Organizational Behavior and Human Decision Processes
Vol. 87, No. 1, January, pp. 136–155, 2002
doi:10.1006/obhd.2001.2959, available online at http://www.idealibrary.com on IDEAL®

Proximity, Compatibility, and Noncomplementarity in Subjective Probability

Kimihiko Yamagishi

Graduate School of Decision Science and Technology,
Tokyo Institute of Technology, Tokyo, Japan

Subjective probability judgments of two mutually exclusive and exhaustive events were investigated. Previous research has documented binary complementarity, that is, such judgments typically sum up to 1.00 in accordance with standard probability theories. However, I argue here that three hypotheses collectively specify conditions wherein systematic binary noncomplementarity is observed. Hypothesis 1 assumes that a possibility that an Example X does not belong to a Category C is assessed by dissimilarity calculation between X and C. According to Hypothesis 2, if X were both similar and dissimilar to C, then the probabilities that X belongs to C and that X does not belong to C would be judged high enough that their sum exceeds 1.00. Hypothesis 3 speculates that such normative contradiction occurs contingent upon task-dependent subjective weighting of relevant features. Experiments 1 through 5 confirmed Hypotheses 1 through 3. Analysis of subjective weight estimates revealed that the compatibility principle (Slovic, Griffin, & Tversky, 1990) provided a coherent account for such violation of probabilistic norms.

Part of this research was supported by Japan Ministry of Education Grant-in-Aid for Encouragement of Young Scientists 10710032. Preliminary reports of this work were presented at the 1997 and 1999 meetings of Subjective Probability, Utility, and Decision Making, and of Japanese Psychological Association, as well as the 1999 annual meeting of Society for Judgment and Decision Making. I thank Maya Bar-Hillel, David Budescu, Craig Fox, Takashi Kusumi, Karl Halvor Teigen, Doug Wedell, and Paul Windschitl for helpful suggestions.

Address correspondence and reprint requests to Kimihiko Yamagishi, Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Ookayama 2-12-1, Meguro, Tokyo 152-8552, Japan. E-mail: kimihiko@ky.hum.titech.ac.jp.

0749-5978/01 $35.00

Human perception of uncertainty has been studied for decades under the rubric of subjective probability (Wright & Ayton, 1994). Among the main findings in the field is binary complementarity, that is, that the probability estimates of two mutually exclusive and exhaustive events sum to 1.00. For more than two events, such complementarity fails to hold in violation of probability theory (Teigen, 1974; Tversky & Koehler, 1994; Koehler, Brenner, & Tversky, 1997). Yet, binary complementarity has been commonly observed (Teigen, 1983; Tversky & Fox, 1995) and is implied by influential theories of subjective probability judgment (e.g., Tversky & Kohler, 1994; Rottenstreich & Tversky, 1997).

Contrary to the theoretical and empirical agreement, however, I argue here that an assumption combined with standard findings in judgmental literature predicts systematic binary noncomplementarity. The purpose of this article is to first test the assumption and then document conditions where judgments would exhibit binary noncomplementarity. Finally, cognitive processes of task-dependent and selective subjective feature weighting are tested as an explanation for such judgmental performance.

EMPIRICAL QUESTIONS

The following three hypotheses constitute the empirical questions pursued in this article. For brevity, I define "affirmation probability" or "Φ_A" as the probability that an Example X belongs to Category C. "Negation probability" or "Φ_N" is defined as the probability that X does not belong to C. A binary noncomplementarity between Φ_A and Φ_N violates probabilistic norms whenever $\Phi_A + \Phi_N > 1$. The rationale for posing these inquiries is discussed before current experiments are introduced. The experiments in this article were conducted to investigate the hypotheses in this order.

Hypothesis 1: Dissimilarity would serve as a heuristic in judging negation probabilities.

Hypothesis 2: If *X* were simultaneously similar and dissimilar *C*, the affirmation and negation judgments of category membership would be high enough to produce $\Phi_A + \Phi_N > 1.00$.

Hypothesis 3: The binary noncomplementarity that $\Phi_A + \Phi_N > 1.00$ occurs contingent upon subjective feature weighting that follows the compatibility principle (Slovic, Griffin, & Tversky, 1990).

PROXIMITY AND LIKELIHOOD JUDGMENT

This section summarizes relevant previous research, introduces an assumption (tested later in this article), and discusses how these ideas would jointly predict binary noncomplementarity. I use the term "proximity" to refer to both similarity and dissimilarity at the same time.

The conjunction fallacy (Tversky & Kahneman, 1982; Tversky & Kahneman, 1983) is a well-known logical contradiction in probabilistic reasoning. "Linda" is described as "31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations."

Tversky and Kahneman's participants judged her as more likely to be a feminist bank teller than a bank teller. This pattern of reasoning exemplifies an error as a probabilistic reasoning because the extension of "bank teller" should be greater than "feminist bank teller." Therefore the likelihood that Linda belongs to the former is greater than that she belongs to the latter.

Tversky and Kahneman's (Tversky & Kahneman, 1982; 1983) account of the conjunction fallacy relies on the notion that participants base their probability judgment on the similarity between Linda's profile and social stereotypes. Because most of Linda's characteristic features are shared commonly with typical feminists, Linda seems more similar to a feminist bank teller than a bank teller. As Tversky (1977) had previously claimed, the existence of such shared features increases the similarity between two objects. The conjunction fallacy indicates that similarity judgments frequently substitute for likelihood judgments (Tversky & Kahneman, 1982). Tversky and Kahneman (1982) argued that this use of similarity in likelihood judgment is a case of a general "representativeness heuristic" in probability judgments (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974).

Although the conjunction fallacy may be coherently explained by Tversky's (1977) feature contrast model, the model was developed to account for a variety of similarity judgment phenomena. Hence, the feature contrast model reveals other interesting aspects in similarity judgments. One aspect includes discrepancies in magnitudes between similarity and dissimilarity. Tversky claimed that proximity calculation between two objects was reached by integrating contributions of common and distinct features in a task-dependent fashion. In the process of integration, common features are weighted heavily in similarity calculation, whereas distinct features are weighted heavily in dissimilarity calculation. The model predicts that a pair of familiar objects (with numerous common features and distinct features) and paired unfamiliar objects (with fewer common features and fewer distinct features) could be rated so that one pair may be judged as both more similar and more dissimilar than the other. For instance, Tversky's participants rated a China/Japan pair as more similar to each other than a Paraguay/Ecuador pair, yet China versus Japan were rated more dissimilar than Paraguay versus Ecuador.

