Cognitive Psychology **38,** 167–189 (1999)
Article ID cogp.1998.0711, available online at http://www.idealibrary.com on IDEAL®

Strength of Evidence, Judged Probability, and Choice Under Uncertainty

Craig R. Fox

Fuqua School of Business, Duke University

This paper traces, within subjects, the relationship between assessed strength of evidence, judgments of probability, and decisions under uncertainty. The investigation relies on the theoretical framework provided by support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997), a nonextensional model of judgment under uncertainty. Fans of professional basketball ($N = 50$) judged the probability that each of eight teams, four divisions, and two conferences would win the National Basketball Association championship. Additionally, participants rated the relative strength of each team, judged the probability that a given team would win the championship assuming a particular pairing in the finals, priced prospects contingent on the winner of the championship, and made choices between chance prospects. The data conformed to the major tenets of support theory, and the predicted relationships between assessed strength of evidence, hypothetical support, judged probabilities, and choices under uncertainty also held quite well.

Nearly three decades of psychological research on judgment under uncertainty has demonstrated convincingly that intuitive judgments of probability systematically violate the calculus of chance (see, e.g., Kahneman, Slovic, & Tversky, 1982). In particular, different descriptions of the same event often give rise to systematically different judgments (e.g., Fischhoff, Slovic, & Lichtenstein, 1978), and the judged probability of the union of disjoint events is generally smaller than the sum of judged probabilities of those events (e.g., Teigen, 1974).

The tendencies to assign higher probabilities to more specific events or more detailed descriptions of events have generally been attributed to heuristic processes of representativeness and availability (see, e.g., Tversky &

The study described in this paper was designed in collaboration with Amos Tversky. Results reported in Figs. 2 and 4 and Table 3 are also reported in Fox and Tversky (1998). Thanks to Lyle Brenner, Rich Gonzalez, Derek Koehler, John Payne, Yuval Rottenstreich, Peter Wakker, and Tom Wallsten for helpful discussions and suggestions. The author is also grateful to Jonathan Levav for assistance compiling individual level analyses.

Address correspondence and reprint requests to the author at Fuqua School of Business, Duke University, Box 90120, Durham, NC 27708. E-mail: cfox@mail.duke.edu.

0010-0285/99 $30.00

Kahneman, 1983). Recently Amos Tversky and his students (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) have developed a more formal model of judgment under uncertainty called *support theory* that accommodates such patterns and provides the theoretical foundation for the present investigation. In particular, support theory conceives of judged probability as the balance of evidence favoring the hypothesis in question and the theory provides a means of predicting judged probabilities from raw assessments of evidence strength. Moreover, it seems that choices under uncertainty can be predicted from judged probabilities that are consistent with support theory (see Fox & Tversky, 1998). However, up to this point there have been no studies that trace the relationship between assessed strength of evidence, judged probability, and choices under uncertainty. The primary purpose of the present article is to provide such an investigation, within subjects, and to simultaneously test the major tenets of support theory using a population of experts. I next turn to a discussion of the major axioms and implications of support theory.

Key Tenets of Support Theory

In support theory probability is not attached to events, as it is in other models, but rather to descriptions of events, called *hypotheses;* hence, two descriptions of the same event may be assigned different probabilities (i.e., the model is nonextensional). Support theory assumes that each hypothesis A has a nonnegative support value $s(A)$ corresponding to the strength of the evidence for this hypothesis. The judged probability $P(A,B)$ that hypothesis A rather than B holds, assuming that one and only one of them obtains, is given by

$$P(A,B) = \frac{s(A)}{s(A) + s(B)}. \tag{1}$$

Hence, judged probability is interpreted as the support of the focal hypothesis A relative to the alternative hypothesis B. The theory further assumes that (i) unpacking a description of an event A (e.g., homicide) into disjoint components $A_1 \vee A_2$ (e.g., homicide by an acquaintance, A_1, or homicide by a stranger, A_2) generally increases its support, and (ii) the sum of the support of the component hypotheses is at least as large as the support of their explicit disjunction, so that

$$s(A) \leq s(A_1 \vee A_2) \leq s(A_1) + s(A_2), \tag{2}$$

provided that (A_1, A_2) is recognized as a partition of A. The rationale for Eq. (2) is that (i) unpacking may remind people of possibilities that they might have overlooked, and (ii) the separate evaluation of hypotheses tends to increase their salience and enhance their support.

Key Implications of Support Theory

There are three major implications of support theory. First, Eq. (1) implies *binary complementarity:* $P(A,B) + P(B,A) = 1$. For instance, if the Atlanta Hawks and Boston Celtics are playing each other on a particular night, the judged probability that Atlanta rather than Boston wins and the judged probability that Boston rather than Atlanta wins should sum to 1. Second, for finer partitions, Eqs. (1) and (2) imply *subadditivity:* the judged probability of A is less than or equal to the sum of judged probabilities of its disjoint components. For example, the judged probability that Atlanta beats Boston should be less than or equal to the judged probability that Atlanta wins by 1 to 10 points plus the judged probability that Atlanta wins by more than 10 points.[1] Binary complementarity and subadditivity have been confirmed in several studies reviewed by Tversky and Koehler (1994). They have been replicated among experienced physicians (Redelmeier, Koehler, Liberman, & Tversky, 1995), lawyers (Fox & Birke, 1998), and options traders (Fox, Rogers, & Tversky, 1996). Similar patterns have also been observed in decision making under uncertainty. In particular, there is evidence that unpacking the description of an event can increase subjects' willingness to pay for insurance policies (Johnson, Hershey, Meszaros, & Kunreuther, 1993; see also Wu & Gonzalez, 1998) and that subadditivity of judged probabilities is associated with subadditivity in subjects' pricing of prospects contingent on sporting events, economic indicators, and other events (Tversky & Fox, 1995; Fox, Rogers & Tversky, 1996; Fox & Tversky, 1998; Fox, 1998).

It is convenient to define the probability (odds) ratio $R(A,B) \equiv P(A,B)/P(B,A)$. One can easily verify that Eq. (1) yields a third implication of support theory, called the *product rule:*

$$R(A,C)R(C,B) = R(A,D)R(D,B), \tag{3}$$

provided all probabilities are nonzero and all four hypotheses are pairwise exclusive. For instance, suppose that $R(A,B)$ is the odds that team A rather than team B leads the league in scoring this season, assuming that one of these teams will lead the league. Equation (3) implies that the product of the odds for Atlanta against Chicago and Chicago against Boston equals the product of the odds for Atlanta against Denver and Denver against Boston.

[1] In addition, Eqs. (1) and (2) imply that the judged probability of the explicit disjunction of component hypotheses lies between these values. I will briefly address this issue in the discussion; see also Rottenstreich and Tversky (1997).

