Cognitive Psychology **38,** 16–47 (1999)
Article ID cogp.1998.0706, available online at http://www.idealibrary.com on IDEAL®

Subjective Probability of Disjunctive Hypotheses: Local-Weight Models for Decomposition of Evidential Support

Lyle A. Brenner

Warrington College of Business Administration, University of Florida

and

Derek J. Koehler

University of Waterloo, Waterloo, Ontario, Canada

When the probability of a single member of a set of mutually exclusive and exhaustive possibilities is judged, its alternatives are evaluated as a composite ''residual'' hypothesis. Support theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994) implies that the process of packing alternatives together in the residual reduces the perceived evidential support for the set of alternatives and consequently inflates the judged probability of the focal hypothesis. Previous work has investigated the *global weights* that determine the extent to which the overall evidential support for the alternatives is discounted by this packing operation (Koehler, Brenner, & Tversky, 1997). In the present investigation, we analyze this issue in greater detail, examining the *local weights* that measure the specific contribution of each component hypothesis included implicitly in the residual. We describe a procedure for estimating local weights and introduce a set of plausible properties that impose systematic ordinal relationships among local weights. Results from four experiments testing these properties are reported, and a local-weight model is developed that accounts for nearly all of the variance in the probability judgments in these empirical tests. Local weights appear to be sensitive both to the individual component with which they are associated and to the residual hypothesis in which the component resides. © 1999 Academic Press

The order of authorship is arbitrary. Research reported in this article was supported by a grant from the Natural Science and Engineering Research Council of Canada (OGP 0183792) to the second author. We thank George Wu, Eldar Shafir, Yuval Rottenstreich, Daniel Osherson, Rick Larrick, and Craig Fox for their comments on an earlier version of this article and Aimee Drolet, Emily Marks, and Jen Pipe for their assistance with data collection.

Address correspondence and reprint requests to Lyle Brenner at Department of Marketing, Warrington College of Business Administration, University of Florida, Gainesville, FL 32611-7155. Fax: (352) 846-0457. E-mail: lyle@dale.cba.ufl.edu.

0010-0285/99 $30.00

1. INTRODUCTION

Intuitive assessments of likelihood, whether given in explicit numerical form or evaluated implicitly in the course of making a decision, depend on the cognitive processes by which uncertain events may be represented. Because of the richness and complexity of the events that typically interest us, there is usually more than one way in which such events can be described in natural language. For example, the same possible outcome of a presidential election could alternatively be described as ‘‘the Democratic candidate wins’’ or ‘‘the incumbent candidate wins.’’ The flexibility with which uncertain events can be described may have important consequences for understanding how judgments of likelihood are made.

Indeed, much research has shown that different descriptions of the same event can evoke substantially different judgments of its probability. Support theory (Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997) accommodates these results by associating subjective probability not with events, but with descriptions of events, called *hypotheses.* In support theory, each hypothesis A has associated with it a degree of support $s(A)$ representing the strength of evidence for that hypothesis. The judged probability that the focal hypothesis A holds, rather than the alternative hypothesis B, assuming one and only one obtains, is given by the proportion of support favoring A:

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

The support scale s allows different descriptions of the same event to evoke different degrees of support. Support theory is thus a nonextensional model; two hypotheses with the same extension (i.e., referring to the same event) may nonetheless have different degrees of perceived support.

In this paper, we use support theory to model the nonextensionality of subjective probability at a finer level of detail than has been done in previous work. We examine a general model in which support for a hypothesis is separated into components, each representing the individual contribution of one of a set of disjoint subhypotheses. A number of qualitative and quantitative principles of this decomposition of support are then proposed and tested. We first review past work on modeling nonextensional judgment in the context of support theory before introducing our decompositional model.

Subadditivity

Support theory makes a distinction between *explicit disjunctions,* which list their individual components, and *implicit disjunctions,* which do not. According to the theory, the process of ‘‘unpacking’’ an implicit disjunction into its components yields greater support for the resulting explicit disjunction. For example, unpacking the hypothesis ‘‘Joe will fly to Miami this year’’ into the explicit disjunction ‘‘Joe will fly to Miami this year for business or Joe will fly to Miami this year for pleasure’’ is assumed to increase perceived support. Unpacking may increase support by bringing to mind otherwise neglected possibilities or by increasing the salience and impact of the unpacked components.

Support theory also assumes that evaluation of an explicit disjunction as a whole tends to yield less support than does separate evaluation of each of its components. Formally, if A is an implicit hypothesis referring to the same event as the explicit disjunction $A_1 \vee A_2$, then

$$s(A) \leq s(A_1 \vee A_2) \leq s(A_1) + s(A_2).$$

The inequality involving the two leftmost expressions refers to implicit subadditivity of support, and the inequality involving the two rightmost expressions refers to explicit subadditivity of support. Both properties imply systematic deviations from extensionality in judgments of probability.

The Evaluation of Residual Hypotheses

In the studies to be reported, we examine situations in which there exists a fixed set of discrete possibilities, only one of which can be true. As an example, consider the task of predicting the winner of the Best Picture Oscar after the nominees have been announced. Denote the five nominated movies A, B, C, D, and E and consider judging the likelihood that movie A will win the award. This judgment pits the elementary hypothesis A against the ‘‘catchall’’ or *residual* hypothesis $\overline{A}$ (‘‘not-A’’), which is an implicit disjunction with the same extension as $B \vee C \vee D \vee E$. The *elementary judgment* of the likelihood of A can be represented in terms of support as

$$P(A,\overline{A}) = \frac{s(A)}{s(A) + s(\overline{A})}.$$

Suppose that elementary judgments for each hypothesis are elicited. Denote the sum of elementary judgments across a partition by T. In the case of a five-way partition, for example,

$$T = P(A,\overline{A}) + P(B,\overline{B}) + P(C,\overline{C}) + P(D,\overline{D}) + P(E,\overline{E}).$$

Probability theory requires that $T = 1$, because the hypotheses are mutually exclusive and exhaustive. According to support theory’s assumptions of implicit and explicit subadditivity, however, each elementary judgment introduces a bias in favor of the focal hypothesis, because its alternatives lose support by being packed together in the residual hypothesis. For instance, because of subadditivity of support, $s(\bar{A})$ typically will be less than the sum of the supports of its components (in this case, B, C, D, and E). Consequently, support theory implies that $T > 1$. This pattern has been consistently observed in a number of studies, in judgments of probability and of relative frequency by both experts and nonexperts (e.g., Teigen, 1974a, 1974b; Peterson & Pitz, 1988; Tversky & Koehler, 1994; Redelmeier, Koehler, Liberman, & Tversky, 1995; Fox, Rogers, & Tversky, 1996; Koehler, Brenner, & Tversky, 1997; Fox & Tversky, 1998; Fox, 1999).

It should be noted that the observation of $T > 1$ cannot be explained by a general bias in favor of the focal hypothesis. Judgments which pit a single elementary hypothesis against another elementary hypothesis typically sum to 1, showing no evidence of focal bias (Tversky & Koehler, 1994; Wallsten, Budescu, & Zwick, 1992; Fox, 1999; but also see Brenner & Rottenstreich, 1998; Brenner & Rottenstreich, 1999; Macchi, Osherson, & Krantz, 1999). Evidently, the excessive total probability is a consequence of the ‘‘packing’’ together of elementary hypotheses included implicitly in the residual hypothesis.

Global Weights: Discounting the Entire Residual Hypothesis

There are several ways to measure the degree of subadditivity of support in the assessment of residual hypotheses. The total probability T introduced above provides a simple overall measure of subadditivity across an entire set of elementary hypotheses and their corresponding residuals. One way to represent the degree of subadditivity at the level of individual residual hypotheses is to use *global weights*. Let the global weight $w_{\bar{A}} < 1$ represent the degree to which the support for the residual hypothesis $\bar{A}$ is discounted relative to the total support for its components. In the case of a five-way partition,

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

Using this global weight representation, the support for a residual hypothesis can be framed in terms of the total support for its components. Koehler, Brenner, and Tversky (1997) proposed that global weights for residual hypotheses may be systematically related to the support for the focal elementary hypothesis. In particular, they suggested that the greater the support for the elementary hypothesis, the greater the degree of discounting for the corresponding residual hypothesis and thus the smaller the associated w. This suggestion is based on the intuition that when the focal hypothesis is very strong, a judge will be less likely to carefully unpack the corresponding residual hypothesis, and thus the specific evidence for each of the individual components of the residual will be evaluated less exhaustively, yielding a low value of w. In contrast, when the focal hypothesis is very weak, a judge will be more likely to unpack the residual, considering evidence supporting each component, yielding a higher value of w.

