[10] LEVINE, J. M. and ELDREDGE, D. (December 1970). The effects of ancillary information upon photo interpreter performance. American Institutes for Research, Institute for Research in Psychobiology, Washington Office, AIR-20131-12/70-FR.
$n = 20$: enlisted U.S. Army image interpreters.

[11] MAPES, R. E. A. (1979). Verbal and numerical estimates of probability in therapeutic contexts. *Social Science and Medicine* **13A** 277–282.
$n_a = 29$: physicians given expression "Side effects with chloramphenicol are *frequent*." $n_b = 33$: physicians given expression "Side effects with neomycin sulphate are *frequent*."

[12] NAKAO, M. A. and AXELROD, S. (1983). Numbers are better than words: Verbal specifications of frequency have no place in medicine. *Amer. J. Med.* **74** 1061–1065.
$n_a = 103$ physicians. $n_b = 106$ physicians. Means read from chart.

[13] REAGAN, R. T., MOSTELLER, F. and YOUTZ, C. (1989). The quantitative meanings of verbal probability expressions. *J. Appl. Psychology* **74** 433–442.
$n = 115$: undergraduates in a psychology course, Stanford University.

[14] REYNA, V. F. (1981). The language of possibility and probability: Effects of negation on meaning. *Memory and Cognition* **9** 642–650.
$n = 41$ adult volunteers.

[15] ROBERTS, D. E. and GUPTA, G. (1987). To the editor. *New England J. Med.* **316** 550.
$n_a = 45$ house staff. $n_b = 24$ attending physicians.

[16] ROBERTSON, W. O. (1983). Quantifying the meanings of words. *J. Amer. Med. Assoc.* **249** 2631–2632.
$n_a = 53$: Seattle physicians. $n_b = 80$: graduate students at the University of Washington's School of Business Administration. $n_c = 40$: Board of Trustees at the Children's Orthopedic Hospital and Medical Center, Seattle.

[17] SELVIDGE, J. (1972). Assigning probabilities to rare events. Ph.D. dissertation, Graduate School of Business Administration, George F. Baker Foundation, Harvard Univ. Subjects were Harvard Business School students in MBA program. $n_a = 59$: Estimates made on basis of a statement without context. $n_b = 127$: Contexts were provided.
Also in Mosteller, F. (1977). Assessing unknown numbers: Order of magnitude estimation. In *Statistics and Public Policy* (W. B. Fairley and F. Mosteller, eds.) 163–184. Addison-Wesley, Reading, Mass.

[18] SIMPSON, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. *Q. J. Speech* **49** 146–151.
1942 study. $n_a = 335$: 86 high school and 249 college students.
1962 study. $n_b = 395$ university students.

[19] TOOGOOD, J. H. (1980). What do we mean by "usually"? *Lancet* **1** 1094.
$n = 51$: physicians, nurses, laboratory technologists, secondary school teachers, and engineers.

[20] Current study, estimates from science writers.
$n \approx 230$: science writers. Varies somewhat from expression to expression, 211–237.

[21] BARTHOLOMEW, D. J. (1961). A method of allowing for "not-at-home" bias in sample surveys. *Appl. Statist.* **10** 52–59.

[22] BEYTH-MAROM, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. *J. Forecasting* **1** 257–269.

[23] BOFFEY, P. M. (1976). Anatomy of a decision how the nation declared war on swine flu. *Science* **192** 636–641.

[24] CLIFF, N. (1959). Adverbs as multipliers. *Psychol. Rev.* **66** 27–44. This paper stimulated a four-paper symposium on quantification of the effect of adverbs on the meaning of adjectives: *Chance* **1** (3) 32–51 (1988).

[25] GRIGORIU, B. D. and MIHAESCU, T. (1988). Evaluarea numerica a expresiilor de probabilitate folosite in limbajul medical. *Rev. Med. Chir. Soc. Med. Nat. Iasi* **92** 361–364.

[26] HANSEN, M. H. and HURWITZ, W. N. (1946). The problem of non-response in sample surveys. *J. Amer. Statist. Assoc.* **41** 517–529.

[27] KAHNEMAN, D. SLOVIC, P. and TVERSKY, A., eds. (1982). *Judgment under Uncertainty: Heuristics and Biases.* Cambridge Univ. Press, Cambridge.

[28] MOSTELLER, F. (1976). Swine flu: Quantifying the "possibility." *Science* **192** 1286, 1288.

[29] MOSTELLER, F. (1978). I. Non-sampling errors. In *International Encyclopedia of Statistics* (W. H. Kruskal and J. M. Tanur, eds.) **1** 208–229. The Free Press, New York.