To see the necessity of the product rule, note that according to Eq. (1) both sides of Eq. (3) equal $s(A)/s(B)$.[2]

Note that each of these implications of support theory has an important role in the interpretation of probability as normalized support (Eq. (1)). Binary complementarity is necessary to establish that support for a hypothesis remains the same whether the hypothesis is in the foreground (focal) or background (alternative). Subadditivity of judged probabilities is necessary to establish that separate evaluation of component hypotheses increases their support (Eq. (2)). Finally, the product rule is necessary to establish that support for a hypothesis is independent of the particular (alternative or focal) hypothesis against which it is being compared.

Strength of Evidence, Judged Probability and Choice under Uncertainty

Although the most important tenets of support theory have been empirically validated separately among different groups of subjects in various studies, there has not yet been a thorough within-subjects test of the implications and assumptions of the theory, nor has there been a study that directly relates assessed strength of evidence for hypotheses to willingness to act on those hypotheses. The purpose of the present article is to provide such a comprehensive investigation using expert subjects. In particular, I set out to test the predictions of binary complementarity, subadditivity, and the product rule and then trace the relationship between raw judgments of evidence strength ($\hat{s}$), hypothetical support (s), judged probability (P), and choices under uncertainty (C). Figure 1 provides a schematic illustration of the relationship between these variables that will be elaborated in subsequent sections of this paper.

In order to accomplish this analysis, I recruited fans of professional basketball and asked them to make several judgments and decisions concerning the winner of the National Basketball Association (NBA) playoffs. This domain has several desirable features: (1) basketball fans have considerable expertise predicting outcomes of games and are typically comfortable betting on these outcomes; (2) the hierarchical structure of conferences, divisions, and teams that could win the tournament provides a test of binary complementarity and several tests of subadditivity; (3) basketball fans are comfortable judging the relative strength of teams, which provides a suitable proxy for strength of evidence that a team will win the tournament; and (4) the tournament structure provides a natural means of eliciting *pairwise conditional probabilities* (‘‘assuming teams A and B reach the finals, what is the

[2] If we add a related condition

$$R(A,B) = R(A,D)R(D,B),$$

then binary complementarity and the product rule are both necessary and sufficient for Eq. (1). See Rottenstreich and Tversky (1997).

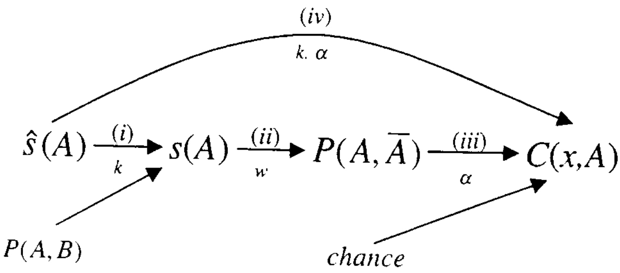

FIG. 1. A schematic representation of the relationship between variables. Note that $\hat{s}(A)$ refers to assessed strength of evidence for hypothesis *A*; $s(A)$ refers to hypothetical support for hypothesis *A*; $P(A,\bar{A})$ refers to judged probability of hypothesis *A* against its complement; $C(x,A)$ refers to the certainty equivalent for the prospect that pays \$*x* if and only if hypothesis *A* obtains; $P(A,B)$ refers to the judged probability of hypothesis *A* against hypothesis *B*, assuming either *A* or *B* obtains (but not both); and *chance* refers to choices between chance prospects that are used to parameterize the utility function for monetary gains. The parameters *k* (the strength-support scaling parameter), *w* (the global weight for discounting support of alternative hypotheses), and α (the utility function parameter), listed below the arrows, are required to estimate the variable to the right of the corresponding arrow from the variable to the left of that arrow.

probability that team *A* wins?''), which are necessary to scale support from strength ratings.

(i) From assessed strength of evidence to support. Previous studies have shown that it is possible to predict probability judgments from direct assessments of evidence strength (Tversky & Koehler, 1994; Koehler, 1996; see also Brenner & Koehler, 1999). To see how strength and support are related, note that in support theory, the support function, *s*, is derived from probability judgments. Let $\hat{s}(A)$ be the rating of the strength of evidence for hypothesis *A*. First, it seems reasonable to assume that the probability of the focal hypothesis will be judged to be at least one-half if and only if the evidence favoring the focal hypothesis is at least as strong as the evidence favoring the alternative hypothesis. For instance, suppose that $\hat{s}(A)$ is the judged strength of a given sports team. In this case, we assume that the judged probability that team *A* beats team *B* will be at least one-half if and only if the assessed strength of team *A* is greater than or equal to the assessed strength of team *B*. In the support theory framework, this assumption suggests that raw strength ratings and hypothetical support (derived from judged probabilities) are related monotonically:

$$\hat{s}(A) \geq \hat{s}(B) \quad \text{iff } s(A) \geq s(B). \tag{4}$$

Second, it seems reasonable to assume that the higher the ratio of judged strength of the focal hypothesis to judged strength of the alternative hypothesis, the higher the odds assigned to the focal hypothesis against the alternative hypothesis (derived from judged probabilities). Suppose again that $\hat{s}(A)$ is the judged strength of a given sports team. We assume that the ratio of the strength ratings of team A to team B will be greater than or equal to the ratio of strength ratings of team C to team D if and only if the odds assigned to team A beating team B are greater than or equal to the odds assigned to team C beating team D (derived from pairwise conditional probability judgments). In the support theory framework, this condition suggests that strength ratios and support ratios are related monotonically:

$$\hat{s}(A)/\hat{s}(B) \geq \hat{s}(C)/\hat{s}(D) \quad \text{iff } s(A)/s(B) \geq s(C)/s(D). \tag{5}$$

It can be shown that if these two conditions hold and both scales are defined on, say, the unit interval, then there exists a constant $k > 0$, such that the two measures of support are related by a power transformation of the form $s(A) = \hat{s}(A)^k$ (for a proof, see Tversky & Koehler, 1994, theorem 2). Using Eq. (1), assuming $P(A,B) > 0$ for all A, B, and defining $R(A,B)$ as before, we get

$$\begin{aligned} R(A,B) &= s(A)/s(B) \\ &= [\hat{s}(A)/\hat{s}(B)]^k. \end{aligned}$$

Taking the logarithm of both sides of the equation yields

$$\ln R(A,B) = k \ln[\hat{s}(A)/\hat{s}(B)]. \tag{6}$$

Hence the parameter k can be estimated from probability judgments and team strength ratings using linear regression, which will allow us to relate team strength ratings to the hypothetical support scale.

The psychological interpretation of the scaling parameter k is an interesting question. Note that as k approaches zero, all probabilities converge to one-half (i.e., $R(A,B)$ goes to 1); as k increases, probabilities diverge to zero (when the strength of evidence for the focal hypothesis is less than the strength of evidence for the alternative hypotheses) and 1 (when the strength of evidence for the focal hypothesis is greater than the strength of evidence for the alternative hypothesis). Hence, k might be interpreted as an index of an individual's sensitivity to differences in assessed support for the focal versus alternative hypothesis when judging probabilities.

