

The Catastrophe of Scientism in Social/Behavioral Science*

Thomas Scheff

Abstract: An ongoing protest about the misuse of statistics and absence of effect sizes may be only a part of a much larger problem. A vast fraud is implied by the large number of reviews of studies of causation that do report effect sizes (such as correlations). In many fields the effect sizes appear to have been near zero for many years, and are not increasing. The main reason seems to be the use of vernacular words rather than clearly defined concepts, a practice that leads to invalid measurement. In attitude research, however, the most common reaction is to blame the subjects for the “inconsistency” between their attitude and their behavior, instead of the invalidity of the measurement instruments. Yet the same studies, because of the prestige of science, continue. It appears the scientism, rather than science, prevails in causal studies. Suggestions are made about procedures that might be more realistic and effective. (2457 words)

Cumming et al (2007), Ziliak and McCloskey (2008), and others have argued that systematic studies in economics, medicine, psychology and social science often misuse statistical tests of significance. The complaint is that these studies report no effect sizes (such as correlation coefficients). In reporting only the likelihood that the results are due to sampling error, they leave out all of the more relevant indicators of actual significance, such as effect size. With only tests of significance, the results are meaningless. This fact has been known but largely ignored for many years. (For sociology, see Selvin (1957) and the reader by Morrison and Henkel (1970; 2006).

Given the seriousness of this problem, it would be important to have it discussed in statistics textbooks. I have looked at four current ones. Weinstein’s elementary text for sociology (2010) and Cohen’s (2008) for psychology ignore it. The more advanced texts by Everitt (2010, 16-19) and Pearson (2010, 196-198) have brief but moderately strong warnings against too much emphasis on tests of significance. None of the four, however, cite Cummings et al or Ziliak and McCloskey, or the 53 year history of protests.

There have been several reviews of the Ziliak/McCloskey book in the social/behavioral sciences. Of the reviews, several have been appreciative, but the rest have been antagonistic, defensive, or neutral. The review in this journal (CS, 2009) is representative of the latter stance. It is judicious, but only two paragraphs long, not a review but only a Briefly Noted. More than words, the brevity implies a minor matter. It is possible, however, that the matter may be major.

In this regard, consider the difficulties encountered by editors of journals in trying to stop the practice of the “sizeless stare” (Cumming et al 2007; Ziliak and McCloskey 2008). The determination of researchers to continue to misuse statistics may point to a broader problem, the abuse of the name of science as a whole in social/behavioral research. (For commentary on the prestige of pseudoscience, see Porter 1995, Agger 2007, and, less directly, Martin and Lynch 2009.)

Reviews of effect sizes in one field, self-esteem studies (some twenty thousand employing standardized scales), provide an example (Scheff and Fearon, 2004). Between 1976 and 2004 there were four reviews of the entire field, and 24 reviews of subfields, a total of 28 assessments. **All of them** found near zero effect sizes for predicting behavior. The lack of any increase in those 28 years also suggests zero progress in these studies.

*I am indebted to David Nasatir for his comments on an earlier draft.

The Baumeister et al review (2003) examined associations between self-esteem scales and behavior in great detail.

...most of the effects are weak to modest. Self-esteem is thus not a major predictor or cause of almost anything (behavioral). Moreover, the effects of self-esteem become weaker as the criteria for evidence become more objective. (Baumeister et al 2003, p. 42).

This is a compelling argument not only because of broad coverage and detail, but also because the first author, Roy Baumeister, is a leading psychologist. The slight difference between their conclusions and mine may be only semantic. Baumeister et al summarize the results as showing only "modest, small, and null correlations". None of the studies that they judge to show "modest" correlation report correlations over .33. Since these correlations translate to less than 10% of the variance, I would call them small, rather than modest: 90% of the variance is left unaccounted for.

Although the number of self-esteem studies has been steady for many years, a ray of hope appeared in 2009, when the number of titles began to decrease, if only slightly. If this trend grows stronger, it means less time and effort wasted on useless studies, and perhaps more exploration of alternatives to pseudoscience, to be discussed below..

In our survey (Scheff and Fearon 2004) we reported that several other fields of study also had near zero results, and were not improving. Self-esteem is just one kind of attitude study, a vast and ever increasing field. Some forty years ago, a review of the field showing near zero effect sizes (Wicker 1969) caused an uproar. Has there been any change since then? More recent reviews claim that what is called the problem of consistency between attitude and behavior is being solved (Albarracin et al 2005), but the arguments seem specious.

