Title
Why is "Quite Certain" More Informative Than "Slight Possibility"? Information Theoretic Analysis of the Informativeness of Probability Statements

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Author
Nakamura, Kunininori

Publication Date
2007

Peer reviewed

Why is "Quite Certain" More Informative than "Slight Possibility"? Information Theoretic Analysis of the Informativeness of Probability Statements

Kunininori Nakamura (knaka@ky.hum.titech.ac.jp)
Japan Society for the Promotion of Science
Graduate School of Decision Science & Technology, Tokyo Institute of Technology
2-12-1, Ohkayama, Meguro-Ku, Tokyo, 152-8552, Japan

Abstract

This research aims to establish a quantitative relation between probability statements and their "informativeness." Keren & Teigen (2001) proposed the "search for definitive predictions" principle. According to this principle, relatively high or low probabilities are preferred to medium ones because high or low probabilities denote the occurrence or nonoccurrence of a single outcome more strongly than they do medium ones. This research formalizes the judgment of the informativeness of probability statements in terms of the information theory and argues that the search for definitive predictions principle can be interpreted as rational information estimation under the rarity assumption (Oaksford & Chater, 1994). The results of two empirical studies supported this argument.

Keywords: probability statement; informativeness; information theory; the rarity assumption

Introduction

Suppose that you have written a research paper. Although you believe your paper to be worthy of publication, you are not confident to submit it. Therefore, you show the paper to five of your colleagues and ask them whether they think your paper would be published in a journal. Their comments are as follows:

"I am quite certain."
"I have some doubts."
"I am not sure."
"About 90%"
"Fifty-fifty"

As is evident, your colleagues' responses vary. Whose comment do you then consider the most *informative*?

This question concerns the manner in which people obtain information from probability statements. In everyday life, we communicate information about uncertainty by using probability statements. Probability refers to a numerical index ranging from 0 to 1, representing the degree of uncertainty of a future event. Probability statements provide information about such uncertainty. Thus, people estimate how informative probability statements are and make decisions based on the most informative probability statement. Then, how do people estimate the "informativeness" of probability statements? The purpose of this research is to answer these questions.

Psychological studies on probability judgment have focused on whether people understand the axioms of probability. A majority of the findings in Tversky and Kahneman's heuristics and biases research program (e.g., Kahneman, Slovic, & Tversky, 1982) suggested that under a wide range of circumstances, human intuitions are incompatible with the formal probability calculus. A series of studies on calibration (Fischhoff, Slovic, & Lichtenstein, 1977; Lichtenstein, Fischhoff, & Phillips, 1982) examined whether people are good probability assessors, or the extent to which people can quantify uncertainty. Although there have been several attempts to detail the psychological aspects associated with probabilistic statements, the main research interest in probability judgment has concerned the formal (normative) probability calculus.

Keren and Teigen (2001) dealt with a different issue in probability judgment; they investigated whether people evaluate the informativeness of a probability statement. In their experiments, they showed participants pairs of probabilities, for example, "20% vs. 40%, and asked them which of the two probabilities in each pair was more informative." Using this procedure, they examined the preference for probability under conditions in which no context was provided (Experiment 1) and wherein both positive and negative contexts were provided (Experiment 2). In addition, they investigated the preference for the probabilities by employing a Bayesian belief revision framework (Experiment 3) and a calibration paradigm (Experiment 4).

They found that participants preferred higher or lower probabilities to medium probabilities. For example, participants judged "20%" and "90%" as more informative than "40%" and "70%," respectively, because "20%" and "90%" are extreme as compared to "40%" and "70%." Second, they found that participants preferred higher probabilities to lower probabilities. For example, when participants were asked to choose the more informative probability between "10%" and "90%," although both probabilities were equally extreme, they perceived "90%" as more informative than "10%."