(ii) From support to judged probability. Equation (6) uses pairwise conditional probability judgments (e.g., the probability that team A wins the championship assuming that team A and team B are in the finals) to determine the relationship between strength of evidence and hypothetical support. Using the parameter k obtained from this analysis we can now relate strength of evidence to the simple judged probability that a particular team wins the championship and simultaneously test for subadditivity of support. Recall

that according to Eq. (1), the probability that a particular team wins the championship depends on the support for the hypothesis that the team wins and the support for the hypothesis that the team does not win. Although we can use team strength ratings as a proxy for support for the hypothesis that a particular team wins, there is no obvious means to elicit raw support for the hypothesis that a particular team *does not win*. However, following Tversky & Koehler (1994; see also Koehler, Brenner, & Tversky, 1997; Brenner & Koehler, 1999), we can estimate support for the hypothesis that team A_i fails to win as a weighted sum of the support of all other teams winning the championship;

$$s(\overline{A}_i) = w_{\bar{A}_i} \sum_{\substack{j=1 \\ j \neq i}}^{n} s(A_j),$$

where $w_{\bar{A}_i}$, called the *global weight*, is less than or equal to 1, as can be derived from Eq. (2). Hence, the global weight provides an index of the general extent to which the support for each alternative elementary hypothesis (e.g., that each team other than the Chicago Bulls wins the championship) is discounted when packed into a residual hypothesis (e.g., that the Bulls do not win the championship). Substituting the above equation into the Eq. (1), and rearranging terms, we get

$$\frac{1 - P(A_i,\overline{A}_i)}{P(A_i,\overline{A}_i)} = w_{\bar{A}_i} \frac{\sum_{j=1,\, j\neq i}^{n} s(A_j)}{s(A_i)}.$$

If we assume that $w_{\bar{A}_i}$ is approximately constant for all i and substitute $\hat{s}^k$ for s, we get

$$\frac{1 - P(A_i,\overline{A}_i)}{P(A_i,\overline{A}_i)} = w_{\bar{A}} \frac{\sum_{j=1,\, j\neq i}^{n} [\hat{s}(A_j)]^k}{[\hat{s}(A_i)]^k}, \qquad (7)$$

so we can estimate $w_{\bar{A}}$ from linear regression using simple judged probabilities, raw strength ratings, and values of k obtained from the previous analysis. This analysis will provide a more direct test of the subadditivity of support and also allow us to predict judgments of the probability that a particular team will win the championship from strength ratings of all teams.

(iii) From judged probability to choice under uncertainty. Now that we have related strength to judged probability, we can relate it to a measure of willingness to act under uncertainty by constructing prospects (x,A_i) that of-

fer \$x if team A_i wins the NBA championship and nothing otherwise. The attractiveness of such a prospect can be measured by eliciting a person's *certainty equivalent* for that prospect, C_i, which is the sure amount of money that the decision maker finds equally attractive to receiving \$x if team A_i wins the tournament (and nothing otherwise). According to expected utility theory,

$$u(C_i) = p(A_i)u(x),$$

where p is an additive subjective probability measure and $u(x)$ is the utility (i.e., subjective value) of receiving \$x. It can be shown that the standard economic model of decision making (expected utility theory with risk aversion) implies that the certainty equivalent for the prospect formed from an event must be at least as large as the sum of certainty equivalents for prospects formed by partitioning that event (i.e., certainty equivalents should be *superadditive* over events).[3] For instance, if I price a prospect that offers \$100 if the Chicago Bulls win their next game by 1–10 points at \$20, and I price a prospect that pays \$100 if the Bulls win by more than 10 points at \$25, the standard model implies that I should price a prospect that pays \$100 if the Bulls win at *no less than* \$45. This result depends on the assumption of risk aversion, and the assumption that subjective probabilities are additive. However, if subjective probabilities are allowed to be *subadditive*, as support theory predicts, then this condition will often be violated. For instance, if I perceive the probability that the Bulls win by 1–10 points to be .30 and the probability that the Bulls win by more than 10 points to be .40, but I perceive the probability that the Bulls win to be only .55 ($<.30 + .40$), then I might price the third prospect lower than the sum of the first two prospects. Such failures of the standard model are especially likely when concavity of the utility function is not very pronounced and the target event is partitioned into many components.

To predict certainty equivalents from judged probabilities, we replace the additive subjective probability, $p(A_i)$, with the judged probability, $P(A_i,\bar{A}_i)$, so that we get[4]

$$u(C_i) = P(A_i,\bar{A}_i)u(x).$$

Previous studies (e.g., Tversky, 1967; Tversky & Kahneman, 1992) have indicated that the utility function for small to moderate gains can be approximated by a power function of the form $u(x) = x^\alpha$, $\alpha > 0$, which gives us

[3] For a proof and fuller discussion, see Fox and Tversky (1998).

[4] Another approach would be to weight support theory probabilities by the S-shaped function of prospect theory (Tversky & Kahneman, 1992). For simplicity I omit this second stage here, but will take up the issue in the discussion section.

$$C_i = [P(A_i, \bar{A}_i)]^{1/\alpha} x. \tag{8}$$

Hence, the certainty equivalents for the prospect that pays \$$x$ if hypothesis A_i obtains (and nothing otherwise) can be predicted from the judged probability that hypothesis A_i obtains and the utility parameter α.

(iv) From assessed strength of evidence to choice under uncertainty. To predict certainty equivalents directly from strength ratings, note that according to support theory

$$P(A_i, \bar{A}_i) = \frac{s(A_i)}{s(A_i) + s(\bar{A}_i)}$$

for every team i. If we assume that $s(A_i) + s(\bar{A}_i)$ is approximately constant for all i, then

$$\frac{P(A_i, \bar{A}_i)}{P(A_j, \bar{A}_j)} \approx \frac{s(A_i)}{s(A_j)}, \tag{9}$$

so

$$\frac{u(C_i)}{u(C_j)} = \frac{P(A_i, \bar{A}_i)\, u(x)}{P(A_j, \bar{A}_j)\, u(x)} \approx \frac{s(A_i)}{s(A_j)}.$$

We can estimate support from assessed strength of evidence using the parameter k obtained from the foregoing analysis, and we can estimate a power utility function using the parameter α as before. Hence, the ratio of certainty equivalents can be predicted from the ratio of assessed strength of corresponding teams:

$$\frac{C(A)}{C(B)} \approx \left[\frac{\hat{s}(A)}{\hat{s}(B)}\right]^{\theta},$$

where $C(A)$ is the certainty equivalent of the prospect that pays \$160 if team A wins the NBA championship and $\theta = k/\alpha$ as previously defined. Taking the logarithm of both sides of this equation yields

$$\ln\left[\frac{C(A)}{C(B)}\right] \approx \theta \ln\left[\frac{\hat{s}(A)}{\hat{s}(B)}\right], \tag{10}$$

so that θ can be estimated by linear regression. Hence, relative certainty equivalents for bets on two different teams can be predicted from judgments of the relative strength of those teams and independent estimations of the parameters k and α.

