# Support Theory: A Nonextensional Representation of Subjective Probability

Amos Tversky and Derek J. Koehler

This article presents a new theory of subjective probability according to which different descriptions of the same event can give rise to different judgments. The experimental evidence confirms the major predictions of the theory. First, judged probability increases by unpacking the focal hypothesis and decreases by unpacking the alternative hypothesis. Second, judged probabilities are complementary in the binary case and subadditive in the general case, contrary to both classical and revisionist models of belief. Third, subadditivity is more pronounced for probability judgments than for frequency judgments and is enhanced by compatible evidence. The theory provides a unified treatment of a wide range of empirical findings. It is extended to ordinal judgments and to the assessment of upper and lower probabilities.

Both laypeople and experts are often called upon to evaluate the probability of uncertain events such as the outcome of a trial, the result of a medical operation, the success of a business venture, or the winner of a football game. Such assessments play an important role in deciding, respectively, whether to go to court, undergo surgery, invest in the venture, or bet on the home team. Uncertainty is usually expressed in verbal terms (e.g., unlikely or probable), but numerical estimates are also common. Weather forecasters, for example, often report the probability of rain (Murphy, 1985), and economists are sometimes required to estimate the chances of recession (Zarnowitz, 1985). The theoretical and practical significance of subjective probability has inspired psychologists, philosophers, and statisticians to investigate this notion from both descriptive and prescriptive standpoints.

Indeed, the question of whether degree of belief can, or should be, represented by the calculus of chance has been the focus of a long and lively debate. In contrast to the Bayesian school, which represents degree of belief by an additive probability measure, there are many skeptics who question the possibility and the wisdom of quantifying subjective uncertainty and are reluctant to apply the laws of chance to the analysis of belief. Besides the Bayesians and the skeptics, there is a growing literature on what might be called revisionist models of subjective probability. These include the Dempster–Shafer theory of belief (Dempster, 1967; Shafer, 1976), Zadeh's (1978) possibility theory, and the various types of upper and lower probabilities (e.g., see Suppes, 1974; Walley, 1991). Recent developments have been reviewed

by Dubois and Prade (1988), Gilboa and Schmeidler (in press), and Mongin (in press). Like the Bayesians, the revisionists endorse the quantification of belief, using either direct judgments or preferences between bets, but they find the calculus of chance too restrictive for this purpose. Consequently, they replace the additive measure, used in the classical theory, with a nonadditive set function satisfying weaker requirements.

A fundamental assumption that underlies both the Bayesian and the revisionist models of belief is the extensionality principle: Events with the same extension are assigned the same probability. However, the extensionality assumption is descriptively invalid because alternative descriptions of the same event often produce systematically different judgments. The following three examples illustrate this phenomenon and motivate the development of a descriptive theory of belief that is free from the extensionality assumption.

1. Fischhoff, Slovic, and Lichtenstein (1978) asked car mechanics, as well as laypeople, to assess the probabilities of different causes of a car's failure to start. They found that the mean probability assigned to the residual hypothesis—"The cause of failure is something other than the battery, the fuel system, or the engine"—increased from .22 to .44 when the hypothesis was broken up into more specific causes (e.g., the starting system, the ignition system). Although the car mechanics, who had an average of 15 years of experience, were surely aware of these possibilities, they discounted hypotheses that were not explicitly mentioned.

2. Tversky and Kahneman (1983) constructed many problems in which both probability and frequency judgments were not consistent with set inclusion. For example, one group of subjects was asked to estimate the number of seven-letter words in four pages of a novel that end with *ing*. A second group was asked to estimate the number of seven-letter words that end with $\_n\_$. The median estimate for the first question (13.4) was nearly three times higher than that for the second (4.7), presumably because it is easier to think of seven-letter words ending with *ing* than to think of seven-letter words with *n* in the sixth position. It appears that most people who evaluated the second category were not aware of the fact that it includes the first.

3. Violations of extensionality are not confined to probabil-

ity judgments; they are also observed in the evaluation of uncertain prospects. For example, Johnson, Hershey, Meszaros, and Kunreuther (1993) found that subjects who were offered (hypothetical) health insurance that covers hospitalization for any disease or accident were willing to pay a higher premium than subjects who were offered health insurance that covers hospitalization for any reason. Evidently, the explicit mention of disease and accident increases the perceived chances of hospitalization and, hence, the attractiveness of insurance.

These observations, like many others described later in this article, are inconsistent with the extensionality principle. We distinguish two sources of nonextensionality. First, extensionality may fail because of memory limitation. As illustrated in Example 2, a judge cannot be expected to recall all of the instances of a category, even when he or she can recognize them without error. An explicit description could remind people of relevant cases that might otherwise slip their minds. Second, extensionality may fail because different descriptions of the same event may call attention to different aspects of the outcome and thereby affect their relative salience. Such effects can influence probability judgments even when they do not bring to mind new instances or new evidence.

The common failures of extensionality, we suggest, represent an essential feature of human judgment, not a collection of isolated examples. They indicate that probability judgments are attached not to events but to descriptions of events. In this article, we present a theory in which the judged probability of an event depends on the explicitness of its description. This treatment, called *support theory*, focuses on direct judgments of probability, but it is also applicable to decision under uncertainty. The basic theory is introduced and characterized in the next section. The experimental evidence is reviewed in the subsequent section. In the final section, we extend the theory to ordinal judgments, discuss upper and lower indicators of belief, and address descriptive and prescriptive implications of the present development.

## Support Theory

Let T be a finite set including at least two elements, interpreted as states of the world. We assume that exactly one state obtains but it is generally not known to the judge. Subsets of T are called *events*. We distinguish between events and descriptions of events, called *hypotheses*. Let H be a set of hypotheses that describe the events in T. Thus, we assume that each hypothesis $A \in H$ corresponds to a unique event $A' \subset T$. This is a many-to-one mapping because different hypotheses, say $A$ and $B$, may have the same extension (i.e., $A' = B'$). For example, suppose one rolls a pair of dice. The hypotheses "The sum is 3" and "The product is 2" are different descriptions of the same event; namely, one die shows 1 and the other shows 2. We assume that H is finite and that it includes at least one hypothesis for each event. The following relations on H are induced by the corresponding relations on T. $A$ is *elementary* if $A' \in T$. $A$ is *null* if $A' = \emptyset$. $A$ and $B$ are *exclusive* if $A' \cap B' = \emptyset$. If $A$ and $B$ are in H, and they are exclusive, then their explicit disjunction, denoted $A \vee B$, is also in H. Thus, H is closed under exclusive disjunction. We assume that $\vee$ is associative and commutative and that $(A \vee B)' = A' \cup B'$.

A key feature of the present formulation is the distinction between explicit and implicit disjunctions. $A$ is an *implicit disjunction*, or simply an implicit hypothesis, if it is neither elementary nor null, and it is not an explicit disjunction (i.e., there are no exclusive nonnull $B$, $C$ in H such that $A = B \vee C$). For example, suppose $A$ is "Ann majors in a natural science," $B$ is "Ann majors in a biological science," and $C$ is "Ann majors in a physical science." The explicit disjunction, $B \vee C$ ("Ann majors in either a biological or a physical science"), has the same extension as $A$ (i.e., $A' = (B \vee C)' = B' \cup C'$), but $A$ is an implicit hypothesis because it is not an explicit disjunction. Note that the explicit disjunction $B \vee C$ is defined for any exclusive $B$, $C \in H$, whereas a coextensional implicit disjunction may not exist because some events cannot be naturally described without listing their components.

An *evaluation frame* $(A, B)$ consists of a pair of exclusive hypotheses: The first element $A$ is the *focal* hypothesis that the judge evaluates, and the second element $B$ is the *alternative* hypothesis. To simplify matters, we assume that when $A$ and $B$ are exclusive, the judge perceives them as such, but we do not assume that the judge can list all of the constituents of an implicit disjunction. In terms of the above example, we assume that the judge knows, for instance, that genetics is a biological science, that astronomy is a physical science, and that the biological and the physical sciences are exclusive. However, we do not assume that the judge can list all of the biological or the physical sciences. Thus, we assume recognition of inclusion but not perfect recall.

We interpret a person's probability judgment as a mapping $P$ from an evaluation frame to the unit interval. To simplify matters we assume that $P(A, B)$ equals zero if and only if $A$ is null and that it equals one if and only if $B$ is null; we assume that $A$ and $B$ are not both null. Thus, $P(A, B)$ is the judged probability that $A$ rather than $B$ holds, assuming that one and only one of them is valid. Obviously, $A$ and $B$ may each represent an explicit or an implicit disjunction. The extensional counterpart of $P(A, B)$ in the standard theory is the conditional probability $P(A' | A' \cup B')$. The present treatment is nonextensional because it assumes that probability judgment depends on the descriptions $A$ and $B$, not just on the events $A'$ and $B'$. We wish to emphasize that the present theory applies to the hypotheses entertained by the judge, which do not always coincide with the given verbal descriptions. A judge presented with an implicit disjunction may, nevertheless, think about it as an explicit disjunction, and vice versa.

Support theory assumes that there is a ratio scale $s$ (interpreted as degree of support) that assigns to each hypothesis in H a nonnegative real number such that, for any pair of exclusive hypotheses $A, B \in H$,

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

If $B$ and $C$ are exclusive, $A$ is implicit, and $A' = (B \vee C)'$, then

$$s(A) \leq s(B \vee C) = s(B) + s(C). \tag{2}$$

Equation 1 provides a representation of subjective probability in terms of the support of the focal and the alternative hypotheses. Equation 2 states that the support of an implicit disjunction $A$ is less than or equal to that of a coextensional explicit disjunction $B \vee C$ that equals the sum of the support of its components.

Thus, support is additive for explicit disjunctions and subadditive for implicit ones.

The subadditivity assumption, we suggest, represents a basic principle of human judgment. When people assess their degree of belief in an implicit disjunction, they do not normally unpack the hypothesis into its exclusive components and add their support, as required by extensionality. Instead, they tend to form a global impression that is based primarily on the most representative or available cases. Because this mode of judgment is selective rather than exhaustive, unpacking tends to increase support. In other words, we propose that the support of a summary representation of an implicit hypothesis is generally less than the sum of the support of its exclusive components. Both memory and attention may contribute to this effect. Unpacking a category (e.g., death from an unnatural cause) into its components (e.g., homicide, fatal car accidents, drowning) might remind people of possibilities that would not have been considered otherwise. Moreover, the explicit mention of an outcome tends to enhance its salience and hence its support. Although this assumption may fail in some circumstances, the overwhelming evidence for subadditivity, described in the next section, indicates that these failures represent the exception rather than the rule.

The support associated with a given hypothesis is interpreted as a measure of the strength of evidence in favor of this hypothesis that is available to the judge. The support may be based on objective data (e.g., the frequency of homicide in the relevant population) or on a subjective impression mediated by judgmental heuristics, such as representativeness, availability, or anchoring and adjustment (Kahneman, Slovic, & Tversky, 1982). For example, the hypothesis "Bill is an accountant" may be evaluated by the degree to which Bill's personality matches the stereotype of an accountant, and the prediction "An oil spill along the eastern coast before the end of next year" may be assessed by the ease with which similar accidents come to mind. Support may also reflect reasons or arguments recruited by the judge in favor of the hypothesis in question (e.g., if the defendant were guilty, he would not have reported the crime). Because judgments based on impressions and reasons are often nonextensional, the support function is nonmonotonic with respect to set inclusion. Thus, $s(B)$ may exceed $s(A)$ even though $A' \supset B'$. Note, however, that $s(B)$ cannot exceed $s(B \vee C)$. For example, if the support of a category is determined by the availability of its instances, then the support of the hypothesis that a randomly selected word ends with $ing$ can exceed the support of the hypothesis that the word ends with $\_n\_$. Once the inclusion relation between the categories is made transparent, the $\_n\_$ hypothesis is replaced by "$ing$ or any other $\_n\_$," whose support exceeds that of the $ing$ hypothesis.

The present theory provides an interpretation of subjective probability in terms of relative support. This interpretation suggests that, in some cases, probability judgment may be predicted from independent assessments of support. This possibility is explored later. The following discussion shows that, under the present theory, support can be derived from probability judgments, much as utility is derived from preferences between options.

## Consequences

Support theory has been formulated in terms of the support function $s$, which is not directly observable. We next characterize the theory in terms of the observed index $P$. We first exhibit four consequences of the theory and then show that they imply Equations 1 and 2. An immediate consequence of the theory is binary complementarity:

$$P(A, B) + P(B, A) = 1. \tag{3}$$

A second consequence is proportionality:

$$\frac{P(A, B)}{P(B, A)} = \frac{P(A, B \vee C)}{P(B, A \vee C)}, \tag{4}$$

provided that $A$, $B$, and $C$ are mutually exclusive and $B$ is not null. Thus, the "odds" for $A$ against $B$ are independent of the additional hypothesis $C$.

To formulate the next condition, it is convenient to introduce the probability ratio $R(A, B) = P(A, B)/P(B, A)$, which is the odds for $A$ against $B$. Equation 1 implies the following product rule:

$$R(A, B)R(C, D) = R(A, D)R(C, B), \tag{5}$$

provided that $A$, $B$, $C$, and $D$ are not null and the four pairs of hypotheses in Equation 5 are pairwise exclusive. Thus, the product of the odds for $A$ against $B$ and for $C$ against $D$ equals the product of the odds for $A$ against $D$ and for $C$ against $B$. To see the necessity of the product rule, note that, according to Equation 1, both sides of Equation 5 equal $s(A)s(C)/s(B)s(D)$. Essentially the same condition has been used in the theory of preference trees (Tversky & Sattath, 1979).

