Journal of Experimental Psychology: Applied
1996, Vol. 2, No. 4, 343–364

1076-898X/96/$3.00

Measuring Psychological Uncertainty: Verbal Versus Numeric Methods

Paul D. Windschitl and Gary L. Wells
Iowa State University

The authors argue that alternatives to the traditional numeric methods of measuring people's uncertainty may prove to hold important advantages under some conditions. In 3 experiments, the authors compared verbal measures involving responses such as *very likely,* and numeric measures involving responses such as *80% chance.* The verbal measures were found to show more sensitivity to various manipulations affecting psychological uncertainty (Experiment 1), to be better predictors of individual preferences among options with unknown outcomes (Experiment 2), and to be better predictors of behavioral intentions (Experiment 3). Results suggest that numeric measures tend to elicit deliberate and rule-based reasoning from respondents, whereas verbal measures allow for more associative and intuitive thinking. Given that there may be many types of situations in which human decisions and behaviors are not based on deliberate and rule-based thinking, numeric measures may misrepresent how individuals think about uncertainty in those situations.

In many theories of behavior, psychological uncertainty is assumed to be an important mediator of human responses in situations with unknown outcomes.[1] In decision theories, a person's estimates of the probabilities of uncertain outcomes are determinants of the person's choice among prospects (e.g., for the prospect theory, see Kahneman & Tversky, 1979). According to theories of health behavior, whether individuals will engage in a health protective behavior is a function of their perceived likelihood of contracting a certain condition (e.g., for the health belief model, see Becker, 1974). Theories of jury behavior use subjective probability notions and probability thresholds to depict the processes involved in judgments of reasonable doubt (see Dane, 1985).

As with any mediated process, it is critical that psychological uncertainty be measured effectively in order to understand and predict the decisions and behaviors being mediated (see Baron & Kenny, 1986). However, in much psychological research, there seems to be an implicit assumption that the consequences of measuring uncertainty one way versus another are generally not significant. Rarely do authors mention their rationale for choosing a Likert-type scale ranging from 1 = *very unlikely* to 7 = *very likely* or a percentage scale ranging from 0% to 100%, or a line segment method anchored by *impossible* and *certain.* The predominant method for the measure-

Paul D. Windschitl and Gary L. Wells, Department of Psychology, Iowa State University.

Portions of this research were presented at the Annual Meeting of the Midwestern Psychological Association, May 1994, Chicago.

Correspondence concerning this article should be addressed to Paul D. Windschitl, Department of Psychology, W112 Lagomarcino Hall, Iowa State University, Ames, Iowa 50011. Electronic mail may be sent via internet to s1.pdw@isumvs.iastate.edu.

[1] In this article, we have often used the term *psychological uncertainty* rather than the shorter term *uncertainty.* As argued by numerous theorists and philosophers, uncertainty is a psychological construct. It exists only in the mind; if a person's knowledge were complete, that person would have no uncertainty. We chose to sometimes use the term *psychological uncertainty* in order to emphasize the psychological nature of the construct.

ment of psychological uncertainty involves the solicitation of subjective probability estimates. Researchers commonly ask people to estimate the probability that a given event has happened, the chance that a given statement is true, or the odds that an event will occur. Regardless of whether the requested response format is a probability estimate (e.g., "Give a number between 0 and 1.0") or a frequency estimate (e.g., "Out of 100 times, how many times would this event occur?"), the assumption is that people's numeric answers are accurate reflections of underlying feelings of uncertainty. A similar assumption seems to play a role in the way theorists conceptualize uncertainty. Human uncertainty is often conceptualized in terms of subjective probabilities rather than in more generic terms such as subjective uncertainty.

In this article, we suggest that there may be good reason to consider alternative methods of measuring and conceptualizing uncertainty. We argue that there may be many situations in which nonnumeric measures might be more informative than numeric measures in regard to how people think about uncertainty information in those situations. Our hypothesis derived from the increasingly accepted observation that people seem to be capable of two systems of reasoning: one based on associative, intuitive, soft-constraint, and automatic processes and the other on rule-based, deliberative, hard-constraint, and controlled processes (see Sloman, 1996, for a discussion of the system distinction). We think that numeric measures of uncertainty might tend to sway people toward more deliberative and rule-based thinking. However, in many situations, people's preferences, decisions, and behaviors are predominantly influenced by the more associative, intuitive system. Consequently, people's responses on a numeric measure of uncertainty can provide a skewed reflection of how people think about uncertainty in some situations. We chose to investigate verbal measures of uncertainty, involving responses such as *very unlikely* or *almost certain,* as an alternative means of assessing people's uncertainty. We hypothesized that verbal measures would be less likely to sway people's processing toward the more deliberative rule-based system.

In this article, we speculate on the possible reasons that researchers have measured and often conceptualized psychological uncertainty in numeric terms. We then briefly review research concerning the verbal communication of uncertainty. Finally, we discuss more fully theoretical reasons why nonnumeric measures of uncertainty might sometimes provide better reflections of people's thoughts regarding uncertainty.

Why Numeric Measures and Representations Have Been Favored

There are many possible reasons why researchers have discussed, thought about, and measured psychological uncertainty in numeric terms. We suspect that one important reason is theorists' usage of the tools-to-theories heuristic in which scientific tools become concepts for scientific theories (Gigerenzer, 1991). Gigerenzer (1991) argued that scientific tools are often used as metaphors for theoretical concepts and that scientists' familiarity with the tools promotes a general acceptance of the theoretical concepts as valid. For example, the deterministic view of the universe as a clock with God as watchmaker flourished after the mechanical clock became indispensable in astronomical research. More relevant to psychological research is the observation that inferential statistics, a tool familiar to every research psychologist, have been used explicitly as an analogy to the operations of the mind (e.g., Kelley, 1967; Tanner & Swets, 1954). We speculate that psychologists' assumptions about mental representations of uncertainty have been influenced by the formal probabilistic systems that they have been trained in and that they use as tools in their own research.

Another factor contributing to the assumption of numeric representations of uncertainty may be a false consensus bias on the part of research psychologists. Psychologists may overlook that they are different from those who are not trained as scientists. Advanced training in psychology has been shown to have substantial effects on graduate students' statistical and methodological reasoning about everyday problems. Lehman, Lempert, and Nisbett (1988) tested 1st- and 3rd-year graduate students' reasoning skills on everyday scenarios involving regression to the mean, the law of large numbers, and confounded variables. The graduate training of psychology students as well as medical students was found to have substantial effects on the tested skills, whereas training in law or chemistry was found to have little or no effect on those skills. These results lead us to speculate that relative to the general population, research psychologists may be more likely to think about everyday uncertainty in numeric terms. If psychologists fail to appreciate that their own way of thinking about uncertainty is unique, they may measure and conceptualize uncertainty in a manner that is appropriate for statistically trained people but not for the general population.

Other possible reasons why psychological uncertainty has been measured in numeric terms are the ease with which these estimates can be elicited and analyzed and can be directly compared with more objective estimates of probability. For example, numeric estimates can be easily added, subtracted, and averaged; thus, they are amenable to many of the analyses that researchers conduct. Questions of accuracy and calibration can be investigated quite easily with numeric estimates of uncertainty. For example, when someone indicates that there is a 75% chance of a "heads" occurring on an upcoming flip of a fair coin, it is easy to conclude that the feeling of uncertainty is off by 25%. These convenient characteristics of numeric measures of uncertainty are not shared by nonnumeric measures. Nevertheless, they do not constitute a compelling reason for assuming that human internal representations of uncertainty are numeric or are best represented in numeric terms. In fact, we argue that these convenient characteristics have unnecessarily led some researchers to use numeric measures of uncertainty when alternative measures would have been more informative.