We have now seen that support theory provides a parsimonious framework for tracing the relationship between assessed strength of evidence, hypothetical support, judged probability, and choice under uncertainty. I now turn to a detailed description of an experiment in which the integrity of this framework is tested.

EXPERIMENT

Method

I recruited fans of professional basketball during the 1995 NBA playoffs and asked them to make several judgments and decisions concerning who wins the championship.

Participants. The participants in this study were 50 students at Northwestern University (46 men, 4 women; median age, 20 years) who responded to flyers calling for fans of professional basketball to take part in a study of decision making. Participants received $10 for completing one 60-min session and were told that some participants would be selected at random to play one of their choices for real money (up to $160). Subjects indicated that they had watched or listened to a large number of NBA games (median, 25) during the regular season.

Procedure. The experiment was run using a computer. All subjects were run on the same day, during the beginning of the quarterfinals. At the time of the study, eight teams remained (Chicago, Indiana, Boston, Orlando, Phoenix, Los Angeles, San Antonio, and Houston) representing four divisions (Midwestern, Eastern, Pacific, and Central) which in turn compose two conferences (Eastern and Western).[5]

The first phase was designed to estimate subjects' certainty equivalents for uncertain prospects. Prospects offered to pay $160 if a particular team, division, or conference would win the 1995 NBA championship. For example, a typical prospect would pay $160 if the Chicago Bulls win the 1995 NBA championship. Each trial involved a series of choices between a prospect and an ascending series of sure payments (e.g., receive $40 for sure). Prospects were presented in an order that was randomized for each subject. Certainty equivalents were inferred from two rounds of such choices. The first round consisted of nine choices between the prospect and sure payments that were spaced evenly from $0 and $160. After completing the first round of choices, a new set of nine sure payments was presented, spanning the narrower range between the lowest payment that the subject had accepted and the highest sure amount that the subject had rejected (excluding the endpoints). The program enforced dominance and internal consistency. For example, the program did not allow a respondent to prefer $30 to a prospect and also prefer the same prospect to $40. The program allowed subjects to backtrack if they felt they had made a mistake in the previous round. The certainty equivalent of each prospect was determined by linear interpolation between the lowest value accepted and the highest value rejected in the second round of choices. This interpolation yielded a margin of error of $\pm$$1.00 for the $160 prospects. Note that although the analysis is based on certainty equivalents, the data consisted of a series of choices between a given prospect and sure outcomes. Thus, respondents were not asked to generate certainty equivalents; these values were inferred from choices.

The second phase was designed to estimate the shape of the utility function for monetary gains (i.e., the mapping from dollars to subjective value). Subjects were presented with a

[5] The nested structure of teams, divisions, and conferences is displayed in Figs. 2 and 4.

''fixed'' prospect of the form (.25, \$*a;* .25, \$*b;* .50, \$0) and a ''variable'' prospect of the form (.25, \$*c*; .25, *x*; .50, \$0). These prospects were presented at ''spinner games'' that paid depending on which region of a circle a spinner was to land. In a given trial, the values of *a, b,* and *c* were fixed, while the value of *x* varied. The initial value of *x* was set so that the expected value of the prospects was equal. Eight such pairs of prospects were constructed, presented below in Table 3, with the analysis of the data. Subjects were asked to indicate their preference between the bets. If a subject began by indicating a preference for the fixed prospect, the value of *x* increased by \$16 each time they did so; if a subject began instead by indicating a preference for the variable prospect, the value of *x* decreased by \$16 each time they did so. When a subject's preference switched from the fixed prospect to the variable prospect or the variable to the fixed, the change in *x* reversed direction and the increment was cut in half (i.e., from \$16 to \$8, from \$8 to \$4, and so forth) until the increment was \$1. This process was repeated until the subject indicated that they found the two prospects equally attractive.

The third phase asked subjects to estimate the probability of each uncertain target event (i.e., that a particular team, division, or conference would win the NBA playoffs). The fourteen events were presented in an order that was randomized for each subject. On each trial, subjects could respond by typing in a number either larger than 0 and smaller than 100, or by clicking and dragging a ''slider'' on a visual scale labeled from 0 to 100.

The fourth phase asked subjects to estimate the probability that a particular team would win the NBA championship, supposing that two particular teams made the finals. Because four teams from each of two conferences were represented, 16 matchups were possible. Each of these sixteen possibilities was presented to subjects in a random order, with the team names in each pairing also presented in a random order. Moreover, the winning team in each possible pairing was selected at random from the two possibilities. For example, one trial might ask, ''Suppose that the Los Angeles Lakers play the Chicago Bulls in the NBA finals. What do you think the probability is that the Los Angeles Lakers win?'' The elicitation mode was otherwise identical to the previous phase.

The fifth phase of the study required subjects to rate the strength of each of the teams on a 100-point scale. The team names were listed together on the same page in a random order, and subjects clicked the mouse on a visual scale to indicate their judgment of the strength of that team. Following Koehler (1996), instructions were as follows:

> First, choose the team you believe is the strongest of the eight, and set that team's strength to 100. Assign the remaining teams ratings in proportion to the strength of the strongest team. For example, if you believe that a given team is half as strong as the strongest team (the team you gave 100), give that team a strength rating of 50.

In addition to these five phases, subjects were asked to make a series of choices from which certainty equivalents for chance prospects were estimated. Results of this task will not be discussed here.

RESULTS

Subadditivity and Binary Complementarity

The median judged probability for each target event is listed in Fig. 2, which shows that the sum of these probabilities is close to 1 for the two conferences, nearly 1.5 for the four divisions, and more than 2 for the eight teams. This pattern is consistent with the prediction of support theory that

