

## ADDITIONAL REFERENCES

- BIERWISCH, M. (1967). Some semantic universals of German adjectives. *Foundations of Language* 3 1-36.
- HÖRMANN, H. (1983a). *Was tun die Wörter miteinander im Satz?*

- oder wieviele sind einige, mehrere und ein paar?* Verlag für Psychologie, Dr. C. J. Hogrefe, Göttingen.
- KATZ, J. J. (1964). Semantic theory and the meaning of "good." *J. Philosophy* 61 739-766.

# Comment

Norman Cliff

The efforts of Mosteller and Youtz to introduce more standardization represent a desirable goal. However, some barriers to such standardization are likely, and the extent to which complete standardization is desirable is questionable. Words are inherently fuzzy and communicating degree of fuzziness is a significant aspect of communication. Studies of ambiguity would do well to focus more on the range of meanings that are communicated and less on measures of central tendency.

One of the social phenomena that can occur in a field is that its terminology can tend toward anarchy, anything meaning anything. This seems to have happened to some degree in the case of probabilistic terms, and Mosteller and Youtz' paper, and the literature it summarizes and will stimulate, may be a useful counter to this tendency. The paper provides a good summary of quite a range of empirical research (one would, however, have hoped to see the pioneering study by Howe, 1962, noted) as well as giving new data of their own; however, there are aspects of the empirical literature, and their own data, to which one could wish they had paid closer attention. The paper's effect should be positive, but it would be optimistic to expect its effect to be great, and from some points of view complete codification has undesirable consequences.

Standardization of terminology has been a goal of individuals and groups since at least the Tower of Babel. While greater uniformity of word usage would seem to have desirable properties, not many such efforts have met with success. One therefore wonders whether Mosteller and Youtz' will be one of the exceptions.

The exceptions that come to mind most easily lie in the sciences, where there are standard terms for the physical units, standard names for the chemical ele-

ments and a standardized procedure for naming chemical compounds, to name a few. A similar situation exists with the Linnaean system for naming organisms. Such standardization also occurs to some extent in other fields as well; a hardware salesperson makes a consistent linguistic distinction between *hardware cloth* and *screening*.

Without posing as more than an interested observer of such phenomena, one can speculate on the variables that lead to standardization. It seems that two important ones are *isolation of communicators* and *specificity of referent*, accompanied by penalty—social or economic—for linguistic error. One can question the extent to which these conditions are present in the case of probability terms. The linguistic community for probability terms does not seem very isolated. Everyone has almost daily necessity of referring to the chances that an everyday event will occur, and people, such as statisticians and various types of data analysts, who have reason to refer to formally estimated probabilities, are frequently faced with referring to more informally defined events. Furthermore, the community of individuals who act as statisticians is not very closed. Thus it seems likely that there is a large degree of interchange, both within and between individuals and between formal within-community usages and informal extra-community ones. This will act to undermine any attempts at standardization.

One can also examine the degree of specificity of referent that characterizes the probability field. The very guidelines suggested by Mosteller and Youtz have themselves a kind of vagueness of boundaries. One can suspect that if boundaries were not vague, there would be endless debates about where such boundaries should lie. It is also hard to see much in the way of direct consequences for violating any linguistic strictures that are developed. Saying "fairly likely" instead of "probable" is unlikely to lead anyone as seriously astray as saying "grams" instead of "dynes" or *Rattus norvegicus* instead of *Rattus rattus*. Thus, trying to keep a writer from using whatever term comes to mind

---

Norman Cliff is Professor of Psychology, University of Southern California, Log Angeles, California 90089.

is likely to prove much more difficult here than in the cases where standardization has been successful.

Mosteller and Youtz perhaps focus too much on averages, making considerable use of the fact that some terms have more stable averages across groups. However, averages have well-known inadequacies. Taking account more formally of the variability of response to a given term would seem particularly relevant in this context.

