Psychological Review
1999, Vol. 106, No. 1, 210–214

0033-295X/99/$3.00

THEORETICAL NOTES

A Note on Superadditive Probability Judgment

Laura Macchi
University of Milan

Daniel Osherson
Rice University

David H. Krantz
Columbia University

Recent studies have demonstrated subadditivity of human probability judgment: The judged probabilities for an event partition sum to more than 1. We report conditions under which people's probability judgments are superadditive instead: The component judgments for a partition sum to less than 1. Both directions of deviation from additivity are interpreted in a common framework, in which probability judgments are often mediated by judgments of evidence. The 2 kinds of nonadditivity result from differences in recruitment of supporting evidence together with reduced processing of nonfocal propositions.

Suppose that an event, *E*, has been partitioned into two or more mutually exclusive subevents and that probability assessments are made for *E* and for each of these subevents. The assessments are said to be *additive* if the probability assigned to *E* is approximately equal to the sum of the probabilities of the subevents. They are *subadditive* if the probability assigned to *E* falls short of the subevent sum, and they are *superadditive* if the assignment to *E* exceeds the subevent sum. Superadditivity is a feature of Shafer's theory of evidence (Shafer, 1976) and has been found previously for evidence judgments (Briggs & Krantz, 1992) but not for probability judgments. Subadditive probability judgment has been widely reported in the literature and helped to motivate Support Theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). The purpose of the present article is to document the existence of superadditive probability judgment in special conditions. Our findings suggest modifications of Support Theory.

Tversky and colleagues (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994) pointed out that subadditivity is common in both nonexpert and expert probability judgments. In a dramatic example (Redelmeier, Koehler, Liberman, & Tversky, 1995), physicians were asked to provide probabilities for the following events, with respect to a particular hospitalized patient whose case had been summarized to them:

a. Dies during the present hospital admission,
b. Discharged alive but dies within 1 year,
c. Lives more than 1 but less than 10 years, and
d. Lives more than 10 years.

Because the four events are exhaustive, additivity entails that their probabilities sum to 1. Each of 52 physicians assessed the probability of exactly one of the four outcomes (a–d), which was randomly assigned to him or her. Under these conditions, the mean judgments of the four component probabilities summed to 1.64, which implies that many of the physicians' assessments were too high.

An appealing explanation for subadditivity was introduced in a seminal article (Fischhoff, Slovic, & Lichtenstein, 1978) and elaborated in Support Theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). It assumes first that explicit details in a description *A* of an event provide cues that lead to recruitment of evidence (or support) for the occurrence of that event. To illustrate, consider a doctor who is asked to evaluate the probability of Outcome (a) from the previous example. The description, "dies during the present hospital admission," suggests ways in which death could occur in the short term. In contrast, its (implicit) negation, "does not die during the present hospital admission," lacks the details that would be contained in the explicit disjunction of the components, Outcomes (b)–(d). To complete this explanation, one further assumes that probability is usually assessed by weighing the evidence that comes to mind for and against a given proposition. This is expressed by the basic equation in Support Theory,

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

Here, *A* and *B* are two descriptions that are understood to be mutually exclusive, and *P(A,B)* is the judged probability of *A*, when the alternative to *A* is *B*. Subadditivity then results from the excess of evidence that is recruited in favor of each description *A*

Laura Macchi, Department of Psychology, University of Milan, Milan, Italy; Daniel Osherson, Department of Psychology, Rice University; David H. Krantz, Department of Psychology, Columbia University.

We thank Donald Redelmeier & Yuval Rottenstreich for helpful discussion.

Correspondence concerning this article should be addressed to David H. Krantz, Department of Psychology, Columbia University, 406 Schermerhorn Hall, 1190 Amsterdam Avenue, New York, New York 10027. Electronic mail may be sent to lmacchi@imiucca.csi.unimi.it, osherson@sanity.hsr.it, or dhk@psych.columbia.edu.

Table 1
The Four Partitions Used in the Experiments

= Statement	> Statement	Reversed > statement
The freezing point of gasoline is equal to that of alcohol.	The freezing point of gasoline is greater than that of alcohol.	The freezing point of alcohol is greater than that of gasoline.
The 1995 birthrate in Burma was equal to that of Thailand.	The 1995 birthrate in Burma was greater than that of Thailand.	The 1995 birthrate in Thailand was greater than that of Burma.
The caloric content of a liter of sunflower oil is equal to that of a liter of corn oil.	The caloric content of a liter of sunflower oil is greater than that of a liter of corn oil.	The caloric content of a liter of corn oil is greater than that of a liter of sunflower oil.
The height of the Duomo in Milan is equal to that of Notre Dame in Paris.	The height of the Duomo in Milan is greater than that of Notre Dame in Paris.	The height of Notre Dame in Paris is greater than that of the Duomo in Milan.