In short, one of Tversky's (1977) contributions that deserves special attention here is that two objects can be similar and dissimilar simultaneously. In this context of proximity, one can pose the following empirical questions. Suppose an Example X and a Category C are simultaneously similar and dissimilar to each other. Numerous replications have observed (e.g., Nisbett, Zukier, & Lemley, 1981; Agnoli & Krantz, 1989; Reeves & Lockhart, 1993) that the likelihood judgment that X belongs to C is judged by how X is similar to C. Assume further that *dis*similarity serves as a heuristic to judge whether X does *not* belong to C. Under this assumption, the prediction holds that the probability that X belongs to C and the probability that X does not belong to C would be simultaneously high. Thus one could expect binary noncomplementarity regarding likelihood judgments between C and X.

A PROBABILITY JUDGMENT MODEL AND THE COMPATIBILITY PRINCIPLE

A Feature Weighting Model

The reader is now introduced to a model of the previously outlined cognitive process. Figure 1 shows a feature set representation. Example X consists of two mutually exclusive sets of features defined in relation to Category C. Each feature belongs to either Sim or DisSim depending on the diagnosticity of its membership of C. The following formulation provides a descriptive model of probability judgment:

$$\Phi_J = \alpha_J f(\mathrm{Sim}) + \beta_J f(\mathrm{DisSim}), \tag{1}$$

where J = A (affirmation) and N (Negation).

Equation (1) is a special case of the feature contrast model in Tversky (1977) with the additional assumptions introduced previously. The function $f(\bullet)$ expresses contribution of features to a judgment, whereas α_J and β_J denote subjective weights associated with each feature set. Implicit in the formulation in Eq. (1) are the following propositions. First, proximity calculation serves as a heuristic in judgments of the affirmation and negation probability. Next, probability judgment is reached as a monotonic function of proximity judgment.

Equation (1) shares close kinship to a class of preferential process models that are derived from Tversky (1977). Particularly, Tversky, Sattath, and Slovic's (1988) Contingent Weighting model, Shafir's (1993) model of choice-rejection asymmetry, and Yamagishi and Miyamoto's focus shift model (Yamagishi & Miyamoto, 1996; Yamagishi, 1996) commonly presume feature-set representations as in Fig. 1, with feature weighting as the explanatory mechanism. For the current purposes, however, it suffices to mention the closeness of these models. A further comparison is developed later.

The Compatibility Principle

To explain judgmental performance, the model in Eq. (1) needs a principle that directs configuration of the subjective weights. The compatibility principle (Slovic, Griffin, & Tversky, 1990; Fisher & Hawkins, 1993) provides such a guiding script. The compatibility principle "states that the weight of a stimulus attribute is enhanced by its compatibility with the response" (Slovic, Griffin, & Tversky, 1990, p. 5). Consequently, in the affirmation judgment, features that

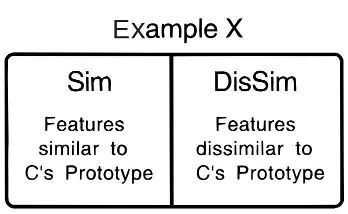

FIG. 1. A feature-set representation of the probability judgment model.

enhance X's similarity to C would be weighted heavily, whereas in the negation judgment, features that strengthen X's dissimilarity would be weighted heavily. The compatibility principle predicts that $\alpha_A > \alpha_N$ and $\beta_N > \beta_A$.

RHONDA AND PHIL

In the experiments reported hereafter, the following personality descriptions were used (except for the omission of the latter in Experiment 5). "Rhonda" and "Phil" were created by adding extra features to the personality descriptions of "Linda" and "Bill," respectively, from the original studies (Tversky & Kahneman, 1982; Tversky & Kahneman, 1983):

> Rhonda is 31 years old, *married young* but currently single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. *She is pro-life, very active in her church, and supports prayer in school.*

> Phil is 29 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities. *He tends to be late for meetings and has little sense of deadlines. He wears tennis shoes at work and his workplace is very disorganized.*

Features added onto the original personality descriptions appear in italics. Rhonda was created by adding features that were contrary to stereotypical "feminist" traits to originally "feministlike" Linda. Phil was created by adding features that were contrary to stereotypical "accountant" traits to originally "accountantlike" Bill. Each feature was selected from a pilot study as indicating characteristics that were dissimilar to those of typical feminists (for Rhonda) or typical accountants (for Phil).

EXPERIMENT 1: CATEGORY RANKING

The purpose of Experiment 1 was to test if dissimilarity assessment mediates judgments of negation probability. I adopted Tversky and Kahneman's (1982) rank ordering task. Tversky and Kahneman's participants rank ordered a set of social categories according to the likelihood that Linda was a member and the similarity of Linda to each category's typical character. The categories used for Linda consisted of Elementary school teacher, Bookstore clerk taking Yoga courses, Feminist, Social worker, League of Women Voters member, Bank teller, Insurance salesperson, and Feminist bank teller. The categories used for Bill included Poker hobbyist Physician, Architect, Accountant, Jazz hobbyist, Surfing hobbyist, Reporter, Jazz hobbyist accountant, and Mountain climber.

Method

Participants. In all experiments reported in this article, participants were undergraduates at the University of Washington. They were enrolled in an introductory psychology course. Each participant was provided with a problem booklet to work on her or his experimental task and other filler items.

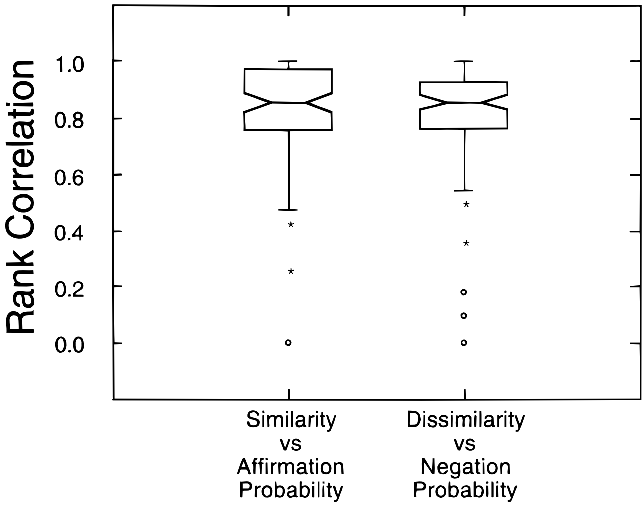

FIG. 2. Notched boxplots between Rhonda's proximity and probability ranks.

Procedure. Four hundred participants were tested in eight groups ranging in size from 5 to 63. They were randomly assigned to four conditions. The recruitment of participants was terminated when the effective sample size of 100 for each condition was collected.

