Memory & Cognition
2005, 33 (6), 1057-1068

Interpersonal comparison of subjective probabilities: Toward translating linguistic probabilities

MANDEEP K. DHAMI
University of Cambridge, Cambridge, England

and

THOMAS S. WALLSTEN
University of Maryland, College Park, Maryland

Interpersonal variability in understanding linguistic probabilities can adversely affect decision making. Using the fact that everyone judges canonical probability events similarly in a manner consistent with axiom systems that yield a probability measure, we developed and tested a method for comparing the meanings of probability phrases across individuals. An experiment demonstrated that despite extreme heterogeneity in participants' linguistic probability lexicons, interpersonal similarity in phrase meaning is well predicted by phrase rank order within the lexicons. Thus, equally ranked phrases have similar meanings, and individual differences in linguistic probabilities may simply be explained by the phrases people use at each rank.

Unlike numerical probabilities that represent precise values on a 0–1 scale, linguistic probabilities such as *improbable* or *almost certain* tend to have imprecise meanings. There is considerable variability in the interpretation and application of probability phrases (for reviews, see Budescu & Wallsten, 1995; Clark, 1990). For example, Reagan, Mosteller, and Youtz (1989) found that different people interpret the phrase *likely* from a probability of .5 to .95. The potential consequences of misinterpretation can be great: Linguistic probabilities used to express the chances of O-rings on the Challenger space shuttle failing at specific ambient temperatures were misunderstood by those deciding to launch the shuttle (Marshall, 1986). We propose that the communication of risk and uncertainty may be improved by translating the meanings of linguistic probabilities from one person's lexicon to another's (see also Karelitz & Budescu, 2004). In this article, we present and test a method for interpersonal comparison of the meanings of linguistic probabilities. First, however, we consider the need for a translation device and how subjective probabilities can be compared across people in a theoretically sound way.

This research was supported by NSF Grant 0196140, awarded to the second author. We thank Megan Bozick and Jared Smith for their research assistance, David Budescu and Tzur Karelitz for many useful conversations on this topic, and Claudia González-Vallejo for comments on an earlier draft. Correspondence should be addressed to M. K. Dhami, University of Cambridge, Institute of Criminology, Faculty of Law, Sidgwick Ave., Cambridge CB3 9DT, England (e-mail: mkd25@cam.ac.uk).

Individual Differences in Interpreting Linguistic Probabilities

Individuals tend to have relatively stable lexicons of probability phrases (Budescu, Weinberg, & Wallsten, 1988). The interpretation of phrases can be meaningfully and reliably scaled via paired-comparison judgments or direct ratings at the individual level, as membership functions over the [0, 1] probability interval (e.g., Budescu & Wallsten, 1990, Experiment 2; Fillenbaum, Wallsten, Cohen, & Cox, 1991; Jaffe-Katz, Budescu, & Wallsten, 1989; Rapoport, Wallsten, & Cox, 1987; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986; Wallsten, Budescu, & Zwick, 1993).[1] Thus, linguistic probabilities can be represented as fuzzy subsets of the probability interval (Zadeh, 1975). The membership function, $\mu_w(p)$, for a phrase w evaluated at probability p equals 0 if the respondent considers p not at all described by w; the function equals 1 if p is considered to be perfectly described by w, and it equals a number between 0 and 1 if p is considered to be described by w to some degree. Studies show that individuals have broad membership functions for most of the phrases in their lexicons [i.e., for the majority of their phrases $0 < \mu_w(p) < 1$ for large ranges of p], which suggests large intrapersonal imprecision in phrase meaning (e.g., Fillenbaum et al., 1991; Jaffe-Katz et al., 1989; Rapoport et al., 1987; Wallsten, Budescu, et al., 1986; Wallsten, Budescu, & Zwick, 1993).

Nevertheless, intraindividual variability is less than interindividual variability. Indeed, whether assessed as point estimates, interval estimates, or membership functions, interpersonal variability is greater than that explained

by intrapersonal variability alone (e.g., Beyth-Marom, 1982; Brun & Teigen, 1988, Study 2; Bryant & Norman, 1980; Lichtenstein & Newman, 1967). There is often little overlap across individuals' full probability lexicons (e.g., Budescu et al., 1988; Erev & Cohen, 1990; Wallsten, Budescu, & Zwick, 1993). For example, the 20 participants in Budescu et al.'s (1988) study generated 111 distinct phrases to describe 11 probabilities. Brun and Teigen (1988, Study 2) found differences in the rank order of probability phrases between parents and physicians. Research has also shown that the location (central tendency) and spread (dispersion) of membership functions for a phrase vary across individuals (e.g., Budescu & Wallsten, 1990; Rapoport et al., 1987; Wallsten, Budescu, et al., 1986). Thus, different people use different phrases to refer to the same probability and the same phrase to refer to different probabilities.

However, people are often unaware of, or they underestimate, the degree of variability in interpretation of probability phrases (Brun & Teigen, 1988). Furthermore, people prefer to communicate risk and uncertainty using linguistic probabilities (Brun & Teigen, 1988, Study 2; Erev & Cohen, 1990; Wallsten, Budescu, Zwick, & Kemp, 1993), often because it is easier, natural, more personal, and allows expression of judgment uncertainty (Wallsten, Budescu, et al., 1993).

There are many situations in which the recipient of a probability phrase must accurately understand the meaning intended by the communicator in order to avoid the negative consequences of a misunderstanding. Indeed, decision makers in high-stakes contexts such as the medical, legal, and financial domains often rely on the probabilistic judgments of experts conveyed in natural language terms. Probability or frequency phrases are also commonly used as response options in surveys (Tourangeau, Rips, & Rasinski, 2000, chap. 2) and in scales measuring health and well-being (Schwarz, 1999). Aggregation of the results of these investigations, whether to inform public policy decisions, plan patient treatments, or test scientific hypotheses, assumes that phrases have specific agreed-upon meanings.

Past theorists have proposed that experts or forecasters should avoid using linguistic probabilities altogether (e.g., von Winterfeldt & Edwards, 1986) or should adopt a standardized quantification of probability phrases (e.g., Hamm, 1991; Mosteller & Youtz, 1990). However, neither proposal seems workable. Banning the use of linguistic probabilities in favor of numerical probabilities is untenable because people prefer to use natural language (Zimmer, 1983), the task may not allow such quantification (Budescu & Wallsten, 1995), and people may prefer language since it communicates lack of confidence in estimates (Wallsten & Budescu, 1990). Standardization is problematic because people find it difficult to suppress their normal meanings of probability phrases (Wallsten & Budescu, 1990) and the meanings of phrases are influenced by the contexts in which they are used (Tanur, 1990). Alternatively, we propose that the probability phrases in one person's lexicon (i.e., forecaster) can be translated to those of similar meaning in another's lexicon (i.e., decision maker). The development of a translation device hinges on our ability to compare the meanings of probability phrases across individuals. Next, we review the theoretical underpinnings of such a comparison.