Equations 1 and 2 together imply the unpacking principle. Suppose $B$, $C$, and $D$ are mutually exclusive, $A$ is implicit, and $A' = (B \vee C)'$. Then

$$P(A, D) \leq P(B \vee C, D) = P(B, C \vee D) + P(C, B \vee D). \tag{6}$$

The properties of $s$ entail the corresponding properties of $P$: Judged probability is additive for explicit disjunctions and subadditive for implicit disjunctions. In other words, unpacking an implicit disjunction may increase, but not decrease, its judged probability. Unlike Equations 3–5, which hold in the standard theory of probability, the unpacking principle (Equation 6) generalizes the classical model. Note that this assumption is at variance with lower probability models, including Shafer's (1976), which assume extensionality and superadditivity (i.e., $P(A' \cup B') \geq P(A') + P(B')$ if $A' \cap B' = \emptyset$).

There are two conflicting intuitions that yield nonadditive probability. The first intuition, captured by support theory, suggests that unpacking an implicit disjunction enhances the salience of its components and consequently increases support. The second intuition, captured by Shafer's (1976) theory, among others, suggests that—in the face of partial ignorance—the judge holds some measure of belief "in reserve" and does not distribute it among all elementary hypotheses, as required by the Bayesian model. Although Shafer's theory is based on a logical rather than a psychological analysis of belief, it has also been interpreted by several authors as a descriptive model. Thus, it provides a natural alternative to be compared with the present theory.

Whereas proportionality (Equation 4) and the product rule (Equation 5) have not been systematically tested before, a number of investigators have observed binary complementarity (Equation 3) and some aspects of the unpacking principle

(Equation 6). These data, as well as several new studies, are reviewed in the next section. The following theorem shows that the above conditions are not only necessary but also sufficient for support theory. The proof is given in the Appendix.

> Theorem 1: Suppose $P(A, B)$ is defined for all exclusive $A, B \in H$ and that it vanishes if and only if $A$ is null. Equations 3–6 hold if and only if there exists a nonnegative ratio scale $s$ on H that satisfies Equations 1 and 2.

The theorem shows that if probability judgments satisfy the required conditions, it is possible to scale the support or strength of evidence associated with each hypothesis without assuming that hypotheses with the same extension have equal support. An ordinal generalization of the theory, in which $P$ is treated as an ordinal rather than cardinal scale, is presented in the final section. In the remainder of this section, we introduce a representation of subadditivity and a treatment of conditioning.

## Subadditivity

We extend the theory by providing a more detailed representation of subadditivity. Let $A$ be an implicit hypothesis with the same extension as the explicit disjunction of the elementary hypotheses $A_1, \ldots, A_n$; that is, $A' = (A_1 \vee \cdots \vee A_n)'$. Assume that any two elementary hypotheses, $B$ and $C$, with the same extension have the same support; that is, $B', C' \in T$ and $B' = C'$ implies $s(B) = s(C)$. It follows that, under this assumption we can write

$$s(A) = w_{1A}s(A_1) + \cdots + w_{nA}s(A_n), \quad 0 \le w_{iA} \le 1,$$

$$i = 1, \ldots, n. \quad (7)$$

In this representation, the support of each elementary hypothesis is "discounted" by its respective weight, which reflects the degree to which the judge attends to the hypothesis in question. If $w_{iA} = 1$ for all $i$, then $s(A)$ is the sum of the support of its elementary hypotheses, as in an explicit disjunction. On the other hand, $w_{jA} = 0$ for some $j$ indicates that $A_j$ is effectively ignored. Finally, if the weights add to one, then $s(A)$ is a weighted average of the $s(A_i)$, $1 \le i \le n$. We hasten to add that Equation 7 should not be interpreted as a process of deliberate discounting in which the judge assesses the support of an implicit disjunction by discounting the assessed support of the corresponding explicit disjunction. Instead, the weights are meant to represent the result of an assessment process in which the judge evaluates $A$ without explicitly unpacking it into its elementary components. It should also be kept in mind that elementary hypotheses are defined relative to a given sample space. Such hypotheses may be broken down further by refining the level of description.

Note that whereas the support function is unique, except for a unit of measurement, the "local" weights $w_{iA}$ are not uniquely determined by the observed probability judgments. These data, however, determine the "global" weights $w_A$ defined by

$$s(A) = w_A[s(A_1) + \cdots + s(A_n)], \quad 0 \le w_A \le 1. \quad (8)$$

The global weight $w_A$, which is the ratio of the support of the corresponding implicit ($A$) and explicit ($A_1 \vee \cdots \vee A_n$) disjunctions, provides a convenient measure of the degree of subadditivity induced by $A$. The degree of subadditivity, we pro-

pose, is influenced by several factors, one of which is the interpretation of the probability scale. Specifically, subadditivity is expected to be more pronounced when probability is interpreted as a propensity of an individual case than when it is equated with, or estimated by, relative frequency. Kahneman and Tversky (1979, 1982) referred to these modes of judgment as singular and distributional, respectively, and argued that the latter is usually more accurate than the former[1] (see also Reeves & Lockhart, 1993). Although many events of interest cannot be interpreted in frequentistic terms, there are questions that can be framed in either a distributional or a singular mode. For example, people may be asked to assess the probability that an individual, selected at random from the general population, will die as a result of an accident. Alternatively, people may be asked to assess the percentage (or relative frequency) of the population that will die as a result of an accident. We propose that the implicit disjunction "accident" is more readily unpacked into its components (e.g., car accidents, plane crashes, fire, drowning, poisoning) when the judge considers the entire population rather than a single person. The various causes of death are all represented in the population's mortality statistics but not in the death of a single person. More generally, we propose that the tendency to unpack an implicit disjunction is stronger in the distributional than in the singular mode. Hence, a frequentistic formulation is expected to produce less discounting (i.e., higher $w$s) than a formulation that refers to an individual case.

## Conditioning

Recall that $P(A, B)$ is interpreted as the conditional probability of $A$, given $A$ or $B$. To obtain a general treatment of conditioning, we enrich the hypothesis set H by assuming that if $A$ and $B$ are distinct elements of H, then their conjunction, denoted $AB$, is also in H. Naturally, we assume that conjunction is associative and commutative and that $(AB)' = A' \cap B'$. We also assume distributivity, that is, $A(B \vee C) = AB \vee AC$. Let $P(A, B \mid D)$ be the judged probability that $A$ rather than $B$ holds, given some data $D$. In general, new evidence (i.e., a different state of information) gives rise to a new support function $s_D$ that describes the revision of $s$ in light of $D$. In the special case in which the data can be described as an element of H, which merely restricts the hypotheses under consideration, we can represent conditional probability by

$$P(A, B \mid D) = \frac{s(AD)}{s(AD) + s(BD)}, \quad (9)$$

provided that $A$ and $B$ are exclusive but $A \vee B$ and $D$ are not.

Several comments on this form are in order. First, note that if $s$ is additive, then Equation 9 reduces to the standard definition of conditional probability. If $s$ is subadditive, as we have assumed throughout, then judged probability depends not only on the description of the focal and the alternative hypotheses but also on the description of the evidence $D$. Suppose $D' = (D_1 \vee D_2)'$, $D_1$ and $D_2$ are exclusive, and $D$ is implicit. Then

---

[1] Gigerenzer (1991) has further argued that the biases observed in probability judgments of unique events disappear in judgments of frequency, but the data reviewed here and elsewhere are inconsistent with this claim.

$$P(A, B | D_1 \vee D_2) = \frac{s(AD_1 \vee AD_2)}{s(AD_1 \vee AD_2) + s(BD_1 \vee BD_2)}.$$

But because $s(AD) \leq s(AD_1 \vee AD_2)$ and $s(BD) \leq s(BD_1 \vee BD_2)$ by subadditivity, the unpacking of $D$ may favor one hypothesis over another. For example, the judged probability that a woman earns a very high salary given that she is a university professor is likely to increase when "university" is unpacked into "law school, business school, medical school, or any other school" because of the explicit mention of high-paying positions. Thus, Equation 9 extends the application of subadditivity to the representation of evidence. As we show later, it also allows us to compare the impact of different bodies of evidence, provided they can be described as elements of H.

Consider a collection of $n \geq 3$ mutually exclusive and exhaustive (nonnull) hypotheses, $A_1 \cdots A_n$, and let $\bar{A}_i$ denote the negation of $A_i$ that corresponds to an implicit disjunction of the remaining hypotheses. Consider two items of evidence, $B, C \in$ H, and suppose that each $A_i$ is more compatible with $B$ than with $C$ in the sense that $s(BA_i) \geq s(CA_i)$, $1 \leq i \leq n$. We propose that $B$ induces more subadditivity than $C$ so that $s(B\bar{A}_i)$ is discounted more heavily than $s(C\bar{A}_i)$ (i.e., $w_{B\bar{A}_i} \leq w_{C\bar{A}_i}$; see Equation 7). This assumption called *enhancement*, suggests that the assessments of $P(A_i, \bar{A}_i | B)$ will be generally higher than those of $P(A_i, \bar{A}_i | C)$. More specifically, we propose that the sum of the probabilities of $A_1 \cdots A_n$, each evaluated by different judges,[2] is no smaller under $B$ than under $C$. That is,

$$\sum_{i=1}^{n} P(A_i, \bar{A}_i | B) \geq \sum_{i=1}^{n} P(A_i, \bar{A}_i | C). \qquad (10)$$

Subadditivity implies that both sums are greater than or equal to one. The preceding inequality states that the sum is increased by evidence that is more compatible with the hypotheses under study. It is noteworthy that enhancement suggests that people are inappropriately responsive to the prior probability of the data, whereas base-rate neglect indicates that people are not sufficiently responsive to the prior probability of the hypotheses. The following schematic example illustrates an implication of enhancement and compares it with other models.

Suppose that a murder was committed by one (and only one) of several suspects. In the absence of any specific evidence, assume that all suspects are considered about equally likely to have committed the crime. Suppose further that a preliminary investigation has uncovered a body of evidence (e.g., motives and opportunities) that implicates each of the suspects to roughly the same degree. According to the Bayesian model, the probabilities of all of the suspects remain unchanged because the new evidence is nondiagnostic. In Shafer's theory of belief functions, the judged probability that the murder was committed by one suspect rather than by another generally increases with the amount of evidence; thus, it should be higher after the investigation than before. Enhancement yields a different pattern: The binary probabilities (i.e., of one suspect against another) are expected to be approximately one half, both before and after the investigation, as in the Bayesian model. However, the probability that the murder was committed by a particular suspect (rather than by any of the others) is expected to increase with the amount of evidence. Experimental tests of enhancement are described in the next section.

## Data

In this section, we discuss the experimental evidence for support theory. We show that the interpretation of judged probability in terms of a normalized subadditive support function provides a unified account of several phenomena reported in the literature; it also yields new predictions that have not been tested heretofore. This section consists of four parts. In the first part, we investigate the effect of unpacking and examine factors that influence the degree of subadditivity. In the second, we relate probability judgments to direct ratings of evidence strength. In the third, we investigate the enhancement effect and compare alternative models of belief. In the final part, we discuss the conjunction effect, hypothesis generation, and decision under uncertainty.

### Studies of Unpacking

Recall that the unpacking principle (Equation 6) consists of two parts: additivity for explicit disjunctions and subadditivity for implicit disjunctions, which jointly entail nonextensionality. (Binary complementarity [Equation 3] is a special case of additivity.) Because each part alone is subject to alternative interpretations, it is important to test additivity and subadditivity simultaneously. For this reason, we first describe several new studies that have tested both parts of the unpacking principle within the same experiment, and then we review previous research that provided the impetus for the present theory.

### Study 1: Causes of Death

Our first study followed the seminal work of Fischhoff et al. (1978) on fault trees, using a task similar to that studied by Russo and Kolzow (1992). We asked Stanford undergraduates ($N = 120$) to assess the likelihood of various possible causes of death. The subjects were informed that each year approximately 2 million people in the United States (nearly 1% of the population) die from different causes, and they were asked to estimate the probability of death from a variety of causes. Half of the subjects considered a single person who had recently died and assessed the probability that he or she had died from each in a list of specified causes. They were asked to assume that the person in question had been randomly selected from the set of people who had died the previous year. The other half, given a frequency judgment task, assessed the percentage of the 2 million deaths in the previous year attributable to each cause. In each group, half of the subjects were promised that the 5 most accurate subjects would receive $20 each.

Each subject evaluated one of two different lists of causes, constructed such that he or she evaluated either an implicit hypothesis (e.g., death resulting from natural causes) or a coextensional explicit disjunction (e.g., death resulting from heart disease, cancer, or some other natural cause), but not both. The

---

[2] Enhancement, like subadditivity, may not hold when a person evaluates these probabilities at the same time because this task introduces additional constraints.