Participants in Condition 1 were presented with Rhonda's personality description and rank ordered the Linda categories according to the affirmation probability that Rhonda was a member. After a 1-week interval, the same participants rank ordered the categories according to Rhonda's similarity. Condition 2 followed the same procedure regarding Rhonda's dissimilarity and negation probability. Participants in Conditions 3 and 4 provided the same kind of rank ordering regarding Phil and the Bill categories. The order of administering the two tasks was counterbalanced across participants.

Results and Discussion

For each participant, a rank correlation coefficient was calculated between rankings of proximity and probability assessments. Figure 2 shows notched boxplots[1] (McGill, Tukey, & Larsen, 1978) of the rank correlation coefficients for Rhonda's case (Conditions 1 and 2). In replication of Tversky and Kahneman's findings, data on the left side show substantial agreement between judgments of affirmation probability and similarity. In turn, data on the right side in Fig. 2 indicate high correlation coefficients between judgments of negation probability and dissimilarity. Likewise, Fig. 3 shows notched boxplots for Phil's case (Conditions 3 and 4). Figure 3 follows the same pattern as Fig. 2, supporting that similarity served as a heuristic for affirmation probability judgments, whereas dissimilarity served as a heuristic for negation probability judgments.

[1] Notched boxplots are drawn in the same way as standard boxplots except for the oblique lines that extend to the upper and lower bounds for the 95% confidence intervals for the median.

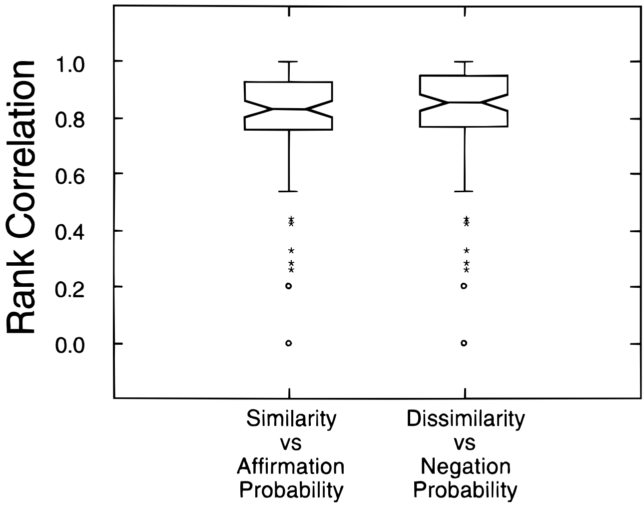

FIG. 3. Notched boxplots between Phil's proximity and probability ranks.

Based on the data on the right side in Figs. 2 and 3, it may be claimed that people substitute dissimilarity judgments for negation probability judgments.

Figure 4 shows scatterplots of mean ranks for the eight descriptions characterizing Rhonda (top) and Phil (bottom). The mean similarity rank shows a positive correlation with the mean rank for the affirmation probability (r = .884 for Rhonda and r = .845 for Phil). Also, the mean dissimilarity rank shows

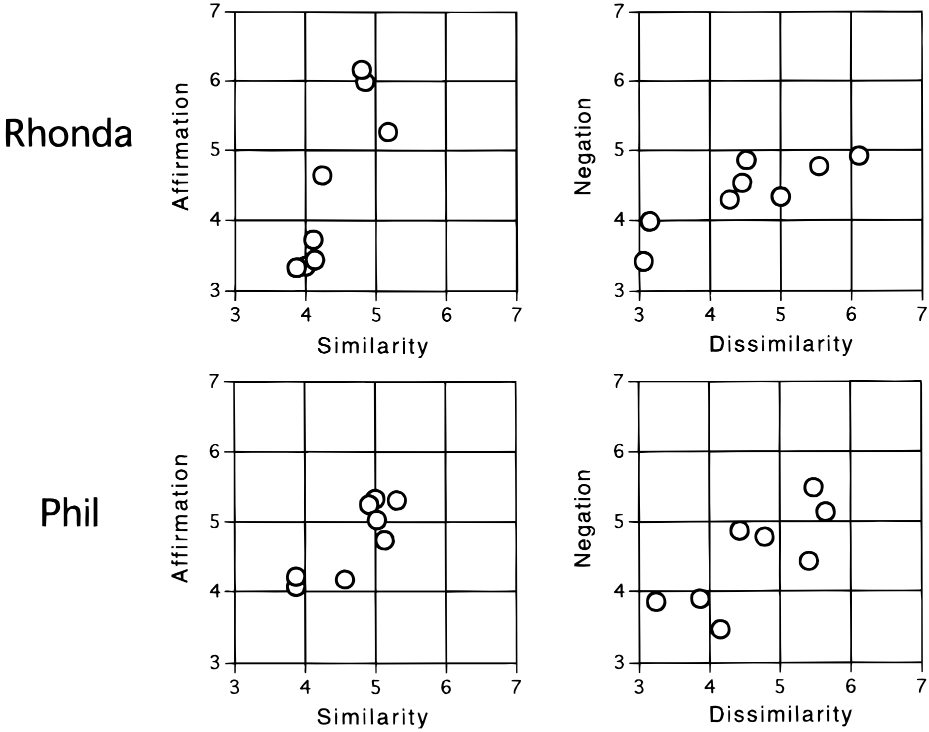

FIG. 4. Scatterplots of mean ranks between probability and proximity judgments.

a positive correlation with the mean rank for the negation probability (r = .854 for Rhonda and r = .768 for Phil). Each of the four correlation coefficients was significantly positive at the .05 level. Figure 4 is informative in two ways: First, in addition to Figs. 2 and 3, Fig. 4 further supports Hypothesis 1. Second, the patterns in the scatterplots suggest linear relationships between judgments of proximity and probability, as the relationships are characterized by Eq. (1).

EXPERIMENT 2: DOES AFFIRMATION RANKING INVERSELY REFLECT NEGATION RANKING?

The results from Experiment 1 may seem trivial if one assumes that, upon reading Rhonda's profile, participants first execute a similarity calculation and then reverse the ordering to produce dissimilarity ranking. Contrarily, the model in Eq. (1) and the compatibility principle postulate that the affirmation and negation probabilities differ as a consequence of task-dependent configuration of subjective weights. Therefore the compatibility principle implies that the ranking of the affirmation and negation probabilities may lack correspondence. Experiment 2 was conducted to contrast these different predictions. If the former interpretation were valid, it should be possible to predict a negation probability ranking based on the information from the affirmation probability (and vice versa). If the model in Eq. (1) and the compatibility principle were in effect, attempts to restore the affirmation probability ranking based on the information from the negation probability would be problematic.