In trying to communicate to an audience of more than trivial size, one would not only want to be confident that an average of the auditors' interpretations was appropriate, but that as many as possible of the interpretations were close to that average. The paper might have paid more attention to measures of inter-individual variability and intra-individual ranges of response or acceptability than to the stability of averages across fairly comparable populations or mild changes of context. Their acceptability functions are useful in this regard, but not enough of this data is presented to permit the full effect. If uniformity is the goal, we would want to choose those terms which have the smallest inter-individual variability, and I wish they had presented more data of the acceptability range type.

In this context I have noticed a perhaps surprising phenomenon. When I first studied word combination principles (Cliff, 1959, 1988), I expected that the combination of a word with a modifier would result in a combination of their respective variabilities, so that combinations would have more variability than single words. This seems not to be the case. The data I have looked at in Gallipeau (1986) as well as my own supports the idea that the reverse is true. There is less variability, both between and within persons, for combinations than for single words. Variabilities are rarely reported in the literature, so it is difficult to validate this finding in other research. It does seem likely, though, that, while substituting "fairly probable" for "probable" would have only a small effect on the average of the numerical probabilities assigned, there would be a reduction in the variance of the responses. In this respect, the recommendation that verbal statements of probabilities be formed by modifying a stem is a good one. It might go a step farther, not using the unmodified form at all.

We also need to consider whether in fact variability of interpretation is undesirable. One of the fruits of the fuzzy set approach to semantic processing is the production of data that indicate the *degree* of fuzziness of different words or phrases. Apparently there is (a) considerable fuzziness for almost all terms, and (b) considerable variation among verbal stimuli in this regard. This is at first puzzling. If the set of possible meanings for a term is so broad, what good is the term? There are probably two answers to this.

The first answer is that the breadth is something that can be decreased by context or modification. Even a term like "probable" has a range of meanings that is rather broad, not only in the numerical referents it could have, but also in a more qualitative sense. The auditor selects specifics out of this breadth according to the context in which the term is used as well as in terms of the modifiers that are applied to it. Even in a context, and even under modification, the term's meaning set can still be rather broad. Requiring it to be even moderately precise is probably hopeless.

The second answer is that one of the things we have to communicate to the auditor is a notion of the degree of fuzziness of the statement. When we translate a numerical probability into a verbal statement, is not one goal to trade the specificity of the number for a generality of the statement, modified by the message that we do not really mean, say exactly 0.3574 . . . , or even  $0.357 \pm 0.064$ , but just some number around the lower middle of the probability scale? That is, the degree of fuzziness is a dimension of the communication. Attempting to standardize at some level of fuzziness would deprive the communicator of this dimension.

There is another issue that has been touched on tangentially above but that deserves its own mention. This is whether any terms have a context-free meaning. Basic issues of the psychology of judgment apply here, if nothing else. There is now much evidence that we have an internal scale of quantity in much the same sense that we have an internal scale of pitch (Moyer and Landauer, 1967; Holyoak and Walker, 1976). All the effects that are known to occur in psychophysics are very likely to occur here as well, including effects of adaptation level, anchoring stimuli, and the like. Indeed, some of these are already demonstrated (Banks, Fujii and Kayra-Stuart, 1976; Holyoak and Mah, 1982). This means that the numerical referent of any probability term is likely to vary, perhaps considerably, no matter how narrowly we restrict usage. One can go further and say that the terminology should be allowed to vary as a function of context if the appropriate interpretation by the auditor is to be made.

The degree to which there are context effects was illustrated by the study by Pepper and Prytulak (1974). They studied frequency adverbs using a probability response scale, and found, for example, that the term "very often" used in one context had a smaller numerical probability referent than "sometimes" had when used in another. That is, there was a strong effect of the prior probability of the event described. Much the same sort of effect can be expected for probability terms themselves. I think Mosteller and Youtz are too ready to dismiss such effects. All of this tends to support the idea that only

modest uniformity at best can be expected for any probability terms.