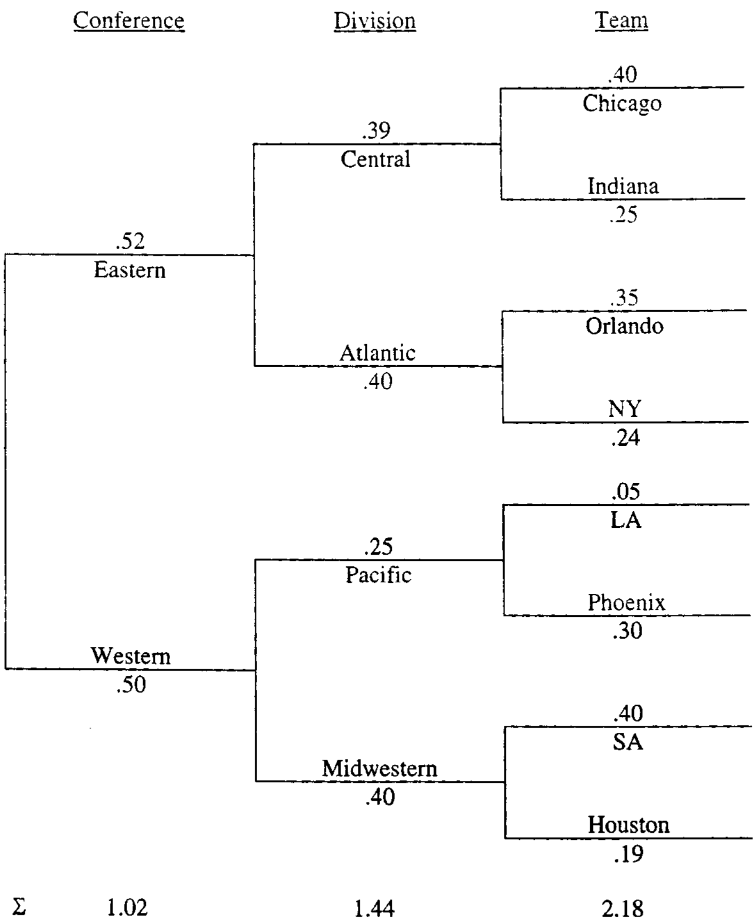

FIG. 2. Median judged probabilities for all target events.

$$\sum_{\text{teams}} P \geq \sum_{\text{divisions}} P \geq \sum_{\text{conferences}} P, \tag{11}$$

and the sum over the two conferences equals 1. Moreover, in every case the sum of probabilities for the individual teams is greater than the probability of the respective division, and the sum of the probabilities for the divisions is greater than the probability of the respective conference, consistent with support theory.[6]

The same pattern holds in the analysis of individual subjects. The median sum of probabilities for the eight teams is 2.40, the median sum for the four

[6] In every case this also holds for a significant majority of subjects ($p < .01$ by sign test).

TABLE 1
Test of Binary Complementarity for Binary Conditional Probability Judgments

Eastern Conf. team (*A*)	Western Conf. team (*B*)	$P(A,B)$	$1 - P(B,A)$	t
Chicago	Los Angeles	78.4	78.8	−0.10
Chicago	San Antonio	48.9	48.3	0.11
Chicago	Houston	61.1	54.2	1.17
Chicago	Phoenix	51.5	54.5	−0.52
Orlando	Los Angeles	78.5	78.6	0.04
Orlando	San Antonio	43.8	48.8	−0.92
Orlando	Houston	65.3	55.1	2.00*
Orlando	Phoenix	61.9	48.2	2.43**
Indiana	Los Angeles	68.8	66.9	0.40
Indiana	San Antonio	35.6	36.4	−0.15
Indiana	Houston	49.5	55.3	−1.19
Indiana	Phoenix	41.5	35.0	1.29
New York	Los Angeles	75.1	67.1	1.60
New York	San Antonio	40.9	36.4	0.81
New York	Houston	44.4	50.1	−1.16
New York	Phoenix	41.6	41.3	0.04

Note. The first and second columns, respectively, list the Eastern Conference and Western Conference teams in a given matchup. The third column lists the mean judged probability that the Eastern Conference team wins conditional on such a matchup. The fourth column lists one minus the mean judged probability that the Western Conference team wins conditional on such a matchup. The last column lists the t statistic for this comparison.

* $.05 < p < .10$.

** $p < .05$.

divisions is 1.44, and the median sum of probabilities for the two conferences is 1.00. Moreover, 41 of 50 respondents satisfied Eq. (11) with strict inequalities, and 49 of 50 respondents reported probabilities for the eight teams that sum to more than one ($p < .001$ by sign test in both cases).

We have now seen that binary complementarity holds nicely for judgments that each conference wins the championship (both within and between subjects). To verify that binary complementarity holds for the conditional probability data (which can only be tested between subjects), we compared $P(A,B)$ and $1 - P(B,A)$ for each of these 16 possible matchups. Recall that each participant judged the probability of one team winning, assuming that a particular pair reaches the finals. Table 1 displays means and t statistics for each of the 16 possible pairings presented. In only one case of 16 is $P(A,B)$ significantly ($p < .05$) different from $1 - P(B,A)$; an acceptable rate considering that there were 16 tests.[7] Moreover, the mean sum of median probabili-

[7] Note that there is a 56% probability that one or more tests of 16 would have come out significant at $p < .05$ by chance.

ties for the 16 pairs of complementary events was 1.008, very close to 1. Analyses that follow therefore assume that binary complementarity holds.

Product Rule

To verify the product rule, recall that odds terms were generated from pairwise conditional probability judgments, which were constructed from four teams belonging to each of two conferences. This design provides 36 tests of Eq. (3) that involve unique quartets of teams. For each of these 36 quartets there are four unique instantiations of the product rule that can be tested using the data collected in this study. For example, given any teams *A* and *B* from the Eastern Conference and teams *C* and *D* from the Western Conference, both Eq. (3) and its reciprocal are valid tests of the product rule,[8] as well as

$$R(C,A)R(A,D) = R(C,B)R(B,D)$$

and its reciprocal. However, if we take the logarithm of both sides of Eq. (3) and subtract the right side from the left, we get

$$\ln R(A,C) + \ln R(C,B) - \ln R(A,D) - \ln R(D,B) = 0, \qquad (12)$$

which yields identical test statistics for all four redundant permutations. For each subject we can calculate the left-hand side of Eq. (12) for each of the 36 tests of the product rule and conduct 36 sign tests. This analysis yields values significantly ($p < .05$) different from zero in only 3 of 36 cases, indicating an adequate fit for the product rule in this between-subject analysis[9].

Next, we can examine the median response to each item in the survey and construct ordered pairs consisting of both sides of Eq. (3) for each of the (36 unique quartets) × (4 permutations) × (2 orders) = 288 data points. Taking the natural log of both terms, these points are plotted in Fig. 3. The result is a very high correlation ($r = .95$) and a very good fit of the data to the identity line ($R^2 = .90$). The same analysis on individual subjects yields a median correlation of .73 ($R^2 = .53$).

The excellent fit of the product rule for these data lends credence to the notion that support for a particular hypothesis is independent of the hypothesis to which it is being compared. The robustness of the product rule is a bit surprising in the context of the present study, as one might expect that the specific pairing of teams would influence experts' predictions of which

[8] Note that $R(Y,X) = [R(X,Y)]^{-1}$ for all X, Y.

[9] Note that there is a 27% probability that three or more tests of 36 would have come out significant at $p < .05$ by chance. For an alternative test of the product rule, see Brenner and Rottenstreich (1998).