Koehler et al. (1997) proposed a linear model of the global weights as a simple form that captures the proposed structure:

$$w_{\bar{A}} = 1 - \beta s(A), \qquad \text{with } \beta > 0.$$

This *linear discounting model* implies a qualitative pattern referred to by Tversky and Koehler (1994) and Koehler et al. (1997) as the *enhancement effect:* As the support for all hypotheses increases by a common factor, the degree of subadditivity also increases, because the global weights associated with the residual hypotheses decrease. As a result, the total probability T is greater for a strongly supported set of hypotheses than for a weakly supported set of hypotheses. For example, in one study by Koehler et al. (1997), subjects read a description of a murder in which there were five possible suspects. Judgments of the probability of guilt for each suspect were elicited under low information (which minimally implicated each suspect) and also under high information (which more strongly implicated each suspect). Consistent with the linear discounting model, subadditivity was more pronounced (i.e., T was greater) under high information—where the support for each hypothesis had increased—than under low information.

The linear discounting model also predicts that, within a given set of hypotheses, subadditivity is more pronounced for residuals of strongly supported elementary hypotheses than for residuals of weakly supported elementary hypotheses. These patterns were observed in the results of several studies by Koehler et al. (1997), in which global weights were derived jointly from probability judgments and direct assessments of support.

Local Weights: Discounting Individual Components of the Residual

A more detailed representation of subadditivity of residual hypotheses, which is the focus of this paper, incorporates *local weights* rather than global weights. Each local weight represents the degree of discounting of an individual component of the residual hypothesis, in contrast to the global weight approach in which the total support for the residual components is discounted by a single factor. Using local weights, the support for a residual hypothesis can be expressed as a weighted sum of its components' supports:

$$s(\bar{A}) = w(B,\bar{A})s(B) + w(C,\bar{A})s(C) + w(D,\bar{A})s(D) + w(E,\bar{A})s(E).$$

The local weight $w(B,\bar{A})$, which can be read as ''the weight of B in not-A,'' represents the weight of the component B within the residual hypothesis $\bar{A}$. Note that the global weight is the support-weighted average of the local weights.

Local weights have a natural interpretation as the degree to which the judge considers or attends to a particular component when evaluating the residual. We should stress that this representation does not entail an explicit and deliberate process of discounting each component of the residual. Indeed, the assumption underlying the subadditivity of the support function is that the residual is typically evaluated as a composite hypothesis, not as a set of components. Rather, the local weight expresses the degree of component-specific discounting resulting from a process in which the overall residual is evaluated without necessarily being unpacked into its components. We elaborate further on issues of interpretation regarding local weights in Sections 3 and 6.

By estimating local weights, in the present treatment we explore the representation of residual hypotheses in terms of the support accounted for by the individual components of the residual. This approach can provide a richer description of how residual hypotheses are mentally represented and evaluated and allows a more detailed formal representation of subadditivity of support. For instance, when judging the likelihood that *As Good As It Gets* will win this year's Best Picture Oscar, how and to what extent are each of the remaining nominated films considered? Do some films contribute more strongly to the impression of the residual than others? Is each film in the residual discounted, or are some actually given extra weight? Does a particular film (e.g., *Titanic*) contribute more to some residuals in which it resides than to others? Our approach can shed light on these and other questions. We are concerned especially with how the support for a particular component hypothesis in the residual may depend on (a) the support for the focal hypothesis and (b) the support for the component itself.

Overview

The remainder of the paper is organized as follows. In Section 2, we introduce several plausible ordinal properties that may characterize local weights. In Section 3, we describe an empirical procedure for estimating local weights. In Section 4, we present the results of several studies using this estimation procedure and test the candidate properties of local weights introduced in Section 2. Section 5 introduces a linear model of the local weights that accounts for most of their variance. Section 6 summarizes the results and concludes with a general discussion of the nature of residual representation and decomposition and the value of considering local weights instead of global weights.

2. POSSIBLE LOCAL WEIGHT PROPERTIES

In this section, we consider a number of plausible properties that local weights may exhibit. For maximal generality, in most cases the properties are ordinal. It is important to note that the properties below do not follow directly from any of the core axioms or assumptions of support theory.

Rather, these properties posit additional, psychologically plausible structure characterizing the degree of discounting of support within the constraints of the theory. Exploration of the structure of local (or global) weights can be seen as analogous to describing the shape of the utility or subjective value function in a model of decision making under uncertainty, such as expected utility theory or prospect theory (Kahneman & Tversky, 1979). The properties we consider, then, do not provide direct tests of support theory, but rather illustrate the power of the theory to aid in the generation and formal description of new substantive hypotheses.

As a running example to clarify the exposition and motivation of these properties, we consider the probability of winning a four-way race for public office involving a popular incumbent (I), who has the highest degree of evidential support, and three challengers of high, medium, and low support (H, M, and L, respectively).

Property 1. Local subadditivity

$$w(B,\overline{A}) \leq 1 \text{ for distinct } A,B.$$

Local subadditivity is a natural generalization of ''global'' subadditivity applied to local weights. Global subadditivity entails that the support of the overall disjunction is discounted relative to the sum of its components' supports; that is, the global weight is less than or equal to 1. Local subadditivity entails that *each individual component* of a disjunction is discounted to some extent. Local subadditivity implies global subadditivity, but the converse is not true. Global subadditivity could hold even if some components have local weights greater than 1, as long as other components have sufficiently low weights so that the support-weighted average of the local weights is not greater than 1.

In our election example, local subadditivity would hold if the support for each remaining candidate in the election was discounted somewhat when judging the likelihood that any particular candidate will win. For example, when judging the likelihood of I winning the race, support for each of H, M and L would be discounted in the residual, relative to the support evoked when they are focal or otherwise unpacked. Similarly, when judging the likelihood of M winning the race, each of I, H, and L would be discounted, and so on.

Property 2. Residual-dependent local weighting

$$w(A,\overline{B}) < w(A,\overline{C}) \text{ iff } w(D,\overline{B}) < w(D,\overline{C}) \qquad \text{for distinct } A, B, C, D.$$

Note that in the statement of Property 2, in each inequality the first argument (the component) is held constant, while the second argument (the residual) is varied. This property requires that for every individual component, the ordering of local weights across different residual hypotheses is the same.

In other words, if one component (e.g., A) carries more weight in residual $\overline{C}$ than in residual $\overline{B}$, then every other component that is shared by both residuals (e.g., D) also carries more weight in $\overline{C}$ than in $\overline{B}$. This property would be satisfied, for example, if local weights were constant for all components within a particular residual, but varied across the different residuals. To a first approximation, this property implies a ‘‘main effect’’ of the particular residual hypothesis in which the component resides—in other words, a main effect of the second argument in the $w(\cdot,\cdot)$ notation.

Returning to the election example, if the medium-support challenger M is discounted to a greater extent when the incumbent I is focal than when the low-support challenger L is focal, then residual-dependent weighting requires that the high-support challenger H is also discounted to a greater extent when I is focal than when L is focal. This may occur if the residual not-I evokes greater discounting for all of its included hypotheses than does not-L. For example, when judging the likelihood of I, one may simply consider evidence for why I may win against evidence for why I may lose. The latter set of evidence may not strongly support any of the individual components L, M, or H. In contrast, when judging the likelihood of L, one may consider evidence why L may win, and also evidence why each of the other candidates may win; in other words, when L is the focal hypothesis, each of the remaining individual possibilities may be considered more explicitly. Property 2 thus requires that local weights for components depend systematically on the residual hypothesis in which they reside—hence the summary label ‘‘residual-dependent local weighting.’’ The next property suggests a more specific form of this dependence.