Table 1

*Mean Probability and Frequency Estimates for Causes of Death in Study 1, Comparing Evaluations of Explicit Disjunctions With Coextensional Implicit Disjunctions*

| Hypothesis | Mean estimate (%) | | Actual % |
|---|---|---|---|
| | Probability | Frequency | |
| Three-component | | | |
| P(heart disease) | 22 | 18 | 34.1 |
| P(cancer) | 18 | 20 | 23.1 |
| P(other natural cause) | 33 | 29 | 35.2 |
| Σ(natural cause) | 73 | 67 | 92.4 |
| P(natural cause) | 58 | 56 | |
| Σ/P | 1.26 | 1.20 | |
| P(accident) | 32 | 30 | 4.4 |
| P(homicide) | 10 | 11 | 1.1 |
| P(other unnatural cause) | 11 | 12 | 2.1 |
| Σ(unnatural cause) | 53 | 53 | 7.6 |
| P(unnatural cause) | 32 | 39 | |
| Σ/P | 1.66 | 1.36 | |
| Seven-component | | | |
| P(respiratory cancer \| natural) | 12 | 11 | 7.1 |
| P(digestive cancer \| natural) | 8 | 7 | 5.9 |
| P(genitourinary cancer \| natural) | 5 | 3 | 2.7 |
| P(breast cancer \| natural) | 13 | 9 | 2.2 |
| P(urinary cancer \| natural) | 7 | 3 | 1.0 |
| P(leukemia \| natural) | 8 | 6 | 1.0 |
| P(other cancer \| natural) | 17 | 10 | 5.1 |
| Σ(cancer \| natural) | 70 | 49 | 25.0 |
| P(cancer \| natural) | 32 | 24 | |
| Σ/P | 2.19 | 2.04 | |
| P(auto accident \| unnatural) | 33 | 33 | 30.3 |
| P(firearm accident \| unnatural) | 7 | 12 | 1.3 |
| P(accidental fall \| unnatural) | 6 | 4 | 7.9 |
| P(death in fire \| unnatural) | 4 | 5 | 2.6 |
| P(drowning \| unnatural) | 5 | 4 | 2.6 |
| P(accidental poisoning \| unnatural) | 4 | 3 | 3.9 |
| P(other accident \| unnatural) | 24 | 17 | 9.2 |
| Σ(accident \| unnatural) | 83 | 78 | 57.9 |
| P(accident \| unnatural) | 45 | 48 | |
| Σ/P | 1.84 | 1.62 | |

*Note.* Actual percentages were taken from the 1990 *U.S. Statistical Abstract.* Σ = sum of mean estimates.

full set of causes considered is listed in Table 1. Causes of death were divided into natural and unnatural types. Each type had three components, one of which was further divided into seven subcomponents. To avoid very small probabilities, we conditioned these seven subcomponents on the corresponding type of death (i.e., natural or unnatural). To provide subjects with some anchors, we informed them that the probability or frequency of death resulting from respiratory illness is about 7.5% and the probability or frequency of death resulting from suicide is about 1.5%.

Table 1 shows that, for both probability and frequency judgments, the mean estimate of an implicit disjunction (e.g., death from a natural cause) is smaller than the sum of the mean estimates of its components (heart disease, cancer, or other natural causes), denoted Σ (natural causes). Specifically, the former equals 58%, whereas the latter equals 22% + 18% + 33% = 73%. All eight comparisons in Table 1 are statistically significant ($p < .05$) by Mann–Whitney $U$ test. (We used a nonparametric test because of the unequal variances involved when comparing a single measured variable with a sum of measured variables.)

Throughout this article, we use the ratio of the probabilities assigned to coextensional explicit and implicit hypotheses as a measure of subadditivity. The ratio in the preceding example is 1.26. This index, called the *unpacking factor,* can be computed directly from probability judgments, unlike $w$, which is defined in terms of the support function. Subadditivity is indicated by an unpacking factor greater than 1 and a value of $w$ less than 1. It is noteworthy that subadditivity, by itself, does not imply that explicit hypotheses are overestimated or that implicit hypotheses are underestimated relative to an appropriate objective criterion. It merely indicates that the former are judged as more probable than the latter.

In this study, the mean unpacking factors were 1.37 for the three-component hypotheses and 1.92 for the seven-component hypotheses, indicating that the degree of subadditivity increased with the number of components in the explicit disjunction. An analysis of medians rather than means revealed a similar pattern, with somewhat smaller differences between packed and unpacked versions. Comparison of probability and frequency tasks showed, as expected, that subjects gave higher and thus more subadditive estimates when judging probabilities than when judging frequencies, $F(12, 101) = 2.03, p < .05$. The average unpacking factors were 1.74 for probability and 1.56 for frequency.

The judgments generally overestimated the actual values, obtained from the 1990 *U.S. Statistical Abstract.* The only clear exception was heart disease, which had an actual probability of 34% but received a mean judgment of 20%. Because subjects produced higher judgments of probability than of frequency, the former exhibited greater overestimation of the actual values, but the correlation between the estimated and actual values (computed separately for each subject) revealed no difference between the two tasks. Monetary incentives did not improve the accuracy of people's judgments.

The following design provides a more stringent test of support theory and compares it with alternative models of belief. Suppose $A_1$, $A_2$, and $B$ are mutually exclusive and exhaustive; $A' = (A_1 \vee A_2)'$; $A$ is implicit; and $\bar{A}$ is the negation of $A$. Consider the following observable values:

$$\alpha = P(A, B);$$

$$\beta = P(A_1 \vee A_2, B);$$

$$\gamma_1 = P(A_1, A_2 \vee B), \gamma_2 = P(A_2, A_1 \vee B), \gamma = \gamma_1 + \gamma_2; \text{ and}$$

$$\delta_1 = P(A_1, \bar{A}_1), \delta_2 = (A_2, \bar{A}_2), \delta = \delta_1 + \delta_2.$$

Different models of belief imply different orderings of these values:

support theory, $\alpha \leq \beta = \gamma \leq \delta$;

Bayesian model, $\alpha = \beta = \gamma = \delta$;

belief function, $\alpha = \beta \geq \gamma = \delta$; and

regressive model, $\alpha = \beta \leq \gamma = \delta$.

Support theory predicts $\alpha \leq \beta$ and $\gamma \leq \delta$ due to the unpacking of the focal and residual hypotheses, respectively; it also predicts $\beta = \gamma$ due to the additivity of explicit disjunctions. The Bayesian model implies $\alpha = \beta$ and $\gamma = \delta$, by extensionality, and $\beta = \gamma$, by additivity. Shafer's theory of belief functions also as-

Table 2

*Mean and Median Probability Estimates for Various Causes of Death*

| Probability judgments | Mean | Median |
|---|---|---|
| $\beta = P(\text{accident or homicide, OUC})$ | 64 | 70 |
| $\gamma_1 = P(\text{accident, homicide or OUC})$ | 53 | 60 |
| $\gamma_2 = P(\text{homicide, accident or OUC})$ | 16 | 10 |
| $\gamma = \gamma_1 + \gamma_2$ | 69 | 70 |
| $\delta_1 = P(\text{accident, OUC})$ | 56 | 65 |
| $\delta_2 = P(\text{homicide, OUC})$ | 24 | 18 |
| $\delta = \delta_1 + \delta_2$ | 80 | 83 |

*Note.* OUC = other unnatural causes.

sumes extensionality, but it predicts $\beta \geq \gamma$ because of superadditivity. The above data, as well as numerous studies reviewed later, demonstrate that $\alpha < \delta$, which is consistent with support theory but inconsistent with both the Bayesian model and Shafer's theory.

The observation that $\alpha < \delta$ could also be explained by a *regressive model* that assumes that probability judgments satisfy extensionality but are biased toward .5 (e.g., see Erev, Wallsten, & Budescu, 1994). For example, the judge might start with a "prior" probability of .5 that is not revised sufficiently in light of the evidence. Random error could also produce regressive estimates. If each individual judgment is biased toward .5, then $\beta$, which consists of a single judgment, would be less than $\gamma$, which is the sum of two judgments. On the other hand, this model predicts no difference between $\alpha$ and $\beta$, each of which consists of a single judgment, or between $\gamma$ and $\delta$, each of which consists of two. Thus, support theory and the regressive model make different predictions about the source of the difference between $\alpha$ and $\delta$. Support theory predicts subadditivity for implicit disjunctions (i.e., $\alpha \leq \beta$ and $\gamma \leq \delta$) and additivity for explicit disjunctions (i.e., $\beta = \gamma$), whereas the regressive model assumes extensionality (i.e., $\alpha = \beta$ and $\gamma = \delta$) and subadditivity for explicit disjunctions (i.e., $\beta \leq \gamma$).

To contrast these predictions, we asked different groups (of 25 to 30 subjects each) to assess the probability of various unnatural causes of death. All subjects were told that a person had been randomly selected from the set of people who had died the previous year from an unnatural cause. The hypotheses under study and the corresponding probability judgments are summarized in Table 2. The first row, for example, presents the judged probability $\beta$ that death was caused by an accident or a homicide rather than by some other unnatural cause. In accord with support theory, $\delta = \delta_1 + \delta_2$ was significantly greater than $\gamma = \gamma_1 + \gamma_2$, $p < .05$ (by Mann–Whitney $U$ test), but $\gamma$ was not significantly greater than $\beta$, contrary to the prediction of the regressive model. Nevertheless, we do not rule out the possibility that regression toward .5 could yield $\beta < \gamma$, which would contribute to the discrepancy between $\alpha$ and $\delta$. A generalization of support theory that accommodates such a pattern is considered in the final section.

### Study 2: Suggestibility and Subadditivity

Before turning to additional demonstrations of unpacking, we discuss some methodological questions regarding the elicita-

tion of probability judgments. It could be argued that asking a subject to evaluate a specific hypothesis conveys a subtle (or not so subtle) suggestion that the hypothesis is quite probable. Subjects, therefore, might treat the fact that the hypothesis has been brought to their attention as information about its probability. To address this objection, we devised a task in which the assigned hypotheses carried no information so that any observed subadditivity could not be attributed to experimental suggestion.

Stanford undergraduates ($N = 196$) estimated the percentage of U.S. married couples with a given number of children. Subjects were asked to write down the last digit of their telephone numbers and then to evaluate the percentage of couples having exactly that many children. They were promised that the 3 most accurate respondents would be awarded $10 each. As predicted, the total percentage attributed to the numbers 0 through 9 (when added across different groups of subjects) greatly exceeded 1. The total of the means assigned by each group was 1.99, and the total of the medians was 1.80. Thus, subadditivity was very much in evidence, even when the selection of focal hypothesis was hardly informative. Subjects overestimated the percentage of couples in all categories, except for childless couples, and the discrepancy between the estimated and the actual percentages was greatest for the modal couple with 2 children. Furthermore, the sum of the probabilities for 0, 1, 2, and 3 children, each of which exceeded .25, was 1.45. The observed subadditivity, therefore, cannot be explained merely by a tendency to overestimate very small probabilities.

Other subjects ($N = 139$) were asked to estimate the percentage of U.S. married couples with "less than 3," "3 or more," "less than 5," or "5 or more" children. Each subject considered exactly one of the four hypotheses. The estimates added to 97.5% for the first pair of hypotheses and to 96.3% for the second pair. In sharp contrast to the subadditivity observed earlier, the estimates for complementary pairs of events were roughly additive, as implied by support theory. The finding of binary complementarity is of special interest because it excludes an alternative explanation of subadditivity according to which the evaluation of evidence is biased in favor of the focal hypothesis.

### Subadditivity in Expert Judgments

Is subadditivity confined to novices, or does it also hold for experts? Redelmeier, Koehler, Liberman, and Tversky (1993) explored this question in the context of medical judgments. They presented physicians at Stanford University ($N = 59$) with a detailed scenario concerning a woman who reported to the emergency room with abdominal pain. Half of the respondents were asked to assign probabilities to two specified diagnoses (gastroenteritis and ectopic pregnancy) and a residual category (none of the above); the other half assigned probabilities to five specified diagnoses (including the two presented in the other condition) and a residual category (none of the above). Subjects were instructed to give probabilities that summed to one because the possibilities under consideration were mutually exclusive and exhaustive. If the physicians' judgments conform to the classical theory, then the probability assigned to the residual category in the two-diagnosis condition should equal the sum of the probabilities assigned to its unpacked components in the five-diagnosis condition. Consistent with the predictions of sup-

port theory, however, the judged probability of the residual in the two-diagnosis condition (mean = .50) was significantly lower than that of the unpacked components in the five-diagnosis condition (mean = .69), $p < .005$ (Mann–Whitney $U$ test).

In a second study, physicians from Tel Aviv University ($N$ = 52) were asked to consider several medical scenarios consisting of a one-paragraph statement including the patient's age, gender, medical history, presenting symptoms, and the results of any tests that had been conducted. One scenario, for example, concerned a 67-year-old man who arrived in the emergency room suffering a heart attack that had begun several hours earlier. Each physician was asked to assess the probability of one of the following four hypotheses: patient dies during this hospital admission ($A$); patient is discharged alive but dies within 1 year ($B$); patient lives more than 1 but less than 10 years ($C$); or patient lives more than 10 years ($D$). Throughout this article, we refer to these as *elementary judgments* because they pit an elementary hypothesis against its complement, which is an implicit disjunction of all of the remaining elementary hypotheses. After assessing one of these four hypotheses, all respondents assessed $P(A, B)$, $P(B, C)$, and $P(C, D)$ or the complementary set. We refer to these as *binary judgments* because they involve a comparison of two elementary hypotheses.

As predicted, the elementary judgments were substantially subadditive. The means of the four groups in the preceding example were 14% for $A$, 26% for $B$, 55% for $C$, and 69% for $D$, all of which overestimated the actual values reported in the medical literature. In problems like this, when individual components of a partition are evaluated against the residual, the denominator of the unpacking factor is taken to be 1; thus, the unpacking factor is simply the total probability assigned to the components (summed over different groups of subjects). In this example, the unpacking factor was 1.64. In sharp contrast, the binary judgments (produced by two different groups of physicians) exhibited near-perfect additivity, with a mean total of 100.5% assigned to complementary pairs.

Further evidence for subadditivity in expert judgment has been provided by Fox, Rogers, and Tversky (1994), who investigated 32 professional options traders at the Pacific Stock Exchange. These traders made probability judgments regarding the closing price of Microsoft stock on a given future date (e.g., that it will be less than $88 per share). Microsoft stock is traded at the Pacific Stock Exchange, and the traders are commonly concerned with the prediction of its future value. Nevertheless, their judgments exhibited the predicted pattern of subadditivity and binary complementarity. The average unpacking factor for a fourfold partition was 1.47, and the average sum of complementary binary events was 0.98. Subadditivity in expert judgments has been documented in other domains by Fischhoff et al. (1978), who studied auto mechanics, and by Dube-Rioux and Russo (1988), who studied restaurant managers.