Verbal Systems of Uncertainty Communication

Alternative systems of uncertainty expression have not gone untested. Verbal systems have received significant research attention. Much of the relevant research has focused on the translation of verbal expressions into numeric values or on the benefits of using numeric or verbal expressions for communicating uncertainty. One common finding from such research has been that the interpretations of verbal expressions of uncertainty vary greatly among individuals. Numerous researchers have shown that there is substantial between-subjects variability in the numeric (probabilistic) interpretations given for words such as *likely* and phrases such as *reasonable chance* and *close to certain* (e.g., Beyth-Marom, 1982; Bryant & Norman, 1980; Budescu & Wallsten, 1985; Sutherland et al., 1991). A second common finding is that interpretations of verbal expressions differ depending on the context to which the expression was referring. For example, the phrase *slight chance* might be interpreted differently depending on whether it was used to refer to the possibility of rain versus the possibility of complications occurring during a routine operation (see Brun & Teigen, 1988; Wallsten, Fillenbaum, & Cox, 1986; Weber & Hilton, 1990).

Although findings such as these might seem to play against the use of verbal systems of uncertainty expression, we note that these findings have come from studies concerning the communication of uncertainty (e.g., between a medical professional and a patient; a forecaster and a decision maker; or a financial consultant and a client). The research findings have not necessarily addressed how psychological uncertainty can best be measured. For instance, issues regarding how uncertainty can best be communicated between two parties have limited relevance to how a researcher should measure consumer perceptions of product reliability for the purpose of predicting purchasing behavior.

Furthermore, studies that have examined decision making and behavior that is based on verbal and numeric uncertainty expressions have not revealed consistent and overwhelming advantages for one mode over the other. It appears that, depending on the circumstances, numeric expressions of probability, relative to verbal expressions, can lead to slightly more consistent and profitable decisions, to less profitable decisions, or to no significant differences in decisions (e.g., Budescu, Weinberg, & Wallsten, 1988; Erev & Cohen, 1990; González-Vallejo, Erev, & Wallsten, 1994; González-Vallejo & Wallsten, 1992; Wallsten, Budescu, & Zwick, 1993; see Budescu & Wallsten, 1995, for a review). Also, there is important work that has documented positive features of communicating uncertainty with words (e.g., see Teigen, 1988). Hence, previous findings regarding verbal systems of uncertainty expression should not lead to the belief that verbal systems of uncertainty expression will not prove to have important merits for measuring people's uncertainty.

Why Verbal Expressions Might Better Reflect Psychological Uncertainty

When describing their own uncertainty, most people in most everyday situations use words rather than numbers. Commonly asked questions (e.g., "Will you be home by 5?"; "Do you think the Bulls won today?"; "Do you think the bookstore has this text in stock?") usually generate responses like "definitely," "probably," and "it's really unlikely." However, there are some domains or situations in which numeric expressions dominate. There is evidence to suggest that people can use words and numbers interchangeably when they think about uncertainty (Jaffe-Katz, Budescu, & Wallsten, 1989). Hence, it can be argued that numbers and words are alternative overt responses of the same internal construct of uncertainty.

Although we do not entirely disagree with such an argument, we posit that the external solicitation of a numeric expression of uncertainty can often have importantly different consequences than a solicitation of a verbal expression.[2] As argued above, numeric measures of uncertainty may tend to sway people toward more deliberative and rule-based thinking, whereas verbal measures of uncertainty tend to allow more intuitive and associative thinking.

The idea that there are two systems of reasoning has a long history in both the fields of psychology and philosophy (e.g., James, 1890/1950; Vygotsky, 1934/1987). There are a large number of current theories in social cognition that have posited dichotomies involving coacting subsystems (see Abelson, 1994; Sloman, 1996). Examples include Langer's (1989) mindful versus mindless, Epstein's (1990) rational versus experimental, Dovidio and Fazio's (1992) deliberate versus spontaneous, and Bruner's (1986) logical versus narrative dichotomies. Common to most of these dichotomies is the general idea that people can and do reason in ways that are relatively automatic, associative, and spontaneous, as well as in ways that are controlled, rule-based, and deliberate. Some theories allow for these systems to operate simultaneously, but some theories do not. However, virtually all of the theories posit that one system or the other can dominate at a given time and that whatever system is dominant has implications for the responses of an individual.

Why might numeric measures of uncertainty prompt more deliberative and rule-based processing, whereas verbal measures seem to be more compatible with intuitive and associative processing? There may be several qualities of numeric uncertainty measures that trigger special thoughts or considerations that would not be triggered by verbal measures of uncertainty. Numeric measures of uncertainty may enhance a person's consideration of accuracy. Whether or not a person is well versed in the formal rules of probability, that person likely knows that objective probabilities and formal rules of probability exist. Numeric response scales may make this knowledge especially salient. People may then be more inclined to work deliberately and to apply their understanding, whether correct or not, of formal rules regarding mathematics and probability. Verbal measures may be less likely to make the applicability of formal rules salient. Also, verbal expressions allow people to be somewhat immune to accuracy checks. When asked about the chances of producing three "heads" on three coin flips, someone saying 20% can be called inaccurate, whereas a person saying *unlikely*

[2] Our ideas are distinct from those of Zimmer (1983), who argued that people's internal representations of uncertainty are verbal. Nevertheless, we think Zimmer did make an important point in arguing that uncertainty is probably not internally represented in terms of numbers. Zimmer noted that it was not until the 17th century that the concept of uncertainty was treated mathematically: "Even then it took place only in the European culture where it was motivated by practical problems, such as gambling and insurance. It seems unlikely that . . . numeric estimates of uncertainty have become automatized since then. It is more likely that people handle uncertainty by customary verbal expression" (p. 161). We agree with Zimmer's view that people have not automatized numeric representations of uncertainty, and we suggest that human intuition in the late 20th century is largely pre-Bernoullian. However, we do not assume that people's internal representations of uncertainty are verbal.

cannot. In other words, people who are thinking about how to verbally express their uncertainty rarely have to concern themselves with what the normatively correct answer is, as there is no normative model to compute a verbal answer. Another related reason why numeric measures might trigger special considerations is the possible belief on the part of the respondent that a numeric response should be defensible with an explicit derivation (see Beyth-Marom, 1982). A verbal uncertainty measure may more likely be construed as an assessment of an opinion. Hence, people may feel less concerned about whether their verbally expressed opinions can be defended with a derivation. Finally, traditional numeric measures of uncertainty do not allow people vagueness in expressing an opinion. Because of the range of interpretation for a given phrase (e.g., Beyth-Marom, 1982; Wallsten, Budescu, Rapoport, Zwick, & Forsyth, 1986), verbal measures allow a person to provide a response, yet still remain vague.

We think that considerations such as these regardless of whether they operate at a fully conscious level may influence the way in which a person processes uncertainty information when preparing a numeric uncertainty response. Of course, the influence of these considerations would be only apparent in situations where people would not normally think in a deliberate, controlled, and rule-based manner. For example, if a researcher were to ask some meteorologists about the numeric probability of snow, meteorologists' thought processes would not be influenced by the question because they would typically think about such uncertainty in a deliberate and rule-based way. However, we think that in many everyday situations, the average person thinks in a more intuitive and associative manner. According to Epstein (1990), most everyday behaviors, emotions, and decisions are driven by information processing that operates in what he calls the *experiential system.* In situations in which more intuitive and associative thinking is the norm, numeric measures of uncertainty may change the manner in which people think about uncertainty information. The numeric measures may yield less informative responses than verbal measures, which would not move people away from the intuitive and associative mode of processing typical for the given situation.

Our analysis suggests that there should be instances in which verbal measures of uncertainty exhibit important advantages over numeric measures. We report three experiments testing the merits of verbal versus numeric methods of measuring people's uncertainty in response to numerous everyday scenarios. In Experiment 1, we compared and contrasted numeric with verbal measures to see if verbal measures were more sensitive to subtle manipulations of context and framing. In Experiment 2, we tested the hypothesis that verbal expressions of uncertainty can be better predictors of people's preferences than numeric expressions. In Experiment 3, we tested the hypothesis that verbal expressions can be more closely related to behavior intentions than numeric expressions.

Experiment 1

In Experiment 1, we tested the hypothesis that participants' verbal reports of uncertainty would show more sensitivity to various manipulations of context and framing than numeric reports. The experiment involved 13 uncertainty scenarios, each with two versions. Across the two versions of each scenario, the objective uncertainty regarding whether a particular target outcome would occur was identical.[3] However, there were context or framing manipulations used between the versions that we thought might affect the psychological uncertainty associated with the target outcome.