Property 3. Local enhancement

$$w(A,\overline{B}) < w(A,\overline{C}) \text{ iff } s(B) > s(C) \qquad \text{for all distinct } A, B, C.$$

Here we propose that not only do local weights depend on the residual hypothesis, they depend specifically on the support for the focal hypothesis that defines the residual. In particular, if B has greater support than C, then the local weight for each component shared by both residuals $\overline{B}$ and $\overline{C}$ is lower in $\overline{B}$ than in $\overline{C}$. This property is an ordinal, local-weight generalization of the global-weight linear discounting model proposed by Koehler et al. (1997). In that model, recall that the global weight associated with a residual hypothesis is lower to the extent that support for the focal hypothesis under evaluation is stronger. That is, subadditivity is enhanced by increased support for the focal hypothesis that defines the residual. In local enhancement, this pattern applies to each individual component's local weight as well. Local enhancement implies global enhancement, but not vice versa.

Local enhancement captures the intuition that residuals may be more likely to be treated in a component-by-component manner when the focal hypothesis is weak than when the focal hypothesis is strong. The example above, which contrasts evaluation of the residual defined by I with evaluation of the residual defined by L, illustrates this notion. If I is a strong hypothesis, in evaluating its residual it may be natural to consolidate the remaining components, rather than evaluating them individually. Again, one might simply consider evidence why I might lose (e.g., he has been the subject of allegations involving shady business dealings), rather than evidence why each of the challengers may win. To the extent that this evaluation ignores relevant evidence supporting individual challengers, the degree of discounting of the residual components may be high, and the local weights low. In contrast, if a weaker hypothesis such as L is focal, when evaluating the residual it may be more natural to consider evidence for each of the other, more compelling candidates in turn (e.g., I has the advantage of incumbency, M has a better funded and organized campaign, and H is the media favorite). Such a pattern would lead to local enhancement.

We must be careful to distinguish between the claim that the degree of discounting depends on the residual hypothesis and the alternative claim that it depends on the focal hypothesis. These two conceptions are naturally confounded in the situations we have considered, in that each residual hypothesis is defined by a particular focal hypothesis. Thus, Properties 2 and 3 could also be framed in terms of focal dependence of local weights, rather than residual dependence. Such a characterization would be substantially more general than ours, however, because it could be applied to all disjunctions, not just to residual hypotheses. To illustrate, note that if local weights for a disjunction were dependent on the focal hypothesis, then the very same disjunction could be weighted differently when pitted against different focal hypotheses. For example, in this account, the support for the disjunction $B \vee C$ in the judgment $P(A, B \vee C)$ could be different from that in the judgment $P(D, B \vee C)$, despite holding fixed the evidence on which the judgments are based. Were this the case, it would represent a violation of one of the central principles of support theory—the assumption that support for a particular hypothesis is the same regardless of the hypothesis with which it is paired. This principle of *context independence* of support (which has been empirically validated in several studies, e.g., Brenner & Rottenstreich, 1998; Fox, 1999; see also Koehler, 1996) is implied by support theory's use of a single-argument support function that associates a unique support value with each hypothesis. Framing Properties 2 and 3 in terms of residual dependence rather than focal dependence avoids any violation of context independence. These properties, then, describe systematic variability in local weights across *different disjunctions*, not variability of support for the same disjunction across different contexts.

Recall that local weights are defined as a joint function of the component hypothesis with which they are associated and the residual hypothesis in which the component resides. Properties 2 and 3 concern residual dependence—that is, systematic variability of local weights across different residual hypotheses. The next two properties concern analogous component dependence—that is, systematic variability of local weights across different components.

Property 4. Component-dependent local weighting

$$w(A,\overline{C}) < w(B,\overline{C}) \text{ iff } w(A,\overline{D}) < w(B,\overline{D}) \text{ for all distinct } A, B, C, D.$$

Recall that Property 2 (residual-dependent local weighting) entailed a main effect of the residual; Property 4 entails an analogous main effect of the individual component. In each inequality in Property 4, the second argument (the residual) is held constant, while the first argument (the component) is varied. This property requires that the ordering of components' local weights within a residual is the same across different residuals. If one component is weighted more heavily than another in one residual, then it is weighted more heavily than the other in every residual that shares both components. For instance, in the election example, the component L could be systematically ignored or forgotten when evaluating the residuals of I, H, and M. In contrast, the component I could always be carefully evaluated in each of the residuals in which it resides. As a result, Property 4 would be satisfied in that component I has a higher local weight than component L in every residual in which they both reside. Essentially, Property 4 requires that some components carry consistently more weight in all residual hypotheses than do others.

Just as local enhancement (Property 3) described a specific form of residual-dependent weighting (Property 2), inverse-support component weighting (Property 5) describes a specific form of component-dependent weighting (Property 4):

Property 5. Inverse-support component weighting

$$w(A,\overline{C}) < w(B,\overline{C}) \text{ iff } s(A) > s(B) \text{ for distinct } A, B, C.$$

According to Property 5, lower-support components are weighted more than higher-support components within a common residual. This property suggests that the differences between the support values of various components are ‘‘smoothed out’’ somewhat by their inclusion in the residual hypothesis, such that strongly supported hypotheses are given relatively less weight, and weakly supported hypotheses are given relatively greater weight.

One possible process by which this property could operate is if consideration of the residual highlights features common to each of the residual's components or otherwise draws attention to evidence that is consistent with each of these components. In the election context, for example, evidence that the focal candidate may lose could be the primary evidence considered when evaluating the residual; this evidence may support each of the remaining candidates roughly equally. By ‘‘blurring’’ distinctions among the components of the residual, this process results in a more equal distribution of support across the components (compared to the case in which each individual component is explicitly considered). Hence, stronger components would receive relatively less weight in the residual, while weaker components would receive relatively more weight. We should stress that ‘‘more’’ or ‘‘less’’ weight is defined relative to the other components within the same residual hypothesis. By focusing on components within the same residual, this property controls for overall differences in weights across different residuals and specifically measures variability of local weights associated with the components' support values.

An opposite version of this property—entailing a positive correlation between a component's support and local weight—is also plausible. Such a property might be satisfied, for example, if attention is focused primarily on the strongest components when considering the residual hypothesis. The example described previously, in which strong components like *I* and *H* are given substantial weight in the residual, while weaker components like *L* and *M* are neglected, would produce results opposite to those predicted by Property 5. Regardless of its predicted direction of influence, at its most general level, Property 5 concerns the correlation between support of a component and the local weight for that component. It turns out, as will be seen shortly, that our empirical results are most consistent with Property 5 as formulated above.

3. ESTIMATING LOCAL WEIGHTS

We now describe a general procedure for estimating local weights empirically. Consider a partition with five elementary hypotheses *A, B, C, D,* and *E*. To assess the local weights, we require judgments of the likelihood of each hypothesis; further, the contribution of each hypothesis must be evaluated within the context of each residual hypothesis that includes it. For instance, evaluations are required of the contribution of hypothesis *B* to the overall support of the residuals $\bar{A}$, $\bar{C}$, $\bar{D}$, and $\bar{E}$.

To accomplish this, one group of participants (the control group) assigns probabilities to all five hypotheses simultaneously. These participants distribute 100% of probability mass across the five possibilities in a single step. The ‘‘neutral’’ setting of these judgments allows estimates of the individual support values $s(A)$, $s(B)$, $s(C)$, $s(D)$, and $s(E)$.

Other groups of participants first judge an arbitrarily assigned target hypothesis (e.g., *A*) against its residual (e.g., $\bar{A}$). This elementary judgment allows an assessment of the relative support for *A* and for $\bar{A}$. To allow estimation of the individual contributions of each of the components within $\bar{A}$, these participants then allocate the remaining probability (i.e., the amount not assigned to *A*) across the components *B, C, D,* and *E*. The elementary hypothesis provided as the initial target is varied across participants. In all experimental conditions, each participant is required to give a set of five judgments that sums to one.