## Review of Previous Research

We next review other studies that have provided tests of support theory. Tversky and Fox (1994) asked subjects to assign probabilities to various intervals in which an uncertain quantity might fall, such as the margin of victory in the upcoming Super Bowl or the change in the Dow–Jones Industrial Average over the next week. When a given event (e.g., "Buffalo beats
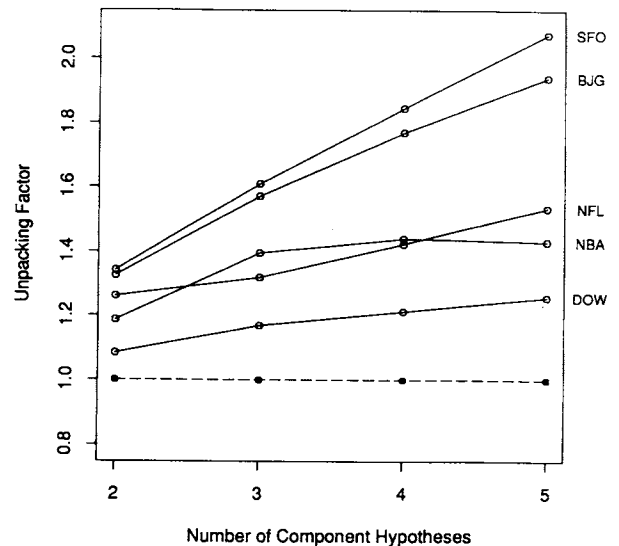


*Figure 1.*   Unpacking factors from Tversky and Fox's (1994) data. SFO = San Francisco temperature; BJG = Beijing temperature; NFL = 1991 National Football League Super Bowl; NBA = National Basketball Association playoff; DOW = weekly change in Dow–Jones index.

Washington") was unpacked into individually evaluated components (e.g., "Buffalo beats Washington by less than 7 points" and "Buffalo beats Washington by at least 7 points"), subjects' judgments were substantially subadditive. Figure 1 plots the unpacking factor obtained in this study as a function of the number of component hypotheses in the explicit disjunction. Judgments for five different types of event are shown: future San Francisco temperature (SFO), future Beijing temperature (BJG), the outcome of the Super Bowl of the National Football League (NFL), the outcome of a playoff game of the National Basketball Association (NBA), and weekly change in the Dow–Jones index (DOW). Recall that an unpacking factor greater than 1 (i.e., falling above the dashed line in the plot) indicates subadditivity. The results displayed in Figure 1 reveal consistent subadditivity for all sources that increases with the number of components in the explicit disjunction.

Figure 2 plots the median probabilities assigned to complementary hypotheses. (Each hypothesis is represented twice in the plot, once as the focal hypothesis and once as the complement.) As predicted by support theory, judgments of intervals representing complementary pairs of hypotheses were essentially additive, with no apparent tendency toward either subadditivity or superadditivity.

Further evidence for binary complementarity comes from an extensive study conducted by Wallsten, Budescu, and Zwick (1992),[3] who presented subjects with 300 propositions concerning world history and geography (e.g., "The Monroe Doctrine was proclaimed before the Republican Party was founded") and asked them to estimate the probability that each was true. True and false (complementary) versions of each proposition were presented on different days. Figure 3 plots the mean probabilities assigned to each of the propositions in both their true and
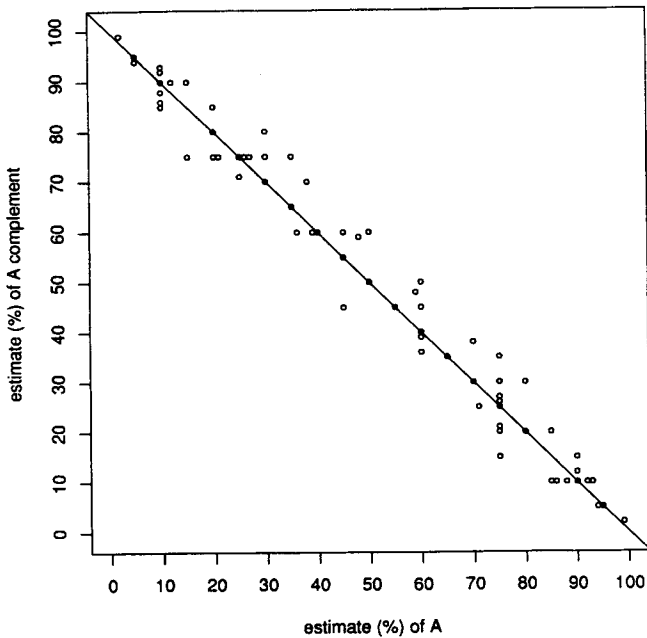
*Figure 2.* A test of binary complementarity based on Tversky and Fox (1994).

false versions using the format of Figure 2. Again, the judgments are additive (mean = 1.02) through the entire range.

We next present a brief summary of the major findings and list both current and previous studies supporting each conclusion.

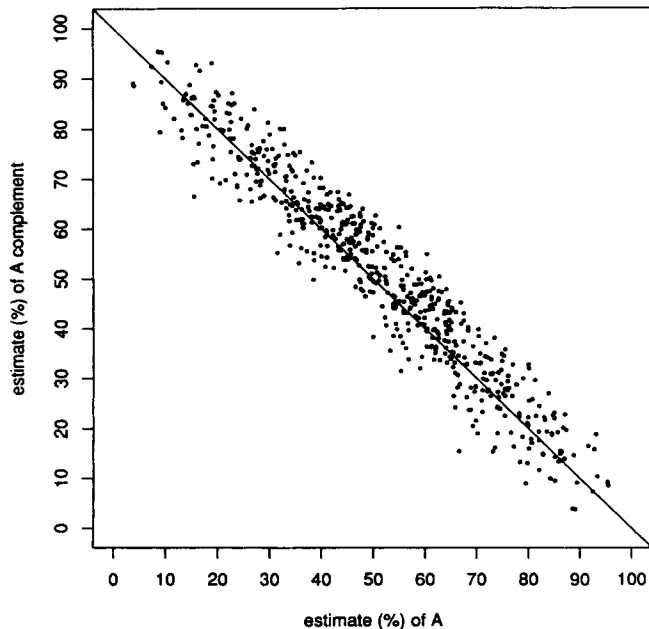*Subadditivity.* Unpacking an implicit hypothesis into its



*Figure 3.* A test of binary complementarity based on Wallsten, Budescu, and Zwick (1992).

component hypotheses increases its total judged probability, yielding subadditive judgments. Tables 3 and 4 list studies that provide tests of the unpacking condition. For each experiment, the probability assigned to the implicit hypothesis and the total probability assigned to its components in the explicit disjunction are listed along with the resulting unpacking factor. All of the listed studies used an experimental design in which the implicit disjunction and the components of the explicit disjunction were evaluated independently, either by separate groups of subjects or by the same subjects but with a substantial number of intervening judgments. The probabilities are listed as a function of the number of components in the explicit disjunction and are collapsed over all other independent variables. Table 3 lists studies in which subjects evaluated the probability of qualitative hypotheses (e.g., the probability that Bill W. majors in psychology); Table 4 lists studies in which subjects evaluated quantitative hypotheses (e.g., the probability that a randomly selected adult man is between 6 ft and 6 ft 2 in. tall).

The tables show that the observed unpacking factors are, without exception, greater than one, indicating consistent subadditivity. The fact that subadditivity is observed both for qualitative and for quantitative hypotheses is instructive. Subadditivity in assessments of qualitative hypotheses can be explained, in part at least, by the failure to consider one or more component hypotheses when the event in question is described in an implicit form. The subadditivity observed in judgments of quantitative hypotheses, however, cannot be explained as a retrieval failure. For example, Teigen (1974b, Experiment 2) found that the judged proportion of college students whose heights fell in a given interval increased when that interval was broken into several smaller intervals that were assessed separately. Subjects evaluating the implicit disjunction (i.e., the large interval), we suggest, did not overlook the fact that the interval included several smaller intervals; rather, the unpacking manipulation enhanced the salience of these intervals and, hence, their judged probability. Subadditivity, therefore, is observed even in the absence of memory limitations.

*Number of components.* The degree of subadditivity increases with the number of components in the explicit disjunction. This follows readily from support theory: Unpacking an implicit hypothesis into exclusive components increases its total judged probability, and additional unpacking of each component should further increase the total probability assigned to the initial hypothesis. Tables 3 and 4 show, as expected, that the unpacking factor generally increases with the number of components (see also Figure 1).

*Binary complementarity.* The judged probabilities of complementary pairs of hypotheses add to one. Table 5 lists studies that have tested this prediction. We considered only studies in which the hypothesis and its complement were evaluated independently, either by different subjects or by the same subjects but with a substantial number of intervening judgments. (We provide the standard deviations for the experiments that used the latter design.) Table 5 shows that such judgments generally add to one. Binary complementarity indicates that people evaluate a given hypothesis relative to its complement. Moreover, it rules out alternative interpretations of subadditivity in terms of a suggestion effect or a confirmation bias. These accounts imply a bias in favor of the focal hypothesis yielding $P(A, B) + P(B, A) > 1$, contrary to the experimental evidence. Alternatively,

Table 3

*Results of Experiments Using Qualitative Hypotheses: Average Probability Assigned to Coextensional Implicit and Explicit Disjunctions and the Unpacking Factor Measuring the Degree of Subadditivity.*

| Study and topic | $n$ | Explicit $P$ | Implicit $P$ | Unpacking factor |
|---|---|---|---|---|
| Fischhoff, Slovic, & Lichtenstein (1978) | | | | |
|   Car failure, Experiment 1 | 4 | 0.54 | .18 | 3.00 |
|   Car failure, Experiment 5 | 2 | 0.27 | .20 | 1.35 |
|   Car failure, Experiment 6 (experts) | 4 | 0.44 | .22 | 2.00 |
| Mehle, Gettys, Manning, Baca, & Fisher (1981): college majors | 6 | 0.27 | .18 | 1.50 |
| Russo & Kolzow (1992) | | | | |
|   Causes of death | 4 | 0.55 | .45 | 1.22 |
|   Car failure | 4 | 0.55 | .27 | 2.04 |
| Koehler & Tversky (1993) | | | | |
|   College majors | 4 | 1.54 | 1.00[a] | 1.54 |
|   College majors | 5 | 2.51 | 1.00[a] | 2.51 |
| Study 1: causes of death | 3 | 0.61 | .46 | 1.33 |
| | 7 | 0.70 | .37 | 1.86 |
| Study 4: crime stories | 4 | 1.71 | 1.00[a] | 1.71 |
| Study 5: college majors | 4 | 1.76 | 1.00[a] | 1.76 |

*Note.* The number of components in the explicit disjunction is denoted by $n$. Numbered studies with no citation refer to the present article.

[a] Because the components partition the space, it is assumed that a probability of 1.00 would have been assigned to the implicit disjunction.

one might be tempted to attribute the subadditivity observed in probability judgments to subjects' lack of knowledge of the additivity principle of probability theory. This explanation, however, fails to account for the observed subadditivity in frequency judgments (in which additivity is obvious) and for the finding of binary complementarity (in which additivity is consistently satisfied).

The combination of binary complementarity and subadditive elementary judgments, implied by support theory, is inconsistent with both Bayesian and revisionist models. The Bayesian model implies that the unpacking factor should equal one because the unpacked and packed hypotheses have the same extension. Shafer's theory of belief functions and other models of lower probability require an unpacking factor of less than one, because they assume that the subjective probability (or belief) of the union of disjoint events is generally greater than the sum of the probabilities of its exclusive constituents. Furthermore, the data cannot be explained by the dual of the belief function (called the plausibility function) or, more generally, by an upper probability (e.g., see Dempster, 1967) because this model requires that the sum of the assessments of complementary events exceed unity, contrary to the evidence. Indeed, if $P(A, B) + P(B, A) = 1$ (see Table 5), then both upper and lower probability reduce to the standard additive model. The experimental findings, of course, do not invalidate the use of upper and lower probability, or belief functions, as formal systems for representing uncertainty. However, the evidence reviewed in this section indicates that these models are inconsistent with the principles that govern intuitive probability judgments.

*Probability versus frequency.* Of the studies discussed earlier and listed in Tables 3 and 4, some (e.g., Fischhoff et al., 1978) used frequency judgments and others (e.g., Teigen, 1974a,

1974b) used probability judgments. The comparison of the two tasks, summarized in Table 6, confirms the predicted pattern: Subadditivity holds for both probability and frequency judgments, and the former are more subadditive than the latter.

### Scaling Support

In the formal theory developed in the preceding section, the support function is derived from probability judgments. Is it possible to reverse the process and predict probability judgments from direct assessments of evidence strength? Let $\hat{s}(A)$ be the rating of the strength of evidence for hypothesis $A$. What is the relation between such ratings and the support estimated from probability judgments? Perhaps the most natural assumption is that the two scales are monotonically related; that is, $\hat{s}(A) \geq \hat{s}(B)$ if and only if (iff) $s(A) \geq s(B)$. This assumption implies, for example, that $P(A, B) \geq \frac{1}{2}$ iff $\hat{s}(A) \geq \hat{s}(B)$, but it does not determine the functional form relating $\hat{s}$ and $s$. To further specify the relation between the scales, it may be reasonable to assume, in addition, that support ratios are also monotonically related. That is, $\hat{s}(A)/\hat{s}(B) \geq \hat{s}(C)/\hat{s}(D)$ iff $s(A)/s(B) \geq s(C)/s(D)$.

