



CAMBRIDGE
UNIVERSITY PRESS

Nonattitudes and American Public Opinion: Comment: The Status of Nonattitudes

Author(s): Philip E. Converse

Source: *The American Political Science Review*, Vol. 68, No. 2 (Jun., 1974), pp. 650-660

Published by: [American Political Science Association](#)

Stable URL: <http://www.jstor.org/stable/1959510>

Accessed: 27-02-2016 23:38 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Political Science Association and Cambridge University Press are collaborating with JSTOR to digitize, preserve and extend access to *The American Political Science Review*.

<http://www.jstor.org>

Comment: The Status of Nonattitudes

PHILIP E. CONVERSE

University of Michigan

This is the fourth set of extended comments I have been asked to make on this evolving manuscript, my third effort having been a 27-page rejoinder prepared for publication. The original version of the Pierce-Rose piece contained, in addition to several interesting and fresh ideas, a lengthy section impugning my methods, common sense, and good faith in reporting the work I did years ago with what I called the "black-and-white model." This polemic rested, as it has turned out, on a labyrinth of factual misconceptions as to what I had actually done or why, underpinned by some embarrassing confusions on the authors' part with respect to elementary distinctions in the philosophy of science, the nature of attitude measurement, types of error structure and the like.

The majority of these flaws, including the *ad hominem* elements and the most egregious factual errors, have been progressively weeded out over three distinct versions of the manuscript. But in each new version new inaccuracies or misleading statements have relentlessly appeared, requiring another round of commentary on my side. As of this writing I have spent well over one hundred hours in these tutorial communications, a matter which may help to explain why my patience has worn thin.

A small fraction of this time has, as it seems to me, been well spent. One of the dozen or more issues covered below—that regarding the effect of response set in these estimates—was left unexplained in my original article because of space pressures, and I am pleased to have an opportunity to discuss it here. The other nine-tenths of the time and space has been spent, however, either on direct misreadings of things made clear in my original text that seem to have caused no confusion among astute readers in the profession, or on matters usually sifted clear in elementary statistics courses. I have doubts as to whether any of this has been worth anybody's time, or surely this much space in the *Review*. However, even in its current version the Pierce-Rose piece seems likely to stir up a maze of unnecessary confusion, and hence requires this fourth rejoinder.

I

It is still not clear to me whether the authors quite grasp the purpose of the black-and-white

model, as used in the original context. Therefore I would like to begin with some background.

The observations that ultimately led me to write those pieces had originated years before in my own interviewing experience. In that period I had been struck by two facts in particular. One was that very many respondents could not understand that a battery of pure opinion items had no objective "right-wrong" scoring, or that don't know responses were not a confession of the most abject ignorance, to be avoided at all cost. The other was the frequency with which respondents chose a response alternative dutifully but accompanied the choice with side cues (shoulder-shrugging, eye-rolling, giggles, and even *sotto voce* comments) indicating that they were very much out of their element and would pick any alternative haphazardly by way of helping me out. I should make very clear that such events were rare to nonexistent when questions dealt with the immediate terms of the respondent's life, such as whether the respondent was satisfied with his housing or how he would feel about blacks moving in next door. But they began to mount significantly in questions concerning politics, and hit something of a maximum on the national issue items of the structured kind being discussed here. I should also be clear that even these episodes involved nothing approaching a majority of respondents: they remained a limited minority. In addition to those who betrayed themselves in such relaxed fashion, however, there were substantial numbers of further respondents of harassed mien who were obviously struggling to pick an answer and who, if there were the embarrassment of follow-up probes, could muster no self-starting comments whatever on the issue, suggesting that their failure to give a high sign as to the absurdity of the situation was mainly an effort to save face. In short, I came to feel that the tiny proportion of cases—usually one or two per cent—in polls and surveys of the period registered as explicit "don't knows" on items of this type was an underestimate of heroic proportions.

I should add that this is not some idiosyncratic experience of my own. Most professional interviewers are painfully aware of the same

phenomenon. Moreover, one of the most persistent themes of disenchantment with survey research that arises among our graduate students who do actual interviewing in connection with the Detroit Area Study practicum is the spectacle of older scholars lacking interviewing experience, who treat tabular data on attitudes as though it were all real and meaningful in the Pierce-Rose mode, whereas the students themselves have become sensitized in the interaction of real interviews to the fact that some fractions of it are simply not meaningful, as they or I would define the term.

Such observations lay behind one major innovation that we introduced in the 1956 study with respect to the structured issue items Pierce and Rose discuss. Instead of browbeating our respondents into giving opinions they did not feel they possessed, as most interviewers were instructed to do in that period in order to minimize the analyst's "missing data," we explicitly invited respondents who had no opinions on a particular issue to report that fact directly, instead of laboring to concoct some kind of meaningless data point for us.

The result of this supplementary operation handsomely confirmed our expectations. The proportion of self-confessed nonattitudes increased by an order of magnitude over the conventionally-trivial residue of "don't knows" left after strong interviewer pressure to force a response. Even on the two most familiar items of the set of issues, the "no opinions" almost reached 10 per cent. On the issue most remote to the average respondent—the power and housing item central to this debate—the proportion passed to 27 per cent, with a further handful of "don't knows" bringing the total to 29 per cent. There was in addition another 7 per cent who said they were not sure how they stood. Such a high tide of null responses was in fact one of several reasons that I had for selecting this item for more concentrated attention as a limiting case of the nonattitude phenomenon. Some 29 to 36 per cent of the adult electorate of 1956 appeared to have nonattitudes by self-confession on this issue, despite its centrality in elite debate of the preceding thirty years, and this much was true without any elaborate chain of statistical inference or fligree of mathematical assumptions whatever.