If the distribution of probability across these components when they make up the residual differs substantially from the distribution of probability in the control condition, we have evidence that the local weights differ from one. Let $P_A(B)$ denote the judgment of hypothesis B when A is the target hypothesis and $P(B)$ denote the judgment of hypothesis B in the control condition. Using the notation of support theory, these judgments can be represented as follows:

$$P(B) = s(B)/[s(A) + s(B) + s(C) + s(D) + s(E)]$$

$$P_A(B) = w(B, \bar{A})s(B)/[s(A) + s(\bar{A})]$$

$$P_A(A) = P(A,\bar{A}) = s(A)/[s(A) + s(\bar{A})].$$

Note that the representations above ensure that $\sum_j P_i(j) = 1$ for all i (including the control condition). Aggregate local weights across a set of judges can then be estimated by comparing the mean judgment of a particular hypothesis when it is in a residual to the mean judgment of that hypothesis in the control condition. Specifically,

$$w(B,\bar{A}) = \frac{P_A(B)/P_A(A)}{P(B)/P(A)}.$$

One can verify this result either via algebra or via the following intuitive argument. The numerator gives the support for B relative to A, when A is the target and B is evaluated within the residual $\bar{A}$. The denominator gives the support for B relative to A, when A and B are on ‘‘equal footing’’ in the control condition. The ratio, then, gives the degree to which B is discounted when it is evaluated within the residual $\bar{A}$, rather than by itself (as in the control condition), which is precisely what is meant by a local weight.

This procedure is attractive in that it allows a completely ‘‘nonparametric’’ estimation of the local weights and the support values. That is, no assumptions are made about the form of the support values and local weights, beyond assuming that they are all nonzero. This can be most clearly seen by noting that the estimation procedure entails a one-to-one transformation from a set of probability judgments (assuming that no judgments are 0 or 1) to a set of support values and local weights. For example, in the case of five hypotheses, there are six conditions in the empirical procedure (the control condition and five conditions where each elementary hypothesis is focal), in which five probability judgments are made. Because the judgments must sum to 1, the total number of freely-varying data points for the probability judgments is $6 * (5 - 1) = 24$. There are also 24 parameters to be estimated:

4 support values (one for each elementary hypothesis, with one arbitrarily set to 1) and 20 local weights (4 weights for each of the 5 residual hypotheses). In general, for k elementary hypotheses, there are $(k - 1) * (k + 1)$ freely varying probability judgments, $k - 1$ support values and $k * (k - 1)$ local weights. It is easily seen that the number of data points is equal to the number of parameters to be estimated. In short, the estimation procedure can be seen as a transformation from a set of probability judgments into an equivalent set of support values and local weights; no information is lost, and no constraints are imposed on the data. The resulting local weights and support values can then be examined to determine if they can be characterized by simple principles of the kind posited in Properties 1–5. (For a related nonparametric approach to the estimation of the probability weighting function in the domain of decision making under risk, see Gonzalez & Wu, 1999, in this issue.)

4. DATA

We now turn to several experiments in which we use this general procedure to estimate local weights. In Experiment 1, participants judged the likelihood that each of the films nominated for the 1998 Best Picture Oscar would win the award. In Experiment 2, participants judged the probability of winning for both the Best Picture Oscar nominees and the Best Actor Oscar nominees. In Experiment 3, basketball fans assessed the probability of the different ‘‘Final Four’’ teams winning the 1998 NCAA Basketball Championship. In Experiment 4, we extend our procedure to judgments of absolute frequency; participants made judgments regarding the popularity of different college majors. In each case, we estimate the local weights and support values and test the properties presented in Section 2.

Experiment 1: Best Picture Oscar Judgments

Participants were 186 students at the University of Waterloo, approached in public areas of campus, who completed a short questionnaire in exchange for $2 Canadian. The questionnaire concerned predictions and evaluations of the five films nominated for the 1998 Best Picture Oscar. Initial instructions read in part:

> The following five films have been nominated for the **Best Picture** Oscar:
>
> *As Good As It Gets, The Full Monty, Good Will Hunting, L.A. Confidential, Titanic*
>
> In this questionnaire you will be asked to make a number of judgments about these films. *We recognize that you are unlikely to have seen all of these films, and may not have seen any of them.* Please make your evaluations on the basis of whatever information you have about each film, such as movie reviews, advertisements, or comments from friends who have seen the film.

Participants in the control condition, who distributed probability across the five candidates, were given the following instructions:

> Please estimate each film's *probability* of winning below. Your probability judgments should each be between 0% (indicating complete certainty that the film will NOT win) and 100% (indicating complete certainty that the film will win). Please make sure that the probabilities you assign to the five films add up to 100%.

These participants then proceeded to rate a subset of the films in terms of several attributes (*entertaining, thought provoking, emotionally moving,* and *funny*).

The remaining participants first judged the probability of a single designated ''target'' film, which varied across participants. After judging the probability of the target film, participants completed attribute ratings for that film, as a filler task. This attribute-rating task was designed to strengthen the manipulation of designating a target film and prevent participants from simply allocating probability across the five films at once, as in the control condition. On the following page, after completing several more attribute ratings, participants distributed the remaining probability among the four remaining films. For example, a participant who initially evaluated the target *As Good As It Gets* saw the following instructions:

> Now, please judge the probability of winning the Best Picture Oscar for each of the remaining nominated films. Make sure that the total you assign to the four of them, plus the probability that you already assigned to *As Good As It Gets,* adds up to 100%. In other words, make sure that the total of the probabilities of all five films adds to 100%.
>
> *The Full Monty* ______%
> *Good Will Hunting* ______%
> *L.A. Confidential* ______%
> *Titanic* ______%

Data from an additional 54 participants were excluded from the analyses that follow, as these participants failed to give probability judgments that added to 100%. Almost of all of these participants were in the noncontrol conditions and misinterpreted the instructions as requesting that the probabilities assigned to the four nontarget films (rather than all five films) add to 100%. Methodological precautions were taken to prevent this problem in Experiment 3.

Results

The mean probability ratings for the five films, by target film, are presented in Table 1. Rows of Table 1 designate different conditions, and columns represent the different judged films. By construction, the sum of the ratings within each row is 100%.

Consistent with subadditivity of support, the mean rating for each of the target films (represented by the bolded diagonal entries of the table) exceeds

TABLE 1
Means (and Standard Errors) of Probability Judgments for Best Picture Oscar in Experiment 1

	Judged hypothesis				
Target hypothesis	*As Good As It Gets*	*The Full Monty*	*Good Will Hunting*	*L.A. Confidential*	*Titanic*
Control	12.3	7.1	14.1	10.6	56.0
(n = 36)	(1.5)	(1.1)	(1.7)	(1.6)	(4.1)
As Good As It Gets	**14.7**	9.8	15.9	13.5	46.0
(n = 28)	(2.5)	(1.6)	(1.7)	(2.3)	(4.5)
The Full Monty	15.3	**8.4**	16.6	13.1	46.6
(n = 31)	(1.9)	(1.7)	(1.8)	(2.2)	(3.8)
Good Will Hunting	12.3	8.7	**21.0**	9.9	48.1
(n = 30)	(2.4)	(1.5)	(3.4)	(1.3)	(4.4)
L.A. Confidential	12.5	9.1	18.0	**14.3**	46.0
(n = 27)	(1.6)	(1.2)	(2.2)	(2.2)	(3.3)
Titanic	6.0	3.5	6.7	6.9	**76.9**
(n = 34)	(1.3)	(0.9)	(1.4)	(2.0)	(4.4)

the mean rating for the corresponding film in the control condition. Indeed, the sum of the target film elementary judgments, T, is 135% (SE = 6.7%), substantially greater than the 100% expected under additivity.

The local weights for each film in each residual were calculated based on the set of mean judgments, using the analysis described in Section 3. The resulting local weights are presented in Table 2, along with the support values for each film derived from the mean ratings in the control condition. In Table 2, the rows containing local weights represent different residual hypotheses, and the columns represent different component hypotheses. Because the support scale has an arbitrary unit, without loss of generality we adopt the convention that the sum of the support values across the partition is 1.

The set of local weights can be tested for correspondence with the properties proposed in Section 2. In accordance with Property 1 (local subadditivity), 80% (16 of 20) of the local weights are less than 1. Furthermore, the few weights that exceed 1 do so only slightly, while many of the weights are quite small.

The local weights appear to be dependent on the residual hypothesis; Property 2 is satisfied in 26 of 30 tests (87%). Indeed, examining Table 2 one can see clearly that the local weights (as well as the global weights) vary substantially across the different residual hypotheses (the rows of the table). The local weights are nearly 1.0 for the complement of *The Full Monty* but less than .5 for the complement of *Titanic.* As further evidence, in a main-effects ANOVA of the local weights by residual and component, the variability accounted for by the residual hypothesis is substantial, $F(4, 11) = 38.8$, $MSE = .007$, $p < .0001$.

