny describes as "incredibly high." In fact, it is easily explained, as stated in footnote 24 on pp. 82–83: children, like adults, do not select television shows on the basis of their violence content. Those who watch a lot of violence watch a lot of television in general, and vice versa. Further, we oversampled violent shows in our lists, so they form a larger percentage of total viewing in our subset of television shows than in the total number of shows a child might watch.

A final reason that Kenny doubts the validity of the exposure measure is that, in his judgment, "factor analyses yield nonsensical viewing factors." We don't

[1] The results of the validity analyses are reported in the book on pp. 85–88 and by Stipp (4).

see it that way. In fact, in our discussion of television viewing factors among boys, we reported information on 42 factors and showed 39 to be clearly identifiable in terms of known patterns of viewing—by part of day, station, and content. Only 3 of the 42 factors were uninterpretable.[2]

On the basis of these considerations, we find nothing to suggest that Kenny's doubts about the validity of the viewing measure are well founded.

Choice of appropriate significance test. Kenny's next criticism touches on an issue that we grappled with for a long time: how to make an overall assessment of statistical significance for a set of partially overlapping regression coefficients. Kenny asserts that the decision rule we developed is too stringent and that a more reasonable one shows the television effect to be significant among girls and nearly significant among boys. We agree that the rule we developed is not the only reasonable one, but we disagree with the interpretation and conclusion Kenny offers in its stead.

Kenny is quite correct that, in a set of fifteen independent coefficients, "the chances of finding three or more consistent effects in either direction are less than one percent," but he is too quick in dismissing the complications introduced by the fact that these fifteen wave-pairs are not independent. His citation of a calculation demonstrating that the coefficients will, on average, be correlated only .35 is really quite beside the point. The average is much less important than the magnitude of the largest correlations. When we look at the latter, it becomes clear that overlap among wave-pairs plays a substantial part in the results that he takes as evidence for a television effect.

The evidence Kenny cites as nearly meeting his criterion for global significance among boys is that television coefficients are significant in two of fifteen wave-pairs and a third closely approaches significance. All three of these coefficients are positive. However, Kenny does not take into account that the two significant coefficients are in wave-pairs 2–4 and 3–4, which overlap considerably. Waves 2 and 3 are separated by only four months, and the adjusted correlation between the television scores in these waves is .96 (as reported on p. 88). When the wave 3 television score is controlled, the significant association between wave 2 television and wave 4 aggression disappears. This overlap is discussed explicitly in our interpretation of Table 6.1. The effect of controlling intervening television is presented on the very next page in Table 6.2. This subsequent table also shows that the coefficient that comes close to significance in Table 6.1 is clearly insignificant when the overlap among waves is controlled. We are left, then, with one significant coefficient out of fifteen, hardly evidence that supports a claim for a consistent influence of television on aggression.

Much the same can be said for Kenny's reinterpretation of the data on girls, where two of the three significant television coefficients are in wave-pairs 2–4 and 3–4. The adjusted correlation between the television scores in waves 2 and 3 is .91 (p. 231). When we control for overlapping waves, a previously insignificant coefficient in wave-pair 1–3 becomes significantly negative. Once again, we find it difficult to draw the conclusion that television is consistently leading to increased aggression.

[2] See pp. 196–197. Among girls (not reported in the book), 30 out of 36 made sense to us; among teens (p. 445), 19 out of 22 did.

Kenny's conclusion. Kenny's overall conclusion is that evidence for an effect of television exists in our elementary school data, but that the effect is weak.[3] In this sense he sees our data as consistent with the literature. He goes on to say, though, that "[the] conclusion drawn by the authors that television has no effect is an oversimplification." We object to this characterization.

Our conclusion was that we could not detect any lagged association between television and aggression that is significant by conventional standards. We did note and report a pattern of small positive coefficients and stated: "This requires the consideration of the possibility that television influences aggression to a small extent" (p. 189). It turned out that, when some controls were introduced into the basic analysis models, the pattern became less pronounced. But, we concluded, "we are dealing here with very small, insignificant associations, which makes it impossible to draw absolutely firm conclusions" (p. 191).

In short, there is no way of knowing, on the basis of these data, if the insignificant positive coefficients found are due to a television effect, to unmeasured common causes of television viewing and aggression, to unconsidered methodological factors, or to some combination of all three.

Kenny may wish to hazard a guess that a regression coefficient of about .04 between television violence viewing and subsequent change in aggression, uncorrected for any control variables, represents a small effect of television, but we do not.

As stated at the outset, we didn't design our study to make sense of a set of coefficients as small as those we found, and to try doing so after the fact is no more than chasing after shadows.

REFERENCES

1. Eron, L. D. "Parent-Child Interaction, Television Violence and Aggression of Children." *American Psychologist* 37(2), 1982, pp. 197–211.
2. Jöreskog, K. G. "A General Method for Estimating a Linear Structural Equation System." In A. S. Goldberger and O. D. Duncan (Eds.) *Structural Equation Models in the Social Sciences.* New York: Academic Press, 1973, pp. 85–112.
3. Singer, J. L. and D. G. Singer. *Television, Imagination, and Aggression.* Hillsdale, N.J.: Lawrence Erlbaum, 1981.
4. Stipp, H. H. "Validity in Social Research: Measuring Children's Television Exposure." Unpublished Ph.D. dissertation, Columbia University, 1975.

A *Response* by David A. Kenny

The reply by Milavsky, Kessler, Stipp, and Rubens to my review raises a number of questions to which I will briefly respond. First, my raising the issue of the corporate sponsorship of their project was not intended as a criticism. A surprising number of otherwise intelligent social scientists have dismissed the NBC study because of that sponsorship; such a dismissal is unfair and invalid. I

[3] Kenny agrees with our finding that there are no significant associations in the teen data and suggests that one should not expect to find an effect among that age group. We know of no empirical support or plausible theoretical rationale for that suggestion in the literature and certainly know there was none in 1969 when we designed the study.