The other reasons why I became engrossed with this item as a limiting case came into view once we had cross-time data over three waves from our panel. Basically, these other reasons involved a perplexing tension between two empirical statistics that could be derived

from the cross-time behavior of the power and housing *content* responses (the high tide of no opinions set aside). One of these two statistics is handsomely represented in Table 4 of the Pierce-Rose text. There we see that the power and housing issue is also a limiting case in terms of what the authors call the "stable attitude share of attitude variance," and one which is used as justification for the fifth assumption desired for the Pierce-Rose model, to wit: "(v) no attitude change occurs." The labelling of this table is quite misleading, and if we were not thoroughly familiar with the data being discussed, we would be highly perplexed at the results it shows.¹ But we had been using the algorithm the authors call "causal path analysis" long before Heise popularized the relevant formulae, and indeed our interpretation of the Heise-form data squares almost exactly with the meaning attributed by Pierce and Rose to the relevant statistics presented in Figure 4: where the limiting case of the power and housing item is concerned (although not for the other issue items), the cross-time data structure implies that "no attitude change occurs." However, we say "almost exactly," rather than "exactly," because what lies between the estimates extracted by the Heise route and the vastly different entries one would make in Table 4 if its labelling meant what it appeared to (see footnote 1) is the amendment to our consensual interpretation "no (*true*) attitude change occurs."

The second datum about the power and housing issue which seems striking from cross-time observation is that the correlation of *content* responses (keeping aside the 29 to 36 per cent of self-confessed null responses) is as miserably low as it is (about .3). This means that massive change is occurring in these responses from administration to administration. Yet we have just finished agreeing with the Pierce-Rose interpretation of the Heise-

¹ I assume, for example, that most readers would take the footnote phrase of Table 4 referring to "correlations between attitudes on a question in the two samplings" to mean a correlation formed on the set of apparent content responses (ranging from strong-agree to strong-disagree) at two successive administrations. The reader should be advised, however, that the square of the latter type of correlation for the public power item is not 99%, but somewhere in the vicinity of 10%. A number of the other items, like school integration, run higher, although probably none exceeds 25% and few exceed 20%. In other words, the quanta presented in Table 4 are not what most readers would believe, and frankly I recognize them only from my own lengthy work on the problem. Ironically, in using these estimates, the authors are much closer to accepting the logic of the black-and-white model than they appear to understand.

form statistics which seem to say that "no (true) attitude change occurs" in the limiting case of the power and housing item. Hopefully the reader will have no trouble in sensing the tension here: there is no change in (true) attitudes, yet massive rates of change in attitude response are registered. We can return to the placid simplicity of the ancient syllogism to resolve this tension. We must conclude that the massive change in responses must arise from something other than "true attitudes." This "something other" in turn must, as the simplest logical matter, be something out in the *terra incognita* of the set of "non-true attitudes," or manifest responses in one way or another disengaged from the realm of true attitudes. This brings us perilously close to the notion of nonattitudes.

It does not bring us all the way, because I felt there should be at least one more demonstrable property of such non-true attitudes to validate my suspicion that there was an integral connection between the haphazard, meaningless response alternatives I felt I had seen respondents adopt in some cases, and the more general realms of "non-true attitudes." What I really wanted to know was whether people whose responses to the power and housing item did change sides from interview to interview despite the evidence that no true attitude change was occurring *also* showed a random time path in their responses, which is to say, a correlation between responses at two time points which approximated .00.

It was this focussed analytic question that the black-and-white model was invented to solve. It appeared to solve it very nicely. Given the evidence of the Heise type that on the power and housing issue no true attitude change occurred, the set of people who changed from side to side in successive trials would have to represent a pure group of the phenomenon of "non-true attitudes."² Such

² Earlier versions of the Pierce-Rose manuscript were extremely exercised over my decision to work with this dichotomization of the sample. The authors felt that once into a partitioning game any change of position, even from "strong" to "weak" on the agree side, must be counted as a nonattitude. They also insinuated that by using a dichotomy in the critical correlational test, I was trying to hide a non-fit that would have shown through because of an edge of stability in the weak-strong choice within a side of the issue. The latter allegation was easy to dispose of, since in fact the correlational test was *not* performed dichotomously, but rather on the four-by-four table using the strength differentiation. This left the authors complaining about the original dichotomy, and in my third review I laid out my several reasons for deciding this was the wisest procedure. However, this current version of the Pierce-Rose manuscript, revised with those reasons in hand,

a pure group could be defined on two waves of the panel, saving the behavior in the third wave as an independent test of their cross-time behavior. The crucial stability correlation arising from this test turned out to be essentially zero, as I had come to suspect it would. The remaining group of people who did not change sides on the issue, had to be an impure group of some "true" attitudes and people who by chance took the same side twice. There is no way of splitting this group into two pure types, but the rest of the logic makes it easy to calculate the proportions of the mix in this group, and hence the stability correlation between the second and third wave that should pertain *if* such a mixture of stable true attitudes and persons showing only a random time path in their responses, in known proportions, were present. The test showed almost exactly this correlation. Hence I was willing to conclude that in addition to the 29 to 36 per cent of the sample with *self-confessed* nonattitudes on the power and housing issue, the remaining majority of content responses were themselves strongly adulterated by a contingent responding "as if" randomly across time, and hence who might reasonably be coded as "nonattitudes" as well. It was when all of these were added in to the self-confessed nonattitudes that my estimates of the prevalence of nonattitudes in this limiting case rose as high as they did.