The very term, inconsistency, assumes that subjects have trouble being consistent, rather than researchers have trouble with invalid measurements (such as self-esteem and all other scales). I know of only one field with substantial behavioral prediction, Asch-like conformity studies, but these are seldom pursued.

A second question involves the increasing number of meta-analyses, which show positive results in fields of study with weak findings. Most of these studies may also be meaningless, because they are based only on positive (published) results, rather than the entire population of studies. A report by Kirsch et al (2002) used both published and unpublished studies furnished by the funding agency (FDA). Although meta-analyses that used only published results showed antidepressants slightly more effective than placebos, the Kirsch et al study called that result in question.

Discussion

The clearest meaning of continuing small and nil effect sizes is that whatever the main causes are, these studies have not found them, and are moving no closer to them. For example, there have been several replications of Durkheim's classic study of suicide rates in different cultures. Like the original study, the replications report an effect, but it is quite small, accounting for less than ten percent of the variance. The major causes of suicide are still unknown. Since Durkheim's study is one of the main bases of modern sociology, the foundation is shaky.

Assuming, for the sake of discussion, that this argument has some validity, what changes might it suggest for research? Perhaps one change in causal studies could be less quantitative research. It doesn't do any good to be scientific if it is going nowhere.

One key problem with existing studies, whether qualitative or quantitative, is that their key variables go undefined. In all the vast armada of quantitative studies of self-esteem, there is not a single attempt to state a conceptual definition of self-esteem: all 200 scales in use are only operational definitions. Even a glance at a dictionary suggests that the word has many meanings,

some of them potentially in conflict. We need clearly defined concepts before we can test causal hypotheses.

It might be useful for the study of self-esteem, for example, to develop two different scales, one for self-evaluation (cognitive), the other for emotions, self-feeling. All of the existing scales confound these two dimensions. Suppose genuine pride vs. shame constitutes the main emotional axis of self-feeling. If many high self-evaluations are defenses against shame, the findings that have been named the dark side of self-esteem (egotism) would be understandable. It also would help explain the low effect sizes, since the two variables might cancel each other out.

There is no evidence of such a move in the existing literature, however. For example, the editor of a recent (2010) collection makes a gesture in that direction when she attempted to define self-esteem in terms of affect: “Self-esteem is the attitudinal, evaluative component of the self, the affective judgments placed on the self concept.... (Guindon, 2010, p. 12, emphasis added). However, neither she nor any of the other chapters take up this idea by spelling which emotions might be considered, and how they might be involved.

However, I do not mean to imply that all quantitative studies of causation have no potential. For example, the experiments conducted by Berkowitz (1962) and others showing that persons who vent anger not only do not feel better, but often feel worse, could be of great importance. The assumption that venting gets it out of your system is practically universal in our society, and for that reason, a threat to survival. Perhaps if future reports included ways of managing anger other than venting, they might get the public attention they deserve.

The aggressive anger studies also have another problem. They all equate venting, acting out anger, with catharsis, the resolution of unresolved emotions. The traditional idea that drama helps the audience feel and resolve their hidden emotions is potentially of great merit (Scheff 1979; The absence of cathartic potential in most Top40 popular songs is described in Scheff 2010). By confounding emotion and behavior the way the public does, the anger experiments are upholding what often turns out to be an excuse for violence: “My anger made me do it!” Although the anger studies could be useful, like the rest of social/behavioral science, they need to be based on a clearly defined concept of anger as a feeling rather than the common misconception that is both feeling and behavior.

Conclusion

There is by now a literature on methods for developing concepts (for example, studies using grounded theory, Charmaz and Bryant 2008; also see Scheff 1997 on part/whole methods). Until clearly defined concepts are devised, we may need more studies of the details of the human condition, the kind provided by studies that are at least partly or entirely descriptive, such as ethnographies. The poet William Blake wrote, “Art and science cannot exist but in minutely organized particulars.”

Human conduct is incredibly complex: most conversations, the minute details, are DNA's at the very least. Without genuine concepts, and preliminary links to minute particulars, we are still in the flat earth stage of research. For causal studies, the particular details should not be ends in themselves, as they sometimes are in linguistics and history, for example, but preliminary tests of the goodness of fit with speculative concepts or hypotheses.

For example, the terms solidarity and alienation are central ideas in sociology, but usually go undefined. Seeman (1975) pointed out that the idea of alienation involves many dimensions that are

confounded in research and in discussion. Schacht (1994) found eight different meanings in the literature.