Based on the results of the four experiments, Keren and Teigen proposed that people's perception of the informativeness of probability statements adhere to "the search for definitive predictions" principle. According to this principle, relatively high or low probabilities are preferred to medium ones, because high or low probabilities denote the occurrence or nonoccurrence of a single outcome more strongly than do medium ones. In addition, high probabilities are often favored over complementary low probabilities based on their capacity to predict the occurrence of single outcomes.

Keren and Teigen (2001) positioned "search for definitive predictions" principle as a lay theory of probability, similar to the lay theories of physics (e.g., McCloskey, 1983) or implicit personality theories (e.g., Chiu, Hong, & Dweck, 1997). However, we present one empirical question: Is this principle really a lay theory of probability?

This paper formalizes the amount of informativeness in terms of information theory. In this formalization, this paper also proposes that the “search for definitive predictions” principle is the result of rational estimation based on the amount of information when one assumes the rarity of the target event (rarity assumption: Oaksford & Chater, 1994). Two empirical studies were performed, and both the studies supported this proposition.

The rarity assumption, Kullback–Leibler divergence, and “search for definitive predictions” principle

Consider the following two scenarios in which you observe the results of a coin being flipped. You assume the coin to be fair and observe ten coin flips. In one scenario, both heads and tails appear five times. In the other scenario, heads appears eight times; tails, twice. Which result do you perceive as more “informative”?

Perhaps the latter is more informative because it does not correspond to your belief that the coin is fair. In the first scenario, the results match your belief perfectly, and thus, you need not change it. Thus, this result is not informative because you already knew the outcome. In the second scenario, however, your belief cannot explain the result. Thus, this result is informative because it reveals the incompleteness of your previous belief.

These two scenarios show that the amount of information can be represented as the difference between one’s previous belief and the actual outcome of events. In fact, the amount of information is quantified as the difference between two probability distributions: one corresponds to prior knowledge; the other, to the distribution of the outcomes. In other words, when outcomes match your expectation, they are not informative because they do not require you to change your prior opinion about them. In contrast, when outcomes contradict your expectation, they are very informative because they point out the inadequacies of your prior belief. Thus, the amount of information can be represented as how “surprising” the information is in terms of prior knowledge.

One of the indexes that can represent this “surprisingness” quantitatively is the Kullback-Leibler divergence (Kullback & Leibler, 1951; hereafter, KL divergence). It is expressed as

$$D(P \,\|\, Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} , \qquad (1)$$

where P and Q represent probability distribution; p_i, and q_i represent the probabilities of the events that P and Q predict; and i denotes the number of events.

Next, consider a situation in which you are gathering information in order to predict the outcome of an event. You have prior knowledge about the occurrence of the event. Some information might indicate a high probability for the occurrence of the event, while other information might denote low probability. How then does the amount of information vary according to your prior knowledge and the probability indicated by the information? In this case, a binomial event can be assumed with regard to your prediction. Thus, KL divergence is expressed as:

$$D(P \,\|\, Q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} , \qquad (2)$$

where p denotes the prior probability for the occurrence of the outcome and q denotes the observed probability.

Figures 1a and b demonstrate how KL divergence would vary with p and q. Figure 1a shows a 3D plot of KL divergence: the x-axis denotes p; the y-axis, q; the z-axis, KL divergence. Figure 1a indicates that KL divergence would increase with the difference between p and q; when $p = q$, the value of KL divergence is zero, and it increases with the difference between p and q.