Interpersonal Comparison of Subjective Probabilities

The basic problem of interpersonally comparing subjective probabilities, as long recognized with perceptual dimensions such as brightness and with subjective values such as utility, is that phenomenology cannot be compared across individuals. Psychophysical, utility, or subjective probability measurements are simple on a within-observer basis in a given context, because individuals can judge relative values or trade-offs. (Measurement may not generalize across contexts, as Laming, 1997, and Stewart, Chater, Stott, & Reimers, 2003, have observed, because the nature of the trade-offs is affected by the composition of the stimulus or choice set.) However, there are no satisfactory means for establishing common measurement units between observers. Proposed solutions to this problem involve strong assumptions (for approaches in psychophysics, see Bartoshuk et al., 2002; Borg, 1982; or Teghtsoonian, Teghtsoonian, & Karlsson, 1981; and in terms of utility, see Narens & Luce, 1983).

We suggest that interpersonal comparison is simpler in the domain of subjective probabilities for two reasons. First, probability is measured on a scale bounded by 0 and 1 with the endpoints of *impossible* and *absolutely certain* well understood by everyone.[2] Second, when the axioms leading to a probability measure are satisfied, the resulting scale is unique, not subject to transformation, and thus comparable from one person to the next. Two different classes of axiom systems lead to a unique probability scale (Wallsten, 1974). One concerns the ordering *at least as likely as* on a σ-algebra of events (e.g., Krantz, Luce, Suppes, & Tversky, 1971, chap. 5).[3] When these axioms are behaviorally satisfied, they yield a unique subjective probability scale over events. The other class of axiom systems concerns the ordering *at least as desirable as* over a suitably rich set of lotteries (e.g., de Finetti, 1937/1964; Fishburn, 1970; Savage, 1972; Tversky, 1967).[4] When these axioms are behaviorally satisfied, they yield a unique subjective probability scale over events and an interval-level scale over outcomes.

Although it is easy to construct situations in which either or both axiom sets are systematically violated, it is well established that they hold when the uncertain events are binary outcomes represented on a probability wheel or spinner. This is a circle radially divided into two sectors of different colors such as red and black, over which a pointer can be spun, coming to a random stop. Data show that participants accurately estimate the probabilities of binary spinner outcomes (e.g., that the pointer will stop over the red rather than the black sector of the wheel; Wallsten, 1971), which is sufficient to demonstrate that

the axioms for *at least as likely as* hold. In addition, when small amounts of money are involved, participants choose between gambles based on binary spinners in accordance with expected-value prescriptions (Wallsten & Budescu, 1983), which is consistent with the axioms for *at least as desirable as*.

Therefore, because individuals treat binary spinner events in a way consistent with a unique probability measure, such events can be used as the means of interpersonal comparison of subjective probabilities. For example, assume that person A considers event α to be more likely than a spinner stopping on red when red constitutes .04% of the area and less likely than a spinner stopping on red when red constitutes .12% of the area. Symbolically, letting $S_A(\alpha)$ be A's subjective probability of α, we can write $.04 < S_A(\alpha) < .12$. Assume further that for person B and event β, $.15 < S_B(\beta) < .25$. It then follows that $S_A(\alpha) < S_B(\beta)$, or that A considers α less likely than B considers β. Importantly, these statements assume that persons A and B compare the likelihoods of spinner outcomes and the events in question with absolute reliability, whereas in actuality this will be unlikely over an interval of spinner settings. This problem is handled (in theory) by establishing psychophysical functions showing the proportion of times that $A(B)$ judges $\alpha(\beta)$ to be less likely than spinner setting p as a function of p. Consequently, one can conclude that $S_A(\alpha) < S_B(\beta)$, whenever person A's function for α lies strictly to the left of person B's function for β. Finally, since this technique maps judgments of events to intervals (or functions), which sometimes overlap, rather than to points, it yields a semiorder (Luce, 1959) rather than a linear order of subjective probabilities both within and between participants.

Wallsten (1990) has discussed some of the factors that might affect the width of the probability intervals, including whether the uncertainty is aleatory or epistemic. Uncertainty is aleatory when it is derived purely from relative frequency considerations and epistemic when it is based on one's theoretical or factual understanding (Hacking, 1975). For example, uncertainty regarding the outcomes of rolling a fair die is aleatory, and uncertainty regarding unique events such as the outcome of the 2008 U.S. presidential election is epistemic. Others have used the terms *external uncertainty* and *internal uncertainty*, respectively, for this distinction (e.g., Kahneman & Tversky, 1982). Although uncertainty in most real-world situations has both aleatory and epistemic components, it is useful for research to render the two components as distinct as possible.

One way to compare meanings of probability phrases across individuals is to use the method just described with phrases directly (e.g., "Which event is more likely to occur, the pointer landing on red on a random spin or an event described as *likely*?") However, spinner outcomes by definition are aleatory (so we call this the *aleatory method*), and applying the technique in this manner begs the question of whether it captures the meanings of phrases when speakers use them to communicate epistemic uncertainty. Therefore, a second two-stage (*epistemic*) method involves asking participants first to use probability phrases to express their uncertainty about unique events, and then to judge these events relative to spinner outcomes. These two sets of judgments can then be combined to induce a relationship between the phrases and the spinners. We used a variation of both methods in the present study.

The Present Study

Participants first selected the probability phrases in their lexicons and ranked them from the phrase representing the lowest to the highest probability. Then, using a variation of the aleatory method, participants saw spinners and for each one selected the phrase in their lexicon that best described the probability that the pointer would land on "red." We term the distribution of probabilities associated with each phrase its *aleatory probability signature.* Following this, participants encoded membership functions for their phrases under the aleatory condition. Next, using a variation of the two-stage epistemic method, participants saw a sequence of events and judged the likelihood of each occurring in the future, using one of their phrases. Then, participants provided a probability estimate from 0 to 1 for each event. We term the distribution of probabilities associated with each phrase its *epistemic probability signature*. Finally, participants encoded membership functions for their phrases under the epistemic condition. Half of the sample performed the tasks in the epistemic condition followed by the aleatory condition.

Therefore, the meanings of an individual's phrases are inferred from their usage. A comparison of the resulting probability signatures may be used as an empirical measure of interpersonal similarity. If the signatures truly capture phrase meaning, we predict that they should be most similar for phrases at equal ranks across individuals, regardless of what those phrases are. Our primary goal was to find an index—for example, derived from phrase membership functions or rank orders that best predicts an empirical measure of interpersonal similarity—under conditions of aleatory as well as epistemic uncertainty. However, our first goal was to establish whether the concordances among the intraindividual probability signatures, rankings, and membership functions of the phrases within the two conditions were satisfactory. Additional exploratory goals were to measure similarities in membership functions and in probability signatures across the aleatory and epistemic conditions, both within and between participants.

METHOD

Participants

Twenty-nine students at the University of Maryland, all native English speakers, volunteered to participate in return for $12. The experiment lasted approximately 1 h.