It can be shown that if the two monotonicity conditions are satisfied, and both scales are defined, say, on the unit interval, then there exists a constant $k > 0$ such that the support function derived from probability judgments and the support function assessed directly are related by a power transformation of the form $s = \hat{s}^k$. This gives rise to the *power model*

$$R(A, B) = P(A, B)/P(B, A) = [\hat{s}(A)/\hat{s}(B)]^k,$$

yielding

Table 4

*Results of Experiments Using Quantitative Hypotheses: Average Probability Assigned to Coextensional Implicit and Explicit Disjunctions and the Unpacking Factor Measuring the Degree of Subadditivity*

| Study and topic | $n$ | Explicit $P$ | Implicit $P$ | Unpacking factor |
|---|---|---|---|---|
| Teigen (1974b) Experiment 1: binomial outcomes | 2 | 0.66 | .38 | 1.73 |
| | 3 | 0.84 | .38 | 2.21 |
| | 5 | 1.62 | 1.00[a] | 1.62 |
| | 9 | 2.25 | 1.00[a] | 2.25 |
| Teigen (1974b) Experiment 2: heights of students | 2 | 0.58 | .36 | 1.61 |
| | 4 | 1.99 | .76 | 2.62 |
| | 5 | 2.31 | .75 | 3.07 |
| | 6 | 2.55 | 1.00[a] | 2.55 |
| Teigen (1974a) Experiment 2: binomial outcomes | 11 | 4.25 | 1.00[a] | 4.25 |
| Olson (1976) Experiment 1: gender distribution | 2 | 0.13 | .10 | 1.30 |
| | 3 | 0.36 | .21 | 1.71 |
| | 5 | 0.68 | .40 | 1.70 |
| | 9 | 0.97 | .38 | 2.55 |
| Peterson and Pitz (1988) Experiment 3: baseball victories | 3 | 1.58 | 1.00[a] | 1.58 |
| Tversky and Fox (1994): uncertain quantities | 2 | 0.77 | .62 | 1.27 |
| | 3 | 1.02 | .72 | 1.46 |
| | 4 | 1.21 | .79 | 1.58 |
| | 5 | 1.40 | .84 | 1.27 |
| Study 2: number of children | 10 | 1.99 | 1.00[a] | 1.99 |

*Note.* The number of components in the explicit disjunction is denoted by $n$. Numbered Study with no citation refers to the peresent article.
[a] Because the components partition the space, it is assumed that a probability of 1.00 would have been assigned to the implicit disjunction.

Table 5

*Results of Experiments Testing Binary Complementarity: Average Total Probability Assigned to Complementary Pairs of Hypotheses, Between-Subjects Standard Deviations, and the Number of Subjects in the Experiment*

| Study and topic | Mean total $P$ | $SD$ | $N$ |
|---|---|---|---|
| Wallsten, Budescu, & Zwick (1992): general knowledge | 1.02 | 0.06 | 23 |
| Tversky & Fox (1994) | | | |
| NBA playoff | 1.00 | 0.07 | 27 |
| Super Bowl | 1.02 | 0.07 | 40 |
| Dow–Jones | 1.00 | 0.10 | 40 |
| San Francisco temperature | 1.02 | 0.13 | 72 |
| Beijing temperature | 0.99 | 0.14 | 45 |
| Koehler & Tversky (1993): college majors[a] | 1.00 | | 170 |
| Study 2: number of children[a] | 0.97 | | 139 |
| Study 4: crime stories[a] | 1.03 | | 60 |
| Study 5: college majors[a] | 1.05 | | 115 |

*Note.* Numbered studies with no citation refer to the present article. NBA = National Basketball Association.
[a] A given subject evaluated either the event or its complement, but not both.

Table 6

*Results of Experiments Comparing Probability and Frequency Judgments: Unpacking Factor Computed From Mean Probability Assigned to Coextensional Explicit and Implicit Disjunctions*

| Study and topic | $n$ | Unpacking factor Probability | Frequency |
|---|---|---|---|
| Teigen (1974b) Experiment 1: binomial outcomes | 2 | 1.73 | 1.26 |
| | 5 | 2.21 | 1.09 |
| | 9 | 2.25 | 1.24 |
| Teigen (1974b) Experiment 2: heights of students | 6 | 2.55 | 1.68 |
| Koehler & Tversky (1993): college majors | 4 | 1.72 | 1.37 |
| Study 1: causes of death | 3 | 1.44 | 1.28 |
| | 7 | 2.00 | 1.84 |

*Note.* The number of components in the explicit disjunction is denoted by $n$. Numbered studies with no citation refer to the present article.

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

We next use this model to predict judged probability from independent assessments of evidence strength obtained in two studies.

## Study 3: Basketball Games

Subjects ($N$ = 88) were NBA fans who subscribe to a computer news group. We posted a questionnaire to this news group and asked readers to complete and return it by electronic mail within 1 week. In the questionnaire, subjects assessed the probability that the home team would win in each of 20 upcoming games. These 20 outcomes constituted all possible matches among five teams (Phoenix, Portland, Los Angeles Lakers, Golden State, and Sacramento) from the Pacific Division of the NBA, constructed such that, for each pair of teams, two games were evaluated (one for each possible game location). Use of this "expert" population yielded highly reliable judgments, as shown, among other things, by the fact that the median value of the correlation between an individual subject's ratings and the set of mean judgments was .93.

After making their probability judgments, subjects rated the strength of each of the five teams. The participants were instructed:

> First, choose the team you believe is the strongest of the five, 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 a 100), give that team a strength rating of 50.

We interpreted these ratings as a direct assessment of support.

Because the strength ratings did not take into account the home court effect, we collapsed the probability judgments across the two possible locations of the match. The slope of the regression line predicting $\log R(A, B)$ from $\log[\hat{s}(A)/\hat{s}(B)]$ provided an estimate of $k$ for each subject. The median estimate of $k$ was 1.8, and the mean was 2.2; the median $R^2$ for this analysis was .87. For the aggregate data, $k$ was 1.9 and the resulting $R^2$ was .97. The scatterplot in Figure 4 exhibits excellent correspondence between mean prediction based on team strength and mean judged probability. This result suggests that the power model can be used to predict judged probability from assessments of strength that make no reference to chance or uncertainty. It also reinforces the psychological interpretation of $s$ as a measure of evidence strength.

## Study 4: Crime Stories

This study was designed to investigate the relation between judged probability and assessed support in a very different context and to explore the enhancement effect, described in the next subsection. To this end, we adapted a task introduced by Teigen (1983) and Robinson and Hastie (1985) and presented subjects with two criminal cases. The first was an embezzlement at a computer-parts manufacturing company involving four suspects (a manager, a buyer, an accountant, and a seller). The second case was a murder that also involved four suspects (an activist, an artist, a scientist, and a writer). In both cases, subjects were informed that exactly one suspect was guilty. In the low-information condition, the four suspects in each case were
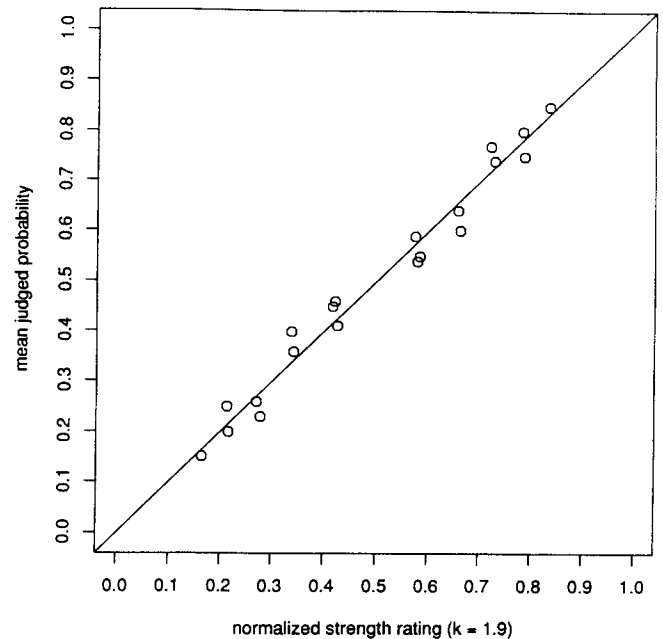


*Figure 4.* Judged probability for basketball games as a function of normalized strength ratings.

introduced with a short description of their role and possible motive. In the high-information condition, the motive of each suspect was strengthened. In a manner resembling the typical mystery novel, we constructed each case so that all the suspects seemed generally more suspicious as more evidence was revealed.

Subjects evaluated the suspects after reading the low-information material and again after reading the high-information material. Some subjects ($N$ = 60) judged the probability that a given suspect was guilty. Each of these subjects made two elementary judgments (that a particular suspect was guilty) and three binary judgments (that Suspect A rather than Suspect B was guilty) in each case. Other subjects ($N$ = 55) rated the suspiciousness of a given suspect, which we took as a direct assessment of support. These subjects rated two suspects per case by providing a number between 0 (indicating that the suspect was "not at all suspicious") and 100 (indicating that the suspect was "maximally suspicious") in proportion to the suspiciousness of the suspect.

As in the previous study, we assumed binary complementarity and estimated $k$ by a logarithmic regression of $R(A, B)$ against the suspiciousness ratio. For these data, $k$ was estimated to be .84, and $R^2$ was .65. Rated suspiciousness, therefore, provides a reasonable predictor of the judged probability of guilt. However, the relation between judged probability and assessed support was stronger in the basketball study than in the crime study. Furthermore, the estimate of $k$ was much smaller in the latter than in the former. In the basketball study, a team that was rated twice as strong as another was judged more than twice as likely to win; in the crime stories, however, a character who was twice as suspicious as another was judged less than twice as likely to be guilty. This difference may be due to the fact that the

judgments of team strength were based on more solid data than the ratings of suspiciousness.

In the preceding two studies, we asked subjects to assess the overall support for each hypothesis on the basis of all the available evidence. A different approach to the assessment of evidence was taken by Briggs and Krantz (1992; see also Krantz, Ray, & Briggs, 1990). These authors demonstrated that, under certain conditions, subjects can assess the degree to which an isolated item of evidence supports each of the hypotheses under consideration. They also proposed several rules for the combination of independent items of evidence, but they did not relate assessed support to judged probability.

## The Enhancement Effect

Recall that assessed support is noncompensatory in the sense that evidence that increases the support of one hypothesis does not necessarily decrease the support of competing hypotheses. In fact, it is possible for new evidence to increase the support of all elementary hypotheses. We have proposed that such evidence will enhance subadditivity. In this section, we describe several tests of enhancement and compare support theory with the Bayesian model and with Shafer's theory.

We start with an example discussed earlier, in which one of several suspects has committed a murder. To simplify matters, assume that there are four suspects who, in the absence of specific evidence (low information), are considered equally likely to be guilty. Suppose further evidence is then introduced (high information) that implicates each of the suspects to roughly the same degree, so that they remain equally probable. Let $L$ and $H$ denote, respectively, the evidence available under low- and high-information conditions. Let $\bar{A}$ denote the negation of $A$, that is, "Suspect $A$ is not guilty." According to the Bayesian model, then, $P(A, B|H) = P(A, B|L) = \frac{1}{2}$, $P(A, \bar{A}|H) = P(A, \bar{A}|L) = \frac{1}{4}$, and so forth.

In contrast, Shafer's (1976) belief-function approach requires that the probabilities assigned to each of the suspects add to less than one and suggests that the total will be higher in the presence of direct evidence (i.e., in the high-information condition) than in its absence. As a consequence, $\frac{1}{2} \geq P(A, B|H) \geq P(A, B|L)$, $\frac{1}{4} \geq P(A, \bar{A}|H) \geq P(A, \bar{A}|L)$, and so forth. In other words, both the binary and the elementary judgments are expected to increase as more evidence is encountered. In the limit, when no belief is held in reserve, the binary judgments approach one half and the elementary judgments approach one fourth.

The enhancement assumption yields a different pattern, namely $P(A, B|H) = P(A, B|L) = \frac{1}{2}$, $P(A, \bar{A}|H) \geq P(A, \bar{A}|L) \geq \frac{1}{4}$, and so forth. As in the Bayesian model, the binary judgments are one half; in contrast to that model, however, the elementary judgments are expected to exceed one fourth and to be greater under high- than under low-information conditions. Although both support theory and the belief-function approach yield greater elementary judgments under high- than under low-information conditions, support theory predicts that they will exceed one fourth in both conditions, whereas Shafer's theory requires that these probabilities be less than or equal to one fourth.

The assumption of equally probable suspects is not essential for the analysis. Suppose that initially the suspects are not equally probable, but the new evidence does not change the binary probabilities. Here, too, the Bayesian model requires additive judgments that do not differ between low- and high-information conditions; the belief-function approach requires superadditive judgments that become less superadditive as more information is encountered; and the enhancement assumption predicts subadditive judgments that become more subadditive with the addition of (compatible) evidence.

## Evaluating Suspects

With these predictions in mind, we turn to the crime stories of Study 4. Table 7 displays the mean suspiciousness ratings and elementary probability judgments of each suspect in the two cases under low- and high-information conditions. The table shows that, in all cases, the sums of both probability judgments and suspiciousness ratings exceed one. Evidently, subadditivity holds not only in probability judgment but also in ratings of evidence strength or degree of belief (e.g., that a given subject is guilty). Further examination of the suspiciousness ratings shows that all but one of the suspects increased in suspiciousness as more information was provided. In accord with our prediction, the judged probability of each of these suspects also increased with the added information, indicating enhanced subadditivity (see Equation 10). The one exception was the artist in the murder case, who was given an alibi in the high-information condition and, as one would expect, subsequently decreased both in suspiciousness and in probability. Overall, both the suspiciousness ratings and the probability judgments were significantly greater under high- than under low-information conditions ($p < .001$ for both cases by $t$ test).

