



RESEARCH DESIGN TO ESTIMATE OFFENSIVENESS OF
PORNOGRAPHIC MATERIAL IN A COMMUNITY

by

Charles H. Proctor

Institute of Statistics Mimeograph Series No. 1987

November 1990

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

MIMEO SERIES # 1987 NOV 1990
RESEARCH DESIGN TO ESTIMATE
OFFENSIVENESS OF PORNOGRAPHIC
MATERIAL IN A COMMUNITY

NAME	DATE
------	------

Department of Statistics Library

Research Design to Estimate Offensiveness of
Pornographic Material in a Community

C. H. Proctor, NCSU, November 1990

1. Background

We were requested to review critically an application of social science methodology in a court case in North Carolina. Researchers had conducted studies using survey interviews in the Charlotte area to see whether certain pornographic material had exceeded community standards of patently offensive. We found that these studies were informative about community viewpoints generally, but seemed to fall short of establishing a standard or reference scale of measurement. Our critique consisted mainly of an outline for the conduct of what we would consider an adequate procedure to measure whether a given pornographic material exceeds community standards for being patently offensive, and we propose here to review this critique.

The legal issue was whether to admit into evidence results of social science investigations of community views on pornography. The jury had the ultimate responsibility for the decision in this case, but the question was whether the social sciences should contribute to the data the jury used. We contended that if the data is of high enough quality it should enter, but otherwise not. More specifically, if the data go beyond description toward measurement then they should be allowed.

Measurement of even such a straightforward characteristic as weight is subject to errors. Use of a bathroom scale may not yield data of sufficient accuracy to be admitted as grounds for accusing the grocery store of overcharging and obviously would not help in detecting, say, traces of lead in

lettuce. On the other hand one can carry out empirical tests of the bathroom scale's performance and may be able to show it to be adequate for many even legal purposes such as providing evidence that a person was, or was not, carrying an infant.

2. Legal Foundations

The opportunity for a social science contribution to this court case was opened by language in North Carolina General Statutes 14-190.1, the North Carolina obscenity law. Material there is defined as obscene if it "... depicts or describes (sexual conduct) in a patently offensive way ...," and also satisfies three other criteria. The law further states "Obscenity shall be judged with reference to ordinary adults" One of the other three criteria is expressed as: "The average person applying contemporary community standards relating to the depiction or description of sexual matters would find that the material taken as a whole appeals to the prurient interest in sex" Just to complete the picture let's note that the third criterion was that "The material lacks serious ... value ..." and the last is "The material is not protected or privileged"

From discussions with lawyers it seems that both sides agreed that local community standards are relevant. That is, one community may be more or less accepting than another, and these differences in viewpoints should enter into the determination in accord with the language of "ordinary adults" and "contemporary community standards." The crucial issue then becomes whether some given material goes beyond patently offensive or falls short of patently offensive or perhaps is not even offensive.

3. Theoretical Schema

What is required in order to answer this or any measurement question is first a conceptual schema for defining the terms and then a description of procedures for applying the definitions. If it were possible to demonstrate that most people were able to agree on the assignment of degrees of offensiveness to standard selected items of pornographic material then it may be possible to determine for each person a "patently" point on this scale such that material more offensive than this point would be unacceptable for display to the person. Both the extent of agreement among persons on the ranking of items and the level of consistency for any one person of the patently point are subject to empirical study. Scientifically agreed upon standards of reliability can be applied here.

The second part of the conceptual scheme concerns the possibility of locating the pornographic materials, at issue in the trial, on the same scale of offensiveness. That is, it must be shown that judgements of community members consistently place the items at certain positions along the scale. The final step involves combining the individual patently points of the members of the community into one overall community standard patently point, and then comparing the positions of the items under trial to this point.

We would propose that, once a measurement scale of offensiveness had been established in focus groups (see below) and a distribution of community members' patently points had been estimated over this scale by a survey, then the 75th percentile of this distribution be taken as the community standard. Possible alternative choices, of a median or of the 90% percentile, could be

argued. The final choice should be stipulated by the parties before data are collected, but the device for conversion of a survey result to the standard would seem to be both necessary and obvious.