TABLE 2
Support Values and Local Weights Derived from Best Picture Oscar Judgments in Experiment 1

	Component					
	As Good As It Gets (A)	*The Full Monty* (B)	*Good Will Hunting* (C)	*L.A. Confidential* (D)	*Titanic* (E)	Global weight
Support	.123	.071	.141	.106	.560	—
Local weight in not-A $w(\cdot, \overline{A})$	—	1.155	.944	1.070	.687	.813
Local weight in not-B $w(\cdot, \overline{B})$	1.051	—	.995	1.045	.703	.833
Local weight in not-C $w(\cdot, \overline{C})$	.671	.823	—	.633	.577	.618
Local weight in not-D $w(\cdot, \overline{D})$	.753	.961	.946	—	.609	.710
Local weight in not-E $w(\cdot, \overline{E})$	.355	.359	.346	.474	—	.381

TABLE 3
Summary of Local Weight Properties: Percentage of Tests Satisfied for Each Property, for Each Data Set

Data	P1	P2	P3	P4	P5
Expt 1: Picture	80	87	90	80	87
Expt 2: Picture	100	73	87	47	43
Expt 2: Actor	80	67	80	67	50
Expt 3: NC Cluster	92	83	75	50	58
Expt 3: KY Cluster	83	67	83	50	42
Expt 4: Majors	100	83	42	67	83
Overall average	89	76	80	63	60

Note. P1 is local subadditivity. P2 and P3 concern residual dependence. P4 and P5 concern component dependence.

Furthermore, the pattern of across-residual variability in the local weights matches the ordering of the support values for the focal hypothesis, consistent with local enhancement (Property 3). To illustrate, note that *Titanic* has high support while *not-Titanic* has very low local weights. In contrast, *The Full Monty* has low support while *not-The Full Monty* has high local weights. Overall, local enhancement is satisfied in 27 of 30 tests (90%). The Pearson correlation between local weights and support values of the negated component ($r = -.78$) can be taken as evidence in favor of a specific form of local enhancement, in which the local weights decrease linearly with increasing support for the focal hypothesis.

Turning to variability of local weights by component (i.e., by column in Table 2), we find that Property 4 is satisfied in 24 of 30 tests (80%). The variability accounted for by component is substantial, as illustrated by a significant main effect of component ($F(4, 11) = 3.6$, MSE $= .007$, $p < .05$), in addition to the main effect of residual noted above. Finally, Property 5 holds in the great majority (87%) of tests; stronger components tend to have smaller local weights. For instance, the *Titanic* column in Table 2 has substantially lower weights than the *Full Monty* column. A summary of the property tests in this and subsequent experiments can be found in Table 3.

Experiment 2: Best Picture and Best Actor Oscar Judgments

UCLA undergraduates ($N = 232$) participated in a replication and extension of Experiment 1; they completed the study within a questionnaire packet containing several other unrelated tasks. Participants assigned probabilities first to the 1998 Best Picture Oscar nominees and then to the 1998 Best Actor Oscar nominees (Matt Damon in *Good Will Hunting*, Robert Duvall in *The Apostle*, Peter Fonda in *Ulee's Gold*, Dustin Hoffman in *Wag the Dog*, and Jack Nicholson in *As Good As It Gets*).

In the control condition, participants gave probability assignments only

TABLE 4
Means (and Standard Errors) of Probability Judgments for Best Picture Oscar in Experiment 2

	Judged hypothesis				
Target hypothesis	*As Good As It Gets*	*The Full Monty*	*Good Will Hunting*	*L.A. Confidential*	*Titanic*
Control	12.4	7.3	17.4	12.9	50.0
(n = 35)	(1.3)	(1.2)	(2.1)	(1.6)	(4.0)
As Good As It Gets	**16.3**	7.9	17.1	11.3	47.4
(n = 31)	(1.8)	(1.0)	(1.8)	(1.4)	(3.7)
The Full Monty	13.2	**9.9**	20.9	14.8	41.2
(n = 32)	(1.4)	(1.4)	(2.4)	(2.3)	(3.5)
Good Will Hunting	10.5	6.2	**27.2**	8.1	48.0
(n = 30)	(1.3)	(1.2)	(3.4)	(1.5)	(4.2)
L.A. Confidential	9.5	6.5	16.9	**18.7**	48.5
(n = 33)	(1.3)	(1.1)	(2.3)	(3.0)	(4.4)
Titanic	7.3	4.2	13.1	9.0	**66.5**
(n = 35)	(1.3)	(0.8)	(1.8)	(2.0)	(4.4)

for each task. Participants not in the control condition (a) assigned a probability to a target film, (b) rated how entertaining that film was, and then (c) distributed the remaining probability among the nontarget films. This sequence of tasks was then repeated for the Best Actor nominees. Again, the rating task placed between the initial and final allocation of probability was designed to prevent participants from assigning all five probabilities at once, as in the control condition.

Results

Our analysis focuses on the 202 participants who provided responses that added to 100% for both tasks. The mean probability judgments and derived local weights and support values are presented in Tables 4 and 5 for Best Picture and in Tables 6 and 7 for Best Actor. Again, we find that total elementary judgments exceed 100% for both tasks; T = 139%, SE = 6.7, for the Best Picture judgments and T = 129%, SE = 7.4, for the Best Actor judgments.

Local weights are again consistently less than 1; Property 1 is satisfied in every case for Best Picture and in 80% of tests for Best Actor. Furthermore, we find that local weights are again highly residual dependent. Property 2 is satisfied in 73 and 67% of tests for Best Picture and Best Actor, respectively. In a main-effects ANOVA, the mean local weights vary by residual hypothesis for both Best Picture (F(4, 11) = 7.2, MSE = .008, p <. 01) and Best Actor (F(4, 11) = 5.7, MSE = .011, p < .01). Again, the variability by residual is consistent with local enhancement; Property 3 is satisfied in 87 and 80% of tests, and the Pearson correlations between local weights and

TABLE 5
Support Values and Local Weights Derived from Best Picture Oscar Judgments in Experiment 2

	Component					
	As Good As It Gets (A)	*The Full Monty* (B)	*Good Will Hunting* (C)	*L.A. Confidential* (D)	*Titanic* (E)	Global weight
Support	.124	.073	.174	.129	.500	—
Local weight in not-A $w(\cdot, \overline{A})$	—	.789	.721	.669	.691	.702
Local weight in not-B $w(\cdot, \overline{B})$	.767	—	.822	.903	.622	.718
Local weight in not-C $w(\cdot, \overline{C})$	.506	.552	—	.402	.623	.565
Local weight in not-D $w(\cdot, \overline{D})$	.560	.634	.689	—	.673	.657
Local weight in not-E $w(\cdot, \overline{E})$	.428	.410	.559	.534	—	.498

TABLE 6
Means (and Standard Errors) of Probability Judgments for Best Actor Oscar in Experiment 2

	Judged Hypothesis				
Target hypothesis	Matt Damon	Robert Duvall	Peter Fonda	Dustin Hoffman	Jack Nicholson
Control	24.5	16.2	13.2	19.1	27.1
(n = 36)	(3.2)	(2.0)	(1.4)	(2.1)	(3.0)
Matt Damon	**28.7**	13.3	13.5	16.0	28.5
(n = 34)	(3.6)	(3.0)	(1.7)	(2.4)	(2.9)
Robert Duvall	18.7	**19.5**	16.8	17.7	27.4
(n = 30)	(2.5)	(2.1)	(2.4)	(2.1)	(2.5)
Peter Fonda	22.2	13.4	**16.0**	14.8	33.7
(n = 31)	(3.3)	(2.4)	(3.3)	(1.6)	(3.1)
Dustin Hoffman	16.2	11.5	16.4	**22.9**	32.9
(n = 32)	(2.8)	(1.4)	(2.3)	(3.6)	(3.6)
Jack Nicholson	16.2	11.7	10.8	16.6	**44.7**
(n = 35)	(2.4)	(1.4)	(1.6)	(1.6)	(3.6)

support values are −.66 and −.36 for Best Picture and Best Actor, respectively.