From a normative standpoint, the support (i.e., suspiciousness) of all the suspects could increase with new information, but an increase in the probability of one suspect should be compensated for by a decrease in the probability of the others. The observation that new evidence can increase the judged probability of all suspects was made earlier by Robinson and Hastie (1985; Van Wallendael & Hastie, 1990). Their method differed from ours in that each subject assessed the probability of all suspects, but this method too produced substantial subadditivity, with a typical unpacking factor of about two. These authors rejected the Bayesian model as a descriptive account and proposed Shafer's theory as one viable alternative. As was noted earlier, however, the observed subadditivity is inconsistent with Shafer's theory, as well as the Bayesian model, but it is consistent with the present account.

In the crime stories, the added evidence was generally compatible with all of the hypotheses under consideration. Peterson and Pitz (1988, Experiment 3), however, observed a similar effect with mixed evidence, which favored some hypotheses but not others. Their subjects were asked to assess the probability that the number of games won by a baseball team in a season fell in a given interval on the basis of one, two, or three cues (team batting average, earned run average, and total home runs during that season). Unbeknownst to subjects, they were asked, over a large number of problems, to assign probabilities to all three components in a partition (e.g., less than 80 wins, between 80 and 88 wins, and more than 88 wins). As the number of cues increased, subjects assigned a greater probability, on average, to all three intervals in the partition, thus exhibiting enhanced

Table 7

*Mean Suspiciousness Rating and Judged Probability of Each Suspect Under Low- and High-Information Conditions*

| Case and suspect | Suspiciousness | | Probability | |
| --- | --- | --- | --- | --- |
| | Low information | High information | Low information | High information |
| Case 1: embezzlement | | | | |
| Accountant | 41 | 53 | 40 | 45 |
| Buyer | 50 | 58 | 42 | 48 |
| Manager | 47 | 51 | 48 | 59 |
| Seller | 32 | 48 | 37 | 42 |
| Total | 170 | 210 | 167 | 194 |
| Case 2: murder | | | | |
| Activist | 32 | 57 | 39 | 57 |
| Artist | 27 | 23 | 37 | 30 |
| Scientist | 24 | 43 | 34 | 40 |
| Writer | 38 | 60 | 33 | 54 |
| Total | 122 | 184 | 143 | 181 |

subadditivity. The unpacking factors for these data were 1.26, 1.61, and 1.86 for one, two, and three cues, respectively. These results attest to the robustness of the enhancement effect, which is observed even when the added evidence favors some, but not all, of the hypotheses under study.

### Study 5: College Majors

In this study, we tested enhancement by replacing evidence rather than by adding evidence as in the previous study. Following Mehle, Gettys, Manning, Baca, and Fisher (1981), we asked subjects ($N = 115$) to assess the probability that a social science student at an unspecified midwestern university majored in a given field. Subjects were told that, in this university, each social science student has one and only one of the following four majors: economics, political science, psychology, and sociology.

Subjects estimated the probability that a given student had a specified major on the basis of one of four courses the student was said to have taken in his or her 2nd year. Two of the courses (statistics and Western civilization) were courses typically taken by social science majors; the other two (French literature and physics) were courses not typically taken by social science majors. This was confirmed by an independent group of subjects ($N = 36$) who evaluated the probability that a social science major would take each one of the four courses. Enhancement suggests that the typical courses will yield more subadditivity than the less typical courses because they give greater support to each of the four majors.

Each subject made both elementary and binary judgments. As in all previous studies, the elementary judgments exhibited substantial subadditivity (mean unpacking factor = 1.76), whereas the binary judgments were essentially additive (mean unpacking factor = 1.05). In the preceding analyses, we have used the unpacking factor as an overall measure of subadditivity associated with a set of mutually exclusive hypotheses. The present experiment also allowed us to estimate $w$ (see Equation 8), which provides a more refined measure of subadditivity because it is estimated separately for each of the implicit hypotheses under study. For each course, we first estimated the support of each major from the binary judgments and then estimated $w$

for each major from the elementary judgments using the equation

$$P(A, \bar{A}) = \frac{s(A)}{s(A) + w_{\bar{A}}[s(B) + s(C) = s(D)]},$$

where $A$, $B$, $C$, and $D$ denote the four majors.

This analysis was conducted separately for each subject. The average value of $w$ across courses and majors was .46, indicating that a major received less than half of its explicit support when it was included implicitly in the residual. Figure 5 shows the
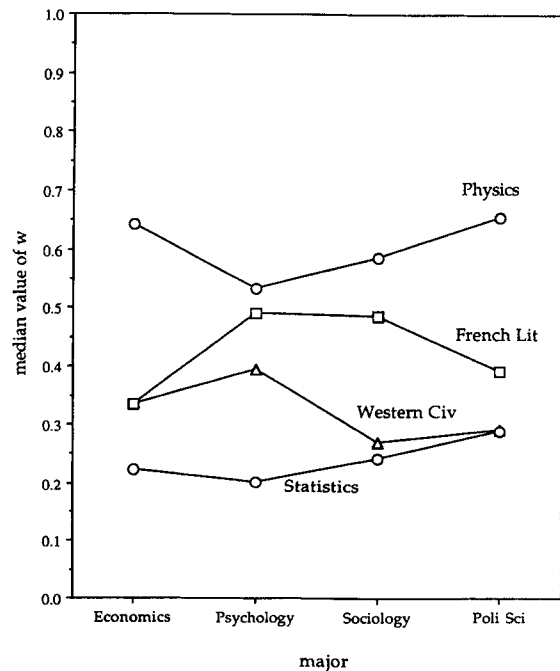


*Figure 5.* Median value of $w$ for predictions of college majors, plotted separately for each course. Lit = literature; Civ = civilization; Poli Sci = political science.

median value of $w$ (over subjects) for each major, plotted separately for each of the four courses. In accord with enhancement, the figure shows that the typical courses, statistics and Western civilization, induced more subadditivity (i.e., lower $w$) than the less typical courses, physics and French literature. However, for any given course, $w$ was roughly constant across majors. Indeed, a two-way analysis of variance yielded a highly significant effect of course, $F(3, 112) = 31.4, p < .001$, but no significant effect of major, $F(3, 112) < 1$.

## Implications

To this point, we have focused on the direct consequences of support theory. We conclude this section by discussing the conjunction effect, hypothesis generation, and decision under uncertainty from the perspective of support theory.

### The Conjunction Effect

Considerable research has documented the conjunction effect, in which a conjunction $AB$ is judged more probable than one of its constituents $A$. The effect is strongest when an event that initially seems unlikely (e.g., a massive flood in North America in which more than 1,000 people drown) is supplemented by a plausible cause or qualification (e.g., an earthquake in California causing a flood in which more than 1,000 people drown), yielding a conjunction that is perceived as more probable than the initially implausible event of which it is a proper subset (Tversky & Kahneman, 1983). Support theory suggests that the implicit hypothesis $A$ is not unpacked into the coextensional disjunction $AB \lor A\bar{B}$ of which the conjunction is one component. As a result, evidence supporting $AB$ is not taken to support $A$. In the flood problem, for instance, the possibility of a flood caused by an earthquake may not come readily to mind; thus, unless it is mentioned explicitly, it does not contribute any support to the (implicit) flood hypothesis. Support theory implies that the conjunction effect would be eliminated in these problems if the implicit disjunction were unpacked before its evaluation (e.g., if subjects were reminded that a flood might be caused by excessive rainfall or by structural damage to a reservoir caused by an earthquake, an engineering error, sabotage, etc.).

The greater tendency to unpack either the focal or the residual hypothesis in a frequentistic formulation may help explain the finding that conjunction effects are attenuated, though not eliminated, when subjects estimate frequency rather than probability. For example, the proportion of subjects who judged the conjunction "$X$ is over 55 years old and has had at least one heart attack" as more probable than the constituent event "$X$ has had at least one heart attack" was significantly greater in a probabilistic formulation than in a frequentistic formulation (Tversky & Kahneman, 1983).

It might be instructive to distinguish two different unpacking operations. In *conjunctive unpacking*, an (implicit) hypothesis (e.g., nurse) is broken down into exclusive conjunctions (e.g., male nurse and female nurse). Most, but not all, initial demonstrations of the conjunction effect were based on conjunctive unpacking. In *categorical unpacking*, a superordinate category (e.g., unnatural death) is broken down into its "natural" components (e.g., car accident, drowning, and homicide). Most of the demonstrations reported in this article are based on categorical unpacking. A conjunction effect using categorical unpacking has been described by Bar-Hillel and Neter (1993), who found numerous cases in which a statement (e.g., "Daniela's major is literature") was ranked as more probable than a more inclusive implicit disjunction (e.g., "Daniela's major is in humanities"). These results held both for subjects' direct estimates of probabilities and for their willingness to bet on the relevant events.

### Hypothesis Generation

All of the studies reviewed thus far asked subjects to assess the probability of hypotheses presented to them for judgment. There are many situations, however, in which a judge must generate hypotheses as well as assess their likelihood. In the current treatment, the generation of alternative hypotheses entails some unpacking of the residual hypothesis and, thus, is expected to increase its support relative to the focal hypothesis. In the absence of explicit instructions to generate alternative hypotheses, people are less likely to unpack the residual hypothesis and thus will tend to overestimate specified hypotheses relative to those left unspecified.

This implication has been confirmed by Gettys and his colleagues (Gettys, Mehle, & Fisher, 1986; Mehle et al., 1981), who have found that, in comparison with veridical values, people generally tend to overestimate the probability of specified hypotheses presented to them for evaluation. Indeed, overconfidence that one's judgment is correct (e.g., Lichtenstein, Fischhoff, & Phillips, 1982) may sometimes arise because the focal hypothesis is specified, whereas its alternatives often are not. Mehle et al. (1981) used two manipulations to encourage unpacking of the residual hypothesis: One group of subjects was provided with exemplar members of the residual, and another was asked to generate its own examples. Both manipulations improved performance by decreasing the probability assigned to specified alternatives and increasing that assigned to the residual. These results suggest that the effects of hypothesis generation are due to the additional hypotheses it brings to mind, because simply providing hypotheses to the subject has the same effect. Using a similar manipulation, Dube-Rioux and Russo (1988) found that generation of alternative hypotheses increased the judged probability of the residual relative to that of specified categories and attenuated the effect of omitting a category. Examination of the number of instances generated by the subjects showed that, when enough instances were produced, the effect of category omission was eliminated altogether.

Now consider a task in which subjects are asked to generate a hypothesis (e.g., to guess which film will win the best picture Oscar at the next Academy Awards ceremony) before assessing its probability. Asking subjects to generate the most likely hypothesis might actually lead them to consider several candidates in the process of settling on the one they prefer. This process amounts to a partial unpacking of the residual hypothesis, which should decrease the judged probability of the focal hypothesis. Consistent with this prediction, a recent study (Koehler, 1994) found that subjects asked to generate their own hypotheses assigned them a lower probability of being true than did other subjects presented with the same hypotheses for eval-

uation. The interpretation of these results—that hypothesis generation makes alternative hypotheses more salient—was tested by two further manipulations. First, providing a closed set of specified alternatives eliminated the difference between the generation and evaluation conditions. In these circumstances, the residual should be represented in the same way in both conditions. Second, inserting a distracter task between hypothesis generation and probability assessment was sufficient to reduce the salience of alternatives brought to mind by the generation task, increasing the judged probability of the focal hypothesis.

## Decision Under Uncertainty

This article has focused primarily on numerical judgments of probability. In decision theory, however, subjective probabilities are generally inferred from preferences between uncertain prospects rather than assessed directly. It is natural to inquire, then, whether unpacking affects people's decisions as well as their numerical judgments. There is considerable evidence that it does. For example, Johnson et al. (1993) observed that subjects were willing to pay more for flight insurance that explicitly listed certain events covered by the policy (e.g., death resulting from an act of terrorism or mechanical failure) than for a more inclusive policy that did not list specific events (e.g., death from any cause).

Unpacking can affect decisions in two ways. First, as has been shown, unpacking tends to increase the judged probability of an uncertain event. Second, unpacking can increase an event's impact on the decision, even when its probability is known. For example, Tversky and Kahneman (1986) asked subjects to choose between two lotteries that paid different amounts depending on the color of a marble drawn from a box. (As an inducement to consider the options with care, subjects were informed that one tenth of the participants, selected at random, would actually play the gambles they chose.) Two different versions of the problem were used, which differed only in the description of the outcomes. The fully unpacked Version 1 was as follows:

| Box A: | 90% white $0 | 6% red win $45 | 1% green win $30 | 1% blue lose $15 | 2% yellow lose $15 |
| Box B: | 90% white $0 | 6% red win $45 | 1% green win $45 | 1% blue lose $10 | 2% yellow lose $15 |

It is not difficult to see that Box B dominates Box A; indeed, all subjects chose Box B in this version. Version 2 combined the two outcomes resulting in a loss of $15 in Box A (i.e., blue and yellow) and the two outcomes resulting in a gain of $45 in Box B (i.e., red and green):

| Box A: | 90% white $0 | 6% red win $45 | 1% green win $30 | 3% yellow/blue lose $15 |
| Box B: | 90% white $0 | 7% red/green win $45 | 1% blue lose $10 | 2% yellow lose $15 |

In accord with subadditivity, the combination of events yielding the same outcome makes Box A more attractive because it packs two losses into one and makes Box B less attractive because it packs two gains into one. Indeed, 58% of subjects chose

Box A in Version 2, even though it was dominated by Box B. Starmer and Sugden (1993) further investigated the effect of unpacking events with known probabilities (which they called an event-splitting effect) and found that a prospect generally becomes more attractive when an event that yields a positive outcome is unpacked into two components. Such results demonstrate that unpacking affects decisions even when the probabilities are explicitly stated.