[30] PEPPER, S. and PRYTULAK, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. *J. Res. in Personality* **8** 95–101.

[31] SLOVIC, P., FISCHOFF, B. and LICHTENSTEIN, S. (1982). Facts versus fears: Understanding perceived risk. In *Judgment under Uncertainty: Heuristics and Biases* (D. Kahneman, P. Slovic and A. Tversky, eds.) 463–489. Cambridge Univ. Press, Cambridge.

[32] TVERSKY, A. and KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science* **211** 453–458.

[33] KENT, S. (1949). *Strategic Intelligence.* Princeton Univ. Press, Princeton, N.J.

Comment

Herbert H. Clark

In the last few years, Mosteller, Youtz and their colleagues have looked at probability and frequency expressions such as *usual, very likely, improbable, frequent* and *as often as not.* Their interest is in how these terms are used in communicating technical information, and their goal is to better that communication, to make it more precise. Their project has two phases. In the first, they will determine what these terms mean to the people who use them. In their own study they have found, for example, that *frequent* is judged to represent an average proportion of about 0.72 of the time with an interquartile range of about 0.15. If you say something is frequent, they claim, you are saying that it occurs about 72% of the time plus

Herbert H. Clark is Professor of Psychology, Stanford University, Stanford, California 94305.

or minus 7.5%. In the second phase, they will codify a selection of these terms so that they can be used and interpreted consistently.

Mosteller and Youtz's goals are admirable. I will argue, however, that the project they describe leaves out one essential step they need for their goals. The problem comes from not distinguishing between word meaning and word use.

MEANING AND USE

We all recognize that words can change interpretations from one situation to the next. We are likely to interpret *good* as "adept" in *good juggler*, "tasty" in *good sauce*, and "healthy" in *good sleep*. We are likely to take *red* to denote different hues in *red car, red hair, red potato, red onion* and *red face*. And we are likely to infer different heights for *tall* when it is used in *tall grass* versus *tall tree*. We would even infer different heights for *tall* in *tall snowman* depending on whether the speaker was thinking about a snowman being built by a couple of kids or a college fraternity.

Yet it would seem wrong to say that the words *good, red* and *tall* actually change meaning from one situation to the next. *Good* still means "greater than some norm on some positive quality" (Katz, 1964), *red* still means "in a blood-colored direction relative to some set of hues," and *tall* still means "greater than some norm in height" (Bierwisch, 1967). What changes is the norm, the quality, the set of hues, and the normal height we infer in order to fill out these meanings. Word meaning is just not the same as word use, where by "word use" I mean what a person means in using a word on a particular occasion.

Tall is typical of words whose interpretations depend on context. It is a so-called relative adjective, one of a large set of adjectives and adverbs that includes *large, small, good, bad, near, far,* and many others. One of their primary properties is that they are two-place relations. One isn't just *tall.* One is *tall relative to C*, where C stands for a comparison set. So of Larry Bird, who plays basketball for the Boston Celtics, we might say "Bird is very tall for a man," "Bird is quite tall for a college basketball player," and "Bird isn't very tall for a professional basketball player." Usually, we leave C up to our listeners to infer. When we say "He certainly is tall," we could mean "for a basketball player," "for a fourth grader," or "for a person to be marrying such a short woman as Jennifer," depending on the situation. There is an explicit or implicit comparison set C for every use of a relative adjective.

Most of the expressions Mosteller and Youtz consider are relative adjectives or adverbs. These include *frequent, infrequent, rare, probable, improbable, likely, unlikely, often, seldom, rarely, usual, unusual, occasionally*, and all of their modified forms. *Frequent* really means "frequent relative to C." When we say "She frequently goes out to eat," we may mean "frequently relative to the frequency with which most people go out to eat" but when we say "She frequently goes to Europe," we may mean "frequently relative to the frequency with which most people go to Europe." Even terms like *probable* and *likely* fit the mold. When we say "He's likely to get a headache if he keeps on drinking," we may mean "likely relative to the likelihood that he would ordinarily get a headache." But when we say "He's likely to break a leg if he keeps on skiing," we may mean "likely relative to the likelihood that he would ordinarily break a leg," which is less probable than getting headaches.

EMPIRICAL RELATIVITY

Several empirical studies show how relative these terms can be. In a study by Pepper and Prytulak (1974), students were shown instances of these sentences, among others:

(1) At a recent press conference, Miss Sweden said she felt that in real life men [very often, frequently, sometimes, seldom, almost never] found her attractive.

(2) The *New York Daily News* reported that in the U.S.A. during February 1966, commercial passenger planes [very often, frequently, sometimes, seldom, almost never] crashed.

In (1) *frequent* was judged to represent about 75% of the time, but in (2), about 28% of the time. There were similar differences for the other four expressions. In a study by Mapes (1979), British physicians were shown one of these two sentences (among others):

(3) In Martindale or some similar text, one might read, "Side effects with chloramphenicol are frequent." What does frequent mean?

(4) In Martindale or some similar text, one might read, "Side effects with neomycin sulphate are frequent" What does frequent mean?

As it happens, chloramphenicol was known to have many side effects, and neomycin sulphate few. In 3 *frequent* was judged to represent a median of 39% of the time, but in 4, only 29% of the time.

For a more systematic study of relativity, let us consider Hörmann's (1983a) investigation of *a few*—or rather its German counterpart *ein paar*. He gave people expressions like "a few crumbs" and asked them to give a range of numbers representing how many objects were being denoted. The median estimates were 8.23 for "a few crumbs," 7.32 for "a few paperclips," 5.61 for "a few pills," but only 5.00 for "a

few children" and 4.58 for "a few mountains." As Hörmann observed, the larger the object, generally the fewer objects inferred.

These judgments depend on more than the noun that *a few* modifies. Consider these sentences (more or less literal translations of the German sentences) and people's median estimates for *a few*:

In front of the hut are standing a few people: 4.55

In front of the house are standing a few people: 5.33

In front of the city hall are standing a few people: 6.34

In front of the building are standing a few people: 6.69

The larger the space, the more people there can be, and the larger the median estimates. The same goes for these:

Out of the window one can see a few people: 5.86

Out of the window one can see a few cars: 5.45

Through the peephole one can see a few people: 4.76

Through the peephole one can see a few cars: 3.95

The estimates can depend on even subtler judgments of the possibilities, as in these pairs:

In front of the city hall there are a few people standing: 6.34

In front of the city hall there are a few people working: 5.14

Out of the window one can see a few people: 5.86

Out of the window one can see a few people arguing: 3.60

In the morning he read a few poems: 4.59

In the morning he wrote a few poems: 3.44

People take into account anything that changes the total number of items one would expect in a situation. The generalization seems to be this. People assess the total number of possibilities that one would expect in the situation described and then interpret *a few* relative to that total. But what precisely is the relation of *a few* to that total? It is probably not a constant proportion, but rather a point on some nonlinear transformation. That is an empirical question.

RECONSTRUCTING WORD MEANINGS

Theoretically, we never see word meanings directly. All we ever see are word uses. We can ask people about what a word denotes when it is used on a particular occasion, but we cannot ask them directly what it means. The meaning of a word is something we, or they, can only abstract from its uses, and even the best lexicographers have trouble doing that. That is, determining the meaning of an expression is a two step process. Step 1: Examine the possible uses of the expression and establish how it is interpreted on each occasion. Step 2: Reconstruct the meaning of the expression from the invariances in these interpretations.

But, you might argue, why can't we get at the meaning of a word directly simply by presenting it without any context? For *a few*, we might ask respondents "How many things does *a few* signify?" Such a procedure, I would argue, is illusory. A person trying to answer this question has to imagine a more or less concrete situation in which *a few* would be used and then estimate the number it would denote in that situation. Different people, of course, will imagine different situations, and as we have seen, different situations can yield very different numbers. If Hörmann had collected responses to this question (which he didn't), he might have got a median estimate of 5.50, but it would have been merely an aggregation over an unknown set of situations. The No Context condition, I suggest, is really the Unknown Context condition.

We can see the problem more clearly with the word *tall*. In the Unknown Context condition, we would ask respondents "What height does the word *tall* signify?" One respondent might think of buildings and reply "200 meters," another of trees and reply "20 meters," and a third of people and reply "2 meters." Or a single respondent might imagine all three types of situations and take the middle figure. In the end we would conclude, say, that *tall* means "10 meters with a range of 0.2 to 400 meters." This just doesn't seem to be the right way to proceed, and it is easy to see why.

1. *Sampling situations.* We have no idea what situations people are imagining for the question "What height does *tall* signify?" These could easily represent a highly skewed, limited, and unrepresentative sample. Further, the situations that come to mind for *tall* may be very different from those that come to mind for, say, *short*. If they are, the height estimates we get for *tall* and for *short* are not comparable. This is a serious worry.

2. *Aggregating over situations.* To be able to aggregate over the individual situations, we have to know how they all work. Do they yield functions that can be linearly combined, or not? It would seem wrong, for example, to average the three height estimates of 200, 20, and 2 meters. Should we take their median? Any answer to this question requires us to make strong assumptions about what we (or the respondents) are aggregating over in coming to the final estimates.

3. *Variation with situations.* The average or median height estimate yielded by this procedure, say 10 meters, will be of little help for any particular use of *tall*.

Estimates for individual situations—from talking about grass to skyscrapers—can vary by four orders of magnitude, and the variation is systematic.

4. *Meaning versus use.* The Unknown Context procedure invites us to think of the meaning of *tall* as a fixed height, our hypothetical 10 meters, located within some range: But *tall* is a relative adjective—*tall relative to C*—and we have not specified its meaning until we have specified that relation. If you asked me "What does *tall* mean?" it would be absurd for me to answer "It means 10 meters with a range of 0.2 to 400 meters." What makes it absurd is that the question is about word meaning, and the answer, about word uses. These are as different as apples and oranges.

THE MOSTELLER AND YOUTZ PROGRAM

By now my objections to Mosteller and Youtz's program should be clear. They assume that the meanings of the probability and frequency expressions can be determined in one step, by eliciting judgments in an Unknown Context condition. Mosteller and Youtz themselves asked science writers to give both the probability "that they personally would attach to each of these expressions" and "the range of probabilities that they thought their readers would associate with that expression." This is an Unknown Context condition. The other 19 studies they summarize relied on similar instructions.

The objections just raised against the Unknown Context condition apply to the Mosteller and Youtz program. Let me illustrate with the word *frequent.*

We have no idea what situations respondents imagine when they estimate proportions for *frequent.* Indeed, there is evidence that suggests that people may well imagine different situations for *frequent* than they do for, say, *rare. Frequent* would bring to mind frequent events, and *rare,* rare events. If so, people would be estimating numbers for *frequent* in such contexts as "TV commercials are frequent" and for *rare* in such contexts as "Airplane crashes are rare." That would make the estimates for *frequent* and *rare* not comparable. To be comparable, the estimates must be drawn from the identical set of contexts, as in "TV commercials are frequent" and "TV commercials are rare."

Whatever situations people imagine, it is misleading to throw the estimates together in a single number and range. *Frequent* is "frequent relative to C." We need more than a single estimate plus range. Recall that the science writers judged *frequent* to represent 72% of the time plus or minus 7.5% in the Unknown Context condition. That is the sort of figure that would be used to codify the meaning of *frequent.* It is troubling that the estimates *frequent* obtained in the contexts specified by 1, 3, and 4 were 28%, 29%, and 39%, and these lie far outside this range. In cases like these, what good is that range?

Ultimately, the problem is again apples and oranges. Just as it is a confusion of meaning and use to say that *tall* means "10 meters with a range of 0.2 to 400 meters," it is also a confusion of meaning and use to say that *frequent* means "72% of the time plus or minus 7.5%." With probability and frequency expressions, it is simply harder to see the confusion.

A practical solution to these problems will take two steps rather than one. Take *tall.* First, we might ask respondents one of many pairs of questions like these: (1) "What height would *tall* signify when speaking of buildings in Chicago?" and (2) "What are the heights of the typical, the tallest, and the shortest buildings in Chicago?" Second, we would look for a way of characterizing the answers of 1 in relation to the answers of 2. What we would like to get at with 2, ideally, is the distribution of assumed building heights in Chicago, but I don't see an easy way of asking people this. We might then be able to say that *tall* means, roughly, "from the 60th to the 75th percentile on the distribution of heights assumed in that situation." Hörmann's findings suggest that *a few* is amenable to this procedure. *Frequent* and its kin may be too.

SUMMARY

Mosteller and Youtz recognize that the interpretations of probability and frequency expressions change with context. Writing about Pepper and Prytulak's study, they say "The paper illustrates that context can push the meaning a good ways but that for ordinary events the differences are modest." But what is an "ordinary event"? It can't be merely an event that matches the Unknown Context condition. That would be circular. Medical and social scientists must consider many of the events they write about "ordinary" despite a wide range in expected frequency of occurrence, and they use these terms. They would be misled if they thought that estimates from the No Context condition were valid for more than a fraction of the uses.

Mosteller and Youtz's goal is the betterment of communication. They will only succeed, I suggest, when they recognize that, in real life, we interpret every expression in a particular context. And for that they must distinguish word meaning from word use.

ACKNOWLEDGMENTS

I am grateful to R. Timothy Reagan for a critical reading of a draft of this commentary. Its preparation was supported in part by Grant BNS-83-20284 from the National Science Foundation.

ADDITIONAL REFERENCES

BIERWISCH, M. (1967). Some semantic universals of German adjectivals. *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 elements 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.