In formulating their recommendations, it would have been useful to pay closer attention to the extensive recent work on very nearly this same topic by Budescu and Wallsten and their collaborators (Wallsten, Budescu, Rapoport, Zwick and Forsyth, 1986; Budescu and Wallsten, 1985; Wallsten, Fillenbaum and Cox, 1986). These provide a broader empirical and conceptual perspective on the topic than I think is provided by the present proposal.

All of this is not to denigrate the value of the present article for the audience to which it is addressed. We all need to apply probability terms judiciously rather than haphazardly or arbitrarily. It is useful to be reminded of the almost inherent ambiguity of the double negative, or that certain phrases bring in connotations that may be unwanted, or that some terms have more ambiguity than others. However, this all comes under the rubric of advice to be careful of what you say and how you say it. A writer of even moderate skill is aware of connotations and contexts and will make use of them in a way that optimizes his or her ability to communicate. Even the double negative "not improbable" has its uses of emphasizing to those who thought something *was* improbable that their expectations were wrong or just as a stylistic device giving the readers a bit of a prod to make them pay attention. I applaud the motives of Mosteller and Youtz toward making us all try to communicate more carefully, without being very sanguine about their chances of

having any major effect. The main barriers to effective communication lie elsewhere, I feel, but that is a different issue.

### ACKNOWLEDGMENT

The author wishes to thank Sumiko Takayanagi for her contribution toward these comments.

### ADDITIONAL REFERENCES

- BANKS, N. P., FUJII, M. and XAYRA-STUART, F. (1976). Semantic congruity in comparative judgments of magnitude of digits. *J. Experimental Psychology: Human Perception and Performance* 2 435-447.
- CLIFF, N. (1988). A model-based analysis. *Chance* 1 (3) 32-36. Response, 50-51.
- GALLIPEAU, D. R. (1986). The nature of adverb-verb and adjective-noun combinations in the formation of connotative meaning. Ph.D. dissertation, Univ. Southern California.
- HOLYOAK, K. J. and MAH, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology* 14 328-352.
- HOLYOAK, K. J. and WALKER, J. H. (1976). Subjective magnitude information in semantic orderings. *J. Verbal Learning and Verbal Behavior* 15 287-299.
- HOWE, E. S. (1962). Probabilistic adverbial qualifications of adjectives. *J. Verbal Learning and Verbal Behavior* 1 225-241.
- MOYER, R. S. and LANDAUER, T. K. (1967). Time required for judgments of numerical inequality. *Nature* 215 1519-1520.
- WALLSTEN, T. S., BUDESCU, D. V., RAPOPORT, A., ZWICK, R., and FORSYTH, B. (1986). Measuring the vague meanings of probability terms. *J. Experimental Psychology: General* 115 348-365.
- WALLSTEN, T. S., FILLENBAUM, S. and COX, J. A. (1986). Base rate effects on the interpretation of probability and frequency expressions. *J. Memory and Language* 25 571-587.

## Comment: Codifying Chance

Joseph B. Kadane

Mosteller and Youtz have given us an interesting meta-analysis of the literature on the probabilities that people associate with various descriptions of uncertainty. They have also added to this literature a study of their own using science writers as subjects.

My only question about their treatment of this data is whether a transformation, such as log-odds or arcsine, would have reduced the boundary effects that are so pronounced in looking at the interquartile range.

---

*Joseph B. Kadane is Leonard J. Savage Professor of Statistics and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213*

They propose that their research be used for codification of at least some probabilistic expressions. One possible application of codification would be for expert witnesses in court. Often when a scientist, for example a statistical scientist, testifies, no one else in the court is comfortable with quantification. It would be very useful to have words in English to express the import of the crucial numbers, and to have the choice of those English words be based on data of the kind provided and reviewed by Mosteller and Youtz.

In order to be an effective tool for this purpose, it is not so important that each possible word be represented, as that each state of uncertainty be represented. Therefore, many less than 52 words, perhaps a dozen, are needed. Accordingly I reorganized some