of a component subevent of a partition, when it serves as the focus, compared with the evidence for its implicit and underdescribed negation $B = -A$. This also explains why subadditivity can be reduced or absent when the partition consists only, or principally, of two complementary and symmetrically described subevents. If A_1 and A_2 are complementary symmetric descriptions, then the (implicit) negative description $-A_1$ may be functionally equivalent to A_2 for many or most people, and similarly for $-A_2$ and A_1. If $s(-A_1) = s(A_2)$ and $s(A_1) = s(-A_2)$, substituting in Equation 1 leads to additivity:

$$P(A_1, -A_1) + P(A_2, -A_2) = 1.$$

In the initial presentations of Support Theory, this last property, called *binary complementarity,* was raised to the status of an axiom. The existing evidence (Tversky & Koehler, 1994; Wallsten, Budescu, & Zwick, 1992) was consistent with this axiom. It would be very much in the spirit of the theory, however, to postulate that additivity for complementary subevents depends on functional symmetry in the descriptions, as noted above. A generalization of Support Theory, in which binary complementarity is no longer an axiom, needs to be developed.

One purpose of this article is to underline the need for generalization by showing that additivity can fail for binary partitions. A more important purpose is to document deviations in the direction of superadditivity when evidence is weak for both A_1 and A_2. This is a new empirical finding,[1] one that suggests complexities in the relationship between evidence and probability judgments.

We suggest that superadditivity should be observed when the recruitment of evidence for each isolated subevent fails to uncover much support, and when the contrary description is not processed equally. Consider, for example, the assertion that the 1995 birthrate in Burma was greater than that in Thailand. A person who knows little about such things may find no clear reason to believe this assertion and may thus judge its support to be weak. The weak support might favor a low assessment of probability, despite the fact that the contrary proposition—namely, the Thai birthrate exceeds the Burmese—also recruits little evidence. The low assessment would result from the implicit character of the contrary proposition; because it is not formulated explicitly, the judge may fail to recognize that it, too, has little support.

In the preceding example, the exhaustive partition elements are as follows: (a) Thailand had a greater 1995 birthrate than Burma, (b) Burma had a greater 1995 birthrate than Thailand, and (c) the birthrates were identical. If (a) and (b) are each judged to have low probability and (c) is judged to have zero probability, the result is superadditivity.

This prediction of superadditivity does not appear to have been tested empirically. Such was the purpose of the experiment we now describe.

Method

On the basis of pilot testing, we constructed four ternary partitions, as shown in Table 1. It was assumed that virtually everyone recognizes the three statements of a given partition to be exhaustive and mutually exclusive. It was also assumed that the credibility of the =-statement of each partition was close to zero for our respondents, but as a precaution it was explicitly set to zero by the wording of our questions. Illustrating with the first partition, the questions corresponding to the >-statements and the reversed >-statements were as follows:

> *>-question:* The freezing point of gasoline is not equal to that of alcohol. What is the probability that the freezing point of gasoline is greater than that of alcohol?
> **Probability:** __
>
> *Reversed >-question:* The freezing point of alcohol is not equal to that of gasoline. What is the probability that the freezing point of alcohol is greater than that of gasoline?
> **Probability:** __

Although the focal proposition varies in these two questions, they are syntactically parallel in all respects. The content of the questions was designed to provoke considerable, but not total, uncertainty in the minds of our respondents and to be associated with few reasons for believing either the >-statement or its reversed variant. The four partitions, each with two distinct possible focal propositions, give rise to eight questions.

The eight questions were translated into Italian and administered to 80 university students in introductory psychology classes in Milan and Padua. Each student received four questions—namely, for each partition, either the >-question or its reversed variant, but not both. The latter choice as well as the order of the four questions was individually randomized under the restriction that exactly 40 students respond to each of the eight questions. The questions prepared for a given student were printed on a single page and administered in a group setting with no time limit. Instructions to students emphasized that the questions do not call for a yes–no answer but rather a percentage representing probability; answers of 0 and 1 were to be entered only in case of certainty.

We subsequently performed an exact replication of the experiment,

[1] One of the reviewers called our attention to the fact that superadditivity has been found under some conditions in children's judgments of frequency of success or failure in a skill task (Cohen, Dearnaley, & Hansel, 1956). However, aspects of the method used in this pioneering study make the findings difficult to interpret, and they may not be related to the prediction we test here.

using a different sample of 80 students from the same population. Finally, to determine the impact of the leading inequality statement in each question, we performed a control experiment with a new group of 80 students. The control was identical to the preceding studies, except that the first sentence was removed from all questions.