The grounds for this proposal come from the disciplines of anthropology, sociology, and social psychology. In these fields one finds definitions and discussions of such concepts as culture, cultural standards, social structure, collective representations, attitudes and so on, that bear on the issue of how the responses of survey participants can be translated into community norms.

Community standards in the present case are the more or less explicitly stated values that members of the community expect to have applied in deciding what materials can and cannot be sold and rented at adult bookstores. As such these rules are a part of community "cultural structure," as one sociologist puts it (Wallace (1988) p. 46 ff.).

The rationale behind such rules would seem to be the general expectation that products offered for public sale or rent should not be harmful or dangerous. Since the viewing of certain scenes could shock and upset certain viewers, it would be wise to remove the material and thus remove the risk. However, the plausibility of this rationale is not at issue, only the existence and consistency of the community standard is.

There are, as one might imagine, differing viewpoints within the social sciences on what evidence is to be used to discover community standards. From social psychology comes the, now outmoded, concept of "group mind" [Allport (1985)]. The more current proposal is to use methods of attitude measurement to establish consensus among the members of the community, [Panel (1989) vol. I, p. 10 ff.].

From anthropology comes a recognition of the importance of culture and of subcultures. That is, there may be distinct groups of persons in the community holding differing sets of standards. Additionally there is quite likely to be spokespersons within these groups whose voices are deemed authoritative so that not everyone's attitude is equally important in setting standards. As an anthropologist has stated "... consensus ... is not the only evidence we should adduce. We should consider also the degree of authoritativeness with which the respective assertions are made." [Nadel (1957), p. 48]

In sociology one finds both of these viewpoints plus discussions concerning procedures for making the measurement [Lundberg (1942)]. As a first step one should recognize that the unit of analysis here is the community as a whole. Thus, in order to judge the bias and precision, or reliability, of say alternative methods of measurement there would need to be repeated determinations made for a variety of communities. The agreement within one method for one community would represent basic measurement reliability. Systematic differences on average between methods would reflect bias, but differences among different communities would be encouraging insofar as they coincided with well recognized differences among the communities.

Before entering into the rather technical detail of measurement protocols a further few words here on the general philosophy of measurement of unusual material in the social ("soft") sciences may be in order. One should never lose sight of the presence of a reality which the method must track. That is, when the reality changes then the method should give a new result. There are in this case two such realities. One is the beliefs and expectations of members of the community and another is the actual concrete pornographic materials

entered into evidence at the trial. Any proposed method must be shown to be capable of yielding a different result if either of these is altered.

In an earlier trial, whose transcript we were shown by the lawyers, an attack was made on the use of the survey method because it did not include the actual materials under trial but used general verbal descriptions of such material. Such an attack has merit in our view, and the following proposal will be seen to include the specific materials.

A shortcoming of the following proposal, however, is its unproved ability to track variability, in the stringency of community standards, among communities or through time. This has already been mentioned and its solution may have to await further institutionalization of social science research centers. Such centers can at present be found at or near universities or in government and even to some extent within the private sector. When a method of measurement can be developed to a point where alternative centers obtain reasonably similar results one can then more confidently admit such evidence into court cases.

4. Field Methods

The research methods that were proposed to be applied consist of two approaches. One is the familiar method of personal interview survey. The other method has been termed that of focus groups, whereby persons of the community are invited to meet together and discuss a topic of interest -- in this case the offensiveness of pornographic material. Such groups are routinely used, for example, in market research. When the same group meets over time they have been termed panels, as in food science research.

Although most of the standardization operations could conceivably be done by persons acting in isolation, the provision of interaction, as in this group setting, would seem to be encouraged by the language in the code stating "community standards," and by the practices of field work in anthropology as were already mentioned.

4.1 Sample Selection. There are at least two steps in this procedure where samples need to be drawn and it may help to discuss these together. One is the selection of a sample of pornographic materials from all such materials to span and to anchor a standard scale of offensiveness. One should here proceed deliberately in searching for a wide range of cases.