Results for variability by component are less consistent. The Best Picture data show no evidence of variability by component: Property 4 is satisfied in only 47% of tests, and the main effect of component is trivial, $F(4, 11) = 1.3$. On the other hand, the Best Actor data does provide evidence for component dependence: Property 4 is satisfied in 67% of tests, and there is a strong main effect of component ($F(4, 11) = 9.8$, $p < .01$) in the ANOVA. This variability by component does not appear to be related to the support values of the components; Property 5 is satisfied at the chance level of 50%.

The extreme discrepancy in support between *Titanic* and the other nominees for the Best Picture category could cause some concern that the structure in the local weights may be driven primarily by this one outlier. The consistent results for the Best Actor category help to dispel this concern, because the variability in support for the Best Actor nominees is not nearly as large as that for the Best Picture nominees. Furthermore, examination of the films other than *Titanic* also reveals roughly the same pattern of results as does the full set of films.

Experiment 3: NCAA Basketball

We now examine a different prediction domain, with judges who are experienced with and inherently interested in the subject matter. Basketball fans ($N = 115$) were recruited from a computer news group devoted to discussion of college basketball and directed to a web site where they assessed the likelihood of different winners of the 1998 NCAA Men's Basketball Cham-

TABLE 7
Support Values and Local Weights Derived from Best Actor Oscar Judgments in Experiment 2

	Component					
	Matt Damon (*A*)	Robert Duvall (*B*)	Peter Fonda (*C*)	Dustin Hoffman (*D*)	Jack Nicholson (*E*)	Global weight
Support	.245	.162	.132	.191	.271	—
Local weight in not-*A* $w(\cdot, \overline{A})$	—	.698	.870	.713	.895	.802
Local weight in not-*B* $w(\cdot, \overline{B})$	.634	—	1.057	.770	.840	.798
Local weight in not-*C* $w(\cdot, \overline{C})$	.748	.682	—	.639	1.026	.799
Local weight in not-*D* $w(\cdot, \overline{D})$	.552	.592	1.036	—	1.013	.793
Local weight in not-*E* $w(\cdot, \overline{E})$	.401	.438	.496	.527	—	.459

pionship. In exchange for their participation, they were entered in a drawing to win one of three commemorative Final Four T-shirts.

At the time of the study, the field had been reduced to the Final Four teams North Carolina, Utah, Kentucky, and Stanford. The design of the experiment on the web site was similar to that of the previous studies. Participants in the control condition simultaneously assigned probabilities to all four teams on a single screen of their web browsers. All other participants assigned a probability to an initial designated target team on the first screen, and then assigned probabilities to the remaining teams on a second screen. The computer program used to run the web site ensured that all fans' probability judgments added to 100% across the four teams; in cases where judgments did not add to 100%, fans were required to reenter their judgments until this condition was met.

Results

The data split naturally into two clusters, independently of the experimental manipulation. Roughly half of the fans ($n = 59$) indicated North Carolina as the favorite, while the other half ($n = 55$) favored Kentucky. As a simple classification rule, those that assigned a larger probability to North Carolina than to Kentucky were placed in the North Carolina cluster; the remainder were placed in the Kentucky cluster. Data from one additional fan whose judgments did not fit either pattern were excluded from the analysis.

As before, the total elementary judgments in each cluster exceeded 100%; $T = 132\%$, SE = 7.5, for the North Carolina cluster, $T = 129\%$, SE = 10.9, for the Kentucky cluster. Table 8 lists the derived local weights for the two clusters. For both clusters, most weights are less than one, satisfying Property 1. The local weights are again strongly residual dependent, satisfying Properties 2 and 3. Consistent with this result, ANOVAs for each cluster indicate a marginal main effect of residual, $F(3, 8) = 3.2$, MSE = .019, $p < .10$, for the North Carolina cluster; $F(3, 8) = 3.0$, MSE = .021, $p < .10$, for the Kentucky cluster. Correlations between local weights and support for the focal hypothesis were also considerable, $-.34$ for the North Carolina cluster and $-.52$ for the Kentucky cluster. For neither cluster is there substantial variability based on component, as postulated by Properties 4 and 5.

Experiment 4: College Majors

The following experiment extends our analysis to judgments of absolute frequency. While effects of unpacking appear to be generally larger for judgments of probability than for judgments of relative frequency (Tversky & Koehler, 1994), judgments of relative and absolute frequency have both been found to exhibit subadditivity (e.g., Rottenstreich & Tversky, 1997). Furthermore, the enhancement effect has also been observed in judgments of relative frequency (Koehler et al., 1997). Examining the applicability of the present analysis to judgments of absolute frequency is worthwhile because some

TABLE 8
Support Values and Local Weights Derived from NCAA Final Four Judgments in Experiment 3

	Component				
	Kentucky (A)	North Carolina (B)	Stanford (C)	Utah (D)	Global weight
Carolina cluster					
Support	.219	.546	.109	.125	—
Local weight in not-A $w(\cdot, \bar{A})$	—	.656	.734	.630	.663
Local weight in not-B $w(\cdot, \bar{B})$	.769	—	.363	.559	.613
Local weight in not-C $w(\cdot, \bar{C})$	1.017	.748	—	.794	.821
Local weight in not-D $w(\cdot, \bar{D})$	.560	.413	.636	—	.477
Kentucky cluster					
Support	.495	.279	.101	.126	—
Local weight in not-A $w(\cdot, \bar{A})$	—	.575	.208	.199	.408
Local weight in not-B $w(\cdot, \bar{B})$	.865	—	.366	1.248	.862
Local weight in not-C $w(\cdot, \bar{C})$	.944	.970	—	2.252	1.13
Local weight in not-D $w(\cdot, \bar{C})$	.707	.536	.518	—	.630

researchers have claimed that commonly observed biases in probability judgments are greatly reduced or eliminated when frequencies are elicited instead (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Gigerenzer, Hoffrage, & Kleinbölting, 1991; see also Kahneman & Tversky, 1996).

Participants were 240 UCLA undergraduates who completed a short questionnaire within a packet containing several other tasks. Initial instructions read, in part:

> At a large Midwestern state university, the social sciences consist of four different majors: **psychology, sociology, political science, and economics.** Suppose there are *1000 students* enrolled at this university who have declared a single major in a social science.

Participants proceeded to allocate these 1000 students to the four possible majors. Participants in the control condition answered the question:

> How many of these students major in each of the individual social science majors? (make sure the total adds up to 1000)
>
> Economics _____ Psychology _____ Political Science _____ Sociology _____

Participants who encountered a designated target major (e.g., Economics) answered questions of the following form:

> How many of these students major in Economics, and how many major in a social science other than Economics? (make sure the total adds up to 1000)
>
> Economics ________ Social science other than Economics ________
>
> Of the number of students you assigned to a *Social science other than Economics*, how many major in each of the remaining individual majors?
>
> Psychology ________ Political Science ________ Sociology ________

Results

We analyze data from the 181 participants who provided judgments that added to 1000 as instructed. For purposes of comparability to the previous experiments, judgments are normalized and represented as proportions. Mean judgments are presented in Table 9; derived local weights are found in Table 10. The total of the elementary judgments (T = 137%, SE = 4.9) is again greater than 100%. Furthermore, all 12 local weights are less than 1, satisfying Property 1.

There is substantial variability of local weights by both residual ($F(3, 5) = 18.0$, MSE = .0006, $p < .01$) and component ($F(3, 5) = 8.5$, $p < .05$), satisfying Properties 2 and 4, respectively. There is no strong evidence of local enhancement in this case; Property 3 is satisfied in only 42% of all tests, and the correlation between local weights and the support for the focal hypothesis is a mere .19. Evidence for inverse-component support weighting is stronger. Property 5 is satisfied in 83% of tests and the correlation between local weights and the support of the particular component modified by the

TABLE 9
Means (and Standard Errors) of Normalized Frequency Judgments for Possible Majors in Experiment 4

	Judged Hypothesis			
Target hypothesis	Economics	Psychology	Political science	Sociology
Control	30.7	24.7	24.9	19.7
(n = 38)	(1.7)	(1.7)	(1.5)	(1.6)
Economics	**41.3**	21.3	20.6	16.8
(n = 34)	(2.7)	(1.5)	(1.7)	(1.1)
Psychology	24.8	**35.2**	21.7	18.3
(n = 33)	(1.8)	(2.6)	(1.6)	(1.5)
Political science	27.2	23.0	**30.9**	19.0
(n = 32)	(1.8)	(1.2)	(2.1)	(1.4)
Sociology	27.6	22.6	20.4	**29.3**
(n = 44)	(1.6)	(1.2)	(1.4)	(2.4)

TABLE 10
Support Values and Local Weights Derived from Normalized Frequency Judgments for Possible Majors in Experiment 4

	Component				
	Economics (A)	Psychology (B)	Political science (C)	Sociology (D)	Global weight
Support	.307	.247	.249	.197	—
Local weight in not-A $w(\cdot, \bar{A})$	—	.640	.613	.632	.628
Local weight in not-B $w(\cdot, \bar{B})$	.571	—	.614	.653	.606
Local weight in not-C $w(\cdot, \bar{C})$	.716	.750	—	.779	.744
Local weight in not-D $w(\cdot, \bar{D})$	.605	.614	.550	—	.591

local weight is $-.31$. This relatively low correlation is primarily due to the constrained range of both the local weights and the support values. The ordinal tests more clearly demonstrate that orderings of local weights for components within a residual are opposite the orderings of the components' support values.