For example, one scenario described a person named Angie who was about to draw for a prize at a charity event. In one version, Angie was drawing from a box containing 10 tickets, 1 of which was a winning ticket. In the other version, Angie was drawing from a box containing 1,000 tickets, 100 of which were winning tickets. Across these versions the objective probability that Angie would win was the same (0.10), but the absolute number of winning tickets was manipulated to affect participant's psychological uncertainty regarding whether Angie would win.

[3] In using the term *objective uncertainty,* we refer to the level of uncertainty that would be calculated with a normative model.

This scenario was adapted from a situation used by Kirkpatrick and Epstein (1992), who showed that people preferred to draw from a bowl containing 100 beans where there were 10 red ones rather than a bowl containing 10 beans where only 1 was red when they had the chance to win money by drawing a red bean.[4] Their preference for the larger bowl was not due to a misperception of the numeric odds; participants knew that the objective probability of drawing a red bean was the same in both cases. Kirkpatrick and Epstein argued that their result is evidence of a difference in subjective probabilities (or psychological uncertainty) associated with picking a red bean from the 1-in-10 bowl versus the 10-in-100 bowl. With Angie, we assumed that participants reading the 1-in-10 version would feel it was less likely that Angie would win than would participants reading the 100-in-1,000 version. After all, there were 100 ways of winning in the 100-in-1,000 version, but only 1 way of winning in the 1-in-10 version we devised.

Although we designed the two versions of each scenario to produce differing levels of uncertainty, we expected nevertheless that numeric measures of participants' uncertainty would not detect differences between versions. We believed that participants giving numeric responses for the scenarios would likely engage in deliberate and rule-based thinking about the scenarios. Rule-based reasoning dictates that there are no differences in objective uncertainty across the scenario versions used in Experiment 1. On the other hand, participants giving verbal responses would not have been prompted to think in a deliberate and rule-based manner. Their responses would be a product of more associative and soft-constraint processes. This type of thinking may be more sensitive to the context and framing manipulations used in the experiment. To test our ideas, we had participants in Experiment 1 read one version of each of the 13 scenarios and provided uncertainty estimates about specified outcomes on either a numeric or verbal uncertainty scale.

Method

Participants. The participants were 148 students in introductory psychology classes at Iowa State University. The students received extra credit for taking part.

Scenarios. Thirteen scenarios each with two versions were used in the experiment. Appendix A contains two example scenarios. Each scenario described circumstances surrounding an event with an unknown outcome. The objective uncertainty associated with a scenario's unknown outcome was calculable from the given information and was equivalent across the two versions. However, the two versions contained context or framing manipulations that were expected to affect the psychological uncertainty associated with the target outcome.[5]

We used various methods of manipulating context or framing without affecting objective uncertainty. We have already described above one such example from the Angie scenario (see other examples in Appendix A). For Scenario 1 in Appendix A, we assumed that a manipulation to the total number of people cleaning classrooms would affect people's uncertainty about whether a specified person named Randy cleaned a particular classroom. In both versions, Randy cleaned 30 of the 50 classrooms, thereby holding constant the objective probability that Randy cleaned any given classroom. However, we suspected that the number of other people who cleaned the remaining 20 classrooms would affect respondents' uncertainty about whether Randy cleaned a given room. Perhaps if there were many cleaners mentioned, Randy would not stand out as the person who necessarily cleaned the room. On the other hand, mentioning many other cleaners might make Randy appear to be a very likely candidate because no one else rivaled him in the number of rooms cleaned. It turned out that the latter appeared to be a more likely explanation. For many of the scenarios, we could make no steadfast assumptions about the direction in which a context or framing manipulation would affect psychological uncertainty. Thus, our treatment of the differences between versions for all scenarios focused on the magnitude of the difference, irrespective of its direction.

[4] Kirkpatrick and Epstein (1992) modeled their scenario from scenarios proposed by Miller, Turnbull, and McFarland (1989).

[5] Copies of the scenarios are available from the authors on request.

In Scenario 2 of Appendix A, we thought that streak performances by a professional baseball player might seem typical. Thus it would seem very possible that the major leaguer would extend the streak (see Gilovich, Vallone, & Tversky, 1985). In reference to a little leaguer, however, two hits in a row might seem to be more a matter of chance or luck. So in accordance with the gambler's fallacy, people would probably perceive that the little leaguer's luck was due to run out.

Uncertainty scales. We constructed a verbal response scale and numeric response scale. Each contained 21 points. Those scales are shown in Appendix B. In constructing the verbal scale, we decided to make consistent the use of the base adjective *likely* except for the scale endpoints and to add various qualifiers of differing strength (e.g., *somewhat* or *extremely*). By making use of the same base adjective, we hoped that participants would have a quick familiarity with the layout and orientation of the scale. Instead of having to read and evaluate every point on the scale, participants could consider the scale as a continuum and thereby mentally move up and down the scale with ease.

The order in which the qualifiers appeared was determined through a translation study conducted with 284 participants. The participants were given a list of 43 verbal uncertainty expressions that were randomized in one of four possible orders. They included the 21 expressions used for the verbal scale in Experiment 1. The participants provided a numeric translation (e.g., 75% chance) for each expression. The mean and median responses were used to order the expressions for the verbal scale of Experiment 1.

Although the data from our translation study would allow us to quantify verbal expressions as probability values (e.g., a *somewhat likely* response could be quantified as .546), the verbal expressions were not treated as though they could be mapped into a numeric point, nor were numeric possibility functions created to represent the verbal expressions. We did not expect that participants would have a numeric value or range in mind when they gave a verbal response. Also, it seemed as though the properties of numeric values did not necessarily apply to verbal responses. For example, although the mean translation of *very likely* was 84.0% and the mean translation of *slightly unlikely* was 42.3%, it seemed odd to consider *very likely* as twice the likelihood of *slightly unlikely.* Thus, it seemed best to not place verbal responses on the formal probability scale of 0.0 to 1.0. In the analyses for Experiment 1, both verbal responses and numeric responses were given a value from 0 to 20 (e.g., both *extremely unlikely* and 10% were scored as 2).

Scenario packets. Four types of scenario packets were constructed. A given packet contained either the Version A or Version B of the 13 scenarios. Half of the packets contained numeric response scales, and half contained verbal response scales. Every scenario was immediately followed by a question about a specified outcome and a response scale. For packets containing numeric response scales, the question took the form of, "What is the chance that [*x* event would occur]?" For packets containing verbal response scales, the question took the form, "How likely is it that [*x* event would occur]?"

Procedure. Participants completed scenario packets in groups of 4 to 12. Each participant was randomly assigned to receive one of the four types of scenario packets mentioned above.

Results and Discussion

The main question driving the following analyses was whether the verbal or the numeric expressions of uncertainty were more sensitive to the context and framing manipulations between the scenario versions. We could not make assumptions about the direction in which the manipulations affected psychological uncertainty between the scenario versions. Thus, tests of interactions between scenario versions and expression types could be misleading. For example, consider a case in which Version A of a scenario evoked a much higher verbal estimate than did Version B, but Version A evoked a much lower numeric estimate than did Version B. This might yield a strong interaction, but it should not be taken to suggest that verbal or numeric estimates were more sensitive to the manipulation. To avoid this problem, we focused our analyses on the absolute values of the effect size estimates for differences (ds) between scenario versions.[6]

Separate effect size estimates for each scenario were computed for participants giving verbal responses and for participants giving numeric responses. Table 1 displays the mean verbal and mean numeric responses for each scenario version, as well as the computed effect size estimates for each scenario. On 11 of the 13 scenarios, the effect size for the difference between versions was larger directionally for participants who gave verbal responses than for participants who gave numeric responses. Overall, the average effect size across scenarios was significantly higher for participants providing verbal responses (d = .38) than for participants providing numeric responses (d = .21), $t(12) = 2.29, p < .05$, two-tailed. These results suggest that use of the verbal uncertainty scale provided more sensitivity to the context and framing manipulations than did use of the numeric uncertainty scale.