The other is the sample of community members used to locate patently points. In this case it may be preferable to use something close to random selection from the adult population of the community. In both cases, however, the results would not be affected by moderate departures either from completeness of the initial collection of pornographic materials or from exactly equal probability sampling of the community.

For example, there will be nonresponse caused by respondents who object to surveys or to the topic of investigation, although the sponsorship of the research by the courts may be able to avoid much of this. Membership in the focus groups that are charged to anchor the extremes of offensiveness may be more self selective. Food panels and other such groups also tend to become self selective over time.

The two major candidates for frames to draw the samples for both the focus groups and the general population are (1) area segments and (2) telephone

numbers. Area frames are still used for nationwide enumerative surveys of unemployment rates, health status and farming, but many major surveys have recently begun to use random digit dialing frames. As far as the general problem of under coverage is concerned, either method would be adequate for present purposes. That is, the sampled population reached by telephone numbers may be a little less than that covered by the area frame, but the difference would be relatively unimportant for the purpose of discovering community standards.

A major advantage of the area frame may be a lower rate of refusals and particularly of item nonresponse. This is conjecture, but I suspect that the ten, or so, standard stimuli will, in the end, have to be a mixture of verbal and pictorial presentations. Thus a telephone interview will not be able to cover the topic. For reasons of measurement therefore the area frame which permits personal interviewing seems almost dictated.

In this case one would use a design having, say, 10 households per segment with geographic stratification into paper zones having, say, 5 segments drawn from each paper zones as 5 replicated subsamples. One of the subsamples could even be used as the basis for invitations to join the focus groups.

4.2 Measurement Methods. The major task in this research is developing the measurement protocols: (1) the one for the focus groups to establish the standard offensiveness scale and to locate the materials on trial on this scale, and (2) the personal interview to establish a patently point for each respondent. Members of the focus groups will examine presentations. A cartoon may serve for pictorial material; a paragraph or sentence, or even a word, may

work as verbal material, while a one-minute scene may work for films. The complexity of the stimuli should be kept within bounds so the judgement of offensiveness need not involve any complicated weighting.

The instructions to these groups also need to be simple. They could read: "You are to consider that you have purchased this material for the purpose of entertainment and are viewing it in your home by yourself or with whatever companions (spouse, roommate, etc.) would ordinarily be there." One would pretest all such arrangements by comparative trials to find which gave most reliable (test-retest) results. The task itself would be to locate the presentations on a standard scale of offensiveness. This scale must also be established by pretest. One should describe or show cartoons of, say 50 instances or kinds of sexual conduct ranging from Marilyn Monroe in a blowing dress to child pornography involving mutilated genitalia. These would be shown to a test panel for ranking. About 10 such presentations would be selected to serve as well-spaced anchor points on an offensiveness scale. Focus groups could then be used to verify the well-ordering of these standard presentations.

Focus groups would also be used to locate the test materials on this standard scale. It will be possible to verify interrater agreement among members of the focus groups and to see with what accuracy it is possible to measure offensiveness of the materials under trial. One would hope that relatively few respondents, less than 100 in total, could give such estimates which would then have a sampling coefficient of variation of around 5%.

Let's now consider the survey interviews. In judging the degree of offensiveness that a survey respondent can tolerate or accept and the amount that is judged patently offensive, it is particularly necessary to establish

the situation of the respondent clearly. That is, if respondents adopt a clinical attitude of collaborators in research one may find their patently point is high, while if they are responding to what they are asked to imagine they just discovered their teenaged daughter to be viewing, their patently point may be low. The interview protocol should be carefully pretested. The end result for each respondent will again be a value on the same standard offensiveness scale. In the focus groups one can validate that the patently points of respondents obtained on the standard scale do or do not agree with their findings of patently offensive and of tolerated on the test materials.

