

THE ECONOMIC
AND SOCIAL
RESEARCH INSTITUTE
MEMORANDUM SERIES
NO. 143

The Validity of Scaling in Social Inquiry

R. C. Geary

Confidential: Not to be quoted
until the permission of the Author
and the Institute is obtained.

The Validity of Scaling in Social Inquiry

R. C. Geary

One assumes that inquiries bear on matters of current interest, on which the population sampled have opinions and perhaps some knowledge and interest. Many investigations have shown, however, that the ordinary public are astonishingly ignorant of even the most commonplace facts, which does not prevent them sometimes from having strong opinions, the more ignorant, indeed, the likely to be the stronger. The great majority of respondents will have but a single view on important issues, yes or no, for or against, like or dislike etc. Indeed the main object of research may be to analyse the public attitude, the reasons which govern it. It will not always be a simple dichotomy. We must also provide for no answer, the reasons for which may include disinterest, refusal, lack of knowledge, inability to decide etc.

It is emphasized that in attitude surveys the statements or questions on which reactions are required should not be factual. It is hard to conceive of any questions, however, without a substratum of fact. Suppose, for instance, I am asked "Do you like the English?" I am supposed to answer quickly, instinctively. I have no personal meter of likes and dislikes; my temperature may not rise measurably when some person or thing I dislike is mentioned. So it is hard to conceive of some direct scale by which I could answer yes or no according to its showing above or below a predetermined value (or with more refined scaling), the same for all respondents. But there are other facts besides measured facts. I may recall Ireland's history, being a strong Nationalist, inclining me to answer no. I may know a few English people whom I like or hear from friends and relations or, being well-informed, know that our prosperity depends largely on the English market for our exports. The last two points are factual also and may lead towards an answer yes. Or a single event may determine my attitude, quite illogically. I may have thought of the problem before. So, a quick answer may not preclude a

large amount of ratiocination, and the more I try to draw on my mental store the more indecisive I may feel, all those yeses and noes leading to the unsatisfactory "I don't know". There must be a tendency on the part of people who can think for themselves on matters not directly affecting them (probably a small proportion of all people) to render an indefinite answer to what might seem a simple question.

In social random sample field surveys, sometimes extending to thousands of respondents, using the personal interview method, it is customary for the interviewer to make a statement (sometimes couched in emphatic terms and not in the form of a question) and asking the person questioned to reply quickly under one of the heads:-

Agree -

Strongly

Moderately

Slightly

Disagree -

Slightly

Moderately

Strongly

There is usually a seventh, more or less neutral category which we discuss later. Some persons might be inclined to ask the interviewer "what do you mean by these categories?" as indeed the present writer asks himself here. The method is to be respected since thousands of persons in many inquiries have been able to answer under one or other of these heads without prompting, and reports do not indicate that particular difficulty was experienced; "don't knows" might show, but there is no evidence that "don't knows" are associated with inability to categorize as above:

Is the rating "strongly" the only definite answer or "strongly" and "moderately" taken together? Yet how can a respondent envisage a difference between "moderately" and "slightly"? Are all ratings except "strongly" to be regarded as degrees of "don't know"?

Suppose that the statement is "Jews are nice people". Immediately a set of more or less relevant aspects may crowd into the respondents mind.

What on earth is meant by 'nice'? Do I know many Jews personally? Have I read about Jews? Do I have to distinguish between Israelis and the majority of Jews elsewhere? Judaism as a religion. Friendship of Ireland and Israel at a low ebb. Hitler's persecution. A recent newspaper account of a cruel Jewish raid into Lebanon. My personal belief may be that all peoples are decent to much the same extent, whatever about their leaders. One could, of course, add to the list. How does one make a rational answer quickly out of this lot alone? Different aspects suggest different ratings. The mildness of the statement increases the difficulty of answering in the form required: a Hitlerian statement would elicit a "Disagree strongly" from most people.

The writer tried out the foregoing statement on a schoolgirl. Asking why she reacted "Agree slightly" she replied "I know very few Jews".

When an inquiry deals with measurable facts relating to the units of inquiry interviewers use the same standard from unit to unit; for instance, for inquiry into the height of schoolchildren of a certain age they use tape measures. It is hard to see how such uniformity can obtain with opinion inquiries: one person's "slightly" may mean the same as the next's "strongly", in view of the very different considerations which may govern individual rating.