Measures and Procedure

Task 1: Selecting the lexicon. The computer-based procedure consisted of three tasks. In Task 1, the participants provided and

rank ordered seven phrases in their probability lexicons. The participants were presented with 18 probability adjectives (e.g., *likely*) and 21 modifiers (e.g., *very*), each in alphabetical order. (These were identified on the basis of a review of past research.) The participants were encouraged to select phrases that they would normally use even if they were not presented. Next to the lists of probability phrases was a vertical scale anchored at 0%, 50%, and 100% at the top, middle, and bottom, respectively. Running down the scale were seven pairs of drop-down menu windows, equally spaced so that the top pair was adjacent to 0%, the middle pair to 50%, and the bottom pair to 100%. The menus in each pair were labeled "modifier" and "adjective," respectively. The participants selected a phrase by clicking on the menus to reveal the modifiers (including the option "no modifier") and the adjectives from the lists, respectively. The participants could select 0, 1, or 2 modifiers and a single adjective, or if they preferred, type in their own phrases. They were encouraged to first provide phrases for the 0%, 100%, and 50% points, in that order, and then to select the other phrases, so that they eventually had a rank-ordered list of phrases covering the probability interval. Finally, the participants had another opportunity to revise their list or rank order.

Task 2: Establishing the probability signatures. The participants were then required to use their phrases to describe the chances of 100 aleatory (Task 2 aleatory) and 100 epistemic (Task 2 epistemic) events occurring. The order of presentation of each set of events was counterbalanced over participants. The individual lexicons were visible on the screen.

On each Task 2 aleatory trial, the participants saw a probability wheel divided into a red and a black area. Their task was to choose the phrase from their personal lexicons that best described the probability that a pointer fixed at the center of the wheel would land on red following a random spin. One hundred proportions of red were equally spaced from 0 to 1. The events were presented in a unique random order for each individual.

On each Task 2 epistemic trial, the participants read descriptions of possible future real-world events. Their task was to choose the phrase from their lexicons to describe their subjective probability that the event would occur. We informally created 100 questions covering current affairs, politics, entertainment, sports, science, and the University of Maryland with the aim of including events for which judgments would span the range from a 0% to 100% chance of occurring. For example, one question was "What are the chances that scientists will find a cure for AIDS in the next 5 years?" In a pilot study, 32 students gave their numerical subjective probability estimates of these events and thus confirmed that the events spanned the full probability range. The events were presented in a unique random order for each individual.

After completing Task 3 (to be described), the participants again saw the same 100 real-world events, but this time they provided numerical subjective probability estimates using a response scale from 0 to 1 in intervals of .1.

Task 3: Encoding membership functions for phrases. Immediately after Task 2 aleatory and Task 2 epistemic trials, the participants provided direct numerical interpretations of each phrase in their lexicons under aleatory (Task 3 aleatory) and epistemic (Task 3 epistemic) conditions, respectively, by encoding membership functions. Membership functions were encoded twice to capture any effect of context on them. The multistimuli membership function technique developed and validated by Karelitz, Budescu, and Wallsten (2000) was used. The participants were presented with each of their phrases separately on the screen along with 11 scales representing the percentage values from 0% to 100% (in 10% intervals). They were asked to move the slider along each scale from *not at all* to *absolutely* to answer the question, "To what extent would each of these numbers substitute for the phrase 'X'?" The phrases were presented in a unique random order for each participant and each task.

RESULTS

We first examine the overlap of individual lexicons in Task 1. Then, at the within-participants level, we compare the aleatory and epistemic probability signatures from Task 2 to each other, to the rank orders from Task 1, and to the aleatory and epistemic membership functions from Task 3. We also relate the aleatory and epistemic functions to each other and to the rank orders. Having established within-participants consistencies among these measures, we derive an empirical measure of interpersonal similarity by comparing the aleatory and epistemic probability signatures across individuals. We then test our prediction by analyzing the similarity of probability signatures for phrases at equivalent ranks in different lexicons. Finally, we turn to our primary goal, which is to find an index based on membership functions and on phrase rank orders that best predicts our empirical measure of interpersonal similarity.

Individual Linguistic Probability Lexicons

The lexicons of 27 (out of 29) participants included phrases with modifiers. A total of 102 distinct phrases were used across the sample. Of these, 38 phrases appeared in 2 or more lexicons, and 64 appeared in single lexicons. Thus, the median number of lexicons within which a phrase appeared was 1 ($M = 1.99$, $SD = 2.03$). *Fifty–fifty chance* was ranked 4 (middle) in the lexicons of all 14 participants who used it. Similarly, *absolutely certain* was used by 11 participants, all of whom ranked it as the highest phrase in their lexicon, and the phrase *absolutely impossible* was used by 7 participants, all of whom ranked it as the lowest phrase in their lexicon. However, two other popular phrases were *good chance* and *almost*

Table 1
Number of Distinct Phrases Assigned to Each Rank Order, With Examples Shown at Each Rank

Phrase Rank Order	No. of Distinct Phrases	Examples of Phrases
1	18	Absolutely impossible, Very unlikely
2	23	Barely certain, Highly doubtful
3	25	Poor chance, Fairly likely
4	13	Fifty–fifty chance, Tossup
5	25	Good chance, Somewhat likely
6	19	Almost certain, Good chance
7	12	Absolutely certain, Definite

certain, which spanned ranks 4–6 and 3–6 across participants, respectively. Overall, of the 38 phrases that were used by more than one person, 25 were assigned different ranks across participants. Table 1 shows the number of distinct phrases used at each rank, along with some examples. Since most phrases appeared in only a single lexicon, phrase heterogeneity was substantial at all ranks, although there was slightly less variability at low, middle, and high ranks. In addition, phrases that did appear in multiple lexicons were often assigned different ranks.

Within-Participants Comparisons

Phrase probability signatures under aleatory and epistemic uncertainty. For each participant, the cumulative distributions of the probabilities associated with the phrases at each rank were derived separately under conditions of aleatory and epistemic uncertainty (see Figure 1). The mean patterns in Figure 1 reflect the individual signatures well. For each phrase, within each participant, a Kolmogorov–Smirnov test was used to determine whether the aleatory and epistemic distributions differed significantly from each other. Using $\alpha = .05$, the two distributions (signatures) differed on 28 of the $29 \times 7 = 203$ tests.[5] The significant differences occurred on zero phrases for 14 participants, one phrase for 7 participants, two phrases for 4 participants, three phrases for 3 participants, and four phrases for 1 participant.

Four features are apparent from an inspection of the probability signatures. First, there is considerable overlap in the probabilities associated with the phrases at each rank. Second, the locations (medians) of the signatures are concordant with the prior self-reported ranks. At the individual level, ranks of the signature medians were perfectly related to ranks of phrases in the lexicons for 26 participants in the epistemic condition and for 28 participants in the aleatory condition ($M\gamma = .97$, $SD = .13$, and $M\gamma = .98$, $SD = .09$, respectively). Third, the epistemic signatures are flatter (have greater variance) than the aleatory signatures. Although the aleatory–epistemic differences are small and significant in only 14% of the 203 comparisons, their consistency suggests that they are real. Fourth, the slopes of the signatures for the lowest, middle, and highest ranked phrases are steeper than for other phrases (i.e., their variances are smaller).