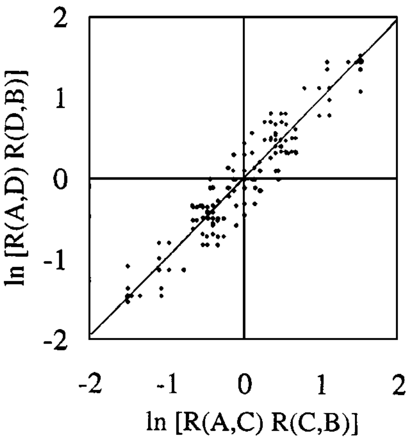

FIG. 3. Tests of the product rule.

team will win. For instance, a participant might think that the San Antonio Spurs match up better against the Orlando Magic than they do against the Chicago Bulls, despite the fact that Orlando is perceived to be a stronger team overall. Apparently such interactions did not affect judged probabilities substantially in the present study.

(i) From assessed strength of evidence to support. Before scaling support using Eq. (6), we can test the necessary and sufficient conditions for the power relation between strength and support (Eqs. (4) and (5)), using raw strength ratings and support inferred from pairwise conditional probabilities. Table 2 lists the median strength rating beside the name of the relevant team in the Eastern Conference (rows) and Western Conference (columns). Each cell lists the median judged probability that the corresponding Eastern Con-

TABLE 2
Median Strength Ratings and Conditional Probability Judgments

		95.0 San Antonio	81.0 Phoenix	67.5 Houston	45.0 Los Angeles
90.0	Orlando	46.5	57.5	60.0	80.0
84.0	Chicago	50.0	50.0	60.0	82.0
75.5	Indiana	35.0	37.5	51.5	70.0
70.0	New York	40.0	40.0	48.0	74.0

Note. Marginal entries are median strength ratings for the adjacent team. Table cells list the median judged probability that the designated Eastern Conference team (row) wins the championship, assuming that they play the designated Western Conference team (column) in the finals.

ference team wins the championship assuming that the corresponding Eastern and Western Conference teams make the finals.[10]

To test the assumption that strength and support are related monotonically (Eq. (4)), one can count in how many cases the stronger rated team of each potential matchup is assigned at least a 50% chance of winning. Using the median response to each item, this pattern was satisfied in 15 of 16 tests. To test the assumption that strength and support *ratios* are related monotonically (Eq. (5)), one can compare the ordering of strength ratios to the ordering of odds ratios for pairs of matchups. Using the median response to each item, Eq. (5) was satisfied for 109 of 120 tests. The median subject satisfied Eq. (4) for 15 of 16 tests, and the median subject satisfied Eq. (5) for 106 of 120 tests.

To scale support[11] we must estimate $R(A,B)$ assuming binary complementarity (i.e., assuming $P(B,A) = 1 - P(A,B)$). Regressing the left-hand term of Eq. (6) on the right-hand term using median strength ratings and median conditional probabilities (with no constant term) yields,[12] $k = 2.03$, $R^2 = .93$. The median estimate of k over the 50 subjects is 2.20, median $R^2 = .75$. The result $k > 1$ suggests that odds ratios (derived from judged probabilities) are more extreme than the corresponding ratios of assessed team strength. The values of k obtained here correspond closely to values obtained by Koehler (1996).

(ii) From support to judged probability. We can estimate the global weight w from Eq. (7) using raw strength ratings and values of k obtained in the previous analysis. Using median responses to each item, and regressing without a constant term, yields $w_{\bar{A}} = .65$, $R^2 = .90$. As predicted, w is substantially less than 1. The median estimate of $w_{\bar{A}}$ over the 50 subjects is .46, and the median $R^2 = .90$. This result suggests that support for the hypothesis that a particular team does *not* win the NBA championship is on average substantially less than the sum of support for the hypotheses that each alternative team wins. Also, it seems that a model using a single global discount factor fits the data reasonably well.

(iii) From judged probability to certainty equivalents. Figure 4 displays the median certainty equivalent for each prospect. The choice data in Fig. 4 echo the judgment data in Fig. 2 and confirm the prediction of subadditivity

[10] These medians are computed over all subjects, assuming binary complementarity.

[11] It is worth noting that because the monotonicity requirements (Eqs. (4) and (5)) are necessary and sufficient for the power relation, the empirically observed goodness-of-fit of the power relation must be closely related to the proportion of tests satisfying the monotonicity requirements.

[12] The R^2 statistics reported for these and subsequent regressions through the origin should be interpreted with caution. They were calculated as the ratio of the sum of squared fitted values to the sum of squared observations. Hence, it should be noted that these statistics compare the fit of the model to the null hypothesis $y_i = 0$ rather than $y_i = \bar{y}$. For a discussion of problems with R^2 statistics in models with zero intercepts, see Aigner (1971), pp. 85–90.

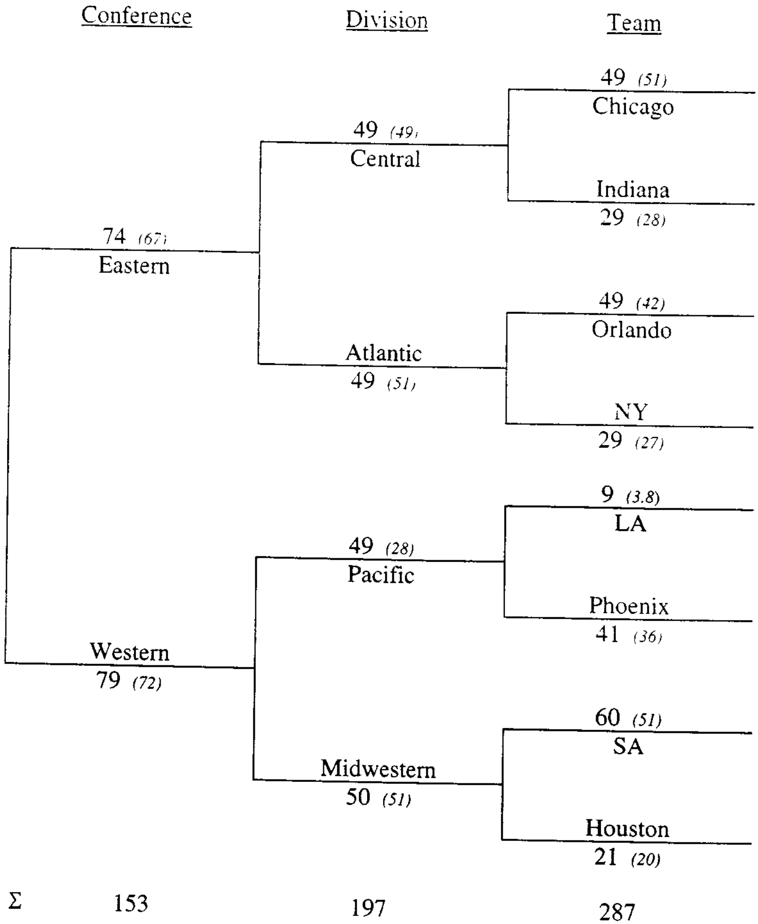

FIG. 4. Median certainty equivalents (in dollars) for all prospects that offer $160 if the designated conference, division, or team wins the NBA playoffs. Values in parentheses are predictions based on median judged probabilities (see Fig. 2) and $\alpha = .80$ (derived from median responses over subjects), using Eq. (8).

that violates the standard model of choice under uncertainty.[13] In every case, the sum of C's for the individual teams is greater than C for the respective division, and the sum of the C's for the divisions is greater than C for the respective conference.[14]

The pattern above can be reconciled with expected utility theory only if subjects are risk seeking. In order to rule out this explanation, we can esti-

[13] For a more detailed discussion of this condition and the data displayed in Fig. 4, see Fox and Tversky (1998).