One of the ways in which the verbal measures might have produced larger effect sizes than the numeric measures is through lower within-cell variance because the effect size d is a difference between means in standard deviation units. However, the standard deviation was directionally smaller for numeric measures than for verbal measures on 16 of the 26 possible comparisons. In addition, for the four scenarios producing the largest effects favoring verbal measures (Scenarios 2, 3, 4, and 7), the standard deviations for verbal responses were smaller than for numeric responses in two cases and larger in the other two. Hence, larger effect sizes were observed for the verbal measures because the manipulations produced greater action on the verbal scales and not because the verbal measures reduced within-cell variance.

Readers might notice from Table 1 that participants using the verbal measure seemed to avoid responses with low likelihoods. Table 1 reveals that the means of the verbal responses and the means of the numeric responses (both scored from 0.0 to 20.0) are in general agreement within scenario versions, except for scenarios in which the means for the numeric responses were very low. On Scenarios 5, 6, and 7, means for the numeric responses fell between 3.5 and 2.1, whereas the means of the verbal responses fell between 8.5 and 6.5. We have no compelling explanation for this pattern of responding, but we believe it might reflect an important difference between the deliberate and rule-based responding by those asked for a numeric estimate and associative and intuitive responding by those asked for a verbal estimate. Specifically, we suspect that events that have very low objective probabilities are overestimated in associative and intuitive processing. Winning the lottery might have a minuscule probability, but associative and intuitive reasoning suggest that it is still possible. Therefore, it allows events of near-zero probabilities to loom larger than they would under more deliberate and rule-based processing. People's fears of objectively improbable events, such as plane crashes, hijackings, and child abductions are consistent with our speculation.

Consider the Randy scenario as an example of the advantage shown by the verbal measure. For the participants providing numeric uncertainty responses, there seemed to be little difference in reported uncertainty between those reading Version A of the scenario (M = 12.3 on the 20-point scale) and those reading Version B (M = 11.7), $t(72) = 1.00$, $p = .31$, $d = .23$. For the participants providing verbal uncertainty responses, however, those reading Version A (M = 13.9) reported substantially higher likelihood estimates than did those reading Version B (M = 11.9), $t(72) = 4.00, p < .001, d = .93$. Apparently, participants had a greater tendency to think Randy cleaned the classroom when it was assumed that no other person cleaned nearly as many classrooms as Randy (see Version A) than when it was assumed that another person cleaned nearly the same amount as Randy (see Version B). This difference in psychological uncertainty was detected by eliciting verbal uncertainty responses but not by eliciting numeric uncertainty responses.

Furthermore, in the scenario about Angie, for participants providing numeric responses, the average uncertainty estimate was not much different for the 100-in-1,000 version (M = 2.5) versus

[6] The effect size statistic used here is Cohen's d, which reflects standardized mean differences. Cohen (1988) has offered conventional values of .20, .50, and .80 as approximates for small, medium, and large effects, respectively.

Table 1
Mean Verbal and Numeric Uncertainty Responses and Standardized Mean Differences Between Scenario Versions for Experiment 1

	Verbal			Numeric		
Version	*M*	*SD*	*d*[a]	*M*	*SD*	*d*[a]
Scenario 1			.16			.10
Version A	10.9	2.9		8.1	1.4	
Version B	10.4	2.7		8.3	2.6	
Scenario 2			.69			.16
Version A	13.3	3.2		12.9	4.1	
Version B	15.2	2.5		13.5	3.0	
Scenario 3[b]			.93			.23
Version A	13.9	2.5		12.3	3.0	
Version B	11.9	1.5		11.7	1.9	
Scenario 4			.48			.12
Version A	10.4	2.0		10.1	1.4	
Version B	11.6	2.9		10.0	0.8	
Scenario 5[c]			.44			.23
Version A	8.1	3.6		2.5	2.0	
Version B	6.5	3.6		2.1	0.8	
Scenario 6			.28			.15
Version A	8.5	4.1		2.9	3.9	
Version B	7.2	4.7		3.5	4.2	
Scenario 7			.48			.18
Version A	8.2	3.5		3.5	2.0	
Version B	6.6	3.3		3.2	1.6	
Scenario 8			.09			.43
Version A	16.0	2.3		16.2	1.5	
Version B	15.8	2.4		15.1	2.9	
Scenario 9			.20			.37
Version A	13.2	2.4		12.7	2.0	
Version B	12.7	2.0		11.9	2.3	
Scenario 10			.18			.05
Version A	14.3	2.6		13.4	4.6	
Version B	14.8	2.7		13.1	4.5	
Scenario 11			.14			.05
Version A	15.0	1.9		13.2	2.2	
Version B	14.7	2.0		13.1	2.9	
Scenario 12[b]			.39			.16
Version A	11.6	4.3		8.0	3.5	
Version B	10.1	3.5		8.6	4.5	
Scenario 13			.48			.46
Version A	14.0	3.6		13.5	3.2	
Version B	15.4	1.8		14.8	2.0	

Note. Each mean is based on $n = 37$. Responses on the scales were scored from 0 to 20 where 20 = *certain* or *100%*.
[a]Reflect the absolute value of the standardized mean differences between scenario versions and within response type. [b]The "Randy" and "Baseball" scenarios in Appendix A. [c]The "Angie" scenario in text.

the 1-in-10 version ($M = 2.1$), $t(72) = .98$, $p = .33$, $d = .23$. However, for participants providing verbal responses, there was a nearly significant difference in estimates for the two versions. Participants who read that there was 1 winning ticket out of a total of 10 tended to report a lesser likelihood of Angie winning ($M = 6.5$) than did participants who read that there were 100 winning tickets out of 1,000 total ($M = 8.1$), $t(72) = 1.90$, $p = .06$, $d = .44$. This was consistent with the idea that people's optimism about winning is greater when there are many ways (e.g., 100 ways) of winning versus only one way of winning.

The results of Experiment 1 demonstrate that verbal measures of uncertainty can be more sensitive to context and framing manipulations than numeric measures. We suggest that participants providing numeric estimates of uncertainty were prompted to think in a deliberate and rule-based manner about the scenario information. Hence, their responses did not differ between scenario versions, which were constructed to have equal objective probabilities. On the other hand, participants providing verbal estimates were less likely to think about the scenario information in a deliberate and rule-based manner. Therefore, they were more sensitive to the manipulations of context and framing, which appear to be more relevant for intuitive and soft-constraint reasoning than for deliberate and rule-based reasoning.

Although Experiment 1 demonstrates that verbal measures can be more sensitive than numeric measures to some variations in context and framing, it did not help us determine whether such variations have significance for judgments, decisions, and behaviors. However, other research provides good reason to assume that such variations in context or framing can have important effects on judgments (Miller, Turnbull, & McFarland, 1989, 1990) and behaviors (Kirkpatrick & Epstein, 1992). As noted, Kirkpatrick and Epstein (1992) found that people preferred to draw from a bowl containing 100 beans in which 10 were winners rather than from a bowl containing 10 beans in which 1 was a winner. Miller et al. (1989) found that people were more suspicious when a child was supposedly lucky enough to blindly draw a favorite type of cookie (a chocolate chip cookie) from a jar containing 1 chocolate chip cookie and 19 oatmeal ones versus a jar containing 10 chocolate chip cookies and 190 oatmeal ones. These findings indicate that the differences in psychological uncertainty detected between the 1-in-10 and 100-in-1,000 versions of the Angie scenario may be quite important in determining related judgments and behavior. More broadly, the findings suggest that verbal measures of uncertainty may hold an important advantage over numeric measures for measuring the psychological uncertainty that mediates many everyday judgments, decisions, and behaviors. The following experiments more directly tested this idea.

Experiment 2

In Experiment 2, we investigated whether verbal uncertainty measures relative to numeric measures could provide a better assessment of the psychological uncertainty that mediates certain preferences or choices involving uncertainty. For this experiment, we made the simple assumption that people tend to prefer a prospect with a more certain reward than a prospect with a less certain reward of equal value. If the verbal responses better reflected people's uncertainty for a given situation, we expected that verbal responses would better predict preferences or choices involving differing levels of uncertainty.