With regard to the last point, it is of considerable interest to note that when the spinner probabilities are transformed to log-odds {i.e., from p to $\ln[p/(1 - p)]$}, the signatures are well represented by equal-variance normal distributions with unique means for each phrase. Figure 2 provides an illustration of this result for the epistemic signatures of a typical participant. The data points represent the empirical distributions (transformed signatures) for the seven phrases, with a different symbol for each phrase. The corresponding solid curves show the best-fitting normal distributions, subject to the equal-variance constraint. We tested the adequacy of the equal-variance assumption by calculating the statistic $G^2 = -2 \ln(L_1/L_0)$ for each participant under each uncertainty condition and referring the result to the chi-squared distribution with 6 *df*. In this formulation, L_0 is the likelihood of the data under the normal distribution with a separate mean and variance for each phrase, and L_1 is the likelihood under the normal distribution with a separate mean for each phrase and a common pooled variance.[6] There are 14 free parameters in fitting L_0 (one mean and one variance for each of the seven phrases) and 8 free parameters in fitting L_1 (one

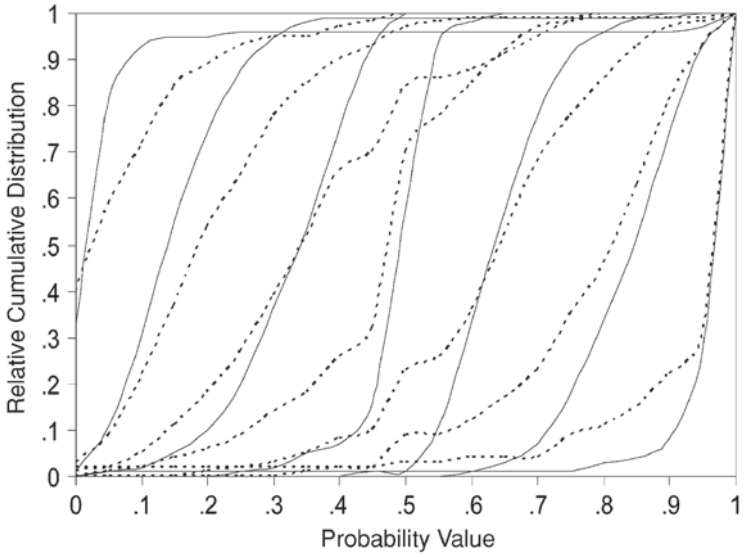

Figure 1. Mean cumulative frequency distributions of probabilities associated with phrases at each rank in the aleatory (solid lines) and epistemic (dotted lines) conditions. The leftmost pair of solid and dotted lines is for the lowest ranked phrases, the next pair for the next-lowest ranked, and so forth.

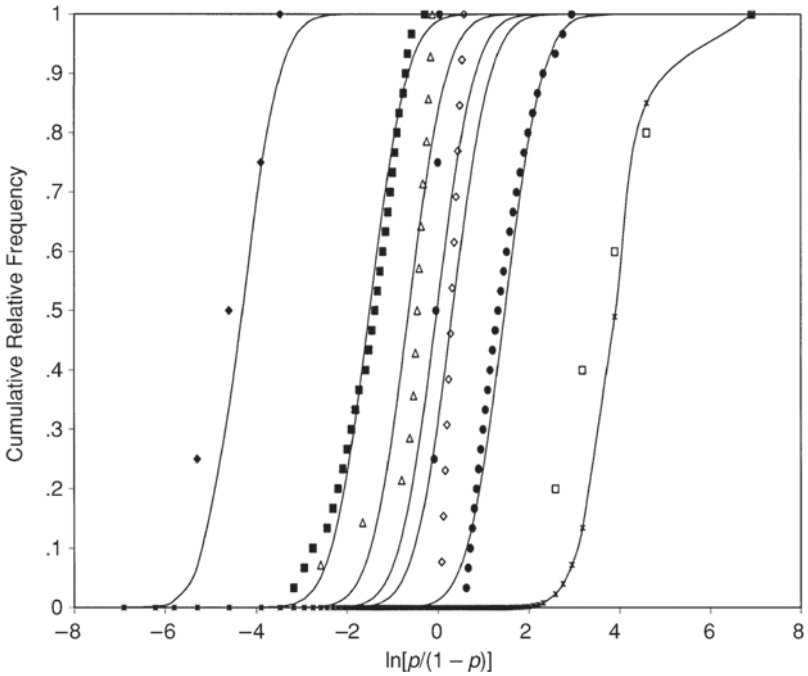

Figure 2. Transformed probability signatures (data points) and fitted equal-variance normal distributions (curves) for the seven phrases ranked from lowest to highest in the lexicon of a typical participant under epistemic uncertainty.

mean per phrase and a single pooled variance), yielding 6 *df* for the test. In none of the 58 tests (2 for each of 29 participants) did the test reach significance at $\alpha = .01$.

Membership functions of probability phrases under aleatory and epistemic uncertainty. Figure 3 shows the mean functions across individuals for phrases at each rank in the aleatory and epistemic conditions. As with Figure 1, these mean patterns reflect the individual functions well. In order to compare the aleatory and epistemic functions for each participant, a repeated measures analysis of variance was computed on the signed differences, where phrase had 7 levels and probability had 11 levels (i.e., the .1 intervals from 0 to 1). There was a significant main effect of probability [$F(10,280) =$

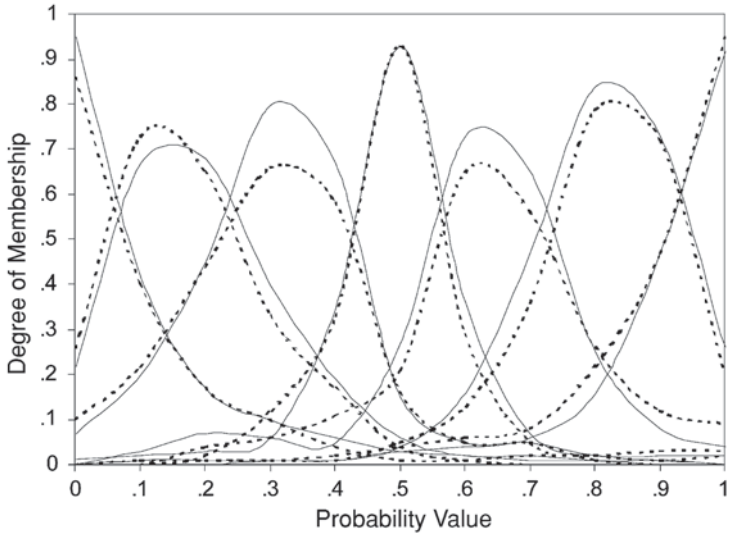

Figure 3. Mean membership functions of probability phrases over participants at each rank. The solid functions were encoded under aleatory uncertainty and the dotted functions under epistemic uncertainty. The leftmost solid and dotted lines are for the lowest ranked phrases, the next pair for the next-lowest ranked, and so forth. The abscissa is scaled as probability instead of percent.