It has been suggested that the validity of scaling of the type indicated might be tested by the following experiment. Pick a random sample of persons and ask them if they consider themselves taller or smaller than average in the subcategories

very much - moderately - slightly, i. e. six decisions in all, not counting "don't knows". The results would be compared with actual heights by calculating actual averages and variances for each of the six scales. If the averages for pairs of scales were found significantly different and rightly ordered (e. g. that the moderately small were not significantly taller than the slightly small) one would decide that the sixfold scaling was valid. One could also study the degree of overlap.

But would this type of experiment validate scaling as applied to pure opinion statements, e. g. as above "Jews are nice people", requiring reactions of agree or disagree each in the grades strongly, moderately, slightly? The difference is that there is an absolute measure in the first case. In fact in this case scaling is quite unnecessary; since scales are numbered (e. g. 1, 2, . . . , 6) and correlations, factors etc., derived from these, it would be far better to use for analysis the actual heights rather than the scale numbers which merely distort mensuration. But the essence of the second type of scaling is the absence of absolute measures: in the example cited there is no meter for accessing "niceness". In fact the main reason for scaling is that by numeration (e. g. 1-6) statistical analysis on the lines indicated may be possible.

To try to find the reasons for responses in scale to a particular statement a questionnaire was handed to a number of friends about one-half of whom were social researchers. The preamble read:-

In attitude research the strongly - moderately - slightly degrees of disagreement by respondents to particular statements is used for statistical analysis. My present object is to ascertain the meaning which respondents attach to their scaling. I would be very much obliged for your cooperation at this pilot inquiry stage, by filling in the following questionnaire, based on the single statement.

The statement was a positive assertion of opinion on foreign politics, very much in the news at present, not relating directly to Ireland. It is not reproduced here lest inferences be drawn from a very small non-random sample on Irish opinion

on an important question and irrelevant to the purpose of inquiry. Number of returns received was 47, summarized as follows under all the heads of the questionnaire

1. I agree or disagree with statement -

	No.
strongly	20
moderately	14
slightly	7
No answer	<u>6</u>
Total	47

2. I consider that the issue is:-

	No.
very important	26
moderately important	11
slightly important	2
unimportant	-
No answer	2

3. I would describe my knowledge of the issue as -

	No.
considerable	3
moderate	30
small or nonexistent	7
No answer	1

4. Do you read regularly about foreign affairs in newspapers or periodicals?

	No.
Yes	36
No	5

5. Would you describe your interest in the issue as -

	No.
considerable	16
moderate	23
slight	2
nil	-

6. If 1 is not answered the reason:-

Difficulty of such discrimination	No. 5
Couldn't be bothered	-
Can't make up my mind	1
Lack of knowledge or interest	<u>-</u> 6

The great majority (in fact 41 out of 47) felt able to react positively to the statement, unprompted; of that number about half (20) felt strongly, half (21) moderately or slightly. It is interesting that so many (36) read about foreign affairs regularly, yet the majority had only a moderate interest in the matter of the questionnaire (head 5).

The rationale of the questionnaire is sense of importance (head 2), knowledge (3, 4), interest (4, 5), in relation to the 41 who answered. Of the 6 who did not (head 6), all could be said to have found answering difficult (or perhaps impossible).

The relationship between head 1 and each of the four heads 2 - 5 was examined using chi-squared technique with the following results:-

<u>Relation</u>	<u>X²</u>	<u>Degrees of freedom</u>	<u>Significance</u>
1 v. 2	2.97	4	Not significant
1 v. 3	3.74	4	" "
1 v. 4	1.17	2	" "
1 v. 5	1.40	4	" "

Significance was assessed in relation to NHP = .05. None of the X² values were anywhere near significance; .05 NHP critical values are 9.49 for 4 d.f. and 5.99 for 2 d.f. According to this experiment scaling did not depend on sense of importance, knowledge or interest.

Respondents who find threefold scaling difficult to conceive but who wish to answer positively may be inclined to opt for "strongly", perhaps because, as a form of words, not much difference is observed between "agree (or disagree)" and "agree (or disagree) strongly". So interest may centre on "moderately, slightly"

by themselves. The following analysis was confined to those who so reacted alone*.