Participants were given information about 12 trivia questions and asked to provide either verbal or numeric uncertainty estimates about whether they could answer each question correctly. We also asked them to provide preference rankings about which questions they would most want to be asked if money were at stake. The trivia questions were never disclosed, but participants were informed about each question's topic (e.g., skiing or Macintosh computers) and supposed base rates for correct responding by other students. Hence, participants' uncertainty estimates and preference rankings were based on this information rather than the question itself. In addition, participants were not requested to answer any trivia questions, and this aspect distinguished this experiment from calibration studies. We expected that both numeric and verbal uncertainty responses would be predictive of participants' preferences about which questions they would like to be asked, but that the predictive utility of the verbal responses would be greater.

Method

Participants. There were 209 different students from the same pool as participants in Experiment 1. They also received the same compensation.

Questionnaire. The instructions for the questionnaire stated that the researchers conducting the study were interested in knowing whether participants thought they could correctly answer trivia questions from various domains and of varying difficulty. The questionnaire made reference to 12 supposed questions. For each supposed question, information about its topic and difficulty (i.e., the base rate for correct responding) was given. Participants were asked about their uncertainty concerning whether they could answer the question correctly. For example, the first question asked, "If we were to ask you a question about Macintosh computers that was answered correctly by 57% of previously asked students, how likely is it that [what is the chance that] you would answer it correctly?" This was immediately followed by a verbal (or numeric) response scale. The second question and all remaining ones took the same form: "If we were to ask you a question about baseball that was answered correctly by 70% of previously asked students, how likely is it that [what is the chance that] you would answer it correctly?" The topics for the 12 questions and the difficulty level or base rates associated with the questions are in Appendix C. The information about these questions was not manipulated, and the order in which they were mentioned did not vary.

On a separate page that followed the uncertainty questions, instructions stated that participants should make preference rankings regarding which trivia questions they most wanted to be asked if money were at stake. These instructions were as follows:

> Now imagine that you will be asked one of the above-mentioned trivia questions and that you would win a money prize if you answered it correctly. Given this situation, please rank order the 12 possible trivia questions from the one you would most want to be asked (1) to the one you would least want to be asked (12). In other words, place a 1 next to the question that you think you would have the best shot at answering correctly, and a 2 next to the second easiest question for you, and so on.

The topics of the 12 questions were listed below these instructions in the same order in which they first appeared. To the left of each topic, participants could insert a rank. To the right, the base rate percentage that had previously accompanied the topic appeared. Appendix C shows how this information appeared.

Uncertainty scales. The verbal response scale and numeric response scale used in Experiment 2 were modified versions of the scales from Experiment 1. Both scales were reduced from 21 to 11 points (see Appendix D).

Procedure. Participants completed scenario packets in groups of 2 to 10. Each participant was randomly assigned to receive a questionnaire containing either verbal or numeric response scales.

Results and Discussion

The primary purpose of the analyses was to determine whether verbal uncertainty responses or numeric uncertainty responses were better predictors of participant preference rankings. For each participant, a Spearman correlation coefficient was computed for the relation between that participant's uncertainty estimates for the trivia questions and that participant's preference rankings for those questions. To aid the interpretations of these correlations, we reverse scored the preference rankings such that the questions they each most wanted to be asked were represented with a 12. The uncertainty responses were scored from 0 to 10. The 10 reflected a response of 100% on the numeric scale and *certain* on the verbal scale. For participants providing numeric responses, the average Spearman coefficient was .69, whereas for participants providing verbal responses the average was .77. Each Spearman coefficient was converted with a Fisher's r to z transformation (for our rationale, see Howell, 1987). Then the coefficients were submitted to a t test that compared participants giving verbal responses with participants giving numeric responses. The test indicated that the correlations between uncertainty estimates and preference rankings were stronger for participants giving verbal responses (the z-transformed $M = 1.02$,

$SD = .42$) than for those giving numeric responses (the z-transformed $M = .84$, $SD = .51$), $t(107) = 2.67$, $p < .01$. In other words, verbal uncertainty estimates were more predictive of participants' preferences than were numeric uncertainty estimates.

One reason for this might be that participants giving numeric responses were prompted to think about uncertainty in a deliberate and rule-based manner. In turn, they provided estimates that were products of such thinking. For instance, those participants may have been especially sensitive to the base rate information associated with each trivia question. Such information had an objective, rule-based quality and was presented in numeric form. Thus, for a person engaged in deliberate and rule-based processing, base rate information might have been heavily used in making the uncertainty estimate. To test this idea, we computed Spearman coefficients (one for each participant) to reflect the relation between the base rates given for the trivia questions and the uncertainty estimates provided by participants. The average of the Spearman coefficients was .68 for participants providing numeric responses and .56 for participants providing verbal responses. A t test analysis indicated that the difference in coefficients was significant. This suggested that uncertainty responses of participants giving numeric uncertainty estimates were more influenced by the base rate information (z-transformed $M = .83$, $SD = .60$) than were the uncertainty responses of participants giving verbal uncertainty estimates (z-transformed $M = .63$, $SD = .53$), $t(107) = 2.54$, $p < .05$.

Initially, the heightened sensitivity to base rates shown by participants giving numeric estimates might seem to be a desirable characteristic of numeric uncertainty measures. However, we saw this sensitivity as artificial. Participants' preferences for trivia questions were not heavily influenced by the base rate information. For each participant, a Spearman coefficient was computed for the relation between the base rates for the trivia questions and each participant's preference rankings for the questions. Across all participants, the average of those Spearman coefficients was only .35, and there was not a significant difference in the correlations from the two groups. This indicates that base rate information had only a moderate influence on participants' preferences. Thus, because numeric responses were quite sensitive to the base rate information, such responses may be less useful for predicting and understanding the thoughts about uncertainty that mediate preferences and decisions. We see verbal uncertainty scales as a way of measuring psychological uncertainty without prompting deliberate and rule-based processes and without sensitizing people to certain forms of uncertainty information such as base rates.

The results of this experiment were replicated in a nearly identical experiment that we conducted with 94 participants. The only difference for the replication experiment was that the base rate information was omitted from the preference-ranking page; thus only the topics appeared. As in the original experiment, the base rate information did appear with each trivia question's initial description. The average of the Spearman correlations between participants' uncertainty estimates and preference rankings was .54 for participants giving numeric responses and .68 for participants giving verbal responses. As in the original experiment, this difference in coefficients from the numeric (z-transformed $M = .60$, $SD = .60$) and verbal groups (z-transformed $M = .83$, $SD = .54$) was significant, $t(92) = 1.94$, $p < .05$, one-tailed. The average of the Spearman correlations between uncertainty responses and base rate information was again found to be higher for numeric responses (.70) than for verbal responses (.64), but the difference was not significant. Also, similar to the original experiment, the overall average of the Spearman correlations between the base rate information and the preference rankings was small (.07) with no significant difference in correlations between groups. The results of this replication experiment support the conclusions of our original experiment.

Thus far, we have focused on the magnitudes of across-item, within-participant correlations found in Experiment 2. That is, we have presented analyses of how well an individual participant's uncertainty responses across items can be used to predict preference rankings across those items. An alternative type of analysis involves within-item, across-participant correlations.[7] Such

[7] In the remainder of the article, we use the terms *across-item* or *across-scenario* instead of the longer term *across-item, within-participant.* Also, we use the terms *within-item* or *within-scenario* instead of the longer term *within-item, across-participant.*

an analysis focuses on how well different participants' uncertainty responses for a particular trivia question can predict the rankings they assign to that trivia question. Given that one of the measured variables in Experiment 2 involved ranking, interpreting results from this approach is somewhat problematic, as the rank assigned to an item is not only a function of an individual's reaction to that item, but it is also a function of the ranks assigned to other items.

In the within-item analyses for Experiment 2, two Spearman coefficients were computed for each of the 12 trivia questions. One of the coefficients for each question reflected the relation between numeric uncertainty responses and preference ranks; the other reflected the relation between verbal uncertainty responses and preferences. These correlations ranged from .26 to .71 in the original experiment and from .32 to .74 in the replication experiment. All of these correlations were significantly different from 0 at the .05 alpha level. Across the original and replication experiments, a total of 24 within-item comparisons were made between correlations for participants providing numeric responses versus correlations for participants providing verbal responses. None of these comparisons indicated that the correlations were significantly different for the two groups of participants. On 8 of the 12 comparisons for the original experiment and on 7 of the 12 comparisons for the replication experiment, the correlations were directionally larger for participants providing verbal responses than for participants providing numeric response.