These results were corroborated by another test of somewhat different form suggested by Wiggins. The Wiggins test also is appropriate only if there is no true change in attitude. It takes the configuration of data from three waves at the same time and estimates the response probability between "agree" and "disagree" of those members of the sample who do not have "true" attitudes leading to probabilities of .00 or 1.00 of choosing the agree response. The probability estimated from the power and housing issue was .52, or very close to the expected 50-50 toss-up. The Wiggins test has one advantage over the black-and-white test, inasmuch as its estimates are invariant whatever the actual temporal sequence of the waves is taken to be. However, like the black-and-white test, the assumption of no true change in attitude means that the Wiggins test is not applicable beyond the single issue of power and housing.

seems to have largely if not completely subsided on this issue. Therefore to save space, I will not reiterate those reasons here, but I would be glad to mail anybody concerned about the matter a copy of my third review.

Two further observations about the test with the black-and-white model deserve stress. This stress is not new: the same emphases were given in the original "Belief Systems" article. But one of the reasons why this private tutorial has been as long and costly as it has is because Pierce and Rose have insisted on ignoring these points. In the original presentation I spent two or more paragraphs trying to make as clear as I knew how that the use of the black-and-white model hinged on (1) side evidence that no change in true attitudes occurred; and (2) goodness of fit of the data when given an independent test with the third wave. Now it is clear from the Pierce-Rose Table 4 that the evidence for the first condition becomes shaky once beyond the power and housing issue. I pointed out that this was so, and counselled against blind application to other items. I did myself experiment with a couple of the further issue items that most nearly approximated condition (1), and found that they looked correspondingly shaky on condition (2); hence I dropped the matter with the extended counsel against stupidly overapplying the model to other items. The problem is a straightforward one: the black-and-white model is extremely simple, and once there is reason to believe that some observed change in responses is true attitude change, the situation is underspecified and blind cranking with the model is likely to produce nonsense estimates.

Pierce and Rose set out originally with this broadcast application of the model, even to panel items beyond the issue set, and took great glee at the nonsense results generated. They attributed these nonsense results to failures in my own thought and methods, and used them to cast doubt on the meaningfulness of the results even in the case of the power and housing issue. Most of these nonsense results have been deleted in shortening the article to its current version, but a few remain and constitute some of what might appear to the casual reader as the more "telling" points the authors muster.

The second observation to be stressed is closely related and involves the power and housing issue as the limiting case of the nonattitude phenomenon. We may recall that it was not only the limiting case for stability of true attitudes (Table 4), but also the limiting case for prevalence of *self-confessed* nonattitudes, across this issue set. Further work with the black-and-white model implied that it was probably also a limiting case of *concealed* nonattitudes as well. I treated it very clearly as such a limiting or extreme case. I pointed out,

and still firmly believe, that the demonstration of substantial numbers of concealed nonattitudes in the power and housing case very likely meant that there were also substantial numbers of concealed nonattitudes on the other issues as well, although I should make clear here that logically they must be fewer. I pointed out that with the underspecification problem the black-and-white model could not provide us with estimates of these lesser numbers, so that the argument had to remain intuitive. However, the possibility that all instances of concealed nonattitudes occurred on the power and housing item and none whatever occurred on any of the other similar items in similar format seemed wildly implausible.

I am a bit sorry that I failed to rough out some estimates as to nonattitude prevalence on the other issues simply by way of making clear my own sense that of data available, the power and housing issue was in fact a limiting or extreme case. However, it is obvious that it was so. For example, some of the more familiar issues had self-confessed nonattitude rates 20 per cent below that for the power and housing issue, so that even if everything else were equal, the estimate of nonattitude prevalence for these items must necessarily be cut back from the extreme in the neighborhood of 70 per cent to something closer to 50 per cent. But everything else was not equal: some of these issues had higher stability correlations across time than the power and housing issue, and hence fewer concealed nonattitudes in addition to the fewer self-confessed ones; also there was for these issues greater evidence of true attitude change, reducing by a still further margin the potential pool of concealed nonattitudes.

While I did not feel empowered to make such estimates beyond the limiting case of power and housing, I did think that I made the limiting status of that item clear by spending large conceptual sections of my article attempting to formulate what accounted for the range of variation in prevalence of nonattitudes as one moved from one type of issue to another. This was the whole motivation for the introduction of explanatory constructs such as the varying centrality of content to the individual, variations in information, group stereotype, and the like. Perhaps I do not write clearly, but it does not seem to me I would have spent so much effort trying to explain the range of variation in nonattitude prevalence below the limiting case set by the power and housing issue if I believed that this range was absent or, for that matter, even

fairly small. Yet this is another point that has consistently eluded the current authors, despite several rounds of my pointing it out, and the obviousness of the original text.