5.38, $p = .001$] and a phrase × probability interaction [$F(60,1680) = 6.33, p = .001$]. Descriptively, the aleatory functions tended to have higher peaks and lower tails, although these differences were smaller for phrases at ranks 1, 4, and 7.

The functions encoded in the present study covered a large range of values, were single-peaked and monotonic at the extremes, tending toward single-peaked symmetry in the middle. The functions were arrayed in the same order as the probability signatures and participants' self-reported ranks. At the individual level, the ranks of the membership function locations (point of maximum membership) were perfectly related to ranks of phrases in the lexicons for 24 participants in the aleatory condition and 27 participants in the epistemic condition ($M\gamma = .96$, $SD = .10$, and $M\gamma = .97$, $SD = .06$, respectively). The membership function locations were also perfectly associated with the ranks of the probability signature medians for 24 participants under both the aleatory and epistemic conditions ($M\gamma = .95$, $SD = .17$, and $M\gamma = .96$, $SD = .13$, respectively). As with the probability signatures, the membership functions were systematically flatter in the epistemic than in the aleatory condition. The differences were small, but unlike with the signatures, they tended to be significant at the individual level.

Between-Participants Comparisons

Empirical measure of interpersonal similarity of probability phrases. Having established that a priori rank orders, probability signatures, and membership functions are internally consistent within participants under both aleatory and epistemic uncertainty conditions, we empirically measure interpersonal similarity of probability phrases by comparing the probability signatures across lexicons. Thus, we assume that phrase i has the same meaning to person j as phrase i' does to person j' if and only if the two phrases have the same probability signature. The fact that the probability signatures are independent relative cumulative distributions means that a measure of the similarity between phrases i and i' from participants j and j' is $s_{ii'} = 1 - d_{ii'}$, where $d_{ii'}$ is the Kolmogorov–Smirnov statistic (i.e., the maximum vertical distance between two signatures ($0 \leq s_{ii'}$, $d_{ii'} \leq 1$). Figure 4 shows, for example, that there is less distance between person *A*'s and person *B*'s third-ranked phrases, than between person *A*'s third-ranked phrase and person *B*'s fourth-ranked phrase. Table 2 presents means of the empirical measure of similarity in the aleatory and epistemic conditions by summarizing over each participant-pair. Tables comparing the statistics of each participant-pair separately conform to this pattern. As predicted, the probability signatures of phrases at equivalent ranks in different lexicons are more similar than are phrases at different ranks (see diagonal).

Testing indices of interpersonal similarity. Finally, assuming that the probability signature is a good measure of the meaning of a phrase and that the Kolmogorov–Smirnov statistic represents a satisfactory measure of interpersonal similarity of phrase meaning, we turn to the main question of which index best predicts our empirical measure of interpersonal similarity.

In order to test how well membership functions predict $s_{ii'}$, we selected the two membership function indices that performed the best out of 19 used in a study of intrapersonal similarity of linguistic probabilities (Zwick, Carlstein, & Budescu, 1987). One index is the absolute

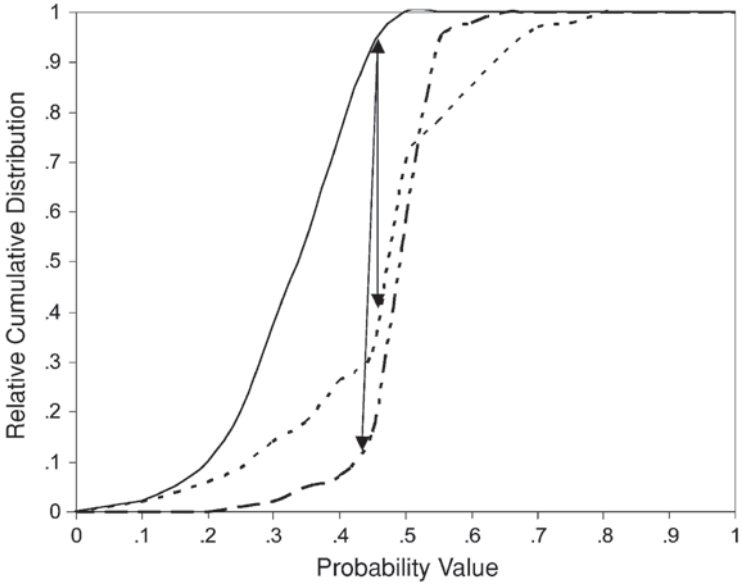

Figure 4. Empirical measure of interpersonal similarity of probability phrases, $s_{ii'}$. The solid line represents person *A*'s third-ranked phrase, and the first dotted line represents person *B*'s third-ranked phrase, followed by the second dotted line, which represents person *B*'s fourth-ranked phrase.

Table 2
Mean $s_{ii'}$ for Ranked Probability Phrases Summarized Across Participant Pairs in the Aleatory and Epistemic Conditions

	1		2		3		4		5		6		7	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Aleatory Condition														
1	.49	.34	.88	.15	.99	.05	1.00	.00	1.00	.00	.99	.07	.96	.17
2			.31	.21	.79	.21	.98	.06	1.00	.01	.99	.02	.99	.02
3					.38	.29	.80	.21	.99	.03	1.00	.01	1.00	.01
4							.31	.23	.84	.16	.99	.02	1.00	.01
5									.30	.18	.77	.21	.98	.06
6											.31	.22	.73	.24
7													.33	.24
Epistemic Condition														
1	.50	.24	.72	.22	.84	.19	.92	.13	.97	.07	.99	.03	.98	.08
2			.40	.18	.58	.21	.76	.16	.90	.10	.96	.06	.98	.09
3					.42	.19	.54	.17	.72	.19	.86	.15	.94	.13
4							.34	.12	.64	.19	.85	.16	.96	.11
5									.37	.17	.56	.20	.87	.18
6											.46	.21	.72	.26
7													.42	.30

difference, $q_{ii'}$, between the peaks of two functions. Thus, letting $\mu_i(p)$ and $\mu_{i'}(p)$ refer to the membership function values of p for phrases i and i' [$0 \leq \mu_{i'}(p), \mu_{i'}(p) \leq 1$],

$$q_{ii'} = \left| p_i^* - p_{i'}^* \right|,$$

where p_i^* is the value of p that satisfies $\mu_i(p^*) = \sup[\mu_i(p)]$.[7] The other index based on membership functions is $1 - m_{ii'}$, where $m_{ii'}$ is the point of greatest membership value in the two functions' intersection.[8] That is,

$$m_{ii'} = \sup_p \left\{ \inf \left[\mu_i(p), \mu_{i'}(p) \right] \right\}.$$

In addition, we tested an index based simply on the absolute difference between the rank orders of phrases in two lexicons. Here, the absolute difference, $r_{ii'} = |r_i - r_{i'}|$, where r_i and $r_{i'}$ are the ranks of phrases i and i' within the lexicons of participants j and j', respectively.