We were not particularly surprised that within-item analyses did not reveal robust differences between verbal and numeric responding, whereas across-item analyses did. Such an analysis in the present experiment is problematic, given that one of the measured variables involved rankings. Also, as many researchers have noted, there are individual differences in the way in which verbal uncertainty phrases can be interpreted and used. Although we note that there are surely important individual differences in the way numeric uncertainty expressions are interpreted and used, we recognize that the statistical noise associated with individual differences in uncertainty responses may be more significant when verbal expressions, rather than numeric expressions, are being used. For instance, in the present experiment, different participants may have used the same phrase (*quite likely*) but it held different meanings; whereas their use of the expression 80% may have been more uniform. Within-item analyses do not control for individual differences, and these individual differences may account for why verbal uncertainty responses were not significantly more predictive of preferences than were numeric responses using the within-item analyses.

On the other hand, across-item analyses controlled for individual differences. Because across-item correlations are based on the relation between each participant's uncertainty responses and his or her preference responses, variability in the way in which different participants use the uncertainty scales does not affect the analyses. However, variability in the way a given participant uses the uncertainty scales (internal consistency) would affect these across-item analyses. Results from several studies have suggested that there is good internal consistency in the interpretation of verbal uncertainty phrases (for a review, see Clark, 1990). In Experiment 2 and its replication, the high across-item correlations between verbal uncertainty responses and preference rankings indicated that participants had used the verbal scales in a reasonably coherent and consistent manner. When the statistical "noise" associated with individual differences was controlled for by the across-item analyses, the results revealed that verbal uncertainty responses can be more useful than numeric responses in predicting a given person's preferences.

Experiment 3

The ultimate goal of research on uncertainty is to understand human behavior in situations involving uncertainty. In Experiment 3, we investigated whether the predictive advantage of verbal measures demonstrated in Experiment 2 would also apply to the prediction of behavior intentions. In addition, we used a nonrank response format for the behavior intention variable because we believed it would better allow for interpretable effects concerning within-item analyses.

Participants read eight scenarios with uncertain outcomes. For all of the scenarios, participants provided either numeric or verbal uncertainty estimates and selected one of five possible options to indicate how they thought they would have behaved in response to the described situation. The scenarios and questions were constructed so that those who expected the key outcome (i.e., the outcome we asked about in the uncertainty question) to occur (rather than not occur) would be favorably disposed toward engaging in the relevant behavior (i.e., the behavior inquired about). For example, one scenario began, "Imagine that you get a letter from an out-of-town prize redemption center informing you that you have won a TV and VCR." This scenario called for the participant to imagine having received information from a knowledgeable friend that about three out of five of these types of claim centers actually do give away the prizes as they indicate; the others require that a person make an outrageous purchase first. The participant's uncertainty was measured with the question, "How likely is it [What is the chance] that you actually won a free TV and VCR?" The behavior question asked, "What is the maximum number of roundtrip miles that you would travel to check into your prize?" The five options were 30, 60, 120, 240, and 480 miles. Our prediction regarding the responses to such questions was that verbal uncertainty responses would be more predictive of behavioral intentions.

Method

Participants. There were 200 students selected from the same pool as participants in Experiments 1 and 2. They also received the same compensation.

Scenario packets. Four types of scenario packets were constructed. A given packet elicited either numeric or verbal responses, and the order of the questions appeared in one of two ways. In half of the packets, the scenarios were first accompanied by behavioral intention questions and were later shown again with accompaniment of the uncertainty questions. This order was reversed in the other half of the packets. When participants were reading the scenarios for the first time and responding to the behavior (uncertainty) questions for each, they were not aware that they would later be rereading the scenarios and answering uncertainty (behavior) questions for each. The scenarios always appeared in the same order, and the information contained in the scenarios did not vary (see Appendix A for examples of the scenarios in the packets).

Uncertainty scales and behavioral intention questions. The verbal and numeric response scales in Experiment 3 were the same as those in Experiment 1 (see Appendix B). As in Experiment 1, the responses on both scales were scored from 0 to 20. Each of the behavior questions was accompanied by five response options. Participant responses to these questions were scored from 1 to 5, where 1 = *always indicating little or none of the relevant behavior* and 5 = *indicating a high degree of the relevant behavior.*

Procedure. Participants completed scenario packets in groups of 2 to 10. Each participant was randomly assigned to receive one of the four types of scenario packets.

Results and Discussion

Preliminary analyses indicated that the order factor did not have a significant effect on any of the correlations reported below and showed no interactions with the verbal versus numeric manipulation. Therefore, in reporting the following analyses, we have collapsed across the order conditions.

For the within-scenario analyses, Spearman coefficients for the relations between the verbal responses and behavioral intentions and between the numeric responses and behavioral intentions were computed for each scenario. These coefficients are displayed in Table 2. For two of the eight scenarios, the correlation for participants giving verbal responses was significantly higher than for those giving numeric responses; for two other scenarios, the difference in correlations approached significance. For none of the scenarios was the correlation significantly higher for participants giving numeric responses. By chance alone, it is expected that less than one (.4) of the eight comparisons would yield significant results.

These results extend those of Experiment 2 and show that even when within-scenario analyses are used verbal measures of uncertainty can hold advantages over numeric measures for predicting behavior. However, it is important to note that

Table 2
Within-Scenario, Across-Participant Rank Correlations Between Verbal or Numeric Uncertainty Responses and Behavior Intentions

Scenario	Verbal	Numeric	Significantly different[a]	p
1	.30	.18	No	.37
2[b]	.21	−.02	No	.10
3	.02	.11	No	.53
4	.34	.40	No	.63
5[c]	.46	.26	No	.10
6	.14	.28	No	.30
7[b]	.66	.37	Yes	<.01
8	.50	.02	Yes	<.001

Note. Each Spearman coefficient is based on $n = 100$. [a]Determined by Fisher r-to-z transformations for the coefficients and comparison of verbal versus numeric coefficients with a z test statistic ($\alpha = .05$, two-tailed). [b]The "Class Presentation" and "Vandal" scenarios in Appendix A. [c]The "Prize Redemption" scenario in text.

verbal responding did not hold a consistent advantage across scenarios. Although we have no compelling explanations as to why the advantage for verbal responding over numeric responding was not more uniform across scenarios, we must note that for Scenarios 1, 2, 3, and 6, there seemed to be only a weak relation between the uncertainty responses and behavior responses. These weak relations may indicate that the uncertainty and behavior questions for those scenarios were only moderately relevant to each other. If that was the case, it is expected that sizable advantages would not be observed for either type of uncertainty responding on those scenarios.

Although the within-scenario analyses revealed a moderate advantage for verbal responding, the results of the across-scenario analyses were much more impressive. For the across-scenario analyses, a Spearman coefficient reflecting the relation between uncertainty responses and behavior responses was computed for each participant. The average of these coefficients was .55 for participants providing verbal uncertainty responses ($n = 100$), but only .15 for participants providing numeric responses ($n = 100$). The difference in coefficients between the two groups of participants was significant, $t(198) = 5.94$, $p < .001$. However, a unique feature of this analysis was that the behavioral intention questions differed across scenarios. This problem was not present in Experiment 2 because the preference measure involved rankings. To compensate for the problem, we transformed participants' behavior responses into z scores that were based on the distribution of responses within a given scenario. Therefore, a participant's z score regarding a particular scenario reflected how a participant answered the behavior question relative to other participants. A high positive z score indicated a relatively high level of the given behavior. Across-scenario Spearman coefficients (one for each participant) were again computed, this time using uncertainty responses and the transformed behavior scores (i.e., the z scores). The average of these coefficients was .36 for participants providing verbal uncertainty responses ($n = 100$), but only .11 for participants providing numeric responses ($n = 100$). The difference in coefficients between the two groups of participants was significant, $t(198) = 3.51, p < .001$.

