

Empirical scaling of common verbal phrases associated with numerical probabilities¹

SARAH LICHTENSTEIN, OREGON RESEARCH INSTITUTE
J. ROBERT NEWMAN, SYSTEM DEVELOPMENT CORPORATION

One hundred eighty-eight Ss gave numerical probability estimates for each of 41 probability-related words and phrases. While the responses were reasonably consistent, asymmetry between mirror-image phrases, e.g., quite likely (.79) and quite unlikely (.11), were found. Thus verbal labels on numerical probability response scales may not be practicable.

Probability theory is one tool we use to deal with the uncertainty of future events. The result of a probability calculation of an event that has not yet occurred is a number varying from 0 to 1. This number expresses the likelihood of that event occurring. There are many verbal phrases that also are used to express intuitively the notion of the likelihood of an event. When getting numerical estimates of probability from Ss, it is often tempting to attach verbal labels to the probability scale used as a response device, in hopes that all Ss would thus be encouraged to think about the scale in similar terms. Little is known, however, about which phrases to use. Stone & Johnson (1959) and Simpson (1963) have scaled a large number of frequentistic phrases (e.g., "often," "seldom"); however, in some contexts (e.g., "what is the likelihood that this patient is psychotic?") such frequentistic phrases are less appropriate than nonfrequentistic expressions of probability (e.g., "likely," "improbable"); this study focuses on the latter. Cohen, Dearnley, & Hansel (1958) scaled a few probability phrases, but not enough to give a full picture: The purpose of the present study is to get an estimate of the correspondence between numerical probabilities and a wide range of associated verbal phrases.

Method

A questionnaire with 41 words and phrases was prepared, requesting respondents to give the probability number from .01 to .99 which most clearly reflected the degree of probability implied by each word or phrase. The questionnaire was sent to a random selection of 225 System Development Corporation male employees listed on the organization chart; 188 scorable replies were received.

The words were chosen to represent a broad range of possible probability phrases. A systematic variation of adverbs were combined with the root "likely" to produce 13 phrases. Included in this group was the phrase "rather unlikely" appearing twice, once near the beginning and once near the end, as a reliability check. Six phrases had the root "probable," five phrases had the root "chance," and three phrases

the root "possible." Six phrases reflected the .40 to .60 range of probability (e.g., "slightly more than half the time"), and four single, commonly used words were included ("rare," "seldom," "tossup," and "usually"). Finally, four words which are not usually associated with particular levels of probability were included ("predictable," "inconclusive," "uncertain," and "rather").

Results

Simple descriptive statistics summarizing the data are shown in Table 1; the phrases are ranked in order of mean response. Adverbs attached to "likely" were ordered the same way as to "unlikely," that is, "very" > "quite" > no adverb > "rather" > "fairly" > "somewhat." However, this symmetric set of phrases showed an asymmetry of response; the more probable phrases like "quite likely" (mean = .79) were about .10 less than what one would expect by looking at their mirror images (e.g., "quite unlikely," mean = .11). This skewness was found for eight of the 11 mirrored pairs.

There was close agreement between means and medians. For only five phrases did the mean and median differ by more than .04. Three of these phrases, "inconclusive," "uncertain," and "possible," had distributions slightly skewed to the left, while the other two, "barely possible" and "faintly possible," were strongly skewed to the right, with modes at .01, .05, and .10.

The item most often omitted, understandably, was "rather." However, the 124 people willing to assign a probability to this ambiguous word showed fair agreement ($\sigma = .11$).

The reliability check on the duplicated entry, "rather unlikely," showed satisfactory stability: Mean = .24, $\sigma = .12$, and mean = .21, $\sigma = .10$, for the two cases.

Discussion

While many of these verbal phrases showed encouraging consistency across Ss, two considerations predicate against their use in labeling probability response scales. The first is discrepancy with other results. Simpson (1963) used three of these phrases, reporting responses within one standard deviation of those reported here ("usually," .85 and .80; "seldom," .10; and "rarely," .05), but Cohen, Dearnley, & Hansel (1958), using three other phrases, report very different results ("likely," .56 and .63; "probable," .54 and .55; and "improbable," .33 and .37). This discrepancy is undoubtedly due to the two special contexts used: "...the Labour Party is likely