We examined how well each of these three indices predicted interpersonal similarity ($s_{ii'}$) across participants separately in the aleatory and epistemic conditions. For this purpose, we paired each participant j ($j = 1, \ldots, 29$) with each of the 28 others j' ($j = 1, \ldots, j-1, j+1, \ldots, 29$). Then, for each of participant j's 7 phrases ($i = 1, \ldots, 7$) we determined the rank of the $s_{ii'}$ for j''s phrases ($i' = 1, \ldots, 7$) and used Spearman's ρ to correlate that with the ranks of each of the three indices. Finally, for each index we took the mean of the 7 correlations for each participant. Thus overall, we have $29 \times 28 / 2 = 406$ nonindependent rank–order correlations per index.

The mean rank order correlations between the probability signatures and the three indices are presented in Table 3. It can be seen that the index based on absolute difference between rank orders of two phrases ($r_{ii'}$) is the best predictor of our empirical measure of interpersonal similarity in both the aleatory and epistemic conditions. The index based on the absolute difference between the peaks of two membership functions ($q_{ii'}$) outperformed the index based on the point of greatest membership value in the intersection of two functions ($m_{ii'}$) when predicting

Table 3
Mean Rank–Order Correlations in the Aleatory and Epistemic Conditions Between the Empirical Measure of Similarity ($s_{ii'}$) and the Three Indices ($r_{ii'}$, $q_{ii'}$, and $m_{ii'}$) and the Percentage of Between-Participants Comparisons Where Each Index Had the Highest Correlation with $s_{ii'}$

$s_{ii'}$ and $r_{ii'}$		$s_{ii'}$ and $q_{ii'}$		$s_{ii'}$ and $m_{ii'}$	
M (*SD*) ρ	%	*M* (*SD*) ρ	%	*M* (*SD*) ρ	%
Aleatory Condition					
.64 (.12)	58	.59 (.14)	2	.54 (.16)	40
Epistemic Condition					
.64 (.10)	67	.55 (.14)	21	.59 (.16)	12

Note—Each mean is based on $N = 406$ nonindependent values.

interpersonal similarity of phrases in the aleatory condition; however, the reverse was true in the epistemic condition. Therefore, whereas all three indices predict interpersonal similarity in probability signatures rather well on average in both the aleatory and epistemic conditions, the simplest measure—namely, the absolute difference in rank order—performs best.

DISCUSSION

The aim of the present study was to develop and test a method for comparing the interpersonal similarity of linguistic probabilities. We first consider implications of the results of the within-participants comparisons. Then, we discuss theoretical and practical implications of the finding that the meanings of linguistic probabilities can be compared across individuals. Finally, we comment on broader issues of interpersonal comparisons of value and meaning.

Within-Participants Comparisons

As expected from past literature (see Budescu & Wallsten, 1995), the membership functions obtained in the present study indicate that probability phrases have broad, overlapping meanings and that their breadth tends to be less at the anchor points (i.e., 0, .5, and 1). An implication of this result is that the probability signatures will show the same patterns, and indeed they do.

In addition, the present study offers three new findings pertinent to a theoretical understanding of how individuals interpret linguistic probabilities. First, to the degree that one accepts the argument that the probability signatures are direct measures of phrase meanings, the data reveal that the broad, overlapping meanings are not an artifact of the rating scales used for converting linguistic probability phrases to numerical scales. Second, the high degree of correspondence at an ordinal level among the explicit rank orders, medians of the probability signatures, and the peaks of the membership functions lends construct validity to all three measures. Finally, the full scope (and not just the central values) of the probability signatures and membership functions coincide to demonstrate that phrase meanings are similar, but also somewhat different under conditions of aleatory and epistemic uncertainty. Specifically, both measures indicate more diffuse phrase meanings in the epistemic than in the aleatory condition. This is consistent with the greater imprecision of epistemic than of aleatory uncertainty (Wallsten, 1990). However, the difference in probability signatures may also have occurred because the epistemic signatures were the products of two responses to events (i.e., assigning both linguistic and numerical probabilities to events), whereas the aleatory signatures resulted only from the choices of linguistic probabilities. By contrast, the membership functions were encoded in an identical manner under both the aleatory and epistemic conditions, and so any differences in functions can only be attributed to the contexts. Thus, people do attribute broader meanings to probability phrases under conditions of epistemic as opposed to aleatory uncertainty, although the difference may not be as great as that implied by the probability signatures.

The fact that the signatures, when represented as distributions over $\ln[p/(1-p)]$, are well described by equal-variance normal distributions, is unexpected but very useful. On theoretical grounds, it suggests a form of regularity in phrase meaning that had not been detected before and that deserves further study. On practical grounds, it suggests that when implementing algorithms for translating phrases across lexicons according to probability signatures, one can use functions fitted to data points rather than the data points themselves, which is a considerable simplification.

Some authors (e.g., Hacking, 1975; Kahneman & Tversky, 1982) have suggested that people tend to use different classes of phrases (referring to relative frequencies, e.g., *likely*, or to confidence, e.g., *confident*) to describe aleatory and epistemic uncertainty, respectively. Yet our procedure required participants to use the same phrases for both. Although it may be worth examining this suggestion empirically in the future, the considerable similarity found in the present study in membership functions and probability signatures of phrases across the two uncertainty conditions demonstrates that this is not necessarily true.

Between-Participants Comparisons

Consistent with past research (see Budescu & Wallsten, 1995), the present results show that individuals have different linguistic probability lexicons. Moreover, when the same phrase appeared in different lexicons, it tended to have a different probability signature, implying that individuals attributed different meanings to a phrase. However, phrases at equal rank tended to have similar signatures, suggesting that they have similar meanings. The latter result supports our prediction that probability signatures at equivalent ranks are similar, and thus $s_{ii'}$ is a valid empirical measure of interpersonal similarity of probability phrases. Finally, the two membership function indices predicted interpersonal similarity relatively well, but not as well as phrase rank order, perhaps being disadvantaged by the fact that individuals differ in how they use the response scale (which is also a problem with numerical translations). A recent study (Karelitz & Budescu, 2004) using different methodology and unlimited lexicon size, but restricted to aleatory uncertainty, obtained results converging with ours.

There are three theoretical implications of the finding that absolute differences in rank order predict interpersonal similarity of probability phrases. First, the ordinal properties of phrases are more important for interpersonal comparison of linguistic probabilities than are the actual phrases. Second, the widely documented individual differences in use of probability phrases seem to be due to the "place holders" that people choose to use in their personal lexicons, rather than to any more fundamental aspect of information processing. Finally, although it has

been argued that probability phrases have "rich semantic structures" such as an affective intensity dimension (Brun & Teigen, 1988) and a directional aspect (Teigen & Brun, 1995) that communicate more information than a numerical probability, it may be unnecessary to take into account such features for interpersonal comparison of linguistic probabilities. Budescu, Karelitz, and Wallsten (2003) demonstrated a strong association between the shape of a membership function and phrase direction. Thus, when translating phrases using an index based on the absolute difference between the rank order of two phrases (r_{ii}) that is well predicted by membership function shape, we can be confident that we are also translating phrase direction. Similarly, the meanings of phrases can be compared interpersonally without any need for encoding membership functions or for considering the aleatory–epistemic distinction.