We have not begun to cover all of the topics on which Pierce and Rose seem reluctant to get beyond the kind of "maximum possible misunderstanding" they originally developed. However, we can deal with further of these if we move to details of their current text.

II

The abstract of the Pierce-Rose article summarizes the main finding of the piece: "The conclusions reject the thesis that only 20 to 30 per cent of the American public has true attitudes (on issues of public policy). . . ." One must read into the text to find out whose thesis this preposterous statement is supposed to be, but it is alleged to be mine. I have tried in the few preceding paragraphs to make clear why such a sweeping statement departs widely from any general thesis I have developed. At first I thought this attribution was an honest product of sloppy scholarship. I now rue the time I have spent trying to explain through this tutorial why it is inaccurate, since it has become clear that the authors are not very interested. I think I have explained my own position well enough above, and shall not belabor the point, although it seems more basic to the debate than it should be.

It will be useful further background for many later points if we review the outlines of the Pierce-Rose conception of attitudes. The key supposition is that a "true and stable attitude" is not a point, but a range on an underlying attitude continuum. In such a conception, different manifest responses to the same item by the same individual are no sign of any attitude unreliability or instability on his part: if anything, they help to show that true attitudes are in fact ranges, not points.

The conception of such a range at a latent level is closely akin to what James S. Coleman has used in his analysis models for panel data, although he has elaborated it further and geared its assumptions more appropriately to survey data than the Pierce-Rose proposal. It is a conception I have always found attractive, within limits, although it figured only indirectly in the black-and-white model. Perhaps its most attractive feature is that it can in principle address the problem of differential crystallization of attitudes. In speaking of nonattitudes, I had in mind a degree of crystallization which would stand at or surely close to zero. At the other extreme, one can imagine persons with atti-

tudes so sharply crystallized on a particular topic that if the truth were known, their attitude range would be shrunken if not all the way to a single point, at least close enough to that state that the coarse response categories of survey instruments could find no range at all. But these are hypothetical extremes, and it would be reasonable to expect most people most of the time to lie somewhere between these poles on issues they know about, and truly to have a meaningful attitude *range* in exactly the Pierce-Rose sense.

When I say I find this conception attractive within limits, one limit comes when one asks whether there is any range so large (from one end of an attitude continuum to its polar opposite, for example) that it may be rather misleading to say the individual has a "true and stable attitude" toward the competing alternatives. Surely the original ascendancy of the attitude construct in social psychology hinged on the assumption that many important attitudes were crystallized to reasonably narrow ranges and were stable within those ranges, so that they could be used as tools to predict other choice behaviors. If any range on a position issue, be it ever so large, is expressive of the same true attitude, then the predictive power which originally justified the construct has evaporated, outside of situations where, empirically, attitudes are in fact of short range. Moreover, the whole rationale for studies of attitude stability is definitely snipped away by such a view, although only in a rather shallow semantic sense. In other words, if one attitude shows high stability within a set of individuals over time and another shows very low stability with the same index, it is not that stability differs; presumably all ranges have the same stability, and it is the *length* of the ranges that is hypothesized to be varying. Perhaps that change of label gains something, although it is not entirely clear to me what it is. Or again, in a practical political context, perhaps if a person with a wide-range attitude on a public referendum eagerly votes pro on Sunday and eagerly con on Tuesday, it is desirable to say that both votes are expressive of the same true and stable attitude. This usage, however, would seem to have obvious drawbacks.

Most of the charm of the range concept, which to our mind lies in variations from person to person in attitude crystallization or attitude range, is however promptly removed in the Pierce-Rose model by the assumption that "(iv) all respondents have the same attitude range." This assumption is quite unsettling ex-

actly because of the evidence on the power issue from the black-and-white test, which made the point that the response-giving part of the population was almost maximally *heterogeneous* in terms of attitude crystallization. One subset was sharply enough crystallized in its opinions not to stray at all from "agree" or "disagree," while the other subset was so uncrystallized as to turn up on any side in repetitive trials, in a pattern of statistical independence over time. Yet the superstructure calculations of the Pierce-Rose model seem to rest on a foundation assumption of perfect population *homogeneity* in these regards, an assumption already known to be thoroughly untrue. Moreover, the authors propose to re-examine the heterogeneity results of my tests with this new model which denies by assumption in advance that such heterogeneity exists. I am open to persuasion that such an exercise has some internal logic, but it continues to escape me.

With the general conception of attitudes involved in the Pierce-Rose model reviewed, we can turn to some related issues that have helped to perpetuate this debate over many months. We have had a dreadful time, for example, with the concept of randomness. My demonstration of a "random" time path for the responses of some respondents led the authors in an earlier version to charge that by imagining these responses were "uncaused" behaviors, I was undermining the deterministic foundations of social science. So I was obliged to explain the obvious: all sorts of processes leave random tracks, including the outcomes of coin flips, without in any ultimate sense being "uncaused." Indeed, the short-term force component figuring so prominently in the Pierce-Rose model has some random properties, as does more conventional measurement error. Yet every error has some proximate cause; the important point is that the list of such proximate causes is enormous and unrelated to the object of measurement itself—just as the Pierce-Rose "short-term forces" are by explicit assumption independent of the attitude itself—so that random properties can often be assumed. The response choices made to survey items by people to whom the content means little or nothing may show random properties across time as well, but I never imagined that they were uncaused in any ultimate sense. The random pattern arises precisely because all sorts of causes irrelevant to the intended content of the item are shaping the response.