Results

Figure 1 presents a histogram of all 960 probability judgments collected in our three experiments. About 34% of the responses were .50, which most likely reflects the cultural convention whereby "50–50" represents ignorance; most of the remaining responses were distributed over the other 10 multiples of .10. The large number of .50 responses evidently reflects the respondents' impression of weak evidence but in a fashion that tends toward additivity for these partitions. In particular, analysis of median responses is unrevealing: Because of the lump at 50–50, the median response in each study was .50. Superadditivity can nonetheless occur if many respondents generate low probabilities in response to statements for which they have little evidence, rather than using the conventional 50–50. The asymmetry between the left and right ends of Figure 1 suggests that this was indeed the case. Table 2 documents this asymmetry, comparing responses at .10 or lower with those at .90 or higher. The table shows that extreme low responses occurred at least twice as often as correspondingly extreme high ones. This pattern was also reflected in the mean probability judgments, which were .423 in the original experiment, .466 in the replication, and .451 in the control. Because these means represent the responses to both directions in each partition, they demonstrate substantial superadditivity. Deviations from the hypothesis of .50 or more (predicted by additivity or subadditivity) are statistically reliable for each experiment ($p < .01$ by one-tailed t-test, $N = 80$). It is interesting that little or no asymmetry is observed when comparing response intervals [.3, .4] versus [.6, .7]. The superadditivity of mean judgments is thus

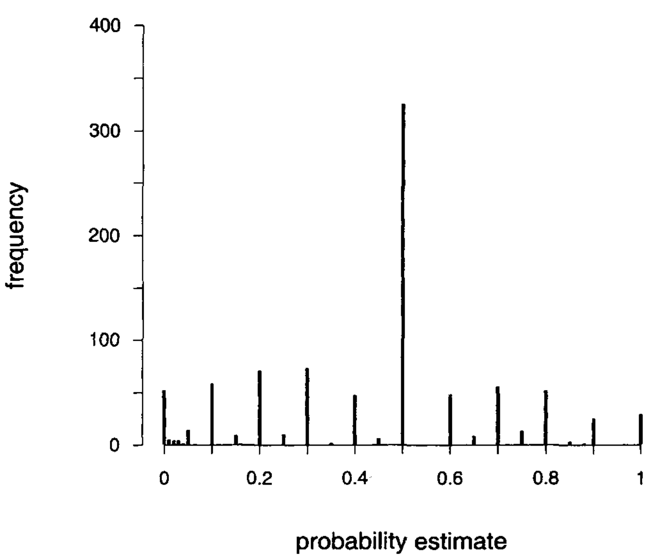

Figure 1. Frequency histogram for all 960 probability judgments collected in the first 3 experiments (4 judgments from each of 240 respondents).

Table 2
Number of Responses (Out of 320) in Each Experiment That Were .10 or Less, Versus .90 or More

	Response range	
Experiment	.00 to .10	.90 to 1.00
Original	52	14
Replication	34	16
Control	62	29

based on responses at the extremes, especially those at or beyond .10 and .90.

The null hypothesis of symmetry between low and high responses should not be tested by a chi-square or binomial test based on Table 2, because the data include up to four responses per respondent, making independence assumptions suspect. Instead, we categorized respondents according to their overall judgment profiles: *low* for respondents who gave one or more responses at .10 or below, but none as high as .90, versus *high* for respondents who showed the reverse profile. As shown in Table 3, the former outnumber the latter approximately three to one, and the null hypothesis of symmetry can be rejected in each experiment.

An analysis of the four partitions, taken separately, supports the conclusion of superadditive judgment. Table 4 shows the means and standard deviations for the eight questions in each experiment. In every case, the sum of the answers given to the >- and reversed >-questions in the same partition sum to less than 1.

We note that the magnitude of superadditivity seen in Table 4 (about 10%) is less than the magnitude of subadditivity reported (Redelmeier et al., 1995; Tversky & Koehler, 1994) for partitions of three or four components (around 50%). We attribute this at least in part to the respondents' mixed response strategies: Absence of evidence is sometimes expressed as a response of .50 and sometimes as a response near 0.

Restoring Symmetry

As explained earlier, our prediction of superadditivity assumes that judges fail to recognize the extent to which nonfocal components of a partition also have slight support. To examine this assumption directly, we replicated the first experiment but added a

Table 3
Number of Respondents (Out of 80) in Each Experiment Classified as Low *versus* High

Experiment	Low	High
Original	27	8
Replication	18	7
Control	29	9

Note. The *low* category consists of all respondents who made at least one judgment of .10 or less and no judgment of .90 or more. The *high* category consists of all respondents who made at least one judgment of .90 or more and no judgment of .10 or less. All differences are significant ($p < .05$) by a two-sided binomial test.