5. A LINEAR MODEL OF LOCAL WEIGHTS

We can test specific versions of Properties 3 and 5 by assuming a simple linear model of the local weights and comparing the predicted probability judgments derived from this model to the observed mean judgments. In the linear local weight model, we represent local weights as

$$w(B, \bar{A}) = 1 - \alpha s(B) - \beta s(A).$$

Property 3 will be satisfied if $\beta > 0$; as support for the focal hypothesis increases, local weights for the components in the residual decrease. Property 5 will be satisfied if $\alpha > 0$; local weights for the particular component decrease as the support for that component increases.

This model of the local weights was fit to the set of mean probability judgments for each of the experiments described above, using SAS's weighted least-squares nonlinear regression procedure PROC NLIN. Table 11 contains summaries of the parameter estimates and goodness of fit measures for each data set. Goodness of fit measures for two comparison models are also provided: (a) an additive model in which all local weights equal 1 and (b) a constant-weight model in which all local weights are equal to each

TABLE 11
Summary of Parameter Estimates (and Standard Errors) and Goodness of Fit for Linear Local-Weight Model: $w(B, \bar{A}) = 1 - \alpha s(B) - \beta s(A)$

Data	Component (α)	Residual (β)	R^2 (%) full model	R^2 (%) all $w = 1$	R^2 (%) all $w = k$
Expt 1: Picture	.059 (.005)*	.113 (.011)*	99.7	89.7	96.8
Expt 2: Picture	.019 (.023)	.124 (.029)*	98.5	92.1	98.6
Expt 2: Actor	−.243 (.201)	.534 (.193)*	93.8	74.3	91.2
Expt 3: NC cluster	.116 (.037)*	.114 (.057)*	97.6	90.3	97.4
Expt 3: KY cluster	−.326 (.294)	.779 (.216)*	95.5	87.5	93.6
Expt 4: Majors	.129 (.120)	.295 (.112)*	97.9	34.5	98.1

other. In the additive model, only the support values are estimated; in the constant-weight model, the (constant) value of the local weights is an additional parameter to be estimated.

Several observations from Table 11 are noteworthy. First, the fit of the linear local weight model is excellent in all cases; R^2 values range from 93.8 to 99.7%. Even though the baseline fit of the additive model is also typically quite good, the linear local weight model improves substantially on the baseline fit in every case and leaves little variability unexplained. It is important to evaluate the improvement of more complex models in terms of the amount of variability left unexplained by simpler models. In Experiment 1, for instance, the linear model explains 97% of the variability left unexplained (10.3%) by the additive model. In other words, although there are substantial deviations from additivity, the pattern of subadditivity can be accounted for almost perfectly by a two-parameter linear model. The linear model also substantially improves on the constant weight model for three of the data sets (Experiment 1 Best Picture, Experiment 2 Best Actor, and Experiment 3 Kentucky cluster), while performing approximately equally well for the remaining three data sets.

Second, the estimates of the parameter β are positive and significantly different from zero in all cases, supporting the linear version of local enhancement. Finally, the estimates of α are significantly different from zero in two cases, tentatively supporting a linear version of inverse-component weighting. In short, in addition to the ordinal tests of Properties 3 and 5, we find solid and consistent evidence for residual-specific local weights that vary linearly with the support for the negated component and some evidence for a comparable form of component-specific weighting.

6. DISCUSSION

Virtually all judgments under uncertainty entail some assessment of a particular target event evaluated against its alternatives taken as a group. Investigation of how residual hypotheses are represented and evaluated is thus an essential aspect of the study of judgment under uncertainty. An appealing aspect of support theory is that it highlights the importance of the representation of residual hypotheses and also provides natural tools for modeling their support. The approach presented in this paper, of modeling support for residual hypotheses with local weights, may shed additional light on this issue.

Patterns of Local Weights

Summarizing the results of the previous experiments, we reach three primary qualitative conclusions about local weights. First, each component of a residual hypothesis appears to be discounted to some extent; Property 1 is overwhelmingly supported in each set of data. Stated differently, individual hypotheses are very rarely given full weight or strengthened by their inclusion within the residual. This is a substantially stronger version of subadditivity than has been considered in previous work on support theory.

Second, local weights vary considerably across different residual hypotheses (Property 2). In particular, the strength of the focal hypothesis appears critically important in determining local weights. Stronger focal hypotheses are associated with smaller local weights for each component in the residual (Property 3). In a sense, strong hypotheses are at a particular advantage when in the focal position—because the alternatives are typically discounted more heavily.

Third, local weights vary based on the particular component of the residual with which they are associated (Property 4). While this finding is less consistent and robust, it appears quite often (in the Best Picture data in Experiment 1, Best Actor data in Experiment 2, North Carolina cluster in Experiment 3, and majors data in Experiment 4). There is some evidence that stronger components of the residual tend to be discounted more than weaker components (Property 5). This can be viewed as a form of ‘‘smoothing’’ of within-residual variability of support.

The present results supplement and extend Koehler et al.’s (1997) analysis of global weights in several ways. First, while Koehler et al. needed to elicit both probability judgments and direct ratings of support to investigate enhancement, using the local-weight approach we can study enhancement using judgments of probability alone. Second, we observe enhancement at the level of local weights rather than global weights. By controlling for common individual components in the local enhancement tests, we can rule out the possibility that variability in global weights is merely due to the different components included in different residuals. For example, consider a model in which only component-wise variability of local weights exists and assume that *A, B,* and *C* always receive local weights of .8, .6, and .4, respectively. Based on these assumptions, *not-A* would have a relatively low global weight, because it contains the lower-weighted components *B* and *C,* while *not-C* would have a relatively high global weight, because it contains the higher-weighted components *A* and *B*. Extending enhancement to local weights as specified by Property 3 allows us to conclude that all local weights tend to be lower for residuals defined by stronger focal hypotheses and thus reject the explanation that global weight differences are merely manifestations of component-wise local-weight differences.

Nonetheless, it should be acknowledged that—in light of the relatively weak influence of component dependence (i.e., Properties 4 and 5)—the global linear discounting model (Koehler et al., 1997) will often provide a good first approximation of subadditivity, without the need to estimate local weights. Identification of the specific conditions under which component dependence does and does not hold will require further research.

Focal, Neutral, and Residual-Resident Hypotheses

Our general pattern of results can be concisely summarized in terms of how components with substantial supporting evidence are affected by either being made focal or being relegated to a residual hypothesis. The local enhancement results (Property 3) suggest that stronger hypotheses are especially helped by being singled out in the focal position, because the residual hypothesis is greatly discounted. In contrast, in accordance with inverse-support weighting (Property 5), a strong hypothesis residing in the residual may be most vulnerable to discounting. We can consider three possible evaluation states for hypotheses: (a) focal; (b) neutral, as in the control condition; or (c) residing within a residual. Our results suggest that hypotheses with a large degree of evidential support show the greatest variability across these different states. Weaker hypotheses, in contrast, appear to be the most stable across these three states. This characterization appears plausible in that only if there is a substantial degree of relevant evidence to be differentially accessed and weighed should there be substantial variability in support across the different states.