When we put together these observations with the earlier discussion of large attitude

ranges, we can begin to see the sense in which much of this debate seems chiefly semantic. I would scarcely feel that I was a traitor to my original argument if I were to rephrase it as follows, trying to adopt the Pierce-Rose vocabulary preferences: "The proportion of Americans whose attitude ranges are incredibly large on this set of issues, and whose manifest responses are strongly influenced by short-term forces with random properties independent of attitude content, is vastly greater than most political scientists have supposed, given the customary tiny per cent of Don't Know answers." Such a change of phrase does nothing, of course, to change the real-world evidence.

No sooner were we over this first small hurdle involving the nature of randomness than I discovered the authors were laboring under the delusion that random time paths (statistical independence, or a correlation of zero between trials) presupposed the equiprobability of states or choices at any given time point. In fact, they do not, as is readily apparent if one contemplates successive flips of a biased coin. Ironically, I had gone to some pains to *avoid* a model which would be tied to an assumption of equiprobability, because I was well aware that other known properties—most especially response set—stood in direct violation of such an assumption. The black-and-white model was useful in this regard, although Pierce and Rose in their several manuscripts have had much fun showing how badly the data meet the equiprobability assumption I succeeded in avoiding, and persist in implying that such demonstrations invalidate the test rather than underscore their own state of confusion.

The problem of response set is a serious one in the 1956 data, as discerning scholars have discovered. Moreover, it creates mischief for both models. I wrestled with the problem for some time before organizing the black-and-white model so that it could at least co-exist with the known presence of these tendencies. The Pierce-Rose model does not recognize the problem at all, and more than one of its key assumptions are thereby compromised.

One demonstration of the seriousness of the problem comes from our most direct experiments with it in 1956. At Richard Christie's suggestion, we included F-scale items of the standard kind in which an "agree" response typically means an "authoritarian" one. We also included some item reversals, such that an "agree" instead meant the nonauthoritarian response. Scores based on the two batteries of items not only failed to show any positive

correlation in content terms, but actually showed a significant *negative* one: people who ranked as relatively authoritarian on one scale ranked as relatively nonauthoritarian on the reversed version. Of course the tendency to agree or disagree to such items regardless of the direction in which the content was keyed thereby was significantly positive: response set actually outweighed the evaluation of content! With such data as these, it would be interesting to know what levels of "true and stable content-related attitudes" the Pierce-Rose model would demonstrate for each authoritarian item in the two batteries, had we repeated them over time. My guess would be that such estimates would be amusingly high, and directly belied by the other facts known as to internal inconsistency.

Space forces us to a very abbreviated discussion of the response set problem. There are in fact several kinds of response set, including a tendency to choose the "strong" alternative rather than less vehement ones. However, the dominant effect is a tendency toward acquiescence or agreement with "agree-disagree" items. One reason why such a response set thrusts itself forcefully on this debate is because it is probable that extraneous factors like question form intrude most sharply on responses where attitudes are least crystallized (or, in Pierce-Rose terms, have the widest true range, even if their model fails to permit such variation). This means in turn that the Pierce-Rose hope that "precision of attitudes is not related to the location of attitudes," necessary for the "convenient" assumption (iv), is probably quite wrong. Moreover, the Pierce-Rose assumptions (ii) and (iii) hinge on the proposition that error terms unrelated to meaningful attitude content are mutually compensating (mean of zero), which of course is exactly what is not true of data vulnerable to response set.

I did not attempt to represent my own thinking on the response set problem in the few brief pages I originally gave to the black-and-white test, so I have no desire to fault Pierce and Rose for their thorough-going misunderstanding of the subject, at least up to the point where I reviewed that thinking for them. I was saved from making obviously inappropriate assumptions by the fact that a model does exist which preserves the independence of successive trials without the assumption of equiprobability at any given trial. This model, in the familiar coin-flipping metaphor, is as I have suggested the one which admits coins of varying bias (p), with the unbiased coin (p of .50) as but

one special case. Thus an item subject to particular response set tendencies favoring an "agree" response, as many of our items show, could be seen as behaving for people without true attitudes like a coin biased toward the "agree" side by some degree (e.g., p of .60 or .70), but still showing a random path of responses over time.

I did not get into these depths in my article chiefly because of an editor already impatient at the article length. But there was another simpler reason. The power and housing issue changers whom I was considering as a successful isolation of respondents with random cross-time responses had initially split about 53–47 in their initial (or, equally, second-wave) choices between agree and disagree. For the N involved, this division could not be reliably distinguished from 50–50, or the simpler unbiased coin model. Therefore it seemed less than imperative to introduce the niceties of the broader model. If the ultimate testing which introduced the third-wave data had obliged me to commit myself as to just how biased the coin-flip was in the specific instance (i.e., how many more nonattitudes were hidden in the stable-agree cell than in the stable-disagree cell), I would have been in trouble. Happily, however, the correlational format of the central test, along with a focus on the crucial set of changers, relieved me of that problem. Given all of these circumstances, I decided that the space cost of introducing response set problems and the biased coin model outweighed the advantages. But contrary to Pierce and Rose, the equiprobability assumption is not any intrinsic assumption of the model.³