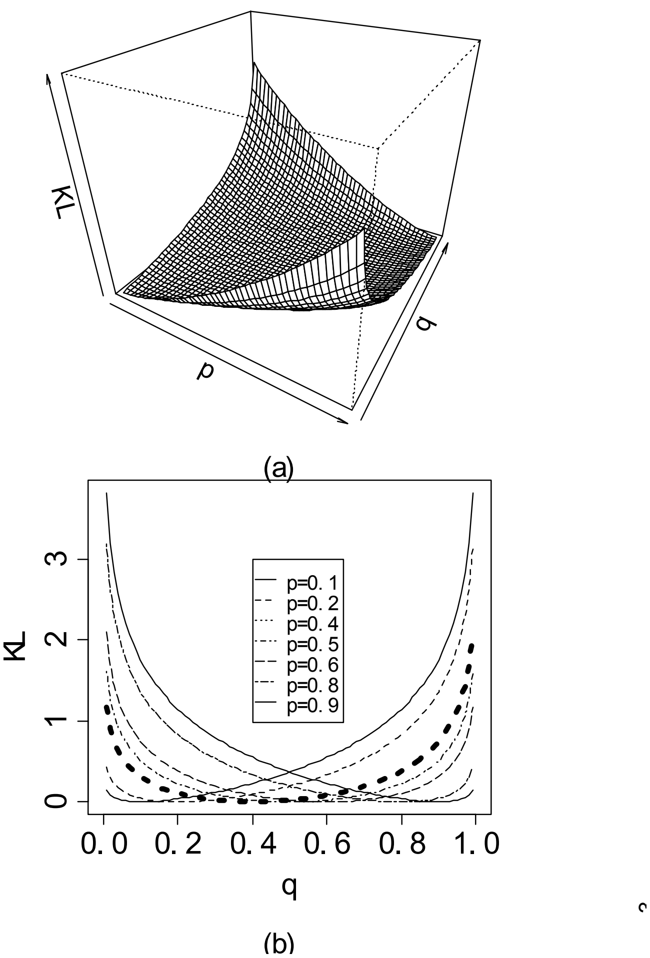


Figure 1: KL divergence as a function of p and q. (a) is a 3D plot of KL divergence, and (b) illustrates a 2D plot of KL divergence; the vertical axis denotes KL divergence and the horizontal axis denotes the q value. The thick dashed line in Figure 1b shows the performance of KL divergence as a function of q when p is 0.4.

Figure 1b displays how the relation between KL divergence and q varies according to the value of p. It also shows that a principle for definitive prediction emerges when p is relatively low. For example, see the line when p is 0.4. When q is 0.4 in this line, the value of KL divergence is zero. In addition, it increases as the value of q deviates from that of p. This trend appears to correspond to the preference for extreme probability. In addition, overall, the value of KL divergence is larger when q is larger than p, than when q is smaller than p. This trend appears to reveal the preference for high probability. Thus, the trend of KL divergence when q is relatively low appears to represent a principle for definitive prediction. This discussion shows that the search for definitive predictions emerges when p is relatively low, that is, below 0.5.

Further, if we assume the rarity assumption proposed by Oaksford and Chater (1994), we can interpret the principle for definitive prediction as a rational information calculation. The rarity assumption entails that people believe in the rarity of the target event. It holds that when estimating the informativeness of a probability statement, one assumes low probability for the occurrence of the target event being referred to. In other words, the rarity assumption denotes a low value of p.

Several empirical studies have supported the psychological validity of the rarity assumption. For example, McKenzie, Ferreira, Mikkelsen, McDermott, and Skrable (2001) provided empirical evidence supporting the rarity assumption. They found that participants tend to phrase conditional statements (or hypotheses) in terms of rare, rather than common, events. Thus, people might regard mentioned confirming observations to be the most informative, or they may consider turning over the mentioned cards most informative, because they usually are most informative, at least from a Bayesian perspective. In addition, Klayman and Ha (1987) proposed the "minority assumption," according to which one presupposes a low probability for the target hypothesis when performing hypothesis testing.

In sum, the above discussion shows that the "search for definitive predictions" principle might result from the normative computation of information under the rarity assumption. The purpose of this research is to test this assertion. We examine whether people's assessment of the informativeness of a probability statement would follow the information theory under the rarity assumption.

Numerical and Verbal probability

We use both numerical and verbal probability statements as stimuli in this research. In daily life, probability information is stated through verbal expression, by using phrases like "very certain" or "quite impossible." Many studies (e. g., Budescu, Karelitz, & Wallsten, 2003) have pointed out that verbal probability possesses features that numerical probability ("50%") does not. Thus, it is possible that the informativeness of verbal probability statements does not follow the same principle as numerical probability statements do.