If such agreement between direct and indirect determination of offensiveness of the test material is established then one can justifiably do a survey of respondents using an indirect determination based on just the, say, ten standard stimuli. It should not be necessary to present all ten stimuli in order to find out a respondent's patently point. Perhaps presenting them more or less in order of offensiveness and stopping when the patently point was passed would be sufficient. This procedure should, of course, be pretested.

5. Data Analysis

Let's say one begins with 50 presentations as candidates for the 10 standard offensiveness stimuli. By preliminary judgement these could be sorted into 5 roughly ordered groups of 10 each. Then, say, 20 sets of 5 ordered presentations could be constructed by random selection in which each presentation would appear twice. Each member of the focus groups would be asked to assign ranks from 1 to 5 to each presentation in the 20 sets.

Suppose there were 15 persons who did all rankings. Thus any one stimulus would be given 30 ranks, 2 from each of 15 persons. The 2 sets of 15 assigned ranks will have different sample means (say \bar{X}_{i1} and \bar{X}_{i2}) as well as sample variances (say, s_{i1}^2 and s_{i2}^2), for the i th presentation, and these can be averaged to get \bar{X}_i and s_i^2 . Roughly speaking, the better stimuli will have a small value of s_i^2 . One would expect the presentations with very high and very low \bar{X}_i to have small s_i^2 values. These stimuli will anchor the ends of the scale, and one will then select the other stimuli on the basis of relatively small s_i^2 values with some attention paid to the consistency of the s_{i1}^2 and s_{i2}^2 values. If pictorial stimuli have been mixed with verbal presentations then the final set should also be a mixture of these types.

Next, the 10 standard stimuli will be used as the anchors of a 10-point scale, and each of the 15 members of a focus group will be asked to locate on this scale a point for each of the materials included in the trial. Each material will then be scored by, say, 15 integers. For the j th material one can compute \bar{X}_j and s_j^2 as before. The value \bar{X}_j estimates how offensive is the material and $\sqrt{s_j^2}/15$ is a rough standard error for this estimate, and combination of estimates over focus groups will further allow for judging uncertainty from group to group.

By now the members of the focus groups will be feeling rather over worked, but it would be helpful if, as their last task, they would pretest the survey questionnaire. This questionnaire will present each of the 10 standard stimuli, roughly in order of offensiveness, and ask if it would be tolerated or would not be accepted in material to be sold in the community. When the respondent has answered "No" to two stimuli the interview can be ended.

The final data will be the determination of patently points for the survey sample (members of the other four replicated subsamples in the earlier example). The estimated quantiles of this distribution would likely be the interesting statistics. The community standard may, for example, be taken as the upper 75th percentile, that is, the level of offensiveness which only 25% of the community can accept. We may call this value X_c and it can be given a standard error, say s_c , to use in judging its usefulness.

The evidence of interest now becomes the \bar{X}_j 's for each of the trial materials in comparison to the one value for X_c . If $\bar{X}_j < X_c$ the j th material is declared not beyond being patently offensive but if $\bar{X}_j > X_c$ it is.

Almost all of the steps and quantities in the procedure are objective and determined by data. The choice of a 75th percentile, or median or 90th percentile, or some other means of choosing X_c from the survey data, should be decided before the survey is done. Also the way in which the standard errors are used to permit a decision to be made that material j is "significantly" larger than X_c must be decided before looking at the data.

References

- Allport, G. W. (1985) "The historical background of social psychology" Chapter 1 in Lindzey, Gardner and Aronson, Elliot (eds) *Handbook of Social Psychology* vol. 1, Random House, NY.
- Lundberg, G. A. (1942) *Social Research*, Longmans, NY.
- Nadel, S. F. (1957) *The Theory of Social Structure*, Cohen and West, London.
- Panel (1984), Turner, C. F. and Martin, Elizabeth (eds) *Surveying Subjective Phenomena*, vol. 1, Russell Sage Foundation, NY.
- Wallace, W. L. (1988) "Towards a disciplinary matrix in sociology," Chapter 1 in Smelser, N. J. (ed) *The Handbook of Sociology*, Sage, Newbury Park, CA.