Quite different evidence may be considered when an element is singled out or relegated to the residual. Furthermore, evidence evoked by a singled-out hypothesis may affect the evaluation of the alternatives. For example, in evaluating evidence for the composite residual of films other than *L.A. Confidential,* one may consider evidence for why the focal film may lose (e.g., the film was not a popular success, it is somewhat cynical and depressing, gritty crime films typically do not win Best Picture, etc.). Much of the evidence relating to *L.A. Confidential*'s chance of losing may support each of the remaining hypotheses nearly equally. For instance, one reason why *L.A. Confidential* is likely to lose—gritty crime films do not often win—could support each of the weaker hypotheses *The Full Monty*, *As Good As It Gets,* and *Good Will Hunting* nearly as much as the stronger hypothesis *Titanic.* Evidence cued or driven by the focal hypothesis may also interfere with the retrieval of evidence that implicates a particular component, but not the entire residual. Focusing on how gritty crime films tend not to win may distract one from the romantic charms of *Titanic* which implicate it as a likely winner. Thus, some of the ''unique'' supporting evidence for the stronger components may be neglected in such an evaluation or may be shifted somewhat in favor of the weaker components in the residual. The reason that local weights are smaller for stronger components may simply be because these components have the most to lose.

Formation and Evaluation of Residual Composites

The observed patterns of local weights are consistent with the general view that the residual is typically not evaluated in a piece-by-piece, extensional form. Rather, the residual may be evaluated as an overall composite (e.g., ''*Titanic* does not win'') which may call to mind or be supported by substantially different evidence than would be evoked by a piece-by-piece evaluation. The view that disjunctions are typically represented as a composite impression has been a central assumption underlying past analyses of support theory (Koehler et al., 1997; Tversky & Koehler, 1994; Rottenstreich & Tversky, 1997). In discussing the representation of residual hypotheses, Koehler et al. (1997) proposed that, rather than first evaluating the evidence for each elementary hypothesis separately, and then aggregating the support of the hypotheses residing in the residual, people instead aggregate the hypotheses in the residual and then evaluate the support of the resulting composite. There are a number of processes by which evaluation of the composite is likely to yield less support than what would have resulted from a piece-by-piece, extensional evaluation. Evidence that supports several components of the residual, for example, may be counted only once when a composite impression of the residual is evaluated, resulting in subadditivity.

Extending this notion somewhat, one partial explanation of local enhancement is that when the focal hypothesis is strongly supported, the tendency to form a composite impression of the residual may be intensified. In this view, local weights are especially low within residuals pitted against strong hypotheses because the residual is not decomposed to the same extent as when the focal hypothesis is weak. Conversely, residuals may be more naturally decomposed when a coherent general impression of the residual is less natural or is harder to form. This is likely to be the case when the focal hypothesis is only weakly supported, and several disparate individual components within the residual may appear plausible, perhaps for quite different reasons.

Local Weight Estimation: Methodology and Interpretation

The conception of residual hypotheses as composites may appear at odds with the ''(de)compositional'' logic of the local weight representation introduced in this article. If the residual is evaluated as a composite, why model it as a weighted sum of the support of its components? We should stress that the local-weight representation is not intended to capture an explicit cognitive process of component-by-component evaluation of the residual. Local weight decompositions of support can be viewed as analogous to attributewise decompositions of utility commonly used to model choice. For example, a consumer's preferences among a set of TVs may be represented parsimoniously by an additive utility model in which a TV's utility is composed of the sum of weighted utilities for its attributes (such as size, price, picture clarity, brand). This type of model is useful for both predictive and explanatory purposes, even though the consumer most likely evaluates each TV in a holistic (rather than an attribute-by-attribute) manner. Just as attribute-specific models of utility can decompositionally represent the structure of preference, local-weight models of support can decompositionally represent the structure of belief.

It is useful, then, to distinguish the method of local-weight estimation from the psychological processes by which these local weights are determined. Methodologically, we can view the determination of local weights in our empirical procedure as having two stages. First, a composite residual is formed, and the support for the explicitly defined focal hypothesis is evaluated against this composite's support. Second, the composite is broken up into its distinct components, which are then individually assessed. We assume that this second process is unlikely to occur in the absence of explicit instructions or other motivation to evaluate the individual components of the residual.

Properties describing how local weights vary across residual hypotheses (e.g., Properties 2 and 3) address the first step, namely, how evidence supporting the composite residual may depend on the focal hypothesis. Roughly, this first step concerns the global weights for each residual, defining an average level for the local weights. Properties describing the distribution of local weights within a particular residual (e.g., Properties 4 and 5), in contrast, address how the evidence that supports the composite is allocated to its individual components in the decomposition step. Thus, analysis of residual dependence can address composite formation, while analysis of component dependence can address composite decomposition.

Strictly speaking, then, it would be a mistake to interpret the local weight associated with a component hypothesis as a direct measure of its salience or contribution to the support assigned to the residual hypothesis as a whole. It is possible, for example, that the judge may not think at all of a particular component of the residual hypothesis when assigning a probability to the focal hypothesis in the first step of the estimation procedure, but may still give that component nonzero weight when asked to explicitly assess each of the individual components of the residual in the second step.

From our view, however, as is suggested by our earlier discussion, imposing this kind of process interpretation of local weights would in any case be misguided, because it implies that support for the residual hypothesis is assessed strictly in terms of its components. We assume, in contrast, that formation of a composite residual hypothesis precedes assessment of its support. Residual hypothesis formation may eventually be best understood in terms of the set of features taken to represent the residual, which may or may not be collected from the residual's component hypotheses. In the residual formation process, certain component hypotheses may be neglected altogether, while factors unrelated to the components (e.g., incorporation of features contrasting those of the focal hypothesis) may play an important role. The likelihood of a component hypothesis making a contribution to the perceived support for the residual as a whole is presumably related to factors such as the hypothesis's distinctiveness and salience.

Nonetheless, there may be circumstances in which the methodological procedure for estimation of local weights bears some resemblance to the psychological processes of residual hypothesis formation. In particular, we suspect that some kind of (de)compositional process may occur when the judge is evaluating one of a small number of competing elementary hypotheses. Under these circumstances the judge may very well implicitly assess the contribution of each component hypothesis to the overall support of the residual. When a large number of discrete hypotheses are under assessment, in contrast, it is highly unlikely that each component will receive individual scrutiny in the process of assessing support for the residual as a whole. Indeed, the functional basis of the process of composite residual hypothesis formation may lie in the computational complexity that would arise if evidence had to be assessed in terms of its implication for each of a large number of alternative hypotheses.

Conclusion

In summary, the evidence considered in evaluating a residual hypothesis appears to be quite different from the evidence considered when one evaluates each of its components individually. Fortunately, how the discrepancies between the two sets of evidence affect judgment under uncertainty can be largely captured with just a few simple principles. The evidence supporting the residual composite is less compelling than the aggregate of the evidence for each individual component; thus, local weights are typically less than 1. Furthermore, consistent with local enhancement, one may form a tighter and more unified composite impression of the residual when the focal hypothesis is strong. Finally, general evidence against the focal hypothesis may be perceived as supporting each of the components in the residual roughly equally; thus, consistent with inverse-support weighting, weaker components in the residual may gain weight at the expense of stronger components. Consistent with the lessons learned from studies of choice and similarity, the evaluation of evidence under conditions of uncertainty appears to be highly dependent on the way in which the judgment task is framed.

REFERENCES

Brenner, L. A., & Rottenstreich, Y. (1998). *Asymmetric support theory: Focus-dependence and context independence in likelihood judgment.* Unpublished manuscript.

Brenner, L. A., & Rottenstreich, Y. (1999). Focus, repacking, and the judgment of disjunctive hypotheses. *Journal of Behavioral Decision Making,* in press.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions of the literature on judgment under uncertainty. *Cognition,* **58,** 1–73.

Fox, C. R., Rogers, B., & Tversky, A. (1996). Options traders exhibit subadditive decision weights. *Journal of Risk and Uncertainty,* **13,** 5–19.

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

Fox, C. R. (1999). Strength of evidence, support, judged probability, and choice under uncertainty. *Cognitive Psychology,* **38,** 167–189.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review,* **102,** 684–704.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review,* **98,** 506–528.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology,* **38,** 129–166.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review,* **103,** 582–591.

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

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

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgments. *Psychological Review,* in press.

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

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

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

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

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

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

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.

Accepted November 18, 1998