To examine this possibility, we measured the informativeness of both numerical and verbal probability statements and examined whether the assessment of informativeness might differ between these two kinds of probability statements.

Overview of this study

This research comprised two empirical studies. Study 1 aimed to establish a quantitative relationship between probability and informativeness. Keren and Teigen (2001) used the pair-comparison method; hence, they could not establish such a relationship between informativeness and probability statements. Study 1 required participants to estimate both the informativeness and subjective probability denoted by probability statements.

Study 2 explored the relation between informativeness and the rarity assumption. This paper states that the "search for the definitive predictions" principle would emerge under the rarity assumption about the outcome. Thus, we can predict that the "search for definitive predictions" principle might not hold when the rarity assumption is manipulated. Study 2 intended to investigate how the judgment of informativeness would vary when the rarity of the outcome was manipulated.

We used three types of response scales as dependent variables throughout the two studies: "how valuable," "how informative," and "how surprising." The first two scales are used in the study by Keren and Teigen (2001). The third scale can also be considered as an index of the informativeness of probability statements because KL divergence is often considered to be a measure of how "surprising" the information is.

Study 1

The purpose of Study 1 is to investigate the relation between the estimated probabilities and the informativeness of the probability statements. In Study 1, participants were required to evaluate the informativeness of the probability statements on 8-point scales. They also simultaneously estimated the subjective probability of the probability statement.

Participants One hundred and forty-nine Japanese undergraduates participated in Study 1 as a course requirement.

Stimuli We used 16 verbal probability expressions and 11 numerical probability statements as stimuli. The verbal probability statements are shown in Table 1; the 11 numerical probabilities statements used were 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%.

Procedure The participants were tested in a group setting. Each participant was given a booklet containing instructions that described the experimental task and the abovementioned probability statements.

Table 1: The 16 verbal probability statements used in Studies 1 and 2.

Faint possibility	Never
Slight possibility	Almost impossible
Slight chance	Not likely
A possibility	Not very likely
Possible	Uncertain
Quite possible	A little uncertain
Good possibility	Not certain
Quite certain	Slightly uncertain

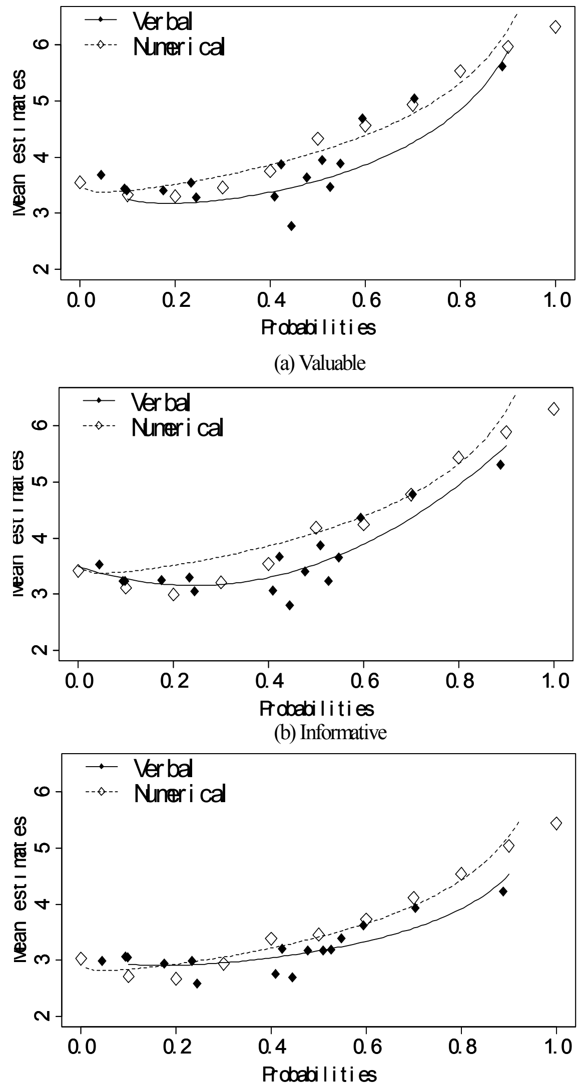
The participants were required to read the verbal probability statements shown in Table 1. The participants were required to read these and answer the following three questions by using an 8-point scale: (1) "How valuable do you consider this statement?"; (2) "How informative is this statement?"; and (3) "How surprising do you find this statement?" After evaluating the 16 verbal probability statements, the participants estimated the probability of each statement by answering the question "In your opinion, what degree of probability does this expression denote?" Finally, the participants estimated 11 numerical probabilities by answering the same three questions, using an 8-point scale for each.