III

I mentioned at the outset that while the current third-round version of the Pierce-Rose

³ The authors' response to criticism of the vulnerability of their model to response set has been mainly made up of the addition of Appendix II, in which they set out to show that the black-and-white model does not perform well with items where sharper response sets than the power and housing item encounters are involved. The Appendix is largely irrelevant, since these other items are the ones where the black-and-white model should not be used in any event, not because of response set intrusion, but because of failure of these items to meet the "no change in true attitude" assumption necessary for the model. Moreover, a crucial proposition in that Appendix—to the effect that the standard form for the five issue items should mean a standard degree of response set—is well-known not to be true among scholars who have looked into these matters at all closely. In general, this Appendix is one of several reasons why I have had trouble building confidence that the authors really have much understanding either of the black-and-white tests or of the several deeper measurement issues involved.

test has purged itself of a number of the most glaring errors that plagued the first version submitted for publication, it remains rather dense in statements which are misleading or even more purely inaccurate. I think I have some responsibility to document this charge, so let me start with the paragraph on page 632 and proceed, as succinctly as possible, sentence by sentence.

"Converse's methods are an insufficient basis for accepting his thesis because they do not lead to an explanation of responses, they do not exclude alternative accounts, and they do not seem to be replicable."

Comment: Parts of the statement are true but misleading. My treatment did not pretend to "explain" responses in any rich sense of the verb, nor did it purport to exclude alternative accounts. The statement is misleading in context, however, as the authors have promised a superior model and the strong implication is that these are invidious comparisons with their own model. Yet where explanation is concerned the texture of their model is exactly the same as mine: a mapping of observed data which leaves a strong fraction of determinants to "short-term forces" with random properties. Where the exclusion of alternative accounts is concerned, no model can claim this under ordinary circumstances, and the Pierce-Rose version surely does not. Where replicability is concerned, the charge seems serious since the authors have already pointed out they are working the same data I used, but later in the paragraph we learn that they do not mean "replicability" in a strict sense, but rather than when the test is extended to other items with properties that do not match its assumptions, it gives nonsense results, as I said it would.

"For example, the 'Black-and-White' consistency test is used in exactly the same form by Philip Converse to demonstrate the *lack of attitudes* and by James Coleman to demonstrate the presence of *attitude and measurement error*."

Comment: This observation makes me despair, because it was I in an earlier review who urged the authors to relate their model to the ground already well-ploughed in this area, mainly by Lee Wiggins and James Coleman, although more recently by Leo Goodman. I am now sorry I asked, because it seems clear that the authors have skimmed Coleman without much understanding. The first problem in the statement is a factual error: the two approaches are in no sense "exactly the same form." The spirit is similar, but the form is dramatically different. Quite apart from this difficulty, the common spirit is a partitioning of data into true attitudes and "noise", and

there is no reason to believe that the two approaches would arrive at substantially different conclusions concerning the power and housing item, save that my model is better suited to smoke out population heterogeneity in these regards than is the Coleman model as commonly used.

"In Converse's approach, the main variable—attitudes—is operationalized as a residual category (what is left after other variables have been accounted for) . . . As Converse neither specifies his other variables nor justifies theoretically the treatment of attitudes as residues, alternative accounts . . . are just as acceptable. . ."

Comment: The simplest reason why I fail to specify "other variables" is that no other variables are involved, any more than they are in any given "running" of the Pierce-Rose model. The two models are entirely equivalent in these terms. Hence the "residue" argument is mainly mystifying.

"The attempts to replicate his test on other questions failed, and the operating criterion for the appropriateness of attempting replication appears to be: if the test shows nonattitudes, it should be used; if not, it is inappropriate."

Comment: The criterion, which the authors have never understood, has gemlike clarity. The model is to be used where its assumptions fit other facts known about the data configuration, and should be avoided where such fit is absent. One assumption is "no change in true attitudes." This is true to all intents and purposes for the power and housing issue, and clearly untrue for the other four issues (see Table 4). Where assumptions have changed from appropriate to inappropriate, the charge of "failure to replicate" borders on the absurd.

Proceeding sentence by sentence is too costly a business, so let us continue more efficiently, point by point. The authors go on to develop "four basic ways" in which my "method" fails to fit the data. The first point elaborated in this bill of particulars hinges on the ascription to me of an equiprobability assumption, and is mainly a throwback to the authors' earlier confusions as to what I was and was not assuming.

The second point is perplexing. The authors show a two-wave cross-tabulation of the power and housing item for the total sample, and claim that it violates my model in its "incremental" variations and evidence of an underlying response continuum. All of Table 1, including the incremental variation, was summarized in my original report as a positive correlation of about .3 for that issue in the total sample, and was one of several pieces of evidence which, when laid side by side,

prompted further thought and ultimately the critical test itself. The evidence in Table 1B seems to be another throwback to the time when the authors assumed I had collapsed "strong" and "weak" categories in making my critical test, an assumption which is simply wrong.