In practical terms, the present results suggest that it may be possible to translate the meanings of linguistic probabilities from forecaster to decision maker on the basis of the rank order of phrases in their respective lexicons. For example, the eighth-ranked phrase in a forecaster's lexicon would be translated into the corresponding ranked phrase in a decision maker's lexicon. Evidence suggests that the interpersonal translation device would be welcomed by users, since Olson and Budescu (1997) found that people prefer to use their own judgments over those of others.

Unexpectedly, our results revealed that it would be possible to translate phrases directly from the probability signatures {via equal-variance normal distributions fitted to $\ln[p/(1-p)]$}, but developing the signatures is an arduous task. In contrast, the rank orders are easy to elicit, and translation would be very simple. Past research has shown that phrase rank order is relatively stable over time (Budescu & Wallsten, 1985; Kong, Barnett, Mosteller, & Youtz, 1986), establishing the reliability of such rankings. Other studies establishing validity include Lichtenstein and Newman (1967), who showed that participants can rank order the phrases in their lexicons in correspondence with their numerical meanings, and Hamm (1991), who showed that participants completed word problems with greater accuracy when phrases in a lexicon were rank ordered according to their numerical meanings.

Several other issues require investigation before a translation device based on phrase rank orders (or probability signatures) can become a reality. For instance, although we have shown that it is straightforward to equate rank orders when lexicons are the same size, it remains necessary to generalize the present results to conditions in which lexicons are of different sizes. Karelitz and Budescu (2004) recently demonstrated the success of several methods (including some similar to ours) in converting phrases across lexicons of different sizes. In addition, it is necessary to examine the effects of context on the meanings and subsequent interpersonal comparisons of phrases. Studies have documented effects of contexts such as outcome severity (Weber & Hilton, 1990), outcome valence (Mullet & Rivet, 1991), and perceived base rate of events (Wallsten, Fillenbaum, & Cox, 1986) on phrase meaning. Context may also affect membership function shape, but it remains to be determined whether it affects the probability signatures, and if so, whether the effect is uniform across participants. Context, however, is unlikely to affect phrase rank order. Finally, if it is possible to create a translation device, it is also necessary to determine how much it improves the quality of decision making. To date, the literature is inconsistent with regard to the effect of numerical versus linguistic probabilities on decision quality. Some studies have found little or no difference between the outcomes of decisions based on numbers and phrases (e.g., Budescu & Wallsten, 1990; Budescu et al., 1988; Erev & Cohen, 1990), some have found numbers to be more efficient (Jaffe-Katz et al., 1989), and others have found phrases to be more effective (González-Vallejo, Erev, & Wallsten, 1994; González-Vallejo & Wallsten, 1992). Olson and Budescu (1997) revealed that good decisions were based on the correspondence between the type of uncertainty (i.e., aleatory vs. epistemic) and the mode of communication (i.e., numerical vs. linguistic). Thus, research could measure the effect of the translation device on decision quality under conditions of aleatory uncertainty.

Limitation

It may be argued that the present study is limited because the size of the lexicon was restricted to 7 phrases. Although people may have large lexicons of linguistic probabilities, other researchers have found that participants provide a relatively small number of phrases. For instance, Zimmer (1983) reported that participants generated an average of 5.44 phrases using a direct report method. Renooij and Witteman's (1999) participants provided a mean of 8.2 phrases when asked which expressions they commonly use. As mentioned earlier, Karelitz and Budescu's (2004) study involved an unlimited lexicon size and yielded findings similar to ours.

Implications for Interpersonal Comparisons of Values and Meanings

Can the techniques developed here be extended to other domains in which vague expressions denote approximate ranges along numerical bases, such as quantity (few, many, etc.), frequency (rarely, often, etc.), and distance (near, far, etc.; see Bradburn & Miles, 1979; Newstead, 1988; Simpson, 1944)? The two properties that allowed interpersonal comparison of probability phrases are (1) the universally accepted end points of *impossible* and *certain* and (2) the manner in which people judge canonical spinner events. Together, they yield a unique probability measure that we can safely assume is comparable over individuals. Corresponding properties must be established in other domains in order to apply the present techniques. Most (maybe all) numerical domains have a universally agreed-upon zero point, but do not seem to have a second universally agreed-upon value (necessary for a common unit) and/or canonical representations. A challenging research question that could lead to the application of the

present technique to another domain is to find such a value or representation.

REFERENCES

Bartoshuk, L. A., Duffy, V. B., Fast, K., Green, B. G., Prutkin, J., & Snyder, D. J. (2002). Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: What we have learned from genetic variation in taste. *Food Quality & Preference*, **14**, 125-138.

Beyth-Marom, R. (1982). How probable is probable? Numerical translation of verbal probability expressions. *Journal of Forecasting*, **1**, 257-269.

Borg, G. (1982). A category scale with ratio properties for intermodal and interindividual comparisons. In H. Geissler, P. Petzold, H. Buffart, & Y. Zabrodin (Eds.), *Psychophysical judgment and the process of perception* (pp. 25-34). Amsterdam: North-Holland.

Bradburn, N. W., & Miles, C. (1979). Vague quantifiers. *Public Opinion Quarterly*, **43**, 92-101.

Brun, W., & Teigen, K. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? *Organizational Behavior & Human Decision Processes*, **41**, 390-404.

Bryant, G. D., & Norman, G. R. (1980). Expressions of probability: Words and numbers. *New England Journal of Medicine*, **302**, 411.

Budescu, D. V., Karelitz, T., & Wallsten, T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making*, **16**, 159-180.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior & Human Decision Processes*, **36**, 391-405.

Budescu, D. V., & Wallsten, T. S. (1990). Dyadic decisions with numerical and verbal probabilities. *Organizational Behavior & Human Decision Processes*, **46**, 240-263.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. In J. Busemeyer, D. L. Medin, & R. Hastie (Eds.), *Decision making from a cognitive perspective* (pp. 275-318). New York: Academic Press.

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 281-294.

Clark, D. A. (1990). Verbal uncertainty expressions: A review of two decades of research. *Current Psychology: Research & Reviews*, **9**, 203-235.

de Finetti, B. (1964). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, **7**, 1-68. (Original work published 1937)

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior & Human Decision Processes*, **44**, 1-18.

Fillenbaum, S., Wallsten, T. S., Cohen, B. L., & Cox, J. A. (1991). Some effects of vocabulary and communication task on the understanding and use of vague probability expressions. *American Journal of Psychology*, **104**, 35-60.

Fishburn, P. (1970). Utility theory for decision making (Vol. 18). New York: Wiley.

González-Vallejo, C., Erev, I., & Wallsten, T. S. (1994). Do decision quality and preference order depend on whether probabilities are verbal or numerical? *American Journal of Psychology*, **107**, 157-172.

González-Vallejo, C., & Wallsten, T. S. (1992). Effects of probability mode on preference reversal. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 855-864.

Hacking, I. (1975). *The emergence of probability.* Cambridge: Cambridge University Press.

Hamm, R. M. (1991). Selection of verbal probabilities: A solution for some problems of verbal probability expression. *Organizational Behavior & Human Decision Processes*, **48**, 193-223.