[14] In every case this also holds for a majority of subjects ($p < .01$).

TABLE 3
Values of *a*, *b*, and *c* Used in the Spinner Games and Median Value of Subjects' Responses

	Fixed prospect			Variable prospect		
Probability outcome	.25 $*a*	.25 $*b*	.50 $0	.25 $*c*	.25 $*x* (median)	.50 $0
1	50	100		25	131	
2	30	60		10	86.5	
3	20	90		40	70	
4	10	110		35	82	
5	85	55		120	31	
6	50	45		75	29	
7	95	25		70	42	
8	115	15		80	43	

mate the utility function for gains assuming $u(x) = x^\alpha$, $\alpha > 0$. To estimate the exponent, we can use data from the ''spinner games'' (phase 2 of the study). If a subject is indifferent between the fixed prospect ($*a*, .25; $*b*, .25; $0, .5) and the variable prospect ($*c*, .25; $*x*, .25; $0, .5) then assuming a power utility function, $a^\alpha + b^\alpha = c^\alpha + x^\alpha$. Because *a*, *b*, and *c* are given and the value of *x* is determined by the subject, one can solve for $\alpha > 0$. The exponent for each subject was estimated using the median value of α over the eight problems listed in Table 3. This analysis showed that participants were generally risk averse: 32 subjects exhibited $\alpha < 1.00$ (risk aversion); 14 exhibited $\alpha = 1.00$ (risk neutrality); and 4 exhibited $\alpha > 1.00$ (risk seeking) ($p < .001$ by sign test). The median response to each of the eight trials yields $\alpha = .80$.

Next, we can predict certainty equivalents from judged probabilities and the parameter α, using Eq. (8). This analysis yields a good fit based on the median response over subjects to each item. Figure 4 displays predicted values in parentheses beside observed values. The mean error over the fourteen $160 prospects is −$4.38, and the mean absolute error is $5.05.[15]

(iv) From assessed strength of evidence to certainty equivalents. We can now bring the foregoing analyses together to predict subjects' choices from their assessments of team strength. First, recall that Eq. (10) assumes that $s(A_i) + s(\bar{A}_i)$ is approximately constant for all *i*, so that

$$\frac{P(A_i,\bar{A}_i)}{P(A_j,\bar{A}_j)} \approx \frac{s(A_i)}{s(A_j)} = R(A_i,A_j).$$

[15] Obviously this result is somewhat sensitive to the estimate of α. Repeating the analysis using the median estimate of α over subjects (.86) and restricting attention to the eight teams yields a remarkably low mean error of $0.10 and a mean absolute error of $2.54.

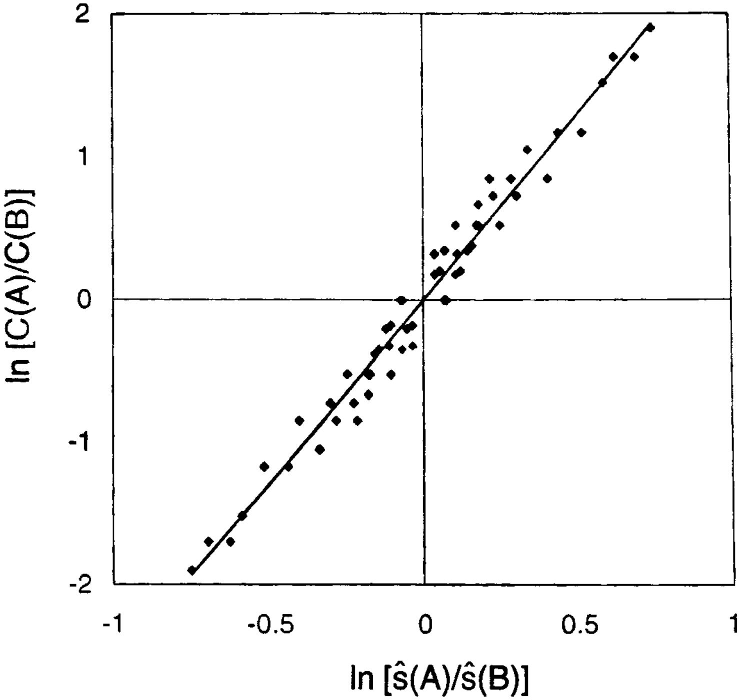

FIG. 5. Natural logarithm of the ratio certainty equivalents for prospects that pay $160 if particular teams win, plotted against the natural logarithm of the ratio of the strength ratings for those teams.

To test this assumption, we compared the median value of $\ln[(P(A_i,\overline{A}_i)/P(A_j,\overline{A}_j))]$ to the median value of $\ln R(A_i,A_j)$ for all 32 comparisons provided by the experiment design. The correlation between these two terms was .964. Fitting the relationship to the identity line yields $R^2 = .73$, a reasonably good fit.

We can now proceed to the prediction of certainty equivalents from strength ratings. Equation (10) implies that certainty equivalent ratios and strength ratios should be linear in a log–log metric. The present design provides 56 such comparisons, which are plotted in Fig. 5.

Using the median response given by our subjects and regressing the left-hand term of Eq. (10) on the right-hand term with no constant yields $\theta = 2.57$, $R^2 = .98$. The fit of the model is excellent, and the estimate of θ is remarkably close to the ratio of k to α, estimated independently from median responses to conditional probability items and choices among chance prospects, respectively: $k/\alpha = 2.03/.80 = 2.54$. Repeating this analysis on individual subjects yields a median $\theta = 1.96$; median $k/\alpha = 2.49$; median $R^2 = .76$.

DISCUSSION

The present study provides a thorough within-subject test of the major assumptions and implications of support theory in judgment and choice. In this investigation support was scaled from raw strength ratings, the major axioms of support theory were verified, and choices under uncertainty were predicted from strength judgments using a support theory framework that held together very nicely. In particular, the judged probabilities that a particular team, division, or conference would win the NBA championship exhibit binary complementarity and consistent subadditivity. Second, both major implications of the support theory representation (i.e., binary complementarity and the product rule) seem to be satisfied reasonably consistently in pairwise conditional probability judgments. Third, both assumptions necessary and sufficient to scale support as a power function of strength hold in the vast majority of cases. Fourth, support for the hypothesis that a particular team fails to win the championship (derived from simple judged probabilities) was far less than the sum of support for the seven remaining teams winning (derived from conditional probabilities and strength ratings), yielding a global weight much less than one. Fifth, the data revealed consistent subadditivity of certainty equivalents that cannot be explained by the standard economic model of choice under uncertainty (expected utility theory with risk aversion) but is consistent with a model that allows subjective probabilities that are subadditive, consistent with support theory. Sixth, certainty equivalents could be predicted to a reasonable degree of accuracy from judged probabilities and independent estimates of the utility function for monetary gains. Finally, the data suggest that relative strength ratings are an excellent predictor of relative certainty equivalents, and regression coefficients correspond closely to those predicted from independent assessments of the utility and strength-support scaling parameters. I close with several caveats concerning the interpretation and generalizability of these results.