Method

Procedure. Two hundred participants were tested in four groups ranging in size from 6 to 59. They were randomly assigned to two conditions. The recruitment of participants was continued until a sample size of 100 for each condition was achieved. Participants in Condition 5 examined the description of Rhonda and rank ordered the Linda categories according to the affirmation probability. A week later, the same participants provided a negation probability ranking regarding Rhonda and the Linda categories. Participants in Condition 6 followed the same procedure for the description of Phil and the Bill categories. The order of executing the tasks was counterbalanced among participants.

Results and Discussion

It follows from the procedure that each participant provided two rank orders of affirmation and negation probabilities. I calculated the "anticipated" affirmation probability ranking by subtracting the negation probability rank from 9 (the category set consisted of eight categories). Figure 5 shows notched boxplots of rank correlation between the affirmation ranking and the anticipated affirmation ranking. The 95% confidence intervals for Rhonda and Phil lie in the vicinity of .3 through .6, suggesting a positive but insignificant correlation. The critical value of Spearman rank correlation coefficient for eight data points

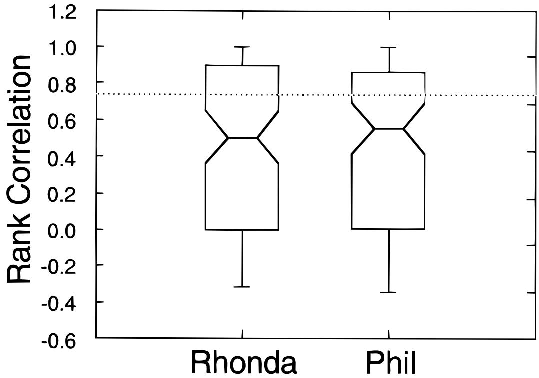

FIG. 5. Notched boxplots of rank correlation between affirmation ranking and anticipated affirmation ranking.

is .738, as shown by the dotted line. Therefore, the correlations between the affirmation and anticipated affirmation rankings suggest that the affirmation ranking cannot be totally restored based on the knowledge of the negation probability. More importantly, the positive correlations provide supportive evidence for the argument that the model in Eq. (1) along with the compatibility principle produced the affirmation and negation probabilities. The data in Fig. 5 support this interpretation because the model postulates that the difference between the affirmation and negation probability judgments stems from the difference in subjective weighting.

The interpretation is supported for the following reason. It holds statistically that two variables generated by different weighting of common variable sets maintain some correlation. Suppose that two variables Q and R are completely orthogonal to each other. In 10 test calculations with 100 data points, the correlation between $2Q + R$ and $Q + 2R$ ranged from .8001 to .8059, with a median of .8012. Analogously, if the cognitive processes underlying the affirmation and negation probability judgments differ in the subjective weighting, an imperfect positive correlation between the two would be expected. Another interpretation of Fig. 5 would be to assume that the affirmation and negation probability judgment processes are “improper linear models” (Dawes, 1988) of each other. Again, two improper linear models typically exhibit a moderate correlation, as shown in Fig. 5.

Yet, an alternative explanation for the result in Fig. 5 needs to be noted. Figure 5 could be observed if the affirmation ranking inversely reflected the negation ranking with low reliability. The current data set does not allow empirical tests of this reliability issue, and Experiment 2 is susceptible to this counterargument until further research is conducted.

EXPERIMENT 3: BINARY CHOICE

Experiments 1 and 2 provided supportive evidence that dissimilarity serves as a heuristic to guide the negation probability judgment, inversely reflecting the standard argument regarding similarity and the affirmation probability judgment (Tversky & Kahneman; 1982, 1983). Experiments 3 through 5 investigated whether the similarity and dissimilarity heuristics jointly produce the binary noncomplementarity as outlined above. Participants in Experiment 3 performed a simple task, namely binary choice. They were asked to compare either Rhonda or Phil to the following: "Pat is 30 years old. He is married with no children. He is well liked by his colleagues. He is generally satisfied with his marriage, but has minor complaints as well."

The description of "Pat" was adopted from Kahneman and Tversky's (1973) "Dick" with a minor modification. Pat (as well as Dick) was designed so that each feature would bear little resemblance to particular societal categories.

Method

Procedure. Experiment 3 was conducted in a large classroom setting for one session. Two hundred fourteen participants were randomly assigned to four conditions. Participants in Condition 7 ($n_7 = 48$) compared Rhonda to Pat and chose who was more likely a feminist. Participants in Condition 8 ($n_8 = 56$) compared Rhonda to Pat and chose who was more unlikely a feminist. In Condition 9, participants ($n_9 = 54$) chose the accountant candidate between Phil and Pat. Finally, in Condition 10 ($n_{10} = 56$) participants chose who was more unlikely an accountant between Phil and Pat. Every participant in Conditions 7 through 10 was instructed to base their judgment on their subjective probability.

Results and Discussion

The majority of participants of Conditions 7 through 9 chose either Rhonda or Phil. Votes from Condition 10 were split into equal halves. The proportion of participants is summarized as follows:

Condition 7 (Affirmation Probability for Feminist) 75.00% chose Rhonda
Condition 8 (Negation Probability for Feminist) 59.26% chose Rhonda
Condition 9 (Affirmation Probability for Accountant) 68.52% chose Phil
Condition 10 (Negation Probability for Accountant) 50.00% chose Phil

Rhonda as a likely and unlikely feminist attracted statistically reliable majority votes from Conditions 7 and 8: The sum of the proportions exceeded 100% ($z = 5.55$, $p < .001$, two-tailed). Correspondingly, the sum of the proportion in favor of Phil as a likely and unlikely accountant exceeded 100% ($z = 2.98$, $p < .01$, two-tailed). Therefore, concerning Hypothesis 2, binary noncomplementarity was noted in the participants' subjective probability, in deviation from a standard probabilistic norm.

EXPERIMENT 4: PROBABILITY ASSESSMENT

Experiment 4 involved a comparison between Rhonda (or Phil) and Pat, as in Experiment 3. Experiment 4 differs from Experiment 3 in two ways. First, judgments of affirmation and negation probabilities were obtained in a within-participant design. Second, participants performed numerical likelihood judgments.

Method

Procedure. One hundred participants were tested in four groups ranging in size from 3 to 42. They were randomly assigned to two conditions. The recruitment of participants was continued until a sample size of 50 for each condition was achieved. Participants in Condition 11 examined the pair of Rhonda and Pat and they were told that either Rhonda or Pat was a feminist. In one session, their task was to indicate the probability that Rhonda would be the feminist. In another session, the same participants provided a negation probability response. An 11-point rating scale was used, wherein the value started at 0% and increased in 10% increments to 100%. Participants in Condition 12 followed an equivalent procedure pertaining to the Phil/Pat pair. The order of the affirmation probability response and the negation probability response was counterbalanced, and there was a 1-week interval between the responses.