The results of the across-scenario analyses clearly reveal a large and reliable advantage for eliciting verbal expressions of uncertainty. In particular, the analyses involving the transformed behavior responses indicated that a given participant's set of verbal responses was much more informative than numeric responses with regard to how that participant answered behavioral intention questions relative to other participants. These results suggest that the processes mediating many behaviors can be better understood through the use of verbal measures of uncertainty rather than numeric measures.

General Discussion

In these three experiments, we have demonstrated that there can be important consequences of measuring psychological uncertainty with verbal versus numeric methods. Relative to numeric responses, verbal responses were found to be more sensitive to manipulations of context and framing (Experiment 1) and better predictors of people's preferences (Experiment 2) and behavior intentions (Experiment 3). These findings clearly suggest that traditional numeric measures do not always provide the adequate means for assessing psychological uncertainty. In the present experiments, numeric measures failed to detect variation in people's uncertainty that was detected by verbal measures. Moreover, that variation was shown to be important for people's preferences and behavior intentions.

These findings are consistent with the idea that people who have been asked to provide a numeric uncertainty estimate think differently about the presented information than those who have been asked to provide verbal uncertainty estimates. We suggest that numeric measures of uncertainty tend to prompt deliberate and rule-based thinking, whereas verbal measures allow for more associative and intuitive thinking. As already noted, there are a number of considerations involved in providing a numeric uncertainty estimate that could engage deliberate and rule-based processing. We assume that there are many situations in which people's decisions, judgments, and behaviors are not products of deliberate and rule-based processing (for a similar contention, see Epstein, Lipson, Holstein, & Huh, 1992). Hence, numeric measures may lead to a somewhat skewed assessment of how people normally think about uncertainty in such situations.

For example, imagine a person who is deciding whether to venture out alone into a poorly lit parking lot. That decision might be more influenced by intuitive and associative thoughts (e.g., involving a story just heard about a recent mugging) than by thoughtful analysis (such as recalling recently seen crime statistics) of the chances of being mugged. However, if a research psychologist were to approach the person and request a numeric estimate of uncertainty, the person's thought process might become more deliberative and rule-based; that is, the person might think more about the base rates for this type of crime and less about recently heard crime stories.

An Analogous Process: The Effect of Self-Reflection on Attitudes

Research conducted by Wilson and colleagues (e.g., Wilson, 1990; Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984; Wilson, Dunn, Kraft, & Lisle, 1989; Wilson & Schooler, 1991) concerning the effects of self-reflection about reasons for individual attitudes may illustrate a process that is analogous to what we think can sometimes occur when numeric measures are used to assess uncertainty. An intriguing consequence of the self-reflection manipulation in their experiments is a drop in attitude–behavior consistency. For example, in one experiment the relation between participants' ratings of various puzzles and how long they subsequently played with those puzzles was examined under two conditions. When participants had been prompted to analyze (prior to giving ratings) why they felt the way they did about each puzzle, the relation between ratings and behavior was much smaller than when participants had not been prompted in such a way. This type of finding has been replicated with various types of attitude objects (e.g., vacation pictures, dating partners, beverages, political candidates; for a summary, see Wilson et al., 1989).

Wilson and colleagues (Wilson et al., 1984, 1989; Wilson & Schooler, 1991) argued that people who are prompted to analyze why they feel the way they do might concentrate on specific justifications for their attitudes or on qualities of the attitude object that would not typically be the subject of close focus. For instance, such individuals may tend to concentrate on clear-cut and easily defined justifications. As a result, the attitudes subsequently expressed might fall in line with the justifications they considered. Yet those attitudes would be poor predictors of decision and behavioral responses that are not affected by the self-reflection. The explanation of Wilson and colleagues about self-reflection effects are somewhat analogous to possible explanations of the effects we found in our Experiments 2 and 3. Numeric measures of uncertainty may have caused people to become concerned with providing logically defensible responses or to focus on information that would not normally be of much interest. Experiment 2 showed some specific evidence of the latter. Numeric expressions of uncertainty were greatly affected by the base rates, whereas people's preferences (and their verbal expressions) were less affected by the base rates. A related finding was reported by Zimmer (1984), who found that bank clerks asked to provide verbal predictions of future monetary exchange rates reported using both qualitative and quantitative variables in making their predictions. However, clerks who were asked to provide numeric predictions tended to report using only quantitative variables.[8] In general terms, we think that prompting people for numeric expressions of uncertainty and prompting people to analyze reasons for their attitudes can have similar effects. That is, both can cause people to think in ways that they normally would not in certain situations. As a result, subsequent responses (i.e., numeric estimates or expressed attitudes) may have a limited ability to predict outcomes (i.e., decisions or behaviors) not influenced by such thinking.

When Numeric Measures Should Be Used

Although the present experiments provided demonstrations that verbal measures of uncertainty can hold advantages over numeric measures, they are far from definitive regarding how uncertainty should be measured. Nor do they suggest that verbal measures should always be used instead of numeric measures. A researcher's decision about how to measure human uncertainty depends on numerous factors.

In fact, there are several instances in which numeric measures of uncertainty might be more appropriate than verbal or other nonnumeric measures. We discuss three such instances below.

Researchers may be interested in assessing the degree to which people are able to think according to logical or normative rules. For example, if a study focuses on whether people are able to solve a particular problem in a way that does not violate a conjunction rule, then a numeric measure of uncertainty can be appropriate.

Researchers may need to study behaviors or decisions that because of situational factors typically involve deliberate and rule-based thinking. Also, there may be various decision situations that require or at least encourage people to think very precisely about their uncertainty (see Budescu & Wallsten, 1995; Wallsten, 1990). For such situations, it may be that verbal measures of uncertainty, because of their vagueness, are less informative of internal representations than numeric measures.

Researchers may prefer numeric measures because they may be more appropriate when study participants are expected to readily use numeric representations of uncertainty themselves. Hence, with statisticians, seismologists, decision scientists, or financial forecasters, it seems reasonable for them to respond with odds ratios or other numeric responses because they readily think in such terms.

Conclusions and Implications

Numeric measures have been a favored method for assessing people's uncertainty. As discussed above, there is a broad spectrum of research for which numeric measures of uncertainty are entirely appropriate. Nevertheless, we consider the present experiments to be important demonstrations that verbal measures of uncertainty can hold advantages over numeric measures. Likewise, we expect that the spectrum of research for which verbal measures or other alternative measures are appropriate may also prove to be quite broad. For example, consider that uncertainty measures are often key means of assessing constructs such as people's perceived vulnerability to disease and misfortune, their illusions of control, their perceptions of product reliability, their expectations regarding the interpersonal behavior of others, their confidence in causal hypotheses, and their expectations of success on challenging academic tasks. Situations in which such constructs are important are often not characterized by deliberate and rule-based thinking, and the participants in research about such constructs are not statisticians, professional forecasters, or psychological scientists, but rather teenagers in high-risk groups, consumers, students, and others in the general population. Hence, for areas of research involving these types of constructs, verbal measures should be considered as an alternative and possibly more informative method of assessing human uncertainty.

Researchers who are testing theories in which psychological uncertainty is construed as a mediating variable should be especially sensitive to the consequences of numeric versus verbal measures. Statistical tests of mediation are signifi-

[8] Zimmer (1984) also reported that clerks giving verbal predictions were more accurate. This is an intriguing result, but it did not address the issue of whether verbal predictions or numeric predictions better reflected the clerks' feelings of uncertainty. Also, it appears that the accuracy issue was complicated by the fact that the accuracy scores of all clerks were dependent on a single economic result.

cantly harmed by any failure of the measure of the mediator to effectively capture the underlying construct. There may be theoretical models of human behavior that become more informative when alternative measures of uncertainty are explored.

Finally, researchers who are manipulating variables that they suspect affect human uncertainty may underestimate or even fail to detect important effects when they rely only on numeric measures. For example, consider a researcher who is using numeric measures of uncertainty to assess the effectiveness of personal testimonials in changing adolescents' perceptions of vulnerability to sexually transmitted diseases. Such testimonials might have a substantial and consequential effect on adolescents' associative and intuitive feelings of vulnerability. However, numeric measures of vulnerability (i.e., uncertainty) might underestimate the effect because they prompt the adolescents to think about the prevalence rates they had previously learned in sex education classes.