First, I have assumed in this paper that relative strength of a particular team is an appropriate proxy for support, which in this case might be interpreted as strength of evidence that a particular team wins the NBA tournament. As mentioned earlier, it is reasonable to expect that sophisticated experts will assign different strength to these hypotheses depending on the specific matchup in question. Moreover, the strength of evidence for a team winning the championship may depend partly on the specific path they face to the finals.[16] For instance, people may view the Eastern Conference as more balanced than the Western Conference. Finally, when we condition probabilities of winning on the assumption that two particular teams reach the finals, it may be reasonable for an expert to reassess the strength of those teams. For

[16] Note that the playoffs are structured so that the Eastern Conference champion faces the Western Conference champion in the final series.

instance, if I am asked to assume that a weak team such as the Los Angeles Lakers reaches the NBA finals, I may augment my assessment of the Lakers' strength accordingly. In light of these subtle distinctions between team strength and support for the specific hypothesis that a particular team wins the championship, it is encouraging that the simplified approach forwarded here fits as well as it does in practice.

Second, in the analysis of the certainty equivalent data I assumed a model of decision under uncertainty in which subadditive judged probabilities, consistent with support theory, replace additive subjective probabilities, consistent with expected utility theory. A further refinement would transform judged probabilities by a weighting function that reflects the commonly observed pattern of diminishing sensitivity to probabilities that depart from zero or 1, as suggested in prospect theory (Tversky & Kahneman, 1992; see also Tversky & Fox, 1995). Indeed, such a model may fit the data better for judged probabilities that are very low or very high, where this distortion is typically most pronounced. A more detailed account of such a ''two-stage model'' is beyond the scope of the present paper, but is discussed at length elsewhere (see Fox & Tversky, 1998; Tversky & Wakker, 1998).

Third, in this study I have chosen to focus attention on the relationship between raw ratings of strength and judged probabilities of simple hypotheses or *implicit disjunctions* (e.g., that the Central Division wins the NBA championship). That is, attention was restricted to the difference between the left-hand term and the right-hand term in Eq. (2). A more thorough investigation would also consider *explicit disjunctions* (e.g., that Chicago or Indiana wins the NBA championship). One challenge for such a study is to determine an appropriate means by which to elicit strength for a hypothesis described as a disjunction of multiple events (see Brenner & Koehler, 1999). However, there is ample evidence consistent with both inequalities in Eq. (2) from studies of judgment (Rottenstreich & Tversky, 1997; Fox & Birke, 1998) and choice (Johnson et al., 1993; Wu & Gonzalez, 1998; Fox & Tversky, 1998; Fox, 1998).

Finally, although there is broad general support for binary complementarity, a few recent studies suggest that this condition may fail under certain conditions (Brenner & Rottenstreich, 1998, in press; Macchi, Osherson, & Krantz, 1998; see also Fox & Levav, 1999). In particular, Brenner and Rottenstreich (1998, in press) provide evidence that the sum of the probability of an explicit disjunction and its complement, $P(A_1 \vee A_2, B) + P(B, A_1 \vee A_2)$ may be systematically less than 1. The investigators attribute this phenomenon to an asymmetric tendency to spontaneously repack the focal hypothesis relative to the alternative hypothesis and suggest that support for a hypothesis may therefore differ systematically depending on whether that hypothesis is in a role as the focal versus the alternative hypothesis. If support for a hypothesis depends on whether the hypothesis is focal or alternative, it will be challenging to devise an appropriate means to assess raw support directly.

Aforementioned caveats notwithstanding, the present study provides strong support for the major axioms and consequences of support theory and demonstrates that this model of intuitive judgment under uncertainty also predicts choices when money is at stake.

REFERENCES

Aigner, D. J. (1971). *Basic econometrics.* Englewood Cliffs, NJ: Prentice Hall International.

Brenner, L., & Koehler, D. J. (1999). Subjective probability of disjunctive hypotheses: Local-weight models for decomposition of evidential support. *Cognitive Psychology,* **38,** 16–47.

Brenner, L., & Rottenstreich, Y. (1998). *Assymetric support theory: Focus, figure, and ground in likelihood judgment.* Unpublished manuscript, University of Florida.

Brenner, L., & Rottenstreich, Y. (in press). Focus, repacking, and the judgment of grouped hypotheses. *Journal of Behavioral Decision Making.*

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance,* **4,** 330–344.

Fox, C. R. (1998). *Evidence for a two-stage model of decision under uncertainty.* Unpublished manuscript, Fuqua School of Business, Duke University.

Fox, C. R., & Birke, R. (1998). *Lawyers' biases in predicting trial outcomes.* Unpublished data, Fuqua School of Business, Duke University.

Fox, C. R., & Levav, J. (1998). *Familiarity bias in relative likelihood judgment.* Unpublished manuscript, Fuqua School of Business, Duke University.

Fox, C. R., Rogers, B. A., & Tversky, A. (1996). Options traders exhibit subadditive decision weights. *Journal of Risk and Uncertainty,* **13,** 5–17.

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science,* **44,** 879–895.

Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty,* **7,** 35–51.

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

Koehler, D. J. (1996). A strength model of probability judgment. *Organizational Behavior and Human Decision Processes,* **66,** 16–21.

Macchi, L., Osheron, D., & Krantz, D. (in press). A note on superadditive probability judgment. *Psychological Review*.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays by F.P. Ramsey* (pp. 158–198). New York: Harcourt, Brace.

Redelmeier, D., Koehler, D. J., Liberman, V., & Tversky, A. (1995). Probability judgment in medicine: Discounting unspecified possibilities. *Medical Decision Making,* **15,** 227–230.

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

Savage, L. J. (1954). *The foundation of statistics*. New York: Wiley.

Teigen, K. H. (1974). Subjective sampling distributions and the additivity of estimates. *Scandinavian Journal of Psychology,* **24,** 97–105.

Tversky, A. (1967). Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology,* **75,** 27–36.

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

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review,* **91,** 293–315.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty,* **5,** 297–323.

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

Tversky, A., & Wakker, P. P. (1998). *On the composition of risk preference and belief.* Unpublished manuscript, University of Leiden.

Wu, G., & Gonzalez, R. (1998). *Dominance violations and event splitting.* Unpublished Manuscript, University of Chicago Graduate School of Business.

Accepted November 18, 1998