Results and Discussion

For each participant, the sum of her or his affirmation probability rating and negation probability rating was calculated. Figure 6 shows notched boxplots of the sum for Rhonda and Phil. For Rhonda, 75% of Condition 11 participants judged her as a likely and unlikely feminist, and the 95% confidence intervals for the sum lie above 100%. Therefore, Condition 11 produced a statistically reliable binary noncomplementarity. With Phil, the confidence intervals included 100%. Still, the majority of Condition 12 participants rated Phil as a likely and unlikely accountant, closely resembling the pattern shown with Rhonda.

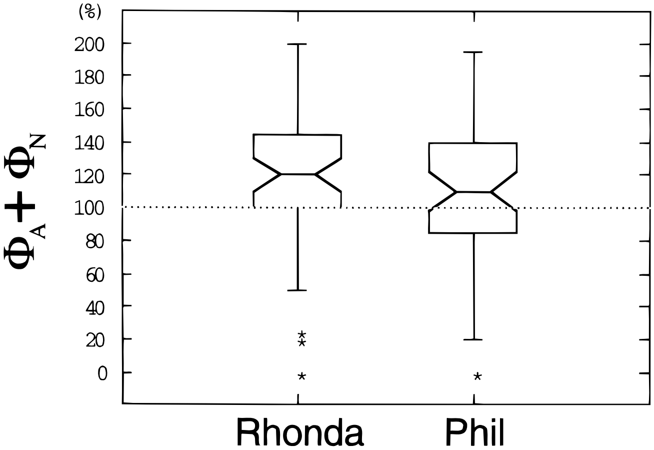

FIG. 6. Notched boxplots of the sum of the affirmation and negation probability rating.

Among Hypotheses 1 through 3, the first gained support from Experiments 1 and 2. Experiments 3 and 4 confirmed the implications that arose from Hypothesis 2. Finally, Experiment 5 was designed to test Hypothesis 3.

EXPERIMENT 5: ESTIMATING SUBJECTIVE WEIGHT PARAMETERS

Experiment 5 aimed at estimating the subjective weighting parameters under judgments of affirmation and negation. I regarded regression coefficients as empirical measurements of the subjective weights. Therefore, Experiment 5 was designed to conduct regression analyses to estimate statistical counterparts for the α and β weights in Eq. (1).

Method

Stimuli and design. Rhonda's stimulus description was broken down into "fixed" and "altered" features. The fixed features consisted of the beginning part of her profile: "Rhonda is 31 years old, married young but currently single, outspoken, and very bright. She majored in philosophy." In contrast, the following feature sets (1) through (4) were treated as the altered features: (1) As a student, she was deeply concerned with issues of discrimination and social justice, (2) She participated in anti-nuclear demonstrations, (3) She is pro-life, very active in her church, and (4) She supports prayer in school.

Clearly, the altered features (1) and (2) belong to the set Sim, whereas the features (3) and (4) to the set DisSim, as in Fig. 1. The fixed features were presented to participants in every trial. The altered features (1) through (4) were either present or absent in the profile and thus generated 16 ($2 \times 2 \times 2 \times 2$) varieties of profiles.

Procedure. Experiment 5 was conducted in a large classroom setting for one session. Experiment 5 required an exact sample size of 224 to assure the predictors in the regression analysis remained orthogonal to each other. Hence, 243 effective samples were collected, and 19 samples, randomly chosen, were discarded from further analysis.

Experiment 5 used a completely between-participant design (16 profile versions $\times$ 2 judgment types). Seven participants were assigned to each profile version. Each participant in Condition 13 ($n_{13} = 7 \times 16 = 112$) read the stimulus description and provided an affirmation probability judgment using the rating scale as in Experiment 4. Participants in Condition 14 ($n_{14} = 112$) provided corresponding negation probability judgments.

Regression Analysis

Estimation of subjective weights. Let α_{A1} and α_{A2} denote the subjective weights for the altered features (1) and (2), respectively, in the affirmation probability judgment. Similarly, let α_{N1} and α_{N2} denote the subjective weights in the negation probability judgment. In parallel, let β_{A3} and β_{A4} express the subjective weights for the altered features (3) and (4) in the affirmation judgment, whereas β_{N3} and β_{N4} express the corresponding weights for the case of negation.

To estimate the parameters, the conceptual model in Eq. (1) takes on the following form:

$$\Phi_J = \alpha_{J1} f(\mathrm{Sim}_1) + \alpha_{J2} f(\mathrm{Sim}_2) + \beta_{J3} f(\mathrm{DisSim}_3) + \beta_{J4} f(\mathrm{DisSim}_4), \quad (2)$$

where J = A (affirmation) and N (Negation). Equation (2) contains four subjective weight parameters and four $f(\bullet)$ variables that correspond to features (1) through (4). In each trial, the altered features (1) through (4) were coded 1 if the feature set was present in the profile and 0 otherwise [i.e., $f(\bullet) = 0$ or 1]. Thusly coded the features served as predictor variables in multiple regression. The predictor variables were regressed on the affirmation responses from Condition 13. This multiple regression produced estimated subjective weight parameters, $\hat{\alpha}_{A1}$, $\hat{\alpha}_{A2}$, $\hat{\beta}_{A3}$, and $\hat{\beta}_{A4}$ (parameter estimates are denoted by "$\wedge$"). The same set of predictor variables was regressed on the negation responses (Condition 14) to obtain $\hat{\alpha}_{N1}$, $\hat{\alpha}_{N2}$, $\hat{\beta}_{N3}$, and $\hat{\beta}_{N4}$.

Calculation of random weights by simulation. The following procedure was carried out to assess how likely it would be that the estimated subjective weights would be obtained by chance alone. I compared the estimated parameters against the distribution of possible weight estimates generated by a Monte Carlo simulation. For the affirmation probability data, every association between the particular configuration of the predictor variables was randomized and reassigned. After this shuffling procedure, a regression analysis calculated a set of regression weights generated by a one-trial simulation. Such regression weights are referred to as "random weights." This simulation procedure was repeated 500 times, generating a set of 500 random weights.

The distribution of random weights indicates the possible range of the corresponding subjective weight if it were obtained solely by chance. Such ranges make it possible to determine if $\hat{\alpha}_{A1}$, $\hat{\alpha}_{A2}$, $\hat{\beta}_{A3}$, and $\hat{\beta}_{A4}$ are distinguishable from the random weights. For the negation probability data set, 500 similar random weights were calculated to compare $\hat{\alpha}_{N1}$, $\hat{\alpha}_{N2}$, $\hat{\beta}_{N3}$, and $\hat{\beta}_{N4}$ against their distributions.