The third point is ingenious, but irrelevant in an important sense, as I have pointed out in earlier reviews. The authors complain that I overestimated the proportion of people with nonattitudes, or that I arrived at my estimates by some vague method of residues which assumed that every response which could possibly be labelled a nonattitude was so considered. Yet my set of identifiable potential nonattitudes numbered 144 cases on the power and housing issue, whereas the authors' purported efforts to examine the same cases for stability and relationship with other issues involves some 348 cases, or a multiplication of nonattitudes by a factor approaching $2\frac{1}{2}$! It is not hard to see what is happening here. The authors draw in a large set of people whom I did not consider as nonattitudes, and then use the traces of stability which these additional people produce (up to a correlation of .17, rather than my original near-zero), by way of showing that my diagnosis fails to fit the data (Table 2). Presumably the same abuses are occurring with respect to Table 3, although in this instance they do not even show the data for the one item appropriate to the black-and-white model, so the display is irrelevant to the discussion in a double sense.

The fourth point suffers many of the same difficulties. Again we are dealing with five issue items, four of which are irrelevant, so that "deductions" and "expectations" purportedly based on my reasoning are necessarily going to prove wrong. In addition, we are dealing with a set of "nonattitude respondents" larger by a factor of two or three even on the single relevant item than any I ever employed, and I presume the same misdefinition is ardently pursued into the four irrelevant questions. Therefore we have already adulterated the effects with which I was working by a factor of about ten ($2\frac{1}{2}$ times 4) to one. And, as though this were not enough, we shall make the crucial statistics depend on the "predominance of 'agree' responses," a tactic which enables the "agree" response set on the four irrelevant items to lend a further compound of negative evidence. At the end, we can talk of "not one chance in a thousand" that an assumption which I cannot recognize in the form stated, is correct. This is surely damning.

I should make clear that the oddities which persist in other parts of the text are not as densely packed as in the few pages I have just discussed. But quite a number remain and we cannot hope to discuss them all. One other matter involving the assumptions of the Pierce-Rose model is the need to attach a scoring system to the response categories in order for the model to operate. The prime demonstrations justifying the particular scoring adopted surround Table 5 and seem enormously impressive. However, the high correlations reported in that table are mainly due to the fact that these issue responses tend to be "lumpy"—certain categories drawing large proportions of responses, and others, small proportions—and these lumps reproduce themselves faithfully from wave to wave. In general, turnover tables will show "category stability" as the authors measure it to be very high for popular categories and much lower for unpopular ones. The standard way of removing this artifact in order to ask the kind of question the authors want to pose is to measure category stability in terms of departures from expectations of the chi-square type. The authors leave the size artifacts in their category stability measure, however, and since their scoring system is mainly a function of category size, seemingly impressive correlations result, although they approach the tautological rather than being any serious test.

The operation of scoring the categories is not as unsettlingly inappropriate to the data as some of the other key model assumptions, and I find such experiments of interest. The scoring does rest on the assumption that underlying true attitudes are always normally distributed, an assumption that flies in the face of many observations of attitude dimensions that polarize in response to relevant social crises. However, such polarizations are not obvious in the data from these five issue items in the 1956–1960 period, and hence I would rate the scoring used as no more than a small strain on other things known about these data.

What is puzzling is that the authors have insisted even through the current version of the manuscript that my black-and-white test requires the same operation of "pinning numbers on categories." Of course it does not, and since the difference has been explained and the authors persist in their contention, I am obliged to conclude that their grasp of "levels of measurement" distinctions is also remarkably weak. In the critical test I used ordinal correlation measures which fit very nicely the uncontested ordinal properties of the attitude categories involved. And the only other part of the opera-

tion involved the initial dichotomization, which requires no manipulation of numbers whatever. The authors will find that my results are invariant whatever cardinal numbers they might wish to assign to the two groups, be they zero and one, one and zero, 347 and -40, or any other conceivable pair of real numbers. Once again, every step of my test was carefully geared to all of the known properties and deficiencies of the data being employed.

IV

Near the close of their discussion, Pierce and Rose draw the following summary comparison between my black-and-white test and their "probability model."

"The probability model possesses more general advantages. It is applicable, with or without modification, to a wide range of questions, as indicated here by the results for the five domestic policy questions. . ."

We should also like to close with a few comments on this proclamation of broad applicability, which we find quite surprising.

At base, I do not understand what "applicability" means to the authors. It is obvious that most models of the type under discussion can accept a batch of empirical data as input and, with a turning of the crank, give some kind of numerical results back. Surely "applicability" means more than this. Some few models—the Coleman is a case in point—give handy warning signals as to their inapplicability in special cases by generating output numbers that are obvious nonsense, such as probabilities that are negative or greater than one. Other models, like the two being compared here, are unlikely to make their own nonsense so obvious, yet both if misapplied can provide results that are nonsense nonetheless. Indeed, it is this kind of hidden nonsense that led me to stress in the original article that the black-and-white model was inappropriate to all but the power and housing item.

To my way of thought, a model becomes "applicable" to a given set of data once it can be shown that (1) its assumptions are tailored to known properties of the data; and (2) there is additional evidence of "goodness of fit" in predictions to further relevant bodies of data. Unfortunately, this is not a situation where half a loaf is better than none: both types of fit must co-occur if a claim of "applicability" is to be certified in any special case.

I have thus far said nothing about the second criterion, or the Pierce-Rose demonstrations of "goodness of fit" for their model. The authors are pleased with these data (Tables 6,

7 and 8), and it is safe to guess that most casual readers will find them quite dazzling. In Table 6, for example, about the poorest prediction is the one associated with the 1958–1960 transition for the power and housing issue. Yet even here we see a value of 90 per cent of "respondents correctly placed," and that seems to leave remarkably few respondents *misplaced*.