<u>Relation</u>	<u>X²</u>	<u>Degrees of freedom</u>	<u>Significance</u>
1 v. 2	0.07	1	Not significant
1 v. 3	1.56	2	" "
1 v. 4	0.05	1	" "
1 v. 5	0.62	2	" "

The results are the same as before: no significant relationship. Again no χ^2 is near significance, removing any doubt about these conclusions because numbers on which based are very small.

In the questionnaire respondents were asked to state briefly "what would determine your scaling of attitude statements of the type indicated?" Following is a summary of the 32 replies received. They have been classified into the four categories of head 1 above, i. e. relating to a single statement. Do respondents react the same way on every issue?

"Strong" reactors How one feels at the time. Reactions mean different things to different people. Comparison of the strength of opinion (i) mine on this compared to mine on other issues, (ii) my opinion compared to others on a particular issue. Importance, whether national, religious or personal, knowledge. Gut feeling. Prior knowledge. Don't know what question means. State of mind at the time, interest, knowledge. Statement unclear. How strongly felt about question. Form badly drafted.

Moderate reactors Knowledge. Effect on me or my lifestyle. Knowledge and interest. Knowledge and importance. Scalings do not always fit one's feelings and so are misleading. Interest, especially on analogy with our own country. Informed knowledge, assessment of meaning of question. Experts differ but I would wish to

* One extra return received late was included. It was a "moderate" under head 1.

test the sensitivity of my conclusions. Cannot answer because key statement is susceptible of two interpretations. The (foreign) country should have reacted but not so much (the basic statement related to the reaction to an event by a foreign country). Action of both countries mentioned in basic statement "despicable", so I scale my opinion "moderate". Question extremely difficult; questionnaires resented because, no matter how many boxes, they fail to represent one's views; rating "moderate" because not favourable to action of both countries mentioned.

"Slight reactors. Difficult; what is difference between "slightly" and "moderately"; mixture of emotion and reason. Unsure, not certain one way or the other. Disagree with action of both countries mentioned in basic statement; balance of considerations on all sides of any issue would determine my scaling in general. Sense of degree of blame attaching to one country in this particular case.

In the case of non-scalers, two stated explicitly that they were prepared to agree or disagree with any statement but not to scale.

Every opinion given is summarized: the "me", "mine" etc. are the respondents'. The reader may form his own opinion as to what they convey. This writer's own impression is the individuality of respondent reaction. There was slight commonalty of the writer's own hypothesized reasons (importance, knowledge, interest) but this may have been suggested by the questionnaire itself. There were two reactions by respondents that the basic statement was an ambiguous one (but only one) that the form was badly drafted. Social researchers tend to fault one another's inquiries.

Conclusion

The writer confesses to a personal difficulty about reacting to any statement in the sixfold scaling, though he has no difficulty with the twofold yes-no etc., when he has any knowledge of the issue. With this bias he must be careful in wording a conclusion to this inquiry. In particular he must take into account the fact that many researchers use this technique and that, without forcing, the majority of respondents seem able to cope.

It is easy to fault the sample inquiry. It was small, non-random and a few (no doubt expert) respondents have criticized the form. If it be small the very insignificance of the values of χ^2 cannot be ignored. If the respondents did not read the statement and marked the form at random, the χ^2 's would be similar in value. The results are consistent with the thesis, surmised before the inquiry, that each respondent has an individual reaction to the sixfold scaling: there is no common measure. To this result it may be objected that when several cognate questions (or statements of attitude) and addressed to the same respondents the results are consistent; when the scaling is numbered (say 1-6) significant ccs are found between pairs of variables. This fact does not negate a fact of individual reaction. In the first place, the ccs, while significant, are rarely large, secondly and more important, a degree of positive relationship could be due to each individual's having a tendency to react to all opinion statements the same way, e.g. if once "strongly", always "strongly", with some of the population sampled. Anyway the dichotomy agree-disagree (not considered in the inquiry) and to which the writer has no objection will impose its own consistency in answering.

To the objection of non-randomness the writer would rejoin that the sample was heavily weighted for intelligence and expertise and, if such persons show inconsistency in answering, can better be expected from ordinary people? As to faults in the questionnaire, objectors numbered only three. Of course they may

be right, but his thinking and the results of the inquiry more firmly entrench the writer in his minority opinion that sixfold scaling of attitude is invalid.