The advantage of the Monte Carlo method is worth noting here. In contrast to the conventional significance test for regression weights, this method allows the researcher to control for the statistical power by the number of calculation of the random weights. Therefore, such an analytic method may be applied to situations where a multicolinearity problem causes sizable lack of statistical power for such significance tests of regression weights. Another possibility might be to examine the relative importance of the individual predictor variables by relying on Dominance Analysis proposed by Budescu (1993). Yet, Dominance Analysis was devised as a strategy for determining relative importance of multicolinear predictor variables. In contrast, Experiment 5 was designed such that all the predictor variables were orthogonal to each other.

Therefore, the current analysis did not call for Dominance Analysis and required no consideration regarding the issues of partial and semipartial correlation.

Results and Discussion

Binary noncomplementarity. Whenever my statistical tests involved means from two conditions, error terms for conventional two independent group *t* tests were used to control for the possibility of Type I errors. In other words, the *MSE* terms from the above-mentioned regression analyses were not used as the variance estimate in constructing the confidence intervals reported hereafter.

The grand mean affirmation probability judgment from Condition 13 was 49.11, whereas the mean negation probability from Condition 14 was 58.75. The 95% confidence intervals for the sum of these two means ranged between 101.51 and 114.21 ($df = 222$, $MSE = 3.22$). Therefore, this main effect showed statistically reliable binary noncomplementarity between the affirmation and negation probabilities.

Figure 7 shows the mean affirmation and negation probabilities in the 16 profile versions. Each row and column indicates the experimental condition

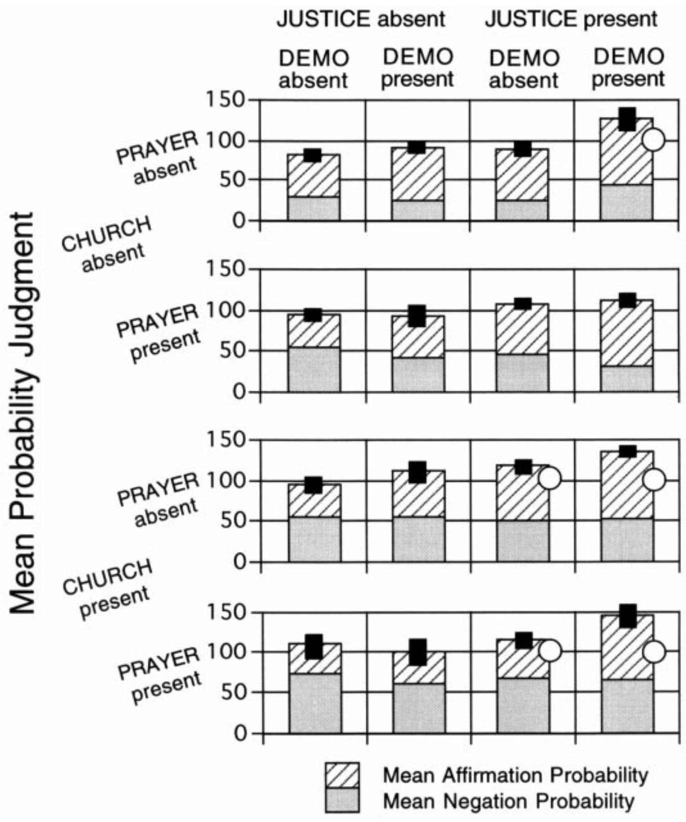

FIG. 7. Mean judged probabilities in the 16 conditions in Experiment 5.

wherein the corresponding feature set was either present or absent. For instance, the rightmost column in the bottom row means that the stimulus used in the cell presented all the features, namely the full description of Rhonda. The solid black bar shows the 95% confidence intervals for the sum of the mean affirmation and negation probabilities within the cell (seven participants from Conditions 13 and 14, $df = 12$ per cell). In five cells, circle symbols appear at the right edge of the 100% line. Each circle indicates that, in the particular cell, the sum of the affirmation and negation probabilities reliably exceeded 100%.

The circles in Fig. 7 concentrate on the lower right quadrangle of the 16 cells. Figure 7 is organized such that there were a greater number of features present with the stimuli used in this quadrangle than were present in other quadrangles. Thus, it may be argued that as more features appear in the profile, the more similar and dissimilar Rhonda sounds as a feminist, hence the more likely and unlikely she is a feminist, producing the binary noncomplementarity.

The Compatibility Principle. The following regression analyses were conducted on the same data set as in the analysis of group means. The compatibility principle predicts that the Rhonda features similar to prototype feminists would attract substantial attention in the affirmation probability but not in the negation probability, therefore the α_A weights would be greater than the α_N weights. Conversely, the features dissimilar to prototype feminists would attract sizable attention in the negation probability but not in the affirmation probability; consequently the β_N weights would be greater than the β_A weights.

The multiple correlation coefficient for the affirmation probability judgment was .59 [$MSE = 333.91$, $F(4, 107) = 14.36$, $p < .001$]. Another regression analysis for the negation probability produced a multiple correlation coefficient of .57 [$MSE = 465.32$, $F(4, 107) = 13.02$, $p < .001$].

The top panel in Fig. 8 shows the estimated and random weights for the affirmation probability judgments. The boxplots show the dispersion of the random weights from 500 simulations, providing the "ground" of the picture. The square symbols constitute the "figure" by indicating the location of $\hat{\alpha}_{A1}$, $\hat{\alpha}_{A2}$, $\hat{\beta}_{A3}$, and $\hat{\beta}_{A4}$. The bottom panel in Fig. 8 shows the estimated and ramdom weights for the negation probability. The square symbols in the bottom panel show the location of $\hat{\alpha}_{N1}$, $\hat{\alpha}_{N2}$, $\hat{\beta}_{N3}$, and $\hat{\beta}_{N4}$. Clearly, the α weights are greater in the affirmation probability than in the negation probability. The weights $\hat{\alpha}_{A1}$ and $\hat{\alpha}_{A2}$ lie above the confidence intervals indicated by the oblique lines, whereas $\hat{\alpha}_{N1}$ and $\hat{\alpha}_{N2}$ are embedded within the confidence limits. Conversely, the β weights are greater in the negation probability than in the affirmation probability. $\hat{\beta}_{N3}$ and $\hat{\beta}_{N4}$ are located above the confidence intervals, whereas $\hat{\beta}_{A3}$ and $\hat{\beta}_{A4}$ are captured within the confidence limits. The data in Figure 8 as a whole are consistent with the compatibility principle by showing the predominance of the α_N and β_N weights, while the incompatible features are essentially ignored in that the α_N and β_N weights are indistinguishable from the random weights.

As a more direct test of the prediction from the compatibility principle,

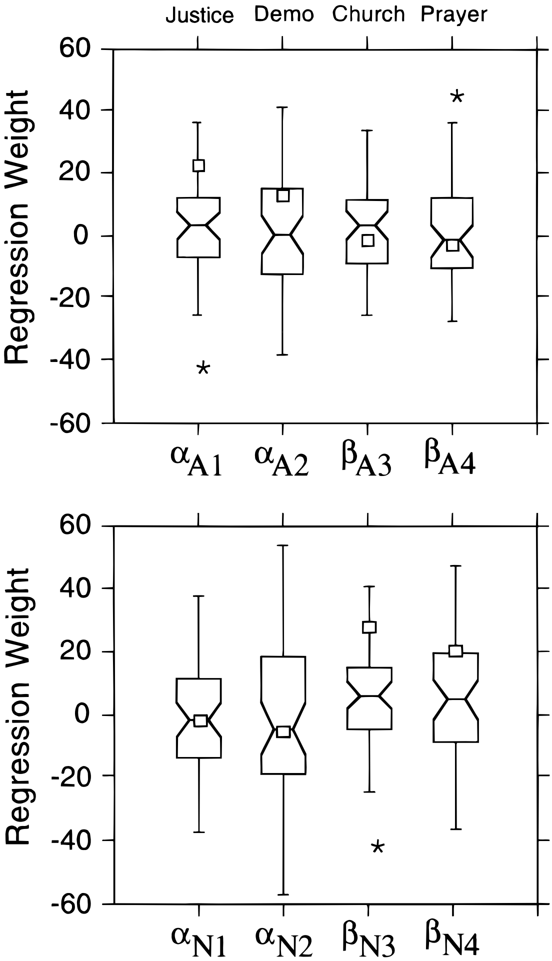

FIG. 8. Regression weights for the affirmation (top) and negation (bottom) conditions.

$\alpha_A > \alpha_N$ and $\beta_N > \beta_A$, I conducted the following analysis of standardized regression weights. The former prediction was tested in the form of $\alpha_A - \alpha_N > 0$, the latter in the form of $\beta_N - \beta_A > 0$. Testing these predictions required comparing corresponding weights from two regression analyses (affirmation and negation). Therefore, standardized regression weights were used to investigate if $\hat{\alpha}_{A1} - \hat{\alpha}_{N1} > 0$, $\hat{\alpha}_{A2} - \hat{\alpha}_{N2} > 0$, $\hat{\beta}_{N3} - \hat{\beta}_{A3} > 0$, and $\hat{\beta}_{N4} - \hat{\beta}_{A4} > 0$ were observed. These differences of standardized weight estimates were compared against differences of standardized random weighs. For every ith simulation $(1 \leq i \leq 500)$, I calculated the differences of random weights for $\alpha_{A1} - \alpha_{N1}$, $\alpha_{A2} - \alpha_{N2}$, $\beta_{N3} - \beta_{N3}$, and $\beta_{N4} - \beta_{A4}$. Figure 9 shows notched boxplots of the differences of the standardized random weights and estimated weights. Like Fig. 8, the boxplots provide the "ground" of the picture by showing the dispersion of the differences of the random weighs. The 95% confidence intervals for the random weights encompass zero. The "X" symbols indicate the location of the differences of the estimated weights. Every "X" symbol lies above the confidence

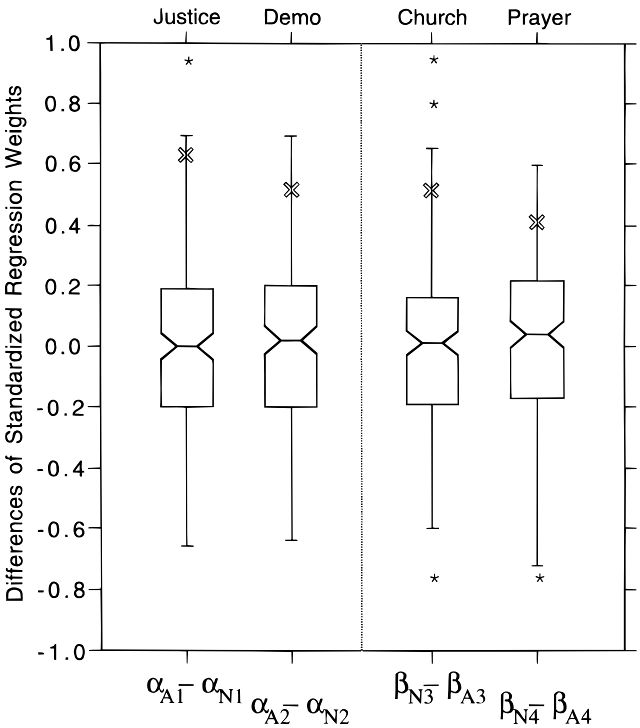

FIG. 9. Differences of the subjective weights and the random weights.

intervals, supporting that $\hat{\alpha}_{A1} - \hat{\alpha}_{N1} > 0$, $\hat{\alpha}_{A2} - \hat{\alpha}_{N2} > 0$, $\hat{\beta}_{N3} - \hat{\beta}_{A3} > 0$, and $\hat{\beta}_{N4} - \hat{\beta}_{A4} > 0$.

In conclusion, Figs. 7, 8, and 9 jointly confirm Hypothesis 3. Systematic binary noncomplementarity such that $\Phi_A + \Phi_N > 1.00$ (Fig. 7) occurred contingent on subjective feature weighting that $\alpha_A > \alpha_N$ and $\beta_N > \beta_A$ (Figs. 8 and 9). It may seem possible to claim that the results in Fig. 9 could be obtained even if the compatibility principle were not in effect; $\alpha_A - \alpha_N > 0$ and $\beta_N - \beta_A > 0$ would follow if α_A and β_N were close to zero and α_N and β_A were very negative. However, the data in Fig. 8 refute this counterargument by showing that the α_A and β_N weights were significantly positive, whereas α_N and β_A were embedded within the confidence intervals that included zero.

GENERAL DISCUSSION

This article started by introducing an assumption that led to the prediction of systematic binary noncomplementarity. Experiments 1 and 2 confirmed the assumption that dissimilarity would serve as a judgmental heuristic for assessment of nonmembership. Experiments 3 and 4 showed that binary noncomplementarity occurs between an example X and Category C. Membership judgments produced $\Phi_A + \Phi_N > 1.00$ if C and X were similar and dissimilar to each other. Experiment 5 investigated subjective weighting of features when $\Phi_A + \Phi_N > 1.00$ to confirm that the weighting pattern was consistent with the compatibility principle.