The role of unpacking in choice was further illustrated by Redelmeier et al. (in press). Graduating medical students at the University of Toronto ($N$ = 149) were presented with a medical scenario concerning a middle-aged man suffering acute shortness of breath. Half of the respondents were given a packed description that noted that "obviously, many diagnoses are possible . . . including pneumonia." The other half were given an unpacked description that mentioned other potential diagnoses (pulmonary embolus, heart failure, asthma, and lung cancer) in addition to pneumonia. The respondents were asked whether or not they would prescribe antibiotics in such a case, a treatment that is effective against pneumonia but not against the other potential diagnoses mentioned in the unpacked version. The unpacking manipulation was expected to reduce the perceived probability of pneumonia and, hence, the respondents' inclination to prescribe antibiotics. Indeed, a significant majority (64%) of respondents given the unpacked description chose not to prescribe antibiotics, whereas respondents given the packed description were almost evenly divided between prescribing (47%) and not prescribing them. Singling out pneumonia increased the tendency to select a treatment that is effective for pneumonia, even though the presenting symptoms were clearly consistent with a number of well-known alternative diagnoses. Evidently, unpacking can affect decisions, not only probability assessments.

Although unpacking plays an important role in probability judgment, the cognitive mechanism underlying this effect is considerably more general. Thus, one would expect unpacking effects even in tasks that do not involve uncertain events. For example, van der Pligt, Eiser, and Spears (1987, Experiment 1) asked subjects to assess the current and ideal distribution of five power sources (nuclear, coal, oil, hydro, solar/wind/wave) and found that a given power source was assigned a higher estimate when it was evaluated on its own than when its four alternatives were unpacked (see also Fiedler & Armbruster, 1994; Pelham, Sumarta, & Myaskovsky, 1994). Such results indicate that the effects of unpacking reflect a general characteristic of human judgment.

## Extensions

We have presented a nonextensional theory of belief in which judged probability is given by the relative support, or strength of evidence, of the respective focal and alternative hypotheses. In this theory, support is additive for explicit disjunctions of exclusive hypotheses and subadditive for implicit disjunctions. The empirical evidence confirms the major predictions of support theory: (a) Probability judgments increase by unpacking the focal hypothesis and decrease by unpacking the alternative hypothesis; (b) subjective probabilities are complementary in the binary case and subadditive in the general case; and (c) subadditivity is more pronounced for probability than for fre-

quency judgments, and it is enhanced by compatible evidence. Support theory also provides a method for predicting judged probability from independent assessments of evidence strength. Thus, it accounts for a wide range of empirical findings in terms of a single explanatory construct.

In this section, we explore some extensions and implications of support theory. First, we consider an ordinal version of the theory and introduce a simple parametric representation. Second, we address the problem of vagueness, or imprecision, by characterizing upper and lower probability judgments in terms of upper and lower support. Finally, we discuss the implications of the present findings for the design of elicitation procedures for decision analysis and knowledge engineering.

## Ordinal Analysis

Throughout the article, we have treated probability judgment as a quantitative measure of degree of belief. This measure is commonly interpreted in terms of a reference chance process. For example, assigning a probability of two thirds to the hypothesis that a candidate will be elected to office is taken to mean that the judge considers this hypothesis as likely as drawing a red ball from an urn in which two thirds of the balls are red. Probability judgment, therefore, can be viewed as an outcome of a thought experiment in which the judge matches degree of belief to a standard chance process (see Shafer & Tversky, 1985). This interpretation, of course, does not ensure either coherence or calibration.

Although probability judgments appear to convey quantitative information, it might be instructive to analyze these judgments as an ordinal rather than a cardinal scale. This interpretation gives rise to an ordinal generalization of support theory. Suppose there is a nonnegative scale $s$ defined on H and a strictly increasing function $F$ such that, for all $A$, $B$ in H,

$$P(A, B) = F\left(\frac{s(A)}{s(A) + s(B)}\right),$$ (11)

where $s(C) \leq s(A \lor B) = s(A) + s(B)$ whenever $A$ and $B$ are exclusive, $C$ is implicit, and $C' = (A \lor B)'$.

An axiomatization of the ordinal model lies beyond the scope of the present article. It is noteworthy, however, that to obtain an essentially unique support function in this case, we have to make additional assumptions, such as the following *solvability condition* (Debreu, 1958): If $P(A, B) \geq z \geq P(A, D)$, then there exists $C \in H$ such that $P(A, C) = z$. This idealization may be acceptable in the presence of a random device, such as a chance wheel with sectors that can be adjusted continuously. The following theorem shows that, assuming the ordinal model and the solvability condition, binary complementarity and the product rule yield a particularly simple parametric form that coincides with the model used in the preceding section to relate assessed and derived support. The proof is given in the Appendix.

*Theorem 2:* Assume the ordinal model (Equation 11) and the solvability condition. Binary complementarity (Equation 3) and the product rule (Equation 5) hold if and only if there exists a constant $k \geq 0$ such that

$$P(A, B) = \frac{s(A)^k}{s(A)^k + s(B)^k}.$$ (12)

This representation, called the power model, reduces to the basic model if $k = 1$. In this model, judged probability may be more or less extreme than the respective relative support depending on whether $k$ is greater or less than one. Recall that the experimental data, reviewed in the preceding section, provide strong evidence for the inequality $\alpha < \delta$. That is, $P(A, B) \leq P(A_1, B) + P(A_2, B)$ whenever $A_1$, $A_2$, and $B$ are mutually exclusive; $A$ is implicit; and $A' = (A_1 \lor A_2)'$. We also found evidence (see Table 2) for the equality $\beta = \gamma$, that is, $P(A_1 \lor A_2, B) = P(A_1, A_2 \lor B) + P(A_2, A_1 \lor B)$, but this property has not been extensively tested. Departures from additivity induced, for example, by regression toward .5 could be represented by a power model with $k < 1$, which implies $\alpha < \beta < \gamma < \delta$. Note that, for explicit disjunctions of exclusive hypotheses, the basic model (Equations 1 and 2), the ordinal model (Equation 11), and the power model (Equation 12) all assume additive support, but only the basic model entails additive probability.

## Upper and Lower Indicators

Probability judgments are often vague and imprecise. To interpret and make proper use of such judgments, therefore, one needs to know something about their range of uncertainty. Indeed, much of the work on nonstandard probability has been concerned with formal models that provide upper and lower indicators of degree of belief. The elicitation and interpretation of such indicators, however, present both theoretical and practical problems. If people have a hard time assessing a single definite value for the probability of an event, they are likely to have an even harder time assessing two definite values for its upper and lower probabilities or generating a second-order probability distribution. Judges may be able to provide some indication regarding the vagueness of their assessments, but such judgments, we suggest, are better interpreted in qualitative, not quantitative, terms.

To this end, we have devised an elicitation procedure in which upper and lower probability judgments are defined verbally rather than numerically. This procedure, called the *staircase method*, is illustrated in Figure 6. The judge is presented with an uncertain event (e.g., an eastern team rather than a western team will win the next NBA title) and is asked to check one of the five categories for each probability value. The lowest value that is not "clearly too low" (.45) and the highest value that is not "clearly too high" (.80), denoted $P_*$ and $P^*$, respectively, may be taken as the lower and upper indicators. Naturally, alternative procedures involving a different number of categories, different wording, and different ranges could yield different indicators. (We assume that the labeling of the categories is symmetric around the middle category.) The staircase method can be viewed as a qualitative analog of a second-order probability distribution or of a fuzzy membership function.

We model $P_*$ and $P^*$ in terms of lower and upper support functions, denoted $s_*$ and $s^*$, respectively. We interpret these scales as low and high estimates of $s$ and assume that, for any $A$, $s_*(A) \leq s(A) \leq s^*(A)$. Furthermore, we assume that $P_*$ and $P^*$ can be expressed as follows:

$$P_*(A, B) = \frac{s_*(A)}{s_*(A) + s^*(B)}$$

and

| Probability (%): | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clearly too high | | | | | | | | | | | | | | | | | | x | x | x | x |
| Slightly too high | | | | | | | | | | | | | | | | x | x | | | | |
| ABOUT RIGHT | | | | | | | | | | | | | | | x | | | | | | |
| Slightly too low | | | | | | | | | | x | x | x | x | x | | | | | | | |
| Clearly too low | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | |

*Figure 6.* Example of the staircase method used to elicit upper and lower probabilities.

$$P^*(A, B) = \frac{s^*(A)}{s^*(A) + s_*(B)}.$$

According to this model, the upper and lower indicators are generated by a slanted reading of the evidence; $P^*(A, B)$ can be interpreted as a probability judgment that is biased in favor of $A$ and against $B$, whereas $P_*(A, B)$ is biased against $A$ and in favor of $B$. The magnitude of the bias reflects the vagueness associated with the basic judgment, as well as the characteristics of the elicitation procedure. Within a given procedure, however, we can interpret the interval $(P_*, P^*)$ as a comparative index of imprecision. Thus, we may conclude that one judgment is less vague than another if the interval associated with the first assessment is included in the interval associated with the second assessment. Because the high and low estimates are unlikely to be more precise or more reliable than the judge's best estimate, we regard $P_*$ and $P^*$ as supplements, not substitutes, for $P$.

To test the proposed representation against the standard theory of upper and lower probability (e.g., see Dempster, 1967; Good, 1962); we investigated people's predictions of the outcomes of the NFL playoffs for 1992–1993. The study was run the week before the two championship games in which Buffalo was to play Miami for the title of the American Football Conference (AFC), and Dallas was to play San Francisco for the title of the National Football Conference (NFC). The winners of these games would play each other two weeks later in the Super Bowl. The subjects were 135 Stanford students who volunteered to participate in a study of football prediction in exchange for a single California Lottery ticket. Half of the subjects assessed the probabilities that the winner of the Super Bowl would be Buffalo, Miami, an NFC team. The other half of the subjects assessed the probabilities that the winner of the Super Bowl would be Dallas, San Francisco, an AFC team. All subjects assessed probabilities for the two championship games. The focal and the alternative hypotheses for these games were counterbalanced. Thus, each subject made five probability assessments using the staircase method illustrated in Figure 6.

Subjects' best estimates exhibited the pattern of subadditivity and binary complementarity observed in previous studies. The average probabilities of each of the four teams winning the Super Bowl added to 1.71; the unpacking factor was 1.92 for the AFC teams and 1.48 for the NFC teams. In contrast, the sum of the average probability of an event and its complement was 1.03. Turning to the analysis of the upper and the lower assessments, note that the present model implies $P_*(A, B) + P^*(B, A) = 1$, in accord with the standard theory of upper and lower probability. The data show that this condition holds to a very close degree of approximation, with an average sum of 1.02.

The present model, however, does not generally agree with the standard theory of upper and lower probability. To illustrate the discrepancy, suppose $A$ and $B$ are mutually exclusive and $C' = (A \lor B)'$. The standard theory requires that $P_*(A, \bar{A}) + P_*(B, \bar{B}) \leq P_*(C, \bar{C})$, whereas the present account suggests the opposite inequality when $C$ is implicit. The data clearly violate the standard theory: The average lower probabilities of winning the Super Bowl were .21 for Miami and .21 for Buffalo but only .24 for their implicit disjunction (i.e., an AFC team). Similarly, the average lower probabilities of winning the Super Bowl were .25 for Dallas and .41 for San Francisco but only .45 for an NFC team. These data are consistent with the present model, assuming the subadditivity of $s_*$, but not with the standard theory of lower probability.

## Prescriptive Implications

Models of subjective probability or degree of belief serve two functions: descriptive and prescriptive. The literature on nonstandard probability models is primarily prescriptive. These models are offered as formal languages for the evaluation of evidence and the representation of belief. In contrast, support theory attempts to describe the manner in which people make probability judgments, not to prescribe how people should make these judgments. For example, the proposition that judged probability increases by unpacking the focal hypothesis and decreases by unpacking the alternative hypothesis represents a general descriptive principle that is not endorsed by normative theories, additive or nonadditive.

Despite its descriptive nature, support theory has prescriptive implications. It could aid the design of elicitation procedures and the reconciliation of inconsistent assessments (Lindley, Tversky, & Brown, 1979). This role may be illuminated by a perceptual analogy. Suppose a surveyor has to construct a map of a park on the basis of judgments of distance between landmarks made by a fallible observer. A knowledge of the likely biases of the observer could help the surveyor construct a better map. Because observers generally underestimate distances involving hidden areas, for example, the surveyor may discard these assessments and compute the respective distances from other assessments using the laws of plane geometry. Alternatively, the surveyor may wish to reduce the bias by applying a suitable correction factor to the estimates involving hidden areas. The same logic applies to the elicitation of probability. The evidence shows that people tend to underestimate the probability of an implicit disjunction, especially the negation of an elementary hypothesis. This bias may be reduced by asking the judge to contrast hypotheses of comparable level of specificity

instead of assessing the probability of a specific hypothesis against its complement.

The major conclusion of the present research is that subjective probability, or degree of belief, is nonextensional and hence nonmeasurable in the sense that alternative partitions of the space can yield different judgments. Like the measured length of a coastline, which increases as a map becomes more detailed, the perceived likelihood of an event increases as its description becomes more specific. This does not imply that judged probability is of no value, but it indicates that this concept is more fragile than suggested by existing formal theories. The failures of extensionality demonstrated in this article highlight what is perhaps the fundamental problem of probability assessment, namely the need to consider unavailable possibilities. The problem is especially severe in tasks that require the generation of new hypotheses or the construction of novel scenarios. The extensionality principle, we argue, is normatively unassailable but practically unachievable because the judge cannot be expected to fully unpack any implicit disjunction. People can be encouraged to unpack a category into its components, but they cannot be expected to think of all relevant conjunctive unpackings or to generate all relevant future scenarios. In this respect, the assessment of an additive probability distribution may be an impossible task. The judge could, of course, ensure the additivity of any given set of judgments, but this does not ensure that additivity will be preserved by further refinement.