Table 4
Mean Probability Provided for Each of the Eight Questions in the Four Experiments

Experiment	>-Statement *M*	>-Statement *SD*	Reversed *M*	Reversed *SD*	Σ *M*
Original					
Partition 1	.457	.211	.427	.287	0.884
Partition 2	.463	.184	.387	.202	0.850
Partition 3	.470	.250	.466	.250	0.936
Partition 4	.436	.293	.279	.221	0.715
Replication					
Partition 1	.439	.238	.518	.227	0.957
Partition 2	.465	.218	.461	.195	0.926
Partition 3	.458	.210	.464	.230	0.922
Partition 4	.436	.252	.484	.261	0.920
Control					
Partition 1	.495	.313	.411	.281	0.906
Partition 2	.552	.265	.353	.214	0.905
Partition 3	.457	.279	.407	.260	0.864
Partition 4	.453	.328	.483	.301	0.936
Restoring symmetry					
Partition 1	.544	.349	.532	.337	1.076
Partition 2	.553	.278	.407	.256	0.960
Partition 3	.491	.283	.443	.278	0.934
Partition 4	.563	.332	.489	.315	1.052

Note. For each cell, $N = 40$. Σ *M* is the sum of the two means in the corresponding row.

reminder about the alternative event. Illustrating with the first partition, the new questions were as follow:

> *>-question:* The freezing point of gasoline is not equal to that of alcohol. Thus, either the freezing point of gasoline is greater than that of alcohol, or the freezing point of alcohol is greater than that of gasoline. What is the probability that the freezing point of gasoline is greater than that of alcohol?
> **Probability:** __
>
> *Reversed >-question:* The freezing point of alcohol is not equal to that of gasoline. Thus, either the freezing point of alcohol is greater than that of gasoline, or the freezing point of gasoline is greater than that of alcohol. What is the probability that the freezing point of alcohol is greater than that of gasoline?
> **Probability:** __

The responses to this version of the questions were highly additive: The 90% confidence interval for the mean of the respondents' average probability judgment is .503 ± .029. The last part of Table 4 shows the means and standard deviations for each partition.

This result suggests that asymmetry of processing the focal proposition and its contrary plays an important role in superadditivity. When respondents are reminded of the contrary, their probability judgments tend to be normalized and, hence, are more nearly additive, as postulated by Support Theory. Note that our findings are consistent with the intuition that additivity will result from having the same person assess the probabilities of all components of a given partition. On the other hand, they are inconsistent with a model that maps low evidence directly into low probability.

Discussion

The simple experiments reported here establish one set of conditions that lead to superadditive probability judgments: low level of knowledge about the questions that are posed, together with asymmetric processing of a description and its negation.[2] Y. Rottenstreich (personal communication, September 9, 1997) has also pointed out that the mechanism suggested here predicts subadditivity for binary partitions when a high level of knowledge is combined with asymmetric processing. Experiments currently in progress (Chen, Krantz, Osherson, & Bonini) indicate this to be the case.

Previous studies of probability judgment for binary partitions have reported additivity. Three possible explanations should be mentioned. First, some or all of these studies may not have satisfied the set of conditions just described. Second, some studies may have used medians as measures of the central tendency of probability judgments. The predominance of 50% responses (see Figure 1) could produce additivity: If our data were summarized by medians rather than means, additivity would be widespread in the item analysis of Table 4. As Figure 1 shows, the median is not a good descriptor of the distribution for probability judgments. Finally, the major study that supports additivity (Wallsten et al., 1992) used participants who made many probability judgments, including both of the complementary descriptions at one time or another during their tasks. This could lead the participants to consider the implicit negation more systematically, as in the experiment on restoring symmetry.

Support Theory (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994) involves two psychological insights: Probability judgments for propositions are mediated by the strength of evidence recruited by the descriptions of those propositions, and mapping onto a scale from 0 to 1 involves normalization of evidence strength, which is well approximated by Equation 1. It seems in the spirit of that theory to assume that recruitment of evidence can be different, depending on whether the description is explicitly presented and whether it is the target for the probability judgment. Such asymmetry between explicit target descriptions and others can occur even where the partition is essentially binary. Thus, our results can be viewed as compatible with the basic insights of Support Theory, although not with its formal statement. Development of a suitable modification of the theory represents an important challenge.

[2] Another set of conditions that produces superadditivity (Y. Rottenstreich, personal communication, April 10, 1996) arises when one of the descriptions is an explicit disjunction that can easily be "repacked" into a single, less detailed description. The greater attention in focal processing makes such repacking more probable and, thus, can reduce the evidