The Knowledge Engineering Review, Vol. 10:1, 1995, 43–62

A review of human linguistic probability processing: General principles and empirical evidence

THOMAS S. WALLSTEN[1] AND DAVID V. BUDESCU[2]

[1]*Department of Psychology, University of North Carolina, Chapel Hill, NC 27599-3270, USA*
[2]*University of Illinois at Urbana-Champaign, USA*

Abstract

This article reviews research on how people use and understand linguistic expressions of uncertainty, with a view toward the needs of researchers and others interested in artificial intelligence systems. We discuss and present empirical results within an inductively developed theoretical framework consisting of two background assumptions and six principles describing the underlying cognitive processes.

1 Overview

This review is based on a much longer chapter (Budescu & Wallsten, 1994) prepared primarily for cognitive psychologists and behavioural decision researchers concerned with how people use and understand linguistic expressions of uncertainty. The version presented here has been condensed and written with a view toward the needs of researchers and others interested in artificial intelligence systems that mimic and interact with human experts. Readers are referred to the original chapter for details and perspectives omitted here.

The standard wisdom among decision theorists (e.g. Behn & Vaupel, 1982; Moore 1977; von Winterfeldt & Edwards, 1986) has been that numerical communication of uncertainty is better than linguistic and, therefore especially in important contexts, it is to be preferred. In contrast, expert systems researchers have recognized the multifaceted nature of uncertainty, and have explored alternative representations including linguistic (e.g. Krause & Clark, 1994). Indeed, a good deal of behavioural evidence suggests that (a) the decision theorists' advice is not uniformly correct, and (b) it is inconsistent with strongly held preferences. In this paper, we review the relevant literature to develop a theory of how humans process linguistic information about imprecise continuous quantities in the service of decision making, judgment and communication. Our theory has evolved over the years as our research has progressed, and pieces of it in one form or another can be found in our and our colleagues' publications.

The remainder of this article is organized as follows. First, we set the stage by presenting a useful three-way taxonomy of sources of vagueness. This typology leads naturally to a pair of background assumptions that underlies our review. Section 2 then summarizes the research on meanings of qualitative probability expressions. Section 3 compares judgments and decisions made on the basis of vague and precise (generally linguistic and numerical) probabilities. Each subsection consists of a brief review of established empirical regularities, followed by a theoretical statement in the form of general principles that can explain the regularities and, when available, a brief summary of research supporting the principles. More complete reviews in all cases can be found in the Budescu and Wallsten (1994) chapter. Section 4 goes beyond the Budescu and Wallsten chapter by reviewing briefly the research on how two linguistic judgments are combined, and suggesting a principle for that process. Section 5, finally, brings the background assumptions and the empirical principles together into a unified theoretical statement and provides some concluding remarks.

2 Setting the stage

2.1 Stimulus and task constraints

Any theory of how people process information about uncertain events must be constrained by stimulus and task considerations. The former refers to the type of event in question, the nature of the uncertainty about it, and the manner in which that uncertainty is represented. The latter concerns the purpose for which the information is being processed.

Zwick and Wallsten (1989) proposed, and Wallsten (1990) somewhat modified, a useful three-way stimulus taxonomy, according to the nature of the event, the uncertainty about that event, and the manner in which the uncertainty is represented. Each of the three ways is a continuum ranging from absolute precision at one end to absolute vagueness at the other. Considering events first, the distinction between those that are precise and those that are vague is straightforward. An event is precise if it is defined such that any outcome in the universe of discourse either is or is not an exemplar of it. For example, the event of noontime temperature exceeding 75°F is precise, whereas the event of a warm noontime is vague. Moreover, it is vague only to some degree, even to an individual, because sufficiently extreme temperatures are definitely outside one's concept of warm, while temperatures within those extremes vary in the degree to which they fit the designation. (Murphy & Brown, 1983, and Murphy et al., 1980, provide additional illustrations of the distinction between vague and precise events.)

The notion that uncertainty varies from vague to precise is more subtle and less easily described. Generally, uncertainty is precise if it depends upon external, quantified random variation; and it is vague if it depends upon internal sources related to lack of knowledge or to judgments about the nature of the database. But regardless of the source of the uncertainty, we say that it is precise if the judge can place the event in question within a total likelihood ordering of a sufficiently rich set of events.[1] Note that this conception of subjective uncertainty differs markedly from a common viewpoint that vague uncertainties (generally misleadingly called "ambiguous probabilities") can be represented by second-order probability distributions, i.e. precise probabilities over all possible probability distributions (see, for example, the recent review by Camerer and Weber, 1992). As Gärdenfors and Sahlin (1982) have suggested, there are psychological and epistemological differences between unspecified, or unspecifiable, probabilities and those that are given in the form of a distribution. The former are vague to some degree and the latter are precise. The distinction matters for some purposes and not for others, but to represent both types of uncertainty by second-order probability distributions is to miss the problems for which it is important.

Finally, representations of uncertainty vary from precise to vague. Probability theory provides a language for precise representation. Imprecise portrayals include numerical intervals, qualified numbers (e.g. *approximately* 0.6), and linguistic probabilities of the sort we are considering in this review. An advantage of numerical intervals is that they signal both location and degree of imprecision. Verbal expressions do not have this benefit, but they compensate by conveying greater nuances of meaning. Their rich semantic structure allows one to convey not only approximate location and degree of imprecision, but also relative weights over levels of uncertainty within an implied range, and perhaps also other aspects of the communicator's knowledge or opinions beyond degrees of uncertainty.

The three continua along which stimulus vagueness can vary, event type, uncertainty type and representation type, are distinct but not fully independent (Budescu & Wallsten, 1987). That is, the precision of an event imposes a bound on that of the uncertainty, which in turn imposes one on that of the representation. The converse, however, does not hold: the uncertainty of any event can be

---

[1] A formal statement of this idea is developed by Budescu and Wallsten (1994).

vague to any degree, as can its representation. Thus it would be perfectly natural for a weather forecaster to report his or her imprecise uncertainty about a precise event (*There is a good chance that the noontime reading will be above* 75°), but not a precise probability about a vague event (*There is a* 70% *chance that the temperature at noon will be warm*).

Tasks are not so easily taxonomized. They may involve making choices, rating or ranking alternatives, forming judgments for later use or for communication to other people, or any of a myriad of other possibilities. The manner in which the task is carried out depends upon the nature of the information and, correspondingly, the way in which the information is processed depends on the purpose for which it is being put.

2.2 Background assumptions

The constraints discussed above are incorporated in two background assumptions that underlie the subsequent theoretical principles. Both assumptions are falsifiable, but they strike us as reasonable and we are treating them without evidence as true. These assumptions have, however, testable corollaries and we will summarize the relevant results below.

- *Background assumption B*1. **Except in very special cases, all representations are vague to some degree in the minds of the originators and in the minds of the receivers.** This assumption builds the logical constraints discussed above into our theory of how humans process information about uncertain events. It implies that both one's own representation of uncertainty, and one's understanding of a representation received from another person or source, depend upon the precision of the event and on the perceived nature of the underlying database.
- *Background assumption B*2. **People use the full representation whenever feasible, but they narrow it, possibly to a single point, if the task requires them to do so.** This assumption expresses the idea that the task determines whether and how an individual resolves representational vagueness. Thus, for example, people treat separate judgments in their vague form when receiving and combining them, but they restrict attention to a narrow range of uncertainty or to a single point value when making specific decisions. Put crudely, one can have imprecise opinions but cannot take imprecise actions.

3 Meanings of qualitative probability expressions

As already mentioned, probability phrases have rich semantic structures, and it is likely that we use them to communicate more information than simply an approximate location on a [0, 1] scale (Moxey & Sanford, 1993; Teigen & Brun, 1993). Nevertheless, most behavioural studies of probability terms have focused on their numerical referents. And on this topic, there is a voluminous literature. We review this work in four subsections, each focusing on distinct issues related to communication with linguistic probabilities: intra-individual vagueness of probability terms; inter-individual variance in the understanding and use of probability phrases; sensitivity to context; and preferences for various modes of communicating degrees of uncertainty.

3.1 Intra-individual vagueness of probability terms

One important aspect of our work in this domain was to establish that a probability term within a context can be modelled by means of membership functions over numerical values within a specific context. We think of a probability phrase as a vague or fuzzy concept whose members are the numerical probabilities in the [0, 1] interval. A membership function assigns to each numerical probability a real number, which we refer to as its "membership" in the concept defined by the phrase (in that context). In principle, these membership values are ratio scaled and range from 0 for probabilities that are absolutely not included in the concept to a maximum value, arbitrarily fixed at 1, for probabilities that are ideal or perfect exemplars of the concept being considered; intermediate values represent intermediate degrees of membership. There are no special constraints on the shape, or any other property, of these functions. In particular, a membership function is not a density function. It need not be continuous and the area under it need not integrate to 1. For example, precise (crisp) terms are characterized by a membership function with only two values (1 for all those probabilities that are ideal exemplars of the concept, and 0 for all other values).

The concept of a membership function provides a useful generalization of simpler, perhaps more intuitive representations of linguistic uncertainties, such as a best probability and a range of probabilities. Presumably, the best point representation of a term characterized by a membership function is some central measure of that function. Of the various measures possible, a natural choice is the probability with the highest membership (or if the value is not unique, a summary such as the mean of all the probabilities with maximal membership). Another measure, analogous to the mean, was proposed by Yager (1981) as

$$W_e = \frac{\int_0^1 \mu_e(p) p \, dp}{\int_0^1 \mu_e(p) \, dp}, \tag{3.1}$$

where $\mu_e(p)$ represents the membership value of p in expression e. W_e is simply an average of the probabilities on the [0,1] interval, weighted by their normalized (forced to sum up unity) membership values. A discrete version of equation (3.1) is straightforward. Other location measures have been proposed in the fuzzy sets literature (see Bortolan & Degani, 1985, for a review), and an interesting empirical question is which, if any, of these measures is the best.

The range of acceptable probabilities for a term within a particular context, usually called the support of the membership function, consists of all the values with positive membership. A convenient measure of spread, which we have used occasionally, is analogous to the variance of a density. Using W_e from equation (3.1), this measure is

$$V_e^2 = \frac{\int_0^1 \mu_e(p)(p - W_e)^2 \, dp}{\int_0^1 \mu_e(p) \, dp} \tag{3.2}$$

Finally, membership functions implicitly define subsets of probabilities that are members of the concept implied by the phrase to specified degrees. This definition is accomplished by restricting subset membership to probabilities with membership values above some threshold, ν, where $0 < \nu < 1$. Subsets monotonically decrease (become narrower) as ν increases. The notion of thresholds provides a convenient quantitative way to theorize about probabilities that are "sufficiently well described by a phrase", and we will use it subsequently. Figure 1 presents hypothetical examples of some membership functions of verbal expressions.

Wallsten et al. (1986) and Rapoport et al. (1987) have shown that membership functions of the sort shown on Figure 1 can be empirically derived and validated at the individual subject level, and that the resulting scales satisfy the ordinal properties of a difference or a ratio representation (see also Norwich & Turksen, 1984). Empirical procedures for establishing the membership function of a phrase include pair-comparison and rating judgments. In the former task, the subject is shown pairs of probabilities represented as spinners[2] and asked which member of the pair better represents the phrase, and to what extent. In the latter, the subject rates the degree to which various probability values are described by the phrase. Most derived membership functions are single-peaked and a sizeable minority are monotonic, decreasing from probabilities close to 0 for low terms and increasing to probabilities close to 1 for high terms. (In many cases, these monotonic functions may actually be single-peaked, but just appear to be monotonic because we failed to include probabilities sufficiently close to the end points. See also Reagan, Mosteller & Youst's,

---

[2] In this, and many other, experiments that we review below, probabilities are represented to subjects as wheels of chance, or spinners divided into two or more radial sectors. People very accurately estimate event relative frequencies in such displays (Wallsten, 1971), and their use makes the task at hand more concrete.

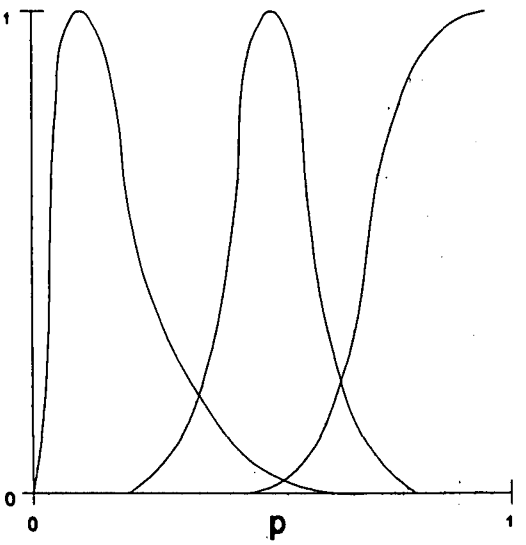

Figure 1 Generic membership functions for three phrases

1989 footnote on this topic.) Single-peaked functions may be considerably skewed in one direction or another. Most functions cover a relatively large range of values, indicating that the terms are vague to individuals. Calculations of the spread measure V^2 in equation (3.2), confirms this impression (e.g. Fillenbaum et al., 1991; Tsao & Wallsten, 1994). Only a small minority of the functions can be classified as relatively "crisp" (i.e. having a narrow support with uniformly high membership). Relatively crisp functions generally represent terms such as *tossup* or *even odds*, which tend to convey specific values.

It is important to realize that because membership functions are derived from temporally stable judgments at the level of individual subjects, they indicate that phrases are *intra-individually vague*. Other procedures, which require fewer judgments than needed to establish membership functions and therefore are simpler for respondents, also indicate that terms are vague to individuals. The simplest procedure is just to ask for a range—a lower and an upper probability—that the phrase represents. Wallsten et al. (1986) posed that question prior to eliciting membership functions. The results are summarized in Fig. 2, reproduced from their article. Median within-subject ranges were substantial, varying from approximately 0.1 for the terms *almost impossible* and *almost certain* to over 0.5 for *possible*. Hamm (1991) also asked subjects to provide lower and upper bounds for 19 terms. The median range (taken across 65 subjects) for 14 of these terms was greater than 0.10, and it was 0 only for the anchor terms *absolutely impossible*, *tossup* and *absolutely certain*.

Other data relevant to intra-individual vagueness come from studies in which subjects are asked for point numerical translations on more than one occasion. A problem with such data from our perspective is that one cannot determine the degree to which non-perfect replicability reflects error variance rather than vagueness. See Budescu and Wallsten (1994) for a review and discussion of that research.

Yet another way to assess the intra-individual vagueness of a phrase is to determine the range of probabilities for which an individual uses it. In the first stage of the experiment of Budescu, Weinberg and Wallsten (1988), subjects assigned verbal or numerical probabilities to spinners on multiple occasions. A variance measure was calculated for each term according to the displays with which it was associated. On the average this variance was larger for the verbal than the numerical judgments, and in about half of the cases the within-subject variance exceeded the between-subject component!

We have assumed (Assumption B1) that all representations are vague to some degree. A somewhat counterintuitive corollary of this statement is that the scaled meanings of numbers should also show imprecision. This prediction has been sustained by at least two studies that have

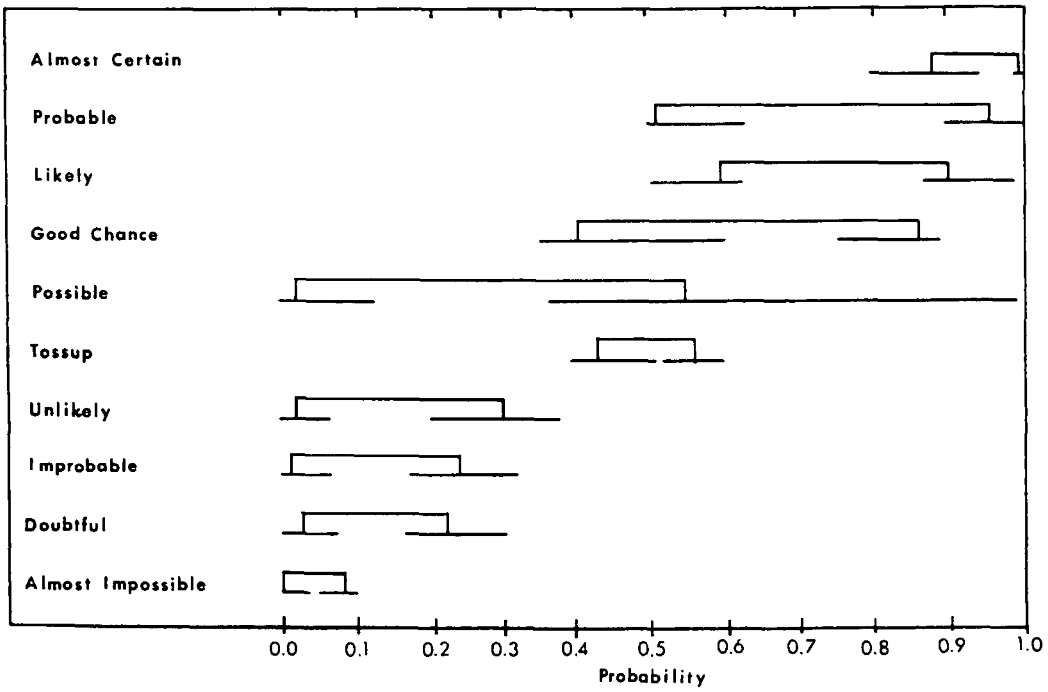

Figure 2 First, second and third quartiles over subjects of the upper and lower probability limits for each phrase in Experiment 1 of Wallsten et al. (1986). (©American Psychology Association.)

established that the meanings of numbers are both imprecise and influenced by context (Mullet & Rivet, 1991; Shapiro & Wallsten, 1994) and by a third study that found both numerical and verbal descriptors applied to a range of graphical displays, although the former to a smaller extent than the latter (Budescu et al., 1988). Despite this evidence, we assume that people consider numbers to be relatively precise when they know the values represent extensive relative frequency data rather than other individuals' judgments. To our knowledge, however, that fact has not been established.

To summarize, a vast array of data at the individual subject level indicate that meanings of phrases (and numbers) extend over probability ranges. Moreover, the data strongly suggest that a given expression represents probabilities within an interval to varying degrees, a point that we formalize below as Principle P1. After considering this principle, we turn to the questions of how these vague meanings differ over people and contexts.

3.2 *Principle P*1

Membership functions can be meaningfully scaled

According to this principle, membership functions, $\mu_e(p)$, over the [0, 1] probability interval meaningfully scale subjective representations of uncertainty, such as numerical or verbal probability expressions. The operative term in this statement is *meaningful*. As reviewed in Section 3.1, numerous procedures have been devised to establish such scales; our assumption is that they capture subjective interpretations in a manner that allows prediction of independent behaviour.

The first test of meaningfulness in this sense was reported by Wallsten et al. (1986). In that study, we scaled membership functions for individual phrases by having subjects judge the degree to which a given phrase better represented one spinner probability than another. We then used those scaled values to successfully predict judgments in which pairs of phrases were shown with single probabilities, and subjects had to indicate how much more descriptive one of the phrases was than the other.

One may consider this a weak test, as the two types of judgments are fundamentally similar. But membership functions have passed stronger hurdles, as well. In one such test, Jaffe-Katz, Budescu and Wallsten (1989) showed subjects pairs of probabilities that on each trial were either both linguistic, both numerical, or one of each type. The subjects' task was to rapidly indicate which member of the pair was the larger (or, on other trials, the smaller) value. A model incorporating empirically derived membership functions predicted the resulting complex set of reaction times 70% better than did its counterpart (Holyoak, 1978) without functions.

In another study, we (Wallsten et al., 1988) used membership functions to predict the stochastic properties of repeated choices between lotteries for equal amounts of money based on the outcomes of events described verbally or numerically. Finally, Budescu and Wallsten (1990, Experiment 2) performed an experiment in which decision makers bid for lotteries with chance events that had been described verbally and numerically by independent forecasters. The decision makers also provided judgments from which their membership functions for the forecasters' phrases could be inferred. Relative membership function values were generally consistent with the pattern of bids.

Given all the results summarized above, we feel justified in concluding that when membership values are properly scaled they provide meaningful representations of an individual's understanding of a phrase within the context that it is being used.Therefore, these membership values can be used in quantitative model testing.

3.3 Inter-individual variance in the understanding and use of probability phrases

A widely accepted generalization is that people differentially understand probability phrases (e.g. Clark, 1990). Consequently, different individuals use diverse expressions to describe identical situations and understand the same phrases differently when hearing or reading them. Data strongly support both of these statements and show that people have surprisingly rich and individualized lexicons of uncertainty.

Consider, first, the range of expressions people use in identical situations. In the Budescu et al. (1988) experiment, 20 subjects spontaneously generated 111 distinct phrases to describe 11 different spinner probabilities. Similar results in a range of other contexts and paradigms and with various types of subjects were obtained by Erev and Cohen (1990), Rapoport et al. (1990). Tsao and Wallsten (1994, Experiment 5) Wallsten et al. (1993) and Zwick and Wallsten (1989). Thus, the evidence is clear that people have different working vocabularies for expressing degrees of confidence or uncertainty and create different lexicons for themselves when given the opportunity to do so.

To investigate the flip side of the problem, that of how people understand phrases when they receive them, it is necessary to compare individuals' responses to distinct expressions. Numerous studies of phrase-to-number conversion have reported very large degrees of between-subject variability in the assessments of the same terms in a fixed context or in the absence of a specified context. The list of studies and replications is very long and covers a wide variety of situations and populations. Readers are referred to the Budescu and Wallsten (1994) chapter for a fuller presentation and discussion.

The evidence of individual differences (as summarized by Budescu & Wallsten, 1994) is overwhelming for tasks in which people are required to translate phrases into numerical equivalents or to indicate which numbers are acceptable translations of expressions. The analysis of individual membership functions carries this result a step further. We have shown (Wallsten et al., 1986; Rapoport et al., 1987; Budescu & Wallsten, 1990) that the location and spread of functions representing any given term vary considerably across subjects. Even more impressive is the fact that the shape of these functions is not universal. For example, Budescu and Wallsten (1990) report that 25% of the function describing *unlikely* are monotonically decreasing and 67% are single peaked; of the functions describing *very good chance*, 44% are single peaked and the same proportion are monotonically increasing. Thus, not only do phrases differ over individuals in their meaning, but they also differ in the extent and nature of their vague referents. To the degree that membership functions carry implications for the semantics of terms, such semantics vary considerably over individuals.

3.4 Intra-individual sensitivity to context

One obvious solution to the potential communication problems raised by intra-individual vagueness and inter-individual variability in understanding probability phrases is to standardize the language. That is, develop a verbal scale by identifying a reasonably small subset (7–13 members) of frequently used terms, impose a ranking and associate a range of probabilities with each of the terms in the list. It is a little known fact that the US National Weather Service (NWS) has done just that with respect to probability of precipitation (POP) forecasts (National Weather Service, 1984, Chapter C-11). In issuing POP forecasts, NWS meteorologists can translate 0.10 or 0.20 only to *slight chance*; 0.30, 0.40 and 0.50 only to *chance*, and 0.60 and 0.70 only to *likely*. Other terms are not allowed with regard to POP forecasts.

The general argument for a standardized probability language is best articulated by Mosteller and Youtz (1990). Comments by other researchers favouring and opposing their proposals plus their rejoinder immediately follow their article, and we refer readers to the entire discussion. Too many scales have been proposed to mention all of them here, but representative examples include Beyth-Marom (1982), who grouped 19 terms into seven categorical ranges; Hamm (1991), who suggested a single best term for each of the 19 intervals of width 0.05 in the [0.05–0.95] range; and Kadane (1990), who proposed 11 terms to label intervals. And, of course, there exists the infamous scale used by NASA engineers, which was linked (Marshall, 1986, 1988) to the space shuttle accident.

In a similar spirit, artificial intelligence researchers who need to incorporate uncertainty in expert systems have suggested that a selected subset of terms be represented by a family of partially overlapping membership functions. For example, Bonissone and Decker (1986) suggested using nine terms from Beyth-Marom's (1982) list and describing each by a trapezoidal function; Degani and Bortolan (1988) proposed triangular membership functions for 13 terms; and Lopez de Mantras et al. (1988) described a medical diagnostic expert system (MILORD) in which uncertainty is captured by nine terms represented by trapezoidal functions.

A standardized scale is feasible if: (1) People can suspend, or suppress, the meanings they normally associate with such terms; and (2) Meanings and representations of selected terms are invariant over all contexts in which they are applied. The former condition has not been studied extensively. In fact, the only study that has addressed this issue was reported by Wallsten et al. (1986, Experiment 1). They showed that weather forecasters were subject to base rate effects[3] in a domain outside of their expertise with the very phrase that had been endowed with standardized meaning in the context of POP forecasts. The effects were substantial. For example, *chance* was interpreted on average as indicating a probability of 0.39 when referring to the likelihood that an ankle twisted in a soccer game was sprained (a very common event) rather than broken, but as 0.18 when referring to the likelihood that severe life-threatening effects would accompany a flu shot (a rare event). This strongly suggests that meanings cannot be legislated!

The second condition is the primary subject of this section. As it turns out, most empirical results obtained to date show that this condition is systematically violated in interesting ways. Just as with terms of frequency (e.g. Pepper, 1981) and of quantity (e.g. Newstead, 1988), probability phrases tend to change their meanings according to the context in which they are used.

The first result of note is simply that **context matters**, probably because different interpretations of the same context yield different interpretations of the phrases. Beyth-Marom (1982), using political forecasters, and Brun and Teigen (1988) using college students, showed that for most

---

[3] We use the term *base rate effect* to refer to the effect of perceived *a priori* event chances on the interpretation of probability expressions.

expressions, inter-individual variance in translating the expressions to numerical values is greater within a specific context than in the absence of one. In a related study with physicians, Mapes (1979) showed that distributions of numerical values assigned to phrases describing the chances of side effects varied over different medications. See Budescu and Wallsten (1994) for more details about these studies.

Context, of course, is not a well defined unidimensional concept. To reach more specific generalizations, we review below a few situational variables that have been shown to systematically affect probability phrase meanings. We first consider perceived base rates. Following Pepper's (1981; see also Pepper & Prytulak, 1974) lead with frequency expressions, Wallsten et al. (1986) showed that the numerical interpretation of a term used to describe the chances of a given event occurring depends upon whether the context implies a high or low base rate. For example, the numerical translation of **probable** when referring to the likelihood of snow in the North Carolina mountains is higher when the statement specifies the month of December than of October. Similar results were obtained by Shapiro and Wallsten (1994) and by Weber and Hilton (1990).

A few studies have looked at phrase meaning directly as a function of the **number of possible alternatives in the sample space**, n.[4] Tsao and Wallsten (1994, Experiments 1–4) found that phrase membership functions, especially for low and medium phrases depended systematically on whether $n = 2$ or 4. Teigen (1988a) obtained parallel results when he examined how the number and relative likelihoods of alternatives affect people's understanding and selection of (Norwegian) probability phrases in real-world contexts. His data also suggested subtle semantic effects not understood simply in terms of implied probabilities, as did related results by Gonzales and Frenck-Mestre (1993).

Weber and Hilton (1990) investigated **outcome severity** in the context of medical scenarios. They found that on average the phrases were interpreted to imply greater probabilities when associated with events of more severe consequence. The authors correctly pointed out that event severity and prevalence are negatively correlated in medical contexts (serious illnesses are generally less common than mild ones) and therefore their separate effects on meanings of probability terms are hard to disentangle. Contrasting results were obtained by Sutherland et al. (1991), who found that cancer patients responded similarly to verbal descriptions of different side effects of blood transfusion (death vs. illness).

Closely related to the notion of outcome severity is that of **outcome valence**, whether the outcome is positively or negatively valued. Mullet and Rivet (1991) compared scale values assigned to 24 French expressions used in predictions of children's chances of passing or failing a test. On average, the positive context induced higher estimates for a given phrase. A similar pattern was found by Cohen and Wallsten (1992) using lotteries.

Yet another context effect relates to **characteristics of the available uncertainty vocabulary** (such as its length, composition and structure) when judging particular expressions. Hamm (1991) found less inter-individual variance and better discrimination among the meanings of 19 phrases when they were presented in ordered rather than random lists. Clarke et al. (1992) obtained similar results. Fillenbaum et al. (1991) found that membership functions for a list of core expressions (e.g. *likely*) were unaffected by the presence or absence of modified expressions (e.g. *very likely*) or anchor terms (e.g. *almost certain*). The additional terms did decrease the degree to which the core words were judged to be most appropriate for certain events.

Finally, a particularly important factor in determining phrase meaning concerns one's **role in a dialogue or exchange**. The recipient of a communication may understand a probability expression differently than intended by the originator, because the two individuals differentially interpret the base rates, valences or severities of outcomes, or because meaning depends more generally upon the direction of communication. We are not aware of research that has compared recipients' and

---

[4] For example, a spinner may be divided into two or more sectors, an election may have two or more candidates, or a set of symptoms may be associated with two or more disease processes.

originators' perceptions of context, but two studies have looked at the overall effects of communication direction. Using very different methodologies, Budescu and Wallsten (1990) and Fillenbaum et al. (1991) showed that recipients of the verbal forecasts generally assigned values closer to 0.5 than originally intended by the communicators. Fillenbaum et al. (1991) also showed that membership functions were broader (in terms of V^2) for phrases selected by others than selected by oneself for communication.

To summarize, context effects on the interpretation of probability terms are pervasive. Individual differences in assigning numerical values to expressions are greater when the expressions are used in discourse than when they appear alone. More specifically, perceived base rate, number of alternatives, valence and severity of outcomes, and direction of communication all have been shown to have large and systematic effects on phrase meaning. To a lesser degree, meanings may also depend upon the available vocabulary and on the order in which phrases are seen.

3.5 Principle P2

The location, spread and shape of membership functions vary over individuals and depend upon context and communication direction.

This principle implies that the individual differences summarized in Section 3.2 and systematic effects on phrase meaning described in Section 3.3 can be represented by membership function location, spread and shape. Individual differences in membership functions are already well documented and were discussed in section 3.3. But the representation of context and communication effects on such terms has not ben tested directly.

3.6 Communication mode preferences

Do people prefer verbal or numerical communications, and to what degree do any such preferences depend upon the properties of language that we have documented above? A corollary of B1, that all representations are vague, is that people generally prefer to communicate their opinions verbally because this mode conveys the underlying vagueness. The general folklore is consistent with this expectation and so are empirical results. Erev and Cohen (1990) found that basketball experts and students preferred to give their opinions of the likelihood of future basketball events in verbal rather than numerical terms. Brun and Teigen (1988) obtained similar results with physicians. These findings were replicated in a survey of 442 undergraduate, graduate nursing and MBA students by Wallsten et al. (1993), of which 77% overall thought most people prefer communicating degrees of uncertainty verbally rather than numerically, and 65% indicated that as their personal preference. An analysis of reasons given for the preferences suggested that verbal communications generally are preferred unless the underlying opinions are based on solid evidence, and therefore relatively precise, or unless the importance of the occasion warrants the effort to attempt greater precision.

Assumption B1 provides less guidance regarding what to expect about people's preferences for receiving judgments of uncertainty. But the data from decision makers in the Erev and Cohen (1990) study and from patients in the Brun and Teigen (1988) study are clear: most people prefer receiving such judgments numerically. In the Wallsten et al. (1993) survey, 70% of the respondents expressed an initial personal preference for receiving information numerically. The same reasons were given as for communicating to others. Those who preferred verbal did so because it is more natural, more personal and easier; and those who preferred numerical did so because it is more precise.

The fact that most people have initial preferences for communicating to others verbally and receiving from others numerically means that there must be some people who have both preferences simultaneously. The Wallsten et al. survey found this pattern for 35% of their respondents, and Erev and Cohen found it for 47% of the decision makers. The reverse preference pattern, to communicate numerically to and receive verbally from others, virtually never appeared.

3.7 Principle P3

Communication mode choices are sensitive to the degrees of vagueness inherent in the events being described, the source of the uncertainty, and the nature of the communication task.

This principle is really a corollary of B2, but it summarizes the main results of Section 3.6. It captures the idea that most people prefer the communication mode that they can use most easily and accurately for the task at hand. Thus, they communicate their options to others in a verbal rather than numerical form, because it is the easiest way to convey different locations and nuances of imprecision. On the other hand, even recognizing that opinions are not precise, people prefer to receive information numerically because (we assume) they know that individuals differ greatly in their phrase selection. Therefore, people believe they can more easily and accurately obtain a fix on the rough uncertainty location another individual intends to communicate when he or she does so numerically rather than verbally.

One implication of this principle was developed and tested by Erev, Wallsten and Neal (1991). They suggested that in many contexts vague communications promote the well being of a group, or society, to a greater extent than do precise ones, and that people will select their communication modes accordingly. Vague language is beneficial because it is more likely to result in heterogeneous rather than homogeneous actions, which in turn increase the chances that at least some individuals will be successful and the group as a whole will benefit. Their experiment supported this idea.

4 Judgment and decision with verbal and numerical probabilities

The picture that emerges from the previous section is that linguistic terms are imprecise and subject to multiple interpretations by different individuals and under different conditions. Therefore, they are problematic in communication and should be avoided for important decisions. We (Budescu & Wallsten, 1985) have referred to the possible "illusion of communication" induced by the use of these terms, and decision analysts routinely have recommended avoiding them (e.g. Behn & Vaupel, 1982; Moore, 1977; von Winterfeldt & Edwards, 1986). In this section we review empirical results comparing judgmental accuracy and decision quality given probabilistic information in verbal and numerical form. Most of these studies involve within-subject comparisons of judgments and decisions performed with the two modes of information. The next subsection focuses on studies of probabilistic judgments and inferences, and the following subsection is concerned with experiments in which subjects made choices, gave bids, or in some way took action on the basis of verbal or numerical information.

4.1 Accuracy of judgments

According to B1 all representations are vague to some degree. On that basis, contrary to intuition, verbal judgments are not necessarily less accurate than their numerical counterparts. Indeed, Zimmer (1983,1984) argued persuasively that they should be more accurate. Beginning with the premise that the mathematics of uncertainty (probability theory) developed only in the seventeenth century, whereas the language of uncertainty is ubiquitous and much older, Zimmer (1983) wrote, "It seems unlikely that the mathematically appropriate procedures with numerical estimates of uncertainty have become automatized since then. It is more likely that people handle uncertainty by customary verbal expressions and the implicit and explicit rules of conversation connected with them." (p.161) He continued the argument in another article by writing, ". . . it seems plausible to assume that the usual way humans process information for predictions is similar to putting forward arguments and not to computing parameters. Therefore, if one forces people to give numerical estimates, one forces them to operated in a 'mode' which requires 'more mental effort' and is therefore more prone to interference with biasing tendencies." (Zimmer, 1984, p. 123) To evaluate these claims, Zimmer (1983,1984) ran experiments in which he tested for biases with verbal responses that typically are found with numerical ones (overconfidence in judgment and underconfidence in revision of opinion). The data appear to substantiate his prediction of greater accuracy in the verbal mode, but as Clark (1990) pointed out, both studies were problematic in certain ways. We tested Zimmer's thesis allowing subjects a much richer vocabulary and comparing the results with closely matched numerical controls.

In one study, Zimmer (1983,1984) had subjects use verbal responses to estimate posterior probabilities in a simple two-alternative Bayesian revision of opinion paradigm.[5] The resulting judgments were more accurate (i.e. closer to the predicted Bayesian response) than generally found in studies using numerical responses (e.g. Edwards, 1968). Rapoport et al. (1990) replicated the study with certain crucial differences. A within-subject design was used, in which each subject provided verbal and numerical judgments in different sessions and a payoff scheme was used to motivate careful responding. Subjects constructed their own vocabularies, and individual membership functions were established for the 14–16 phrases each subject used most frequently. The results were complicated, but indicated near equivalence in the quality of verbal and numerical judgments. See, also, related studies by Hamm (1991) and Timmermans (1994). Thus, our data do not support Zimmer's strong claim that reasoning is better in the revision of opinion context when people can respond verbally rather than numerically, neither do they show the opposite to be true. Rather, it appears that reasoning is about equivalent in these tasks given the two modes of responding.

The same seems to be true in the case of judgment, where the usual finding with numerical responses is that people are overconfident. Zimmer (1983) reported results of a study in which soldiers answered political science questions and assessed their chances of being correct by means of verbal expressions. He interpreted the data as suggesting that subjects are less overconfident and more accurate when their judgments are verbal than when they are numerical. We recently replicated and extended this study and arrived at somewhat different conclusions. Wallsten et al. (1993) asked subjects to report their confidence in the truth of factual items (in the domains of geography, demography and history), consisting of equal numbers of true statements and their semantically identical false complements. Subjects selected their own vocabularies and subsequently encoded membership functions for each phrase. They provided verbal and numerical judgments in separate sessions. The only differences we observed between the two modes were that the central numerical response was used more frequently than its verbal counterpart and that the level of overconfidence (excluding the middle category) was higher for the verbal terms. A possible interpretation of these differences is that judges may use the central category more frequently in the numerical than the verbal mode to represent imprecise judgments, because that category is equally defensible regardless of the outcome. In contrast, the verbal mode allows more honest representations of vaguely formed opinions without resort to the central category. If so, then overconfidence, as usually measured, may actually be greater than observed when subjects respond numerically.

Two points must be added in summarizing this evidence and its relation to B1. First, recent developments by Erev et al. (1994) suggest that underconfidence common in opinion revision studies and overconfidence common in judgment studies may both arise from one set of errorful processes leading to data that researchers analyze differently in the two paradigms. The argument is complex and we refer readers to the Erev et al. (1994) article or to the summary by Budescu and Wallsten (1994). The second point is that although no systematic evidence has accrued thus far, indicating that one mode of responding is more accurate than another, many avenues remain to be explored. For example, Zimmer (1983) suggested that the type of information subjects think about

---

[5] This is a task in which subjects observe data drawn from one of two (or more) sources with known prior probabilities. They know the conditional likelihoods of the data under each source and must estimate the posterior probabilities over the sources.

in making forecasts or predictions depends upon the mode in which they must respond. He claims that people focus more strongly on qualitative data when they can respond verbally and on quantitative data when they can respond numerically. This is an intriguing idea that is consistent with the "compatabily effect" (Tversky et al., 1988), and which deserves follow-up.

Nevertheless, the available data provides no indication that one mode of representation systematically leads to more accurate judgments than does the other. This conclusion is not implied by assumption B1, but it is eminently consistent with it and no further summary principle is required.

4.2 Choice and decision quality

Knight (1921) and Keynes (1921) distinguished between various types of uncertainty, but it was Ellsberg's (1961) famous paradox that drew decision theorists' attention to the effects of vagueness (or ambiguity as it is often, but incorrectly, described) on preference (see Budescu & Wallsten, 1987, and Camerer & Weber, 1992 for reviews). Ellsberg's and subsequent results (e.g. Curley & Yates, 1985; Einhorn & Hogarth, 1985) suggest that, everything else being equal, most people prefer precise probabilities over vague (ambiguous) representations of uncertainty. Since phrases are relatively more vague than numbers, one would expect choices and overt decisions (as operationalized by bids, attractiveness ratings, or rankings or risky options) to reflect this preference. On the other hand, Zimmer's argument (1983,1984) regarding the superiority of the verbal mode makes just the opposite predictions. Next we describe a series of studies in which these two conflicting predictions were tested empirically.

Budescu et al. (1988) compared the quality of decisions based on verbal and numerical probabilities. Subjects in one experiment bid for (and in another, rated the attractiveness of) lotteries for gains or for losses, in which the probabilities were represented either numerically, graphically or verbally (using terms the subjects had previously supplied). On average, subjects won 1.2% more and lost 4.7% less with numerical probabilities; but the bids, attractiveness ratings and decision times (after eliminating the possibility of calculations) were almost identical under the three presentation modes. Budescu and Wallsten (1990) replicated the study with dyads, in which forecasters saw probability spinners and communicated the probabilities of the target events in either numerical or verbal form to decision makers who then bid for gambles. The results showed identical mean bids, and expected gains, under the two modes of communication.

Erev and Cohen (1990) used a more realistic version of this dyadic paradigm: sports writers and sportscasters provided numerical and verbal forecasts for events (e.g. one player will score more points than another) in a randomly chosen basketball game from an identified group of games. Students subsequently ranked gambles under a payoff scheme, knowing only the experts' forecasts (the actual events were disguised) and the monetary outcomes. Subjects' rankings and expected payoffs were well above chance level, but did not vary as a function of the mode of communication. Similar results were obtained by González-Vallejo et al. (1994) under more controlled conditions. The single exception to this line of results is Experiment 5 of Tsao and Wallsten (1994). Their study contrasted numerical and verbal communications from a forecaster to a decision maker when the number of possible outcomes was $n = 2$ or $n = 4$. The decision makers' task was to estimate (under a payoff scheme designed to promote accuracy) how many times out of 100 the event would occur. Estimates based on verbal and numerical judgments were equally accurate when $n = 2$, but not when $n = 4$. In the latter case, quality suffered in the verbal mode. This task differed from the other decision tasks in that it required the estimate of a sample statistic, rather than a choice or a bid for a gamble. Nevertheless, it suggests that more careful work is needed in contexts where $n > 2$.

To summarize, with the single exception just noted, all the experiments described above agree that, on the average, decision quality is unaffected by the mode in which the probability information is provided. To say the least, the result is puzzling in light of the material discussed in Section 3.2. In part, it may be due to the fact that judgmental accuracy is roughly equivalent given verbal and numerical information, as described in Section 4.1. But that explanation alone will not do, as it still does not explain the approximately equal decision quality given the wide interpersonal variability in interpreting phrases and the considerably greater imprecision of verbal than numerical expressions. However, the next principle provides the link that in conjunction with the findings of Section 4.1 brings together the two otherwise contradictory sets of results.

4.3 Principle P4

The question we must address is, how are imprecise assessments of uncertainty resolved for the purpose of taking action? (Assumption B2 asserts that such resolution takes place, but does not specify the mechanism.) We propose that: **when combining, comparing, or trading-off information about uncertainty with information about other dimensions, such as outcome values, the uncertainty representation, $\mu_e(p)$, is converted from a vague interval to a point value by restricting attention only to values of p with membership above a threshold, v, i.e. for which $\mu_e(p) \geq v$. A specific value, p^*, is then selected probabilistically according to a weighting function proportional to the $\mu_e(p) \geq v$.**

The principle was first proposed in the form of a mathematical model and empirically bolstered by Wallsten et al. (1988). Readers are referred to the original article or to Budescu and Wallsten's (1994) chapter for a statement of the model and a summary of the supporting data. Principle P4 provides a qualitative explanation of Budescu and Wallsten's (1990) dyadic decision results, which were discussed above, but has not been further tested. Clearly, additional evaluation is required.

Nevertheless, assuming the principle's validity, we are in a position to understand why average decision quality is unaffected by probability mode. According to the model (not shown here) when a phrase is converted to a probability value, p^*, for decision purposes, its expected value, $E_c(p^*)$, is equivalent to W_e defined for $\mu_c(p)$ restricted to values greater than or equal to v. Moreover, for symmetric single-peaked functions, W_e equals the peak probability, i.e. the value p^{**} such that $\mu_c(p^{**}) = 1$. However, regardless of the shape of the membership function, as v increases $E_c(p)$ approaches p^{**} (or the mean of the p^{**}, should the value not be a single point). Therefore, we can claim that in general, when making decisions, people interpret phrases as equivalent to probabilities in the neighbourhood ranging from their central value, W, to their peak value(s), p^{**}.

Two more steps are needed to complete the explanation of why average decision quality is unaffected by probability mode. Recall, first, that people treat numbers received from others as vague at least to some degree. Therefore, we can assume that just as with phrases, decision makers interpret numbers somewhat more broadly and centrally than the forecaster intended. Second, recall from Section 4.1 that verbal and numerical information is processed with roughly equivalent accuracy. On that basis, we can deduce that for a given decision maker, there is no systematic difference between W_e or p^{**} of the phrase selected by a forecaster and the interpreted meaning of the numerical probability communicated by that person. Given this entire train of argument, it is not surprising that average decision quality was unaffected by probability mode in the experiments reviewed in Section 4.2, while simultaneously decision variance was somewhat greater when communication was verbal than when it was numerical.

4.4 Principle P5

Decision patterns differ in more than just variance given verbal and numerical information, although the additional difference is sufficiently subtle that it alluded us for some time. We express this result as the following principle: **when combining, comparing, or trading-off information across dimensions, the relative weight accorded to a dimension is positively related to its precision.** In other words, the narrower is $\mu_c(p)$, the more weight is accorded to p^*.

The finding leading to this principle was first evident in the data of a pilot study run by González-Vallejo et al. (1994). The effect was replicated in a second, much more substantial experiment. In both cases, decision makers used forecasters' verbal or numerical probability judgments to rank gambles within sets constructed such that the outcomes and probabilities were negatively correlated. The gambles' expected values (EV) agreed in rank order with the outcomes for half the sets and with the probabilities for the other half. When the probabilities were expressed verbally, subjects' rankings correlated positively with the payoffs. In contrast, when the probabilities were expressed numerically, the subjects' ranking were positively related to the probabilities. Thus, *on average* subjects did equally well under either mode, but not given a particular stimulus structure. Further support for P5 is provided by González-Vallejo and Wallsten (1992) in a study on preference reversals given verbally and numerically described probabilities.

5 Combining linguistic probabilities

Generally when we seek expert or lay opinions relevant to decisions that we will make, we obtain advice from more than a single person. Missing in the developments above, however, is any discussion of how individuals combine multiple linguistic inputs into a single, generally imprecise judgment prior to taking action. Wallsten et al. (1994) addressed this issue and proposed a sixth principle to complement the five developed in the Budescu and Wallsten (1994) chapter. We summarize that material here.

5.1 Imprecise judgments

On the assumption that two linguistic probabilities are combined into one judgment by the logical operation, *and*, fuzzy set theory (see for example, Smithson, 1984,1987) provides a family of ready models for describing the process. Members of the family differ in their mathematical combination rules for *and* (and *or*), but all agree that they are non-compensatory. In contrast, and based on considerable behavioral research regarding information integration (see for instance, Anderson, 1991), one might expect that judgment resulting from two linguistic inputs is best represented by their average, where many forms of averaging are possible. Budescu et al. (1990) and Zwick et al. (1988) contrasted these two sets of models in two experiments involving membership functions for phrases and their combinations. The results unambiguously ruled out all fuzzy set models and supported the averaging representations. Unfortunately, however, no single type of averaging proved superior over subjects; rather, individual differences were large and systematic.

5.2 Principle P6

On the basis of these results, Wallsten et al. (1994) proposed that: **when combining separate components of information about a single dimension, the resulting judgment reflects some sort of average of their values**. Note that this principle is consistent with the background assumption, B2, that whenever feasible people use the full representation of the available information. P6 was established as a summary of membership function judgment data, but was intended to predict actual decision behaviour. Wallsten et al. conducted such a test and the principle held up well, but further research is needed here.

6 Recapitulation and conclusions

Fore ease of reference, it is useful to repeat here the two background assumptions and six principles. We then consider the epistemological status of each and their joint implications:

B1. Except in very special cases, all representations are vague to some degree in the minds of the originators and in the minds of the receivers.
B2. People use the full representation whenever feasible, but they narrow it, possibly to a single point, if the task requires them to do so.
P1. Membership functions can be meaningfully scaled.

P2. The location, spread, and shape of membership functions vary over individuals and depend upon context and communication direction.

P3. Communication mode choices are sensitive to the degrees of vagueness inherent in the events being described, the sources of the uncertainty, and the nature of the communication task.

P4. When combining, comparing or trading-off information about uncertainty with information about other dimensions, such as outcome values, the uncertainty representation, $\mu_e(p)$, is converted from a vague interval to a point value by restricting attention only to values of p with membership above a threshold, v, i.e. for which $\mu_e(p) \geq v$. A specific point value, p^*, is then selected probabilistically according to a weighting function proportional to the $\mu_e(p) \geq v$.

P5. When combining, comparing or trading-off information across dimensions, the relative weight accorded to a dimension is positively related to its precision.

P6. When combining separate components of information about a single dimension, the resulting judgment reflects some sort of average of their values.

Our story is straightforward: rarely is one's opinion or judgment precise (B1), although one acts upon precise values when action is called for (B2). The nature and extent of one's vague representation depends upon various individual, situational and contextual factors (P2), and is measurable in a meaningful fashion (P1). One converts vague opinion to a point value for purposes of action by a probabilistic process that is sensitive to the form of the opinion and to the task (P4). However, despite this conversion, the degree of attention one accords to a dimension depends upon its underlying vagueness (P5). Somewhat outside this stream that travels from opinion to action is the issue of communication mode preferences, which are systematically affected by degree of vagueness (P3). Also important, but not formulated as a principle because it does not have theoretical content, is the fact that judgment is approximately equally accurate given verbal and numerical expressions of probability. Finally, multiple inputs typically lead to judgments that are the average of the separate components (P6).

These assumptions and principles are not epistemologically equivalent. We have simply asserted B1, but numerous corollaries that can be inferred from it are empirically supported. Similarly, there are no data to sustain assumption B2 in its full generality, but particular instantiations of it, principles P4 and P6, are sustained. Principles P1 and P5 are well buttressed by data; whereas in contrast, principle P3 describes a range of results, but does not have independent *post hoc* support. Finally, principle P2 is at this point a reasonable conjecture that is consistent with a good deal of related evidence, and P6 requires more extensive empirical evaluation. Nevertheless, taken together, the principles and background assumptions provide a coherent theory that may be useful to the designers of expert systems as they seek ways to represent various types of uncertainty in forms most natural and comprehensible to human operators.

Acknowledgements

The order of authors is arbitrary; both contributed equally. Preparation of this review was facilitated by an Arnold O. Beckman Research Award from the Research Board of the University of Illinois and by NSF Grant SBR-9222159. We wish to acknowledge the contribution of many students and colleagues who have been involved in various stages and aspects of the work summarized in this article. In particular, we thank Brent Cohen, Jim Cox, Ido Erev, Sam Fillenbaum, Claudia González-Vallejo, Amnon Rapoport, Chen-Jung Tsao, Andrea Shapiro, and Rami Zwick for their help in shaping our thinking about these issues over the years.

References

Anderson, NH, 1991. *Information Integration Theory*, Erlbaum.

Bass, BM, Cascio, WF and O'Connor, EJ. 1974. "Magnitude estimation of expressions of frequency and amount" *Journal of Applied Psychology* **59** 313–320.

Behn, RD and Vaupel JW, 1982. *Quick Analysis for Busy Decision Makers*, Basic.

Beyth-Marom, R, 1982. "How probable is probable? A numerical translation of verbal probability expressions" *Journal of Forecasting* **1** 257–269.

Bonissone, PP and Decker, KS, 1987. "Selecting uncertainty calculi and granularity: An experiment in trading off precision and complexity" In: LN Kanal and JF Lemmer (eds.), *Uncertainty and Artificial Intelligence*, Elsevier.

Bortolan, G and Degani, R, 1985. "A review of some methods for ranking fuzzy subsets" *Fuzzy Sets and Systems* **15** 1–19.

Brackner, JW, 1985. "How to report contingent losses in financial statements?" *Journal of Business Forecasting* **4** 13–18.

Bradburn, NM and Miles, C, 1979. "Vague quantifiers" *Public Opinion Quarterly* **43** 92–101.

Brun, W and Teigen, KH, 1988. "Verbal probabilities: Ambiguous, context-dependent, or both?" *Organizational Behavior and Human Decision Processes* **41** 390–404.

Bryant and Norman, 1980. "Expressions of probability: Words and Numbers" *New England Journal of Medicine* **302** 411.

Budescu, DV and Wallsten, TS, 1985. "Consistency in interpretation of probabilistic phrases" *Organizational Behavior and Human Decision Processes* **36** 391–405.

Budescu, DV and Wallsten, TS 1987. "Subjective estimation of precise and vague uncertainties" In: G Wright and P. Ayton (eds.), *Judgmental Forecasting*, pp 63–81, Wiley.

Budescu, DV and Wallsten, TS, 1990. "Dyadic decisions with verbal and numerical probabilities" *Organizational Behavior and Human Decision Processes* **46** 240–263.

Budescu, DV and Wallsten, TS (in press). "Processing linguistic probabilities: general principles and empirical evidence" In: JR Busemeyer, R Hastie and DL Medin (eds.), *Decision Making from the Perspective of Cognitive Psychology*, Academic Press.

Budescu , DV, Weinberg, S and Wallsten, TS, 1988. "Decisions based on numerically and verbally expressed uncertainties" *Journal of Experimental Psychology: Human Perception and Performance* **14** 281–294.

Budescu, DV, Zwick, R, Wallsten, TS and Erev, I, 1990. "Integration of linguistic probabilities" *International Journal of Man–Machine Studies* **33** 657–676.

Camerer, C and Weber, M, 1992. "Recent developments in modeling preferences: Uncertainty and ambiguity" *Journal of Risk and Uncertainty* **5** 325–370.

Chesley, GR, 1985. "Interpretation of uncertainty expressions" *Contemporary Accounting Research* **2** 179–199.

Clark, DA, 1990. "Verbal uncertainty expressions: A review of two decades of research" *Current Psychology: Research and Reviews* **9** 203–235.

Clarke, VA, Ruffin, CL, Hill, DJ and Beamen, AL, 1992. "Ratings of orally presented verbal expressions of probability by a heterogeneous sample" *Journal of Applied Social Psychology* **22** 638–656.

Cliff, N, 1959. "Adverbs as multipliers" *Psychological Review* **66** 27–44.

Cohen, BL and Wallsten, TS, 1992. "The effect of constant outcome value on judgments and decision making given linguistic probabilities" *Journal of Behavioral Decision Making* **5** 53–72.

Curley, SP and Yates, JF, 1985. "The center and range of the probability interval as factors affecting ambiguity preferences" *Organizational Behavior and Human Decision Processes* **36** 273–287.

Degani, R and Bortolan, G, 1988. "The problem of linguistic approximation in clinical decision making" *International Journal of Approximate Reasoning* **2** 143–162.

DeGroot, MH, 1970. *Optimal Statistical Decisions*, McGraw-Hill.

Einhorn, HJ and Hogarth, RM, 1985. "Ambiguity and uncertainty in probabilistic inference" *Psychological Review* **92** 433–461.

Ellsberg, D, 1961. "Risk, ambiguity, and the Savage axioms" *Quarterly Journal of Economics* **75** 643–669.

Erev, I and Cohen, BL, 1990. "Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox" *Organizational Behavior and Human Decision Processes* **45** 1–18.

Erev, I, Wallsten, TS and Budescu, DV, 1994. "Simultaneous over- and underconfidence: The role of error in judgment processes" *Psychological Review* **101** 519–527.

Erev, I, Wallsten, TS and Neal, M, 1991. "Vagueness, ambiguity, and the cost of mutual understanding" *Psychological Science* **2** 321–324.

Farkas, A and Makai-Csasar, M, 1988. "Communication or dialogue of deafs: Pitfalls of use of fuzzy quantifiers" In: *Second Network Seminar of the International Union of Psychological Science*, North-Holland.

Fillenbaum, S, Wallsten, TS, Cohen, B and Cox, JA, 1991. "Some effects of vocabulary and communication task on the understanding and use of vague probability expressions" *American Journal of Psychology* **104** 35–60.

Gärdenfors, P and Sahlin, NE, 1982. "Unreliable probabilities, risk taking, and decision making" *Synthese* **53** 361–386.

González-Vallejo, CC, Erev, I and Wallsten, TS, 1994. "Do decision quality and preference order depend on whether probabilities are verbal or numerical?" *American Journal of Psychology* **107** 157–172.

González-Vallejo, CC and Wallsten, TS, 1992. "Effects of probability mode on preference reversal" *Journal of Experimental Psychology: Learning, Memory, & Cognition* **18** 855–864.

Grether, DM and Plott, CR, 1979. "Economic theory of choice and the preference reversal phenomenon" *American Economic Review* **69** 623–638.

Hamm, RM, 1991. "Selection of verbal probabilities: A solution for some problems of verbal probability expression" *Organizational Behavior and Human Decision Processes* **48** 193–223.

Hammerton, M, 1976. "How much is a large part?" *Applied Ergonomics* **7** 10–12.

Holyoak, KJ, 1978. "Comparative judgments with numerical reference points" *Cognitive Psychology* **10** 203–243.

Holyoak, KJ and Glass, AL, 1978. "Recognition confusions among quantifiers" *Journal of Verbal Learning and Verbal Behavior* **17** 249–264.

Jaffe-Katz, A, Budescu, DV and Wallsten, TS, 1989. "Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty" *Memory and Cognition* **17** 249–264.

Johnson, EM, 1973. "Numerical encoding of qualitative expressions of uncertainty" (Tech. paper No. 250), U.S. Army Research Institute for the Behavioral and Social Sciences.

Johnson, EM and Huber, GP, 1977. "The technology of utility assessment" *IEEE Translations on Systems, Man and Cybernetics* **7** 311–325.

Kadane, JB, 1990. "Comment: Codifying chance" *Statistical Science* **5** 18–20.

Keynes, JM, 1921. *A Treatise on Probability*, Macmillan.

Knight, FH, 1921. *Risk, Uncertainty, and Profit*, University of Chicago Press.

Kong, A, Barnett, GO, Mosteller, F and Youtz, C, 1986. "How medical professionals evaluate expressions of probability" *The New England Journal of Medicine* **315** 740–744.

Lichtenstein, S, Fischhoff, B and Phillips, LD, 1982. "Calibration of probabilities: The state of the art to 1980" In: D Kahneman, P Slovic and A Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases*, pp 306–334, Cambridge University Press.

Lichtenstein, S and Newman, JR, 1967. "Empirical scaling of common verbal phrases associated with numerical probabilities" *Psychonomic Science* **9** 563–564.

López de Mántaras, R, Meseguer, P, Sanz, F, Sierra, C and Verdaguer, A, 1988. "A fuzzy logic approach to the management of linguistically expressed uncertainty" *IEEE Transactions on Systems, Man and Cybernetics* 144–151.

Krause, P and Clark, D, 1993. *Representing Uncertain Knowledge: An artificial intelligence approach*, Intellect.

Mapes, 1979. "Verbal and numerical estimates of probability terms" *Journal of General Internal Medicine* **6** 237.

Marshall, E, 1986. "Feynman issues his own shuttle report, attacking NASA's risk estimates" *Science* **232** 1596.

Marshall, E, 1988. "Academy panel faults NASA's safety analysis" *Science* **239** 1233.

Merz, JF, Druzdzel, MJ and Mazur, DJ, 1991. "Verbal expressions of probability in informed consent litigation" *Journal of Medical Decision Making* **11** 273–281.

Moore, PG, 1977. "The manager's struggle with uncertainty" *Journal of the Royal Statistical Society* **140** 129–165.

Moore, PG and Thomas, H, 1975. "Measuring uncertainty" *The International Journal of Management Science* **3** 657.

Mosier, CI, 1941. "A psychometric study of meaning" *Journal of Social Psychology* **39** 31–36.

Mosteller, F and Youtz, C, 1990. "Quantifying probabilistic expressions" *Statistical Science* **5** 2–16.

Moxey, LM and Sanford, AJ, 1993. *Communicating Quantities: A Psychological Perspective*, Lawrence Erlbaum.

Mullet, E, and Rivet, I, 1991. "Comprehension of verbal probability expressions in children and adolescents" *Language and Communication* **11** 217–225.

Murphy, AH and Brown, BG, 1983. "Forecast terminology: Composition and interpretation of public weather forecasts" *Bulletin of the American Meteorological Society* **64** 13–22.

Murphy, AH, Lichtenstein, S, Fischhoff, B and Winkler, RL, 1980. "Misinterpretations of precipitation probability forecasts" *Bulletin of the American Meteorological Society* **61** 695–701.

Nakao and Axelrod, 1983. "Numbers are better than words: Verbal specifications of frequency have no place in medicine" *American Journal of Medicine* **74** 1061.

National Weather Service, 1984. *Weather Service Operations Manual.*

National Oceanic and Atmospheric Administration.

Newstead, SE, 1988. "Quantifiers as fuzzy concepts" In: T Zetenyl (eds.), *Fuzzy Sets in Psychology*, pp 51–72, Elsevier Science.

Norwich, AM and Turksen, IB, 1984. "A model for the measurement of membership and the consequences of its empirical implementation" *Fuzzy Sets and Systems* **12** 1–25.

Parducci, A, 1968. "How often is often" *American Psychologist* **23** 828.

Pepper, S, 1981. "Problems in the quantification of frequency expression" In: DW Fiske (ed.), *New Directions for Methodology of Social and Behavioral Science*, Jossey-Bass.

Pepper, S and Prytulak, LS, 1974. "Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions" *Journal of Research in Personality* **8** 95–101.

Rapoport, A, Wallsten, TS and Cox, JA, 1987. "Direct and indirect scaling of membership functions of probability phrases" *Mathematical Modelling* **9** 397–417.

Rapoport, A, Wallsten, TS, Erev, I and Cohen, BL, 1990. "Revision of opinion with verbally and numerically expressed uncertainties" *Acta Psychologica* **74** 61–79.

Reagan, RT, Mosteller, F and Youtz, C, 1989. "Quantitative meanings of verbal probability expressions" *Journal of Applied Psychology* **74** 433–442.

Reyna, VF, 1981. "The language of possibility and probability: Effects of negation on meaning" *Memory & Cognition* **9** 642–650.

Rubin, DC, 1979. "On measuring fuzziness: A comment on 'a fuzzy set approach to modifiers and vagueness in natural language'" *Journal of Experimental Psychology: General* **108** 486–489.

Schkade, DA and Kleinmuntz, DN, 1994. "Information displays and choice processes: Differential effects of organization, form, and sequence" *Organizational Behavior and Human Decision Processes* **57** 319–337.

Shapiro, AJ and Wallsten, TS, 1994. "Base rate effects on interpreting verbally and numerically communicated probabilities" Working paper.

Simpson, RH, 1944. "The specific meanings of certain terms indicating differing degrees of frequency" *Quarterly Journal of Speech* **30** 328–330.

Stone, DN and Schkade, DA, 1991. "Numeric and linguistic information representation in multiattribute choice" *Organizational Behavior and Human Decision Processes* **49** 42–59.

Sutherland, HJ, Lockwood, GA, Tritchler, DL and Sem, F, 1991. "Communicating probabilistic information to cancer patients: Is there 'noise' on the line?" *Social Science and Medicine* **32** 725–731.

Svenson, O and Karlson, G, 1986. "Attractiveness of decision alternatives characterized by numerical and non-numerical information" *Scandinavian Journal of Psychology* **27** 74–84.

Teigen, KH, 1988a. "When are low-probability events judged to be 'probable'? Effects of outcome-set characteristics on verbal probability estimates" *Acta Psychologica* **67** 157–174.

Teigen, KH, 1988b. "The language of uncertainty" *Acta Psychologica* **68** 27–38.

Teigen, KH and Brun, W, 1993. "Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms" *SPUDM-14*, Aix-en Provence, France.

Tsao, CJ and Wallsten, TS, 1994. "Effects of the number of outcomes on the interpretation and selection of verbal and numerical probabilities in dyadic decisions". Working paper.

Tversky, A, Sattath, S and Slovic, P, 1988. "Contingent weighting in judgment and choice" *Psychological Review* **95** 371–384.

Vesely, WE and Rasmuson, DM, 1984. "Uncertainties in nuclear probabilistic risk analysis" *Risk Analysis* **4** 313–322.

von Winterfeldt, D and Edwards, W, 1986. *Decision Analysis and Behavioral Research*, Cambridge University Press.

Wallsten, TS, 1971. "Subjectively expected utility theory and subjects' probability estimates: Use of measurement-free techniques" *Journal of Experimental Psychology* **88** 31–40.

Wallsten, TS, 1990. "Measuring vague uncertainties and understanding their use in decision making" In: GM von Furstenberg (ed.), *Acting under Uncertainty* pp 377–398, Kluwer.

Wallsten, TS, Budescu, DV and Erev, I, 1988. "Understanding and using linguistic uncertainties" *Acta Psychologica* **68** 39–52.

Wallsten, TS, Budescu, DV, Rapoport, A, Zwick, R and Forsyth, BH, 1986. "Measuring the vague meanings of probability terms" *Journal of Experimental Psychology: General* **115** 348–365.

Wallsten, TS, Budescu, DV and Tsao, CJ, 1994. "Combining linguistic probabilities" *Symposium on Qualitative Aspects of Decision Making*, Regensburg, Germany, July 20 1994.

Wallsten, TS, Budescu, DV and Zwick, R, 1993. "Comparing the calibration and coherence numerical and verbal probability judgments" *Management Science* **39** 176–190.

Wallsten, TS, Budescu, DV, Zwick, R and Kemp, SM, 1993. "Preference and reasons for communicating probabilistic information in numerical or verbal terms" *Bulletin of the Psychonomic Society* **31** 135–138.

Wallsten, TS, Filenbaum, S and Cox JA, 1986. "Base rate effects on the interpretations of probability and frequency expressions" *Journal of Memory and Language* **25** 571–587.

Weber, EU, 1988. "A descriptive measure of risk" *Acta Psychologica* **69** 185–203.

Weber, EU and Bottom, WP, 1989. "Axiomatic measures of perceived risk: Some tests and extensions" *Journal of Behavioral Decision Making* **2** 113–131.

Webber, EU and Hilton, DJ, 1990. "Contextual effects in the interpretations of probability words: Perceived base rate and severity of events" *Journal of Experimental Psychology: Human Perception and Performance* **16** 781–789.

Wright, GN, Phillips, LD, Whalley, PC, Choo, GT, Ng, KO, Tan, I and Wisudha, A, 1978. "Cultural differences in probabilistic thinking" *Journal of Cross-Cultural Psychology* **9** 285–299.

Wyden, P, 1979. *Bay of Pigs*, Simon & Schuster.

Yager, RR, 1981. "A procedure for ordering fuzzy sets on the unit interval" *Information Sciences* **24** 143–161.

Yates, JF, Zhu, Y, Ronis, DL, Wang, DF, Shinotsuka, H and Toda, M, 1989. "Probability judgment accuracy: China, Japan, and the United States" *Organizational Behavior and Human Decision Processes* **43** 145–171.

Zimmer, AC, 1983. "Verbal vs. numerical processing of subjective probabilities" In: RW Scholz (ed.), *Decision Making Under Uncertainty*, Elsevier.

Zwick, R, 1987. "Combining stochastic uncertainty and linguistic inexactness: Theory and experimental evaluation". Doctoral dissertation, University of North Carolina at Chapel Hill.

Zwick, R, Budescu, DV and Wallsten, TS, 1988. "An empirical study of the integration of linguistic probabilities" In: T Zetenyi (ed.), *Fuzzy Set Theory in Psychology*, pp 91–125, North Holland.

Zwick, R, Carlstein, E and Budescu, DV, 1987. "Measures of similarity among fuzzy concepts: A comparative analysis" *International Journal of Approximate Reasoning* **1** 221–242.

Zwick, R and Wallsten, TS, 1989. "Combining stochastic and linguistic inexactness: Theory and experimental evaluation of four fuzzy probability models" *International Journal of Man–Machine Studies* **30** 69–111.