References

Abelson, R. P. (1994). A personal perspective on social cognition. In P. G. Divine, D. L. Hamilton, & T. M. Ostrom (Eds.), *Social cognition: Impact on social psychology* (pp. 15–37). San Diego, CA: Academic Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Becker, M. H. (Ed.). (1974). The health belief model and personal health behavior. *Health Education Monograph, 2,* 324–508.

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

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

Bruner, J. (1986). *Actual minds, possible worlds.* Cambridge, MA: Harvard University Press.

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

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

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation, 32,* 275–318.

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

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

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dane, F. C. (1985). In search of reasonable doubt: A systematic examination of selected quantification approaches. *Law and Human Behavior, 9,* 141–158.

Dovidio, J., & Fazio, R. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about survey questions.* New York: Russell Sage.

Epstein, S. (1990). Cognitive–experiential self-theory. In L. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 165–192). New York: Guilford Press.

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology, 62,* 328–339.

Erev, I., & Cohen, B. L. (1990). Verbal versus numerical probabilities: Efficiency, biases, and the preference paradox. *Organizational Behavior and Human Decision Processes, 5,* 1–18.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review, 98,* 254–267.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17,* 295–314.

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, and Cognition, 18,* 855–864.

Howell, D. C. (1987). *Statistical methods for psychology* (2nd ed.). Boston: PWS-KENT.

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.

James, W. (1950). *The principles of psychology.* New York: Dover. (Original work published 1890)

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263–291.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive–experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology, 63,* 534–544.

Langer, E. J. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43,* 431–442.

Miller, D. T., Turnbull, W., & McFarland, C. (1989). When a coincidence is suspicious: The role of mental simulation. *Journal of Personality and Social Psychology, 57,* 581–589.

Miller, D. T., Turnbull, W., & McFarland, C. (1990). Counterfactual thinking and social perception: Thinking about what might have been. *Advances in Experimental Social Psychology, 23,* 305–331.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22.

Sutherland, H. J., Lockwood, G. A., Tritchler, D. L., Sem, F., Brooks, L., & Till, J. E. (1991). Communicating probabilistic information to cancer patients: Is there "noise" on the line? *Social Science and Medicine, 32,* 725–731.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401–409.

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica, 68,* 27–38.

Vygotsky, L. S. (1987). Thinking and speech. In R. W. Rieber & A. S. Carton (Eds.)., *The collected works of L. S. Vygotsky. Vol. 1: Problems of general psychology.* New York: Plenum Press. (Original work published 1934)

Wallsten, T. S. (1990). Measuring vague uncertainties and understanding their use in decision making. In G. M. von Furstenberg (Ed.), *Acting under uncertainty* (pp. 377–398). Norwell, MA: Kluwer Academic.

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

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

Wilson, T. D. (1990). Self-persuasion via self-reflection. In J. M. Olson & M. P. Zanna (Eds.), *Self-inference process: The Ontario Symposium* (pp. 43–67). Hillsdale, NJ: Erlbaum.

Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude–behavior consistency. *Journal of Personality and Social Psychology, 47,* 5–16.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude–behavior consistency: The disruptive effects of explaining why we feel the way we do. *Advances in Experimental Social Psychology, 22,* 287–343.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60,* 181–192.

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

Zimmer, A. C. (1984). A model for the interpretation of verbal predictions. *International Journal of Man–Machine Studies, 20,* 121–134.

(Appendixes follow on next page)

Appendix A

Scenarios and Their Versions

Randy Scenario From Experiment 1

Version A

The 50 classrooms at Williams University are cleaned each week by work–study students. Randy cleans an average of 30 per week, Amy cleans 7, Laura cleans 5, Matt cleans 5, and Sylvia cleans 3. They continually rotate who cleans what rooms. A teacher is trying to find out who cleaned a particular lab classroom several weeks ago. How likely is it that Randy cleaned the lab classroom in question?

Version B

The 50 classrooms at Williams University are cleaned each week by 2 work–study students. Randy cleans an average of 30 per week, Sylvia cleans an average of 20. They continually rotate who cleans what rooms. A teacher is trying to find out who cleaned a particular lab classroom several weeks ago. How likely is it that Randy cleaned the lab classroom in question?

Baseball Scenario From Experiment 1

Version A

Suppose that in his 20th game of the 1994 major league baseball season, Thomas Whitmore of the Houston Astros is hitting .334. He also has gotten a hit for his last two times at bat. How likely is it that Whitmore would get a hit at his next at-bat?

Version B

Suppose that in his 20th game of the 1994 little league baseball season, Tommy Whitmore of the Subway Astros is hitting .334. He has also gotten a hit for his last two times at bat. How likely is it that Tommy would get a hit at his next at-bat?

Class Presentation Scenario From Experiment 3

Imagine that in one of your classes, each student is required to give one class presentation. At the beginning of the semester, the professor explained that everyone should have their presentation ready to give on the same due date. For each class period, starting on that due date, 2 of the class's 20 students will be randomly selected to give their presentation. On the day before the due date, you have not even prepared for your presentation.

Uncertainty Question

How likely is it that you will be selected to give your presentation on the due date?

Behavioral Intention Question

How many hours would you spend preparing for the presentation?

0 1 2 3 4

Vandal Scenario From Experiment 3

Imagine that you are a high school principal. One of your school's brightest students has been accused of vandalizing the school. The student's homeroom teacher has told you that the student isn't the type to do such a thing. The police have concluded, based on techniques that are 96% accurate, that the student's fingerprints match those found on many pieces of vandalized material. The student has only a weak alibi for where he was on the night of the vandalization. Punishment for vandalizing is expulsion.

Uncertainty Question

How likely is it that the student did vandalize the school?

Behavioral Intention Question

How much more evidence would you need before you would consider him a vandal?

a. No more evidence
b. A bit more evidence
c. Some more evidence
d. A lot more evidence
e. Much much more evidence

Appendix B

Verbal and Numeric Uncertainty Scales as They Appeared in Experiments 1 and 3

Verbal	Numeric (%)
__ Certain	__ 100
__ Almost totally certain	__ 95
__ Extremely likely	__ 90
__ Very likely	__ 85
__ Quite likely	__ 80
__ Likely	__ 75
__ Rather likely	__ 70
__ Fairly likely	__ 65
__ Somewhat likely	__ 60
__ Slightly likely	__ 55
__ As likely as is unlikely	__ 50
__ Slightly unlikely	__ 45
__ Somewhat unlikely	__ 40
__ Fairly unlikely	__ 35
__ Rather unlikely	__ 30
__ Unlikely	__ 25
__ Quite unlikely	__ 20
__ Very unlikely	__ 15
__ Extremely unlikely	__ 10
__ Almost totally impossible	__ 5
__ Impossible	__ 0

Each participant saw only one of the two scales displayed above.

Appendix C

Trivia Topics and Base Rates From the Ranking Task of Experiment 2

Topic	Base rate (%)
__ Macintosh computers	57
__ Baseball	70
__ U.S. capitals	24
__ Famous artists	44
__ U2	32
__ Literature	80
__ Bicycles	67
__ American history	92
__ Famous women	58
__ Fitness	19
__ Iowa	13
__ Skiing	77

In the replication of Experiment 2, the base rates were not displayed on the ranking task.

Appendix D

Verbal and Numeric Uncertainty Scales as They Appeared in Experiment 2

Verbal	Numeric (%)
__ Certain	__ 100
__ Extremely likely	__ 90
__ Quite likely	__ 80
__ Rather likely	__ 70
__ Somewhat likely	__ 60
__ As likely as is unlikely	__ 50
__ Somewhat unlikely	__ 40
__ Rather unlikely	__ 30
__ Quite unlikely	__ 20
__ Extremely unlikely	__ 10
__ Impossible	__ 0

Each participant saw only one of the two scales displayed above.

Received December 4, 1995
Revision received May 10, 1996
Accepted May 17, 1996 ■