Results and discussion Figure 2 shows the results of Study 1. It illustrates that the relation between the estimated probabilities and each of the three scales appears to be nonlinear. This trend appears to correspond to the prediction from information theoretic analysis. To test our prediction, we constructed a descriptive model of the informativeness judgment expressed by the following equation:

$$Inf = a + b\left(p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}\right) + e, \quad (3)$$

where *Inf* denotes the estimated informativeness of the probability statement; *a*, *b*, and *e* are free parameters—each denoting the intercept and weighting parameter, respectively; *p* is the prior probability; and *q* indicates the estimated probability of the probability statement. We fit equation (3) to the results of both the numerical and verbal probability and estimated *p* and the fit of the model. When we fit equation (3) to the verbal probability data, we input the mean estimated probabilities as *q* and input the values of the numerical probabilities themselves (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1) as *q* for the numerical probability data.

The results shown in Table 2 supported our hypothesis that informativeness judgment can be interpreted as a computation of the amount of the information under the rarity assumption. The *p* parameters for both the numerical (0.00–0.05) and verbal probabilities (0.17–0.19) indicate relatively low values. Although the *p* values differed between the numerical and verbal probability statements, these results correspond to the rarity assumption that people



(a) Valuable

(b) Informative

(c) Surprising

Figure 2: Results of Study 1. (a), (b), and (c) show the results for "valuable," "informativeness," and "surprising," respectively.

Table 2: Results of the parameter estimations in Study 1.

	Verbal			Numerical		
	Valuable	Informative	Surprising	Valuable	Informative	Surprising
a	3.18	3.36	2.91	3.38	3.06	2.82
b	1.94	1.84	1.10	1.45	1.38	1.20
p	0.19	0.19	0.17	0.05	0.00	0.05
R^2	0.74	0.77	0.76	0.96	0.94	0.96

believe in the rarity of the target event. In addition, the descriptive model achieved a good fit of the data for all three scales. Thus, we can conclude that the results of Study 1 consistently supported our hypothesis that the informativeness judgment can be interpreted as a computation of the amount of the information under the rarity assumption.

Study 2

We formalize informativeness judgment as a computation of the information number represented by KL divergence and argue that the principle of definitive prediction emerges under the rarity assumption of the event. Our formalization entails that informativeness judgment depends on the belief about the target event. Thus, if the background belief about the target event changes, the informativeness judgment might also vary according to the change in the belief.

Study 2 aims to test this possibility. To accomplish this, the participants in Study 2 were assigned one of two conditions in which the prior probabilities of the target events were manipulated. The first condition was the "minor disease" condition, wherein which participants estimated the informativeness of the probability statement, supposing that they were told about the possibility of their recovering from a cold. The second condition was the "serious disease" condition," in which they were required to imagine the possibility of their recovering from the Gerstmann-Sträussler-Scheinker syndrome (GSS), which is known to be a serious disease. We can assume that the first condition implies a high probability of recovery, whereas the second condition implies a low probability of recovery. Study 2 examines the effect of this manipulation of the prior probability on the judgment of informativeness.

Participants and procedure Two hundred and fifty-one undergraduates participated in Study 2. The procedure was almost the same as that in Study 1, except that Study 2 comprised two conditions. In both conditions, the participants were required to perform the same task as in Study 1. However, the 121 participants assigned to the "easy disease" condition were instructed to imagine that the verbal probability statements referred to the possibility of their recovering from a cold, and the 130 participants assigned to the "hard disease" condition were told that the probability statements referred to the possibility of their recovering from GSS.

Result and discussion Figure 3 presents the results of Study 2. We performed the same analysis as in Study 1 and examined whether the p values would differ between the two conditions. The results shown in Table 3 demonstrate that the estimated p values were lower in the serious disease condition (0.00–0.53) than in the minor disease condition (0.00–0.93). Specifically, the result of the surprisingness judgment matched our prediction very well. These results supported the prediction that the judgment of informativeness reflects the prior probability of the occurrence of the target event.

These results reveal two interesting findings. The first is that there is a difference in the shape of the function among the dependent variables. The p value estimated from the surprisingness data was remarkably different from those for informativeness and valuableness. The second is a difference in the trend between the verbal and numerical probabilities in the judgment for valuableness and informativeness. These findings suggest the uniqueness of the surprisingness judgment and a difference between the verbal and numerical probability statements. However, the effect of the manipulation of the rarity of the target event itself appeared consistently through all the conditions. Thus,

Table 3: Results of the parameter estimations in Study 2.

		a		b		p		R^2	
		Minor	Serious	Minor	Serious	Minor	Serious	Minor	Serious
Verbal	Valuable	3.39	3.57	2.41	2.60	0.41	0.31	0.83	0.51
	Informative	2.93	3.26	2.12	2.33	0.30	0.23	0.82	0.53
	Surprising	2.43	3.98	2.42	2.17	0.91	0.53	0.81	0.80
Numerica	Valuable	3.96	3.89	0.34	0.92	0.00	0.00	0.62	0.94
	Informative	3.13	3.46	0.75	1.12	0.00	0.00	0.95	0.93
	Surprising	3.70	4.67	0.72	0.57	0.96	0.42	0.76	0.94

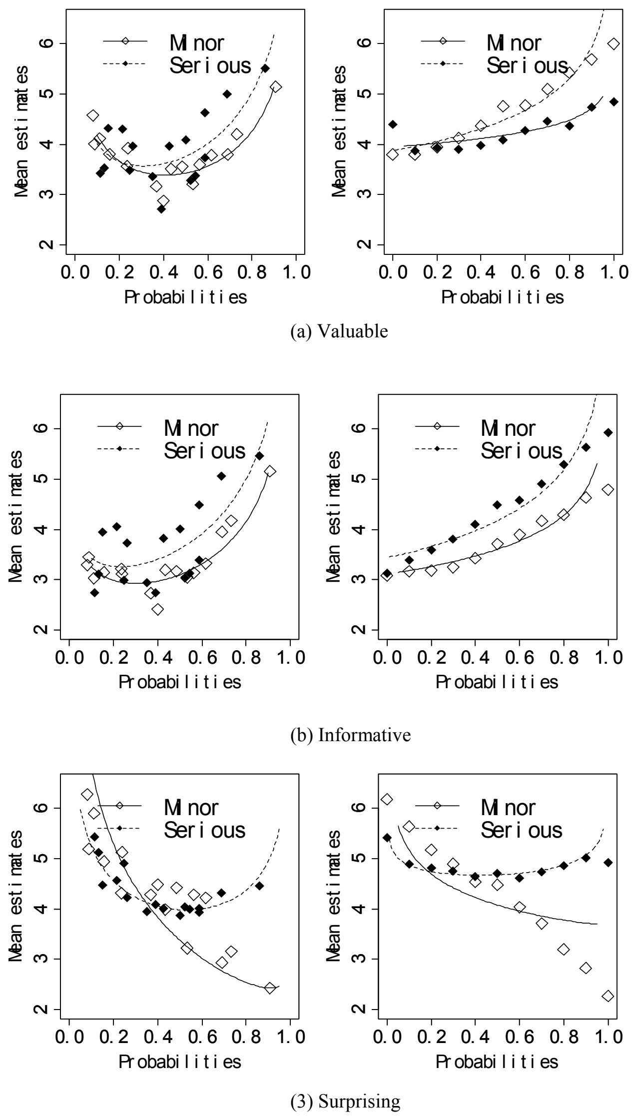


Figure 3: Results of Study 2. The plot shows the mean rating for the informativeness judgment measured by the three questions. The graphs on the left-hand side show the results of the verbal probabilities; those on the right show the results of the numerical probabilities.