Can basic ideas be defined? Elsewhere I attempted conceptual and operational definitions of the solidarity/alienation dimension that propose three steps of recursion in mutual understanding (I know that you know that I know... Scheff 1967, Corballis 2007, Scheff forthcoming) This model generates three degrees of partial understanding/misunderstanding that fall between complete solidarity and complete alienation. These three degrees seem to correspond to pluralistic ignorance, false consensus, and aesthetic distance in three separate literatures. It remains to be seen whether this or other precise definitions will ever come into use.

While we are developing clearly defined concepts, what else might be needed in social/behavioral science? Following the ancient adage "Know thyself," one helpful step might be to learn more about ourselves as individuals. It may be that we can reach our main goal, understanding others, only to the extent that we understand ourselves.

References

- _____. 2009. Review of Ziliac and MacCloskey. Contemporary Sociology, 38, 5, p. 496.
- Agger, Ben (2007). Public Sociology: From Social Facts to Literary Acts (2nd edition). Lanham, MD: Rowman & Littlefield.
- Albarracin, Dolores, Blair Johnson, and Mark Zanna. 2005. The Handbook of Attitudes. Mahwah, N. J.: Lawrence Erlbaum Associates.
- Baumeister, Roy F. Jennifer D. Campbell, Joachim I. Krueger, and Kathleen D. Vohs. 2003. Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles? Psychological Science in the Public Interest, 4, 1, 1 – 44.
- Berkowitz, Leonard. 1962. Aggression: A Social Psychological Analysis. New York: McGraw-Hill
- Charmaz, Kathy and Anthony Bryant. 2008. The SAGE Handbook of Grounded Theory. Newbury Park, Ca: Sage Publications
- Cohen, Barry. 2008. Explaining Psychological Statistics. (Third edition) Hoboken, N.J.: John Wiley.
- Corballis, Michael. 2007. The uniqueness of human recursive thinking. American Scientist 95, May-June, 240-248.
- Cumming, Geoff, Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., Lo, J., McMenamin, N., & Wilson, S. (2007). Statistical reform in psychology: Is anything changing? Psychological Science, 18 s (3), 230-232.
- Everitt, Brian. 2010. Multivariable modeling and multivariate analysis for the behavioral sciences Boca Raton, Fl.: CRC Press
- Guindon, Mary H. 2010. What is Self-Esteem? Pp.3-24, in Self-Esteem Across the Lifespan. Edited by Mary H. Guindon. New York: Routledge.
- Kirsch, Irving, Thomas Moore, Alan Scoboria and Sarah Nicholls. 2002. The Emperor's New Drugs: An Analysis of Antidepressant Medication Data Submitted to the U.S. Food and Drug Administration. Prevention & Treatment, 5, 1-11

- Porter, Theodore M. 1995. Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton: Princeton University Press.
- Martin, Aryn, and M. Lynch. 2009. Counting Things and People: The Practices and Politics of Counting. Social Problems 56, 2, 243 -266
- Morrison, Denton and Ramon Henkel. 1970. The Significance Test Controversy: A Reader. Chicago: Aldine. Republished in 2006. AldineTransaction
- Pearson, Robert. 2010. Statistical Persuasion. Los Angeles: Sage.
- Schacht, Richard. 1994. The Future of Alienation. Champaign, Ill.: U. of Illinois Press.
- Scheff, Thomas. 1967. Toward a Sociological Model of Consensus. American Sociological Review, 32 (February) pp. 32-46.
1979. Catharsis in Healing, Ritual and Drama. Berkeley: University of California Press (Reissued by iUniverse 2001)
1997. Emotions, the Social Bond, and Human Reality: Part/Whole Analysis. Cambridge: Cambridge University Press.
2010. What's Love Got to Do with It?: The Emotional World of Popular Songs. Denver: Paradigm Publishers.
- Forthcoming: Social-Emotional World: Mapping a Continent. Current Sociology
- Scheff, Thomas and David Fearon. 2004. Social and Emotional Components in Self-Esteem. Journal of the Theory of Social Behavior. 34: 73-90
- Seeman, Melvin. 1975. Alienation Studies. Annual Review of Sociology 1: 91-124.
- Selvin, Hanan. 1957. A Critique of Tests of Significance in Survey Research. American Sociological Review, 22, (Oct), 519-527
- Weinstein, Jay. 2010. Applying social statistics: an introduction to quantitative reasoning in sociology. Lanham, Md.: Rowman & Littlefield
- Wicker, A.W. 1969. Attitudes vs. actions. Journal of Social Issues. 25, 41-78.
- Ziliak, Stephen and Deidre McCloskey. 2008. The Cult of Statistical Significance. Ann Arbor: University of Michigan Press.