Jaffe-Katz, A., Budescu, D. V., & Wallsten, T. S. (1989). Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition*, **17**, 249-264.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, **11**, 143-157.

Karelitz, T., & Budescu, D. V. (2004). You say probable and I say likely: Improving interpersonal communication with verbal probability phrases. *Journal of Experimental Psychology: Applied*, **10**, 25-41.

Karelitz, T., Budescu, D. V., & Wallsten, T. S. (2000, November). *Validation of a new technique for eliciting membership functions of probability phrases*. Poster presented at the meeting of the Society for Judgment and Decision Making, New Orleans.

Kong, A., Barnett, G. O., Mosteller, F., & Youtz, C. (1986). How medical professionals evaluate expressions of probability. *New England Journal of Medicine*, **315**, 740-744.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement 1: Additive and polynomial representation*. New York: Academic Press.

Laming, D. R. J. (1997). *The measurement of sensation*. London: Oxford University Press.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, **9**, 563-564.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

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

Mosteller, F., & Youtz, C. (1990). Quantifying probabilistic expressions. *Statistical Science*, **5**, 2-34.

Mullet, E., & Rivet, I. (1991). Comprehension of verbal probability expressions in children and adolescents. *Language & Communication*, **11**, 217-225.

Narens, L., & Luce, R. D. (1983). How we may have been misled into believing in the interpersonal comparability of utility. *Theory & Decision*, **15**, 247-260.

Newstead, S. E. (1988). Quantifiers as fuzzy concepts. In T. Zetenyi (Ed.), *Fuzzy sets in psychology* (pp. 51-72). Amsterdam: Elsevier, North-Holland.

Olson, M. J., & Budescu, D. V. (1997). Patterns of preference for numerical and verbal probabilities. *Journal of Behavioral Decision Making*, **10**, 117-131.

Rapoport, A., Wallsten, T. S., & Cox, J. A. (1987). Direct and indirect scaling of membership functions of probability phrases. *Mathematical Modeling*, **9**, 397-417.

Reagan, T. R., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, **74**, 433-442.

Renooij, S., & Witteman, C. (1999). Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, **22**, 169-194.

Savage, L. J. (1972). *The foundations of statistics* (2nd rev. ed.). New York: Dover.

Schwarz, N. (1999). Frequency reports of physical symptoms and health behaviors: How the questionnaire determines the results. In D. C. Park, R. W. Morrell, & K. Shifren (Eds.), *Processing of medical information in aging patients* (pp. 93-108). Mahwah, NJ: Erlbaum.

Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degrees of frequency. *Quarterly Journal of Speech*, **30**, 328-330.

Stewart, N., Chater, N., Stott, H. P., & Reimers, S. (2003). Prospect relativity: How choice options influence decision under risk. *Journal of Experimental Psychology: General*, **132**, 23-46.

Tanur, J. M. (1990). Comment. *Statistical Science*, **5**, 21-22.

Teghtsoonian, R., Teghtsoonian, M., & Karlsson, J. G. (1981). The limits of perceived magnitude: Comparison among individuals and among perceptual continua. *Acta Psychologica*, **49**, 83-94.

Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, **88**, 233-258.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey responses*. New York: Cambridge University Press.

Tversky, A. (1967). Additivity, utility and subjective probability. *Journal of Mathematical Psychology*, **4**, 175-201.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

Wallsten, T. S. (1971). Subjectively expected utility theory and subjects' probability estimates: Use of measurement-free techniques. *Journal of Experimental Psychology*, **88**, 31-40.

Wallsten, T. S. (1974). The psychological concept of subjective probability: A measurement theoretical view. In C. A. S. Stael von Holstein (Ed.), *The concept of probability in psychological experiments* (pp. 49-72). Dordrecht: Reidel.

Wallsten, T. S. (1990). Measuring vague uncertainties and understanding their use in decision making. In G. M. Von Furstenberg (Ed.), *Acting under uncertainty: Multidisciplinary conceptions* (pp. 377-398). London: Kluwer.

Wallsten, T. S., & Budescu, D. V. (1983). Encoding subjective probabilities: A psychological and psychometric review. *Management Science*, **29**, 151-173.

Wallsten, T. S., & Budescu, D. V. (1990). Comment on Mosteller and Youtz' "Quantifying probabilistic expressions." *Statistical Science*, **5**, 23-26.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, **115**, 348-365.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, **39**, 176-190.

Wallsten, T. S., Budescu, D. V., Zwick, R., & Kemp, S. M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society*, **31**, 135-138.

Wallsten, T. S., Fillenbaum, S., & Cox, A. (1986). Base-rate effects on the interpretations of probability and frequency expressions. *Journal of Memory & Language*, **25**, 571-587.

Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 781-789.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning [Parts 1, 2, 3]. *Information Sciences*, **8**, 199; **8**, 301; **9**, 43.

Zimmer, A. C. (1983). Verbal versus numerical processing of subjective probabilities. In R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 159-182). Amsterdam: North-Holland.

Zwick, R., Carlstein, E., & Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, **1**, 221-242.

NOTES

1. Phrase meaning can also be elicited by asking participants to provide a lower and upper probability that the phrase represents, as well as point numerical translations over occasions, and by inferring the range of probabilities for which a participant uses the phrase.

2. We recognize that the semantics of the terms *impossible* and *absolutely certain* are rich and that they are influenced by context and the speaker's intention. Our reference here is not to the terms per se, but to the constructs to which the probabilities of 0 and 1 refer.

3. An algebra of events (or sets) is a set that is closed under complementation and union. In the former, for every event A in the set, the complement not-A is also in. In the latter, for every pair of events A and B in the set, the union A-and-B is also in. A σ-algebra is closed under countable unions (i.e., for all events A_i, $i = 1, 2, \ldots$, in the set, their union is also in), and is the domain over which the ordering *at least as likely as* is usually axiomatized.

4. A *lottery* is a probability distribution over outcomes. "Suitably rich" implies sufficient overlap among the elements of the lotteries such that the pattern of preferences among the lotteries can be examined for violations of axioms and the probability–outcome trade-offs can be estimated with reasonable precision. Different axiom systems assure this condition differently.

5. Type I probability corrections to account for multiple tests are not appropriate here, because these tests are nonindependent and so none of the established corrections apply. Also, reducing the α-level would only serve to mask possible aleatory–epistemic differences.

6. In taking the log-odds transformation, we converted $p = 0$ to ln(.005/.995) and $p = 1$ to ln(.995/.005) to avoid undefined numbers. The choice of .005 as a correction factor rather than .001 or any other reasonable value had some impact on the pooled variance and on the locations of the functions for phrases 1 and 7, but did not substantially affect the outcomes of the statistical tests. We thank Yoonhee Jang for assistance with this analysis.

7. When an interval of probabilities satisfies this condition, p_i^* is the center of that interval.

8. *Sup* and *inf* refer, respectively, to the maximum and minimum of a set of discrete numbers. We used these operators instead of *max* and *min* because membership functions were encoded at 11 distinct probabilities.

(Manuscript received February 26, 2004;
revision accepted for publication September 23, 2004.)