The model in Eq. (1) is built on the same conception of feature-set representation and task-dependent feature weighting central to other models of preferential performance. Specifically, Tversky *et al.*'s (1988) Contingent Weighting model, Shafir's (1993) model, and Yamagishi and Miyamoto's (1996) focus shift model commonly adopt the feature weighting notion and use the compatibility principle as the explanatory mechanism. The former three models were developed to explain violations of procedure invariance in preferential judgments (preference reversals in Tversky *et al.*, asymmetry between choice and rejection in Shafir, and asymmetric strength of preference in Yamagishi and Miyamoto). This article may be regarded as extending the applicability of the feature weighting approach to a wider variety of logical incoherence beyond preference.

In particular, profound parallelism exists between Experiment 3 and Shafir's (1993) choice-rejection asymmetry. Shafir investigated binary choice where multiattribute alternatives consisted of an "enriched" option (included strongly aversive and strongly favorable features) and a mediocre alternative (made of weakly aversive and weakly favorable features). Shafir noted that the "enriched" option was preferred in choice and was abandoned in rejection. Thus, the "enriched" option was superior and inferior to its rival (see also Houston, Sherman, & Baker, 1989). Shafir predicted this result from the compatibility principle, arguing that favorable features would be weighted heavily in choice, whereas aversive features would be weighted heavily in rejection. I note a correspondence between Shafir's result and the result in Experiment 3 in that the "enriched" options (Rhonda and Phil) attracted the majority votes under logically opposite criteria (choice versus rejection in Shafir and affirmation versus negation in Experiment 3). Moreover, the equivalence between Shafir (1993) and Experiment 3 may be claimed to lie at a theoretical level, whereupon the compatibility principle provides the logical basis of explaining such seemingly paradoxical performance.

A resemblance may be noticed between the current experiments and stimuli in Nisbett, Zukier, and Lemley's (1981) clinical probability judgment task. Nisbett *et al.* examined the effects of diagnostic, nondiagnostic, and counterdiagnostic evidence in assessing the probability that a "client" was a child molester. Diagnostic features included "He was sexually assaulted by his stepfather." Nondiagnostic and counterdiagnostic features included "He has an IQ of 110" and "He would like to adopt a second child," respectively. Unfortunately, the issue of binary complementarity cannot be addressed by the data of Nisbett *et al.* for two reasons. First, counterpart data of negation probability judgment were not collected. Second, probability judgment was provided on a 11-point Likert scale, and the verbal labels expressing the belief strength do not specify what set of responses might violate probabilistic norms. Although personality descriptions that consisted of weakly diagnostic and weakly counterdiagnostic features did appear in the experiment of Nisbett *et al.*, such trials were treated as filler tasks. Therefore, the findings of Nisbett *et al.* and those of the present study are consistent to the extent that feature contribution affected perceived proximity and probability judgment was predictable from proximity judgment.

In sensitive real-life issues, an uncertain prospect may involve one set of reasons to believe that it would materialize and another set of reasons to expect the opposite. Consider the possibility that a free election will take place in Kosovo by the end of the year 2004. Such a possibility involves various relevant factors, some encouraging and others discouraging, of any actualization of the prospect. Hence, a scenario involving inhibiting and energizing factors may be judged as probable and improbable. Probability judgments of affirmation and negation may produce probabilistic contradictions similar to the results from Experiments 3 through 5. Preliminary data were collected from Japanese undergraduates regarding Hideo Nomo, a baseball pitcher in the American Major Leagues. As a Japanese national, Nomo enjoyed popularity on Japan's sports newscasts at the period of data collection. Upon his migration from the Los Angeles Dodgers to the New York Mets in 1998, participants made probability judgments as to whether he could win eight games in the coming season. They were presented with the following known facts: "Nomo lately lost 7 games yet won 2," "Nomo has recovered from his elbow problems," "Nomo's former favorite catcher Mike Piazza belongs to the Mets," and "National League batters are getting accustomed to Nomo's pitching." The four facts represent encouraging and discouraging evidence toward Nomo's success. One group rated the chance that Nomo would win more than seven games as 74.28%. Another group rated the chance that Nomo would win fewer than eight games as 36.67%. Thus, a binary noncomplementarity was observed, and more investigations are being conducted to further pursue this issue.

One possible criticism is that the widespread applicability of the feature weighting approach may indicate that such formulation internalizes a grand theory that lacks constraints in its limitation. It is an open and empirical question as to where such boundaries may be observed, and only systematic research will bring insights.

REFERENCES

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, 21, 515–550.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114, 542–551.

Dawes, R. M. (1988). Proper and improper linear models. In R. M. Dawes (Ed.), *Rational choice in an uncertain world* (pp. 201–227). Orlando, FL: Harcourt Brace Jovanovich.

Fisher, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 580–597.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison on preferences. *Journal of Experimental Social Psychology*, 25, 121–141.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.

Koehler, D., Brenner, L. A., & Tversky, A. (1997). The enhance effect in probability judgment. *Journal of Behavioral Decision Making*, 10, 293–313.

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32, 12–16.

Nisbett, R., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, 13, 248–277.

Reeves, T., & Lockhart, R. S. (1993). Distributional versus singular approaches to probability and errors in probabilistic reasoning. *Journal of Experimental Psychology: General*, 122, 207–226.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in Support theory. *Psychological Review*, 104, 406–415.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition*, 21, 546–556.

Slovic, P., Griffin, D., & Tversky, A. (1990). Compatibility effects in judgment and choice. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel Einhorn* (pp. 5–27). Univ. of Chicago Press.

Teigen, K. H. (1974). Subjective sampling distributions and the additivity of estimates. *Scandinavian Journal of Psychology*, 15, 50–55.

Teigen, K. H. (1983). Studies in subjective probability III: The unimportance of alternatives. *Scandinavian Journal of Psychology*, 24, 97–105.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 453–458.

Tversky, A., & Fox, C. R. (1995). Weighting risk and uncertainty. *Psychological Review*, 102, 269–283.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). New York: Cambridge Univ. Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.

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

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, 101, 547–567.

Yamagishi, K. (1996). Strength of preference and effects of valence in the domains of grains and losses. *Organizational Behavior and Human Decision Processes*, 66, 290–306.

Yamagishi, K., & Miyamoto, J. M. (1996). Asymmetries in strength of preference: A focus shift model of valence effects in difference judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 493–509.

Wright, G., & Ayton, P. (Eds.). (1994). *Subjective probability*. New York: Wiley.

Received February 14, 2000; published online June 19, 2001