we can conclude that the informativeness of the probability statements depends on the prior belief about the occurrence of the target event.

Conclusion

The main contributions of this paper can be summarized in the following two points. First, it establishes the quantitative relationship between probability statements and their informativeness. Although Keren and Teigen (2001) proposed the principle of definitive predictions, their proposition was based on the results of a pair-wise comparison of the probability statements; thus, the amount of informativeness that people estimate for each probability statement is not so clear. This study measured the informativeness of each probability statement by using a Likert-type response scale and showed that the relationship between informativeness and probability appears to be nonlinear. In addition, this research shows that informative judgment for both the verbal and numerical probability statements follow the same principle.

Second, this paper demonstrates that people's perception of the informativeness of the probability statements obeys information theory. This research fitted the model that represents a computation of KL divergence, and the results consistently provided a good fit to the data. In addition, we found that the pattern of the informative judgment varies according to the rarity of the event. The results of this research consistently support the proposition that participants' judgments of the informativeness of probability statements followed information theory.

An important assumption in this research is the rarity assumption (Oaksford & Chater, 1994). Various studies on human reasoning, such as hypothesis testing (Klayman & Ha, 1987), deduction (Oaksford & Chater, 1994), and covariation assessment (Mckenzie & Mikkelsen 2007) have supported the validity of the rarity assumption. This research also reveals the normative sense underlying the principle of definitive predictions under the rarity assumption. It is clear that, in future research, the validity of the rarity assumption must be considered more precisely. However, this research can be cited as another example that supports the rarity assumption.

In conclusion, this paper uncovers the rationality underlying people's intuition toward probability information. Previous studies on judgment and decision-making (e.g., Gilovich, Slovic, & Kahneman, 2002) have emphasized people's deviation from the normative axiom of probability. However, recent approaches have revealed that seemingly irrational behavior makes normative sense when the usual environmental contexts are taken into account (e.g., Mckenzie, 2005). This research is consistent with these approaches. Informativeness judgment is justifiable as optimal data selection (Oaksford & Chater, 1994). We are certainly not claiming that people do not deviate from normative rationality, nor are we providing a process model of informativeness judgment. However, we propose that the normative theory, if it is combined with background knowledge such as the rarity assumption, can help to explain the rationale behind people's behavior.

Acknowledgments

This study was supported by a grant from the Japan Society for the Promotion of Science. I would like to thank Akiko Sakamoto for her help in collecting the data, and the members of Kimihiko Yamagishi's lab for their helpful comments.

References

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

Chiu, C., Hong, Y., & Dweck, C. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality & Social Psychology, 73*, 19–30.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 552–564.

Gilovich, T., Slovic, P., & Kahneman, D. (2002) *Heuristics and biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Keren, G., & Teigen, K. H. (2001). Why is p = .90 better than p = .70? Preference for definitive predictions by lay consumers of probability judgments. *Psychonomic Bulletin and Review, 8*, 191–201.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22* 79–86.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge: Cambridge University Press.

McCloskey, M. (1983, February). Intuitive physics. *Scientific American,* 122–130.

McKenzie, C. R. M. (2005). Judgment and decision making. In K. Lamberts & R. L. Goldstone (Eds.), *Handbook of cognition* (pp. 321–338). London: Sage Publications.

McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P. (2001). Do conditional hypotheses target rare events? *Organizational Behavior and Human Decision Processes, 85*, 291–309.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*, 33–61.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608–631.