The evidence reported here and elsewhere indicates that both qualitative and quantitative assessments of uncertainty are not carried out in a logically coherent fashion, and one might be tempted to conclude that they should not be carried out at all. However, this is not a viable option because, in general, there are no alternative procedures for assessing uncertainty. Unlike the measurement of distance, in which fallible human judgment can be replaced by proper physical measurement, there are no objective procedures for assessing the probability of events such as the guilt of a defendant, the success of a business venture, or the outbreak of war. Intuitive judgments of uncertainty, therefore, are bound to play an essential role in people's deliberations and decisions. The question of how to improve their quality through the design of effective elicitation methods and corrective procedures poses a major challenge to theorists and practitioners alike.

## References

Aczel, J. (1966). *Lectures on functional equations and their applications.* San Diego, CA: Academic Press.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in stereotype judgments. *Journal of Personality and Social Psychology, 65,* 1119–1131.

Briggs, L. K., & Krantz, D. H. (1992). Judging the strength of designated evidence. *Journal of Behavioral Decision Making, 5,* 77–106.

Debreu, G. (1958). Stochastic choice and cardinal utility. *Econometrica, 26,* 440–444.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics, 38,* 325–339.

Dube-Rioux, L., & Russo, J. E. (1988). An availability bias in professional judgment. *Journal of Behavioral Decision Making, 1,* 223–237.

Dubois, D., & Prade, H. (1988). Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems. *Acta Psychologica, 68,* 53–78.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519–527.

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole. *Journal of Personality and Social Psychology, 66,* 633–645.

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., Rogers, B., & Tversky, A. (1994). *Decision weights for options traders.* Unpublished manuscript, Stanford University, Stanford, CA.

Gettys, C. F., Mehle, T., & Fisher, S. (1986). Plausibility assessments in hypothesis generation. *Organizational Behavior and Human Decision Processes, 37,* 14–33.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroche & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 83–115). New York: Wiley.

Gilboa, I., & Schmeidler, D. (in press). Additive representations of non additive measures and the Choquet integral. *Annals of Operation Research.*

Good, I. J. (1962). Subjective probability as the measure of a nonmeasurable set. In E. Nagel, P. Suppes, & A. Tarski (Eds.), *Logic, methodology, and philosophy of sciences* (pp. 319–329). Stanford, CA: Stanford University Press.

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 University Press.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science, 12,* 313–327.

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

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 461–469.

Koehler, D. J., & Tversky, A. (1993). *The enhancement effect in probability judgment.* Unpublished manuscript, Stanford University, Stanford, CA.

Krantz, D. H., Ray, B., & Briggs, L. K. (1990). *Foundations of the theory of evidence: The role of schemata.* Unpublished manuscript, Columbia University, New York.

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

Lindley, D. V., Tversky, A., & Brown, R. V. (1979). On the reconciliation of probability assessments. *Journal of the Royal Statistical Society, 142,* 146–180.

Mehle, T., Gettys, C. F., Manning, C., Baca, S., & Fisher, S. (1981). The availability explanation of excessive plausibility assessment. *Acta Psychologica, 49,* 127–140.

Mongin, P. (in press). Some connections between epistemic logic and the theory of nonadditive probability. In P. W. Humphreys (Ed.), *Patrick Suppes: Scientific philosopher.* Dordrecht, Netherlands: Kluwer.

Murphy, A. H. (1985). Probabilistic weather forecasting. In A. H. Murphy & R. W. Katz (Eds.), *Probability, statistics, and decision making in the atmospheric sciences* (pp. 337–377). Boulder, CO: Westview Press.

Olson, C. L. (1976). Some apparent violations of the representativeness

heuristic in human judgment. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 599–608.

Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The easy path from many to much: The numerosity heuristic. *Cognitive Psychology, 26,* 103–133.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 85–92.

Redelmeier, D., Koehler, D. J., Liberman, V., & Tversky, A. (in press). Probability judgment in medicine: Discounting unspecified alternatives. *Medical Decision Making.*

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

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 443–456.

Russo, J. E., & Kolzow, K. J. (1992). *Where is the fault in fault trees?* Unpublished manuscript, Cornell University, Ithaca, NY.

Shafer, G. (1976). *A mathematical theory of evidence.* Princeton, NJ: Princeton University Press.

Shafer, G., & Tversky, A. (1985). Languages and designs for probability judgment. *Cognitive Science, 9,* 309–339.

Starmer, C., & Sugden, R. (1993). Testing for juxtaposition and event-splitting effects. *Journal of Risk and Uncertainty, 6,* 235–254.

Statistical abstract of the United States. (1990). Washington, DC: U.S. Dept. of Commerce, Bureau of the Census.

Suppes, P. (1974). The measurement of belief. *Journal of the Royal Statistical Society, B, 36,* 160–191.

Teigen, K. H. (1974a). Overestimation of subjective probabilities. *Scandinavian Journal of Psychology, 15,* 56–62.

Teigen, K. H. (1974b). 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., & Fox, C. (1994). *Weighing risk and uncertainty.* Unpublished manuscript, Stanford University, Stanford, CA.

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

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. Part 2. *Journal of Business, 59,* 251–278.

Tversky, A., & Sattath, S. (1979). Preference trees. *Psychological Review, 86,* 542–573.

van der Pligt, J., Eiser, J. R., & Spears, R. (1987). Comparative judgments and preferences: The influence of the number of response alternatives. *British Journal of Social Psychology, 26,* 269–280.

Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition, 18,* 240–250.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities.* London: Chapman & Hall.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1992). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science, 39,* 176–190.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets & Systems, 1,* 3–28.

Zarnowitz, V. (1985). Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics, 3,* 293–311.

# Appendix

*Theorem 1:* Suppose $P(A, B)$ is defined for all disjoint $A, B \in H$, and it vanishes if and only if (iff) $A' = \varnothing$. Equations 3–6 (see text) hold iff there exists a nonnegative ratio scale $s$ on H that satisfies Equations 1 and 2.

*Proof:* Necessity is immediate. To establish sufficiency, we define $s$ as follows. Let $E = \{A \in H: A' \in T\}$ be the set of elementary hypotheses. Select some $D \in E$ and set $s(D) = 1$. For any other elementary hypothesis $C \in E$, such that $C' \neq D'$, define $s(C) = P(C, D)/P(D, C)$. Given any hypothesis $A \in H$ such that $A' \neq T$, $\varnothing$, select some $C \in E$ such that $A' \cap C' = \varnothing$ and define $s(A)$ through

$$\frac{s(A)}{s(C)} = \frac{P(A, C)}{P(C, A)};$$

that is,

$$s(A) = \frac{P(A, C)P(C, D)}{P(C, A)P(D, C)}.$$

To demonstrate that $s(A)$ is uniquely defined, suppose $B \in E$ and $A' \cap B' = \varnothing$. We want to show that

$$\frac{P(A, C)P(C, D)}{P(C, A)P(D, C)} = \frac{P(A, B)P(B, D)}{P(B, A)P(D, B)}.$$

By proportionality (Equation 4), the left-hand ratio equals

$$\frac{P(A, C \vee B)P(C, D \vee B)}{P(C, A \vee B)P(D, C \vee B)}$$

and the right-hand ratio equals

$$\frac{P(A, B \vee C)P(B, D \vee C)}{P(B, A \vee C)P(D, B \vee C)}.$$

Canceling common terms, it is easy to see that the two ratios are equal iff

$$\frac{P(C, D \vee B)}{P(B, D \vee C)} = \frac{P(C, A \vee B)}{P(B, A \vee C)},$$

which holds because both ratios equal $P(C, B)/P(B, C)$, again by proportionality.

To complete the definition of $s$, let $s(A) = 0$ whenever $A' = \varnothing$. For $A' = T$, we distinguish two cases. If $A$ is explicit, that is, $A = B \vee C$ for some exclusive $B, C \in H$, set $s(A) = s(B) + s(C)$. If $A$ is implicit, let $s(A)$ be the minimum value of $s$ over all explicit descriptions of T.

To establish the desired representation, we first show that for any exclusive $A, B \in H$, such that $A', B' \neq T$, $\varnothing$, $s(A)/s(B) = P(A, B)/P(B, A)$. Recall that T includes at least two elements. Two cases must be considered.

First, suppose $A' \cup B' \neq T$; hence, there exists an elementary hypothesis $C$ such that $A' \cap C' = B' \cap C' = \varnothing$. In this case,

$$\frac{s(A)}{s(B)} = \frac{P(A, C)/P(C, A)}{P(B, C)/P(C, B)} = \frac{P(A, C \vee B)/P(C, A \vee B)}{P(B, C \vee A)/P(C, B \vee A)} = \frac{P(A, B)}{P(B, A)}$$

by repeated application of proportionality.

Second, suppose $A' \cup B' = T$. In this case, there is no $C' \in T$ that is not included in either $A'$ or $B'$, so the preceding argument cannot be applied. To show that $s(A)/s(B) = P(A, B)/P(B, A)$, suppose $C, D \in E$ and $A' \cap C' = B' \cap D' = \varnothing$. Hence,

$$\frac{s(A)}{s(B)} = \frac{s(A)s(C)s(D)}{s(C)s(D)s(B)}$$

$$= \frac{P(A, C)P(C, D)P(D, B)}{P(C, A)P(D, C)P(B, D)}$$

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

$$= R(A, B) \text{ (by the product rule [Equation 5])}$$

$$= P(A, B)/P(B, A) \text{ (as required).}$$

For any pair of exclusive hypotheses, therefore, we obtain $P(A, B)/P(B, A) = s(A)/s(B)$, and $P(A, B) + P(B, A) = 1$, by binary complementarity. Consequently, $P(A, B) = s(A)/[s(A) + s(B)]$ and $s$ is unique up to a choice of unit, which is determined by the value of $s(D)$.

To establish the properties of $s$, recall that unpacking (Equation 6) yields $P(D, C) \le P(A \lor B, C) = P(A, B \lor C) + P(B, A \lor C)$ whenever $D' = A' \cup B'$, $A$ and $B$ are exclusive, and $D$ is implicit. The inequality on the left-hand side implies that

$$\frac{s(D)}{s(D) + s(C)} \le \frac{s(A \lor B)}{s(A \lor B) + s(C)};$$

hence, $s(D) \le s(A \lor B)$. The equality on the right-hand side implies that

$$\frac{s(A \lor B)}{s(A \lor B) + s(C)} = \frac{s(A)}{s(A) + s(B \lor C)} = \frac{s(B)}{s(B) + s(A \lor C)}.$$

To demonstrate that the additivity of $P$ implies the additivity of $s$, suppose $A$, $B$, and $C$ are nonnull and mutually exclusive. (If $A' \cup B' = T$, the result is immediate.) Hence, by proportionality,

$$\frac{s(A)}{s(B)} = \frac{P(A, B)}{P(B, A)} = \frac{P(A, B \lor C)}{P(B, A \lor C)} = \frac{s(A)/[s(A) + s(B \lor C)]}{s(B)/[s(B) + s(A \lor C)]}.$$

Consequently, $s(A) + s(B \lor C) = s(B) + s(A \lor C) = s(C) + s(A \lor B)$.

Substituting these relations in the equation implied by the additivity of $P$ yields $s(A \lor B) = s(A) + s(B)$, which completes the proof of Theorem 1.

*Theorem 2:* Assume the ordinal model (Equation 11) and the solvability condition. Binary complementarity (Equation 3) and the product rule (Equation 5) hold iff there exists a constant $k \ge 0$ such that

$$P(A, B) = \frac{s(A)^k}{s(A)^k + s(B)^k}.$$

*Proof:* It is easy to verify that Equations 3 and 5 are implied by the power model (Equation 12). To derive this representation, assume that the ordinal model and the solvability condition are satisfied. Then there exists a nonnegative scale $s$, defined on H, and a strictly increasing function $F$ from the unit interval into itself such that for all $A, B \in$ H,

$$P(A, B) = F\left[\frac{s(A)}{s(A) + s(B)}\right].$$

By binary complementarity, $P(A, B) = 1 - P(B, A)$; hence, $F(z) = 1 - F(1 - z)$, $0 \le z \le 1$. Define the function $G$ by

$$R(A, B) = \frac{P(A, B)}{P(B, A)} = \frac{F\{s(A)/[s(A) + s(B)]\}}{F\{s(B)/[s(B) + s(A)]\}} = G[s(A)/s(B)], B' \ne \varnothing.$$

Applying the product rule, with $s(C) = s(D)$, yields $G[s(A)/s(B)] = G[s(A)/s(C)]G[s(C)/s(B)]$; hence, $G(xy) = G(x)G(y)$, $x, y \ge 0$. This is a form of the Cauchy equation, whose solution is $G(x) = x^k$ (see Aczel, 1966). Consequently, $R(A, B) = s(A)^k/s(B)^k$ and, by binary complementarity,

$$P(A, B) = \frac{s(A)^k}{s(A)^k + s(B)^k}, k \ge 0 \text{ (as required).}$$

---

## New *Publication Manual* for Preparation of Manuscripts

APA has just published the fourth edition of the *Publication Manual of the American Psychological Association*. The new manual updates APA policies and procedures and incorporates changes in editorial style and practice since 1983. Main changes cover biased language, presentation of statistics, ethics of scientific publishing, and typing instructions. Sections on references, table preparation, and figure preparation have been refined. (See the June 1994 issue of the *APA Monitor* for more on the fourth edition.)

All manuscripts to be published in the 1995 volumes of APA's journals will be copyedited according to the fourth edition of the *Publication Manual*. This means that manuscripts now in preparation should follow the guidelines in the fourth edition.

The fourth edition of the *Publication Manual* is available in softcover for $19.95 or in hardcover for $23.95 (members) or $29.95 (nonmembers). Orders must be prepaid, and a charge of $3.50 (U.S.) or $5 (non-U.S.) is required for shipping and handling. To order the fourth edition, write to the APA Book Order Department, P.O. Box 2710, Hyattsville, MD 20784-0710, or call 1-800-374-2721.