TABLE 1
 Statistics on Probability Phrases

| Rank | | N | Mean | Median | σ | Range |
|------|-------------------------------------|-----|------|--------|----------|---------|
| 1 | Highly probable | 187 | .89 | .90 | .04 | .60-.99 |
| 2 | Very likely | 185 | .87 | .90 | .06 | .45-.99 |
| 3 | Very probable | 187 | .87 | .89 | .07 | .60-.99 |
| 4 | Quite likely | 188 | .79 | .80 | .10 | .30-.99 |
| 5 | Usually | 187 | .77 | .75 | .13 | .15-.99 |
| 6 | Good chance | 188 | .74 | .75 | .12 | .25-.96 |
| 7 | Predictable | 146 | .74 | .75 | .20 | .01-.99 |
| 8 | Likely | 188 | .72 | .75 | .11 | .25-.99 |
| 9 | Probable | 188 | .71 | .75 | .17 | .01-.99 |
| 10 | Rather likely | 188 | .69 | .70 | .09 | .15-.99 |
| 11 | Pretty good chance | 188 | .67 | .70 | .12 | .25-.95 |
| 12 | Fairly likely | 188 | .66 | .70 | .12 | .15-.95 |
| 13 | Somewhat likely | 187 | .59 | .60 | .18 | .20-.92 |
| 14 | Better than even | 187 | .58 | .60 | .06 | .45-.89 |
| 15 | Rather | 124 | .58 | .60 | .11 | .10-.80 |
| 16 | Slightly more than half the time | 188 | .55 | .55 | .06 | .45-.80 |
| 17 | Slight odds in favor | 187 | .55 | .55 | .08 | .05-.75 |
| 18 | Fair chance | 188 | .51 | .50 | .13 | .20-.85 |
| 19 | Tossup | 188 | .50 | .50 | .00 | .45-.52 |
| 20 | Fighting chance | 186 | .47 | .50 | .17 | .05-.90 |
| 21 | Slightly less than half the time | 188 | .45 | .45 | .04 | .05-.50 |
| 22 | Slight odds against | 185 | .45 | .45 | .11 | .10-.99 |
| 23 | Not quite even | 180 | .44 | .45 | .07 | .05-.60 |
| 24 | Inconclusive | 153 | .43 | .50 | .14 | .01-.75 |
| 25 | Uncertain | 173 | .40 | .50 | .14 | .08-.90 |
| 26 | Possible | 178 | .37 | .49 | .23 | .01-.99 |
| 27 | Somewhat unlikely | 186 | .31 | .33 | .12 | .03-.80 |
| 28 | Fairly unlikely | 187 | .25 | .25 | .11 | .02-.75 |
| 29 | Rather unlikely | 187 | .24 | .25 | .12 | .01-.75 |
| 30 | Rather unlikely | 187 | .21 | .20 | .10 | .01-.60 |
| 31 | Not very probable | 187 | .20 | .20 | .12 | .01-.60 |
| 32 | Unlikely | 188 | .18 | .16 | .10 | .01-.45 |
| 33 | Not much chance | 186 | .16 | .15 | .09 | .01-.45 |
| 34 | Seldom | 188 | .16 | .15 | .08 | .01-.47 |
| 35 | Barely possible | 180 | .13 | .05 | .17 | .01-.60 |
| 36 | Faintly possible | 184 | .13 | .05 | .16 | .01-.60 |
| 37 | Improbable | 187 | .12 | .10 | .09 | .01-.40 |
| 38 | Quite unlikely | 187 | .11 | .10 | .08 | .01-.50 |
| 39 | Very unlikely | 186 | .09 | .10 | .07 | .01-.50 |
| 40 | Rare | 187 | .07 | .05 | .07 | .01-.30 |
| 41 | Highly improbable | 181 | .06 | .05 | .05 | .01-.30 |

to win the next election," and "...we are likely to have a fine summer" (in England!). Such discrepancies might be expected in using the phrases in other specific contexts.

The asymmetry found between mirror-image pairs also creates problems for practical use. If this asymmetry is taken into account in labeling a probability scale, the resulting scale, although true to Ss' reported impressions, would look so peculiar it might confuse or mislead Ss.

References

- COHEN, J., DEARNLEY, E. J., & HANSEL, C. E. M. A quantitative study of meaning. *Brit. J. educ. Psychol.*, 1958, 28, 141-148.
 SIMPSON, R. H. Stability in meanings for quantitative terms: A comparison over 20 years. *Quart. J. Speech*, 1963, 49, 146-151.
 STONE, D. R., & JOHNSON, R. T. A study of words indicating frequency. *J. educ. Psychol.*, 1959, 50, 224-227.

Note

1. This work was supported by the Advanced Research Project Agency under Contract SD-97 with the System Development Corporation.