I do not find these demonstrations at all impressive, and it is important to convey why. What most readers will fail to understand about a datum like the 90 per cent is that (a) it involves a type of aggregative prediction that does not purport to assign individual respondents to any particular time path of responses; and (b) these items are remarkably inert in their aggregative contours over time. When these two facts are combined, it means that seemingly "bad" predictions would be almost impossible to generate in this context.

Where the former is concerned, the authors are merely trying to predict the aggregate shape of responses for a certain time period, given knowledge of their shape at earlier periods. This is perfectly legitimate in itself, although readers not close to these problems would be astonished to know how numerically inflated such aggregate predictions will necessarily be, relative to their individual-level counterparts. Although the Pierce-Rose text is very clear as to the aggregate nature of these demonstrations, the Table 6 heading "respondents correctly placed" (a phrase usually reserved for exercises like discriminant function analysis where individual assignments are actually made), serves to blur the issue.

The second point, which involves the difficulty of getting poorer absolute results from this kind of operation with data as aggregatively inert as these turn out to be, may be made vivid with a simple experiment. I have taken the 1956–1958 cross-tabulations on the power and housing issue in the same five-by-five form erected by the authors for the purposes of Table 6, and used these twenty-five percentages to "predict" to the aggregate contours of the parallel 1958–1960 version of the table for the same issue. Instead of the 90 per cent predictive success cited for the 1958–1960 test of power and housing in Table 6 when the full Pierce-Rose model is brought to bear, I find 91½ per cent of "respondents correctly placed" using, as it were, no model at all!!

This is a far cry from the closeness of predictive fit achieved by the black-and-white test for a much more surprising prediction. Hence we conclude that it is charitable to say that the Pierce-Rose demonstrations of "goodness of fit"

are indifferent or mediocre.⁴ However, as always with such tests, there is a zone of ambiguity as to how "good" any goodness of fit must be to legitimate a claim of applicability. Therefore let us return to the first criterion—the appropriateness of initial assumptions—where there seems to be no ambiguity at all.

It has been almost impossible to convey to the authors that the black-and-white model was not suggested as a general analytic routine for panel data in the sense the authors claim for their own model. The current text continues to misuse it beyond its proper limits for purposes of derision, all the while developing the new theme that it must be inferior because it is not more broadly applicable. It was not suggested as a method, but rather was developed as a finely-honed tool to answer a very specific question that could be asked of a particular cross-time data structure displaying certain peculiar properties. Since I anticipated intuitively surprising results, I labored very hard to stay within a set of assumptions that violated no known properties of the issue item involved. Only one decision necessary for the operational implementation of the model (*not* a model assumption as such) was even particularly arbitrary, and that involved an *a priori* commitment that changes of strength of agreement or disagreement on the power and housing issue did not have the same significance, at least for people with true attitudes on that polarized elite issue, as a full change from *pro* to *con*. All other assumptions and procedures were unquestionably well tailored to the known properties of the data.

This cannot be said of the Pierce-Rose model. We have seen that in numerous ways the assumptions of their model strain the data, and at a few palpable points stand in the most direct violation of known properties of the data, including, where the power issue is concerned, the heterogeneity property demonstrated by the careful black-and-white test. Actually, in such component facts as respect for the ordinal character of the data, the black-and-white model is *better* suited to the texture of *all* of these issue items than is the Pierce-

Rose model. The only reason why I took pains to limit its use to the power and housing item alone was because the model assumed no true attitude change, an assumption clearly violated by the other items. What is perplexing is that the Pierce-Rose model makes exactly the same assumption, yet this does not deter the authors from proclaiming their model to be more broadly applicable! Clearly the authors and I are operating on dramatically different wave lengths.

None of the above should be taken to mean that I consider the Pierce-Rose model as useless. I think it is worth experimenting with more powerful measurement assumptions, and I suspect that some of the differential statistics that the Pierce-Rose model may generate from item to item are worth heuristic attention. On the other hand, where its results appear to collide in significant degree with the conclusions from the black-and-white test on the power and housing issue, I do not have much trouble knowing which outcomes I would prefer to trust.

The broader point of the Pierce-Rose essay is to "rehabilitate the voter." If the authors truly believed me to be saying that only 20 or 30 per cent of the American populace has real attitudes on any political issue, as opposed to the limiting case of power and housing in 1956, then the voter would indeed be in direst need of such rehabilitation. But this was not the point even for the 1956 materials, and a broad range of data from the later 1960's and early 1970's seem to imply that attention to politics and the crystallization of attitudes have advanced markedly in recent years. Thus it seems likely that the power and housing item was not only a limiting case in the 1956–1960 period, but probably holds some kind of record over a broader period as well. This is not to say that my original interpretation of the 1956–1960 issue responses was in any way defective. Nor is it to say that survey research analysts are not well advised, even in the current period, to keep in mind the likelihood that some fraction of apparent content responses are not very meaningful. However, the range and central tendency of nonattitude prevalence on such items has probably receded significantly. In short, the voter has been rehabilitating himself, and this is much the better way.

⁴ Given what we have said about the "lumpiness" and cross-time inertia of these items, the reader should have no trouble seeing why the other tests of "goodness of fit" (Tables 7 and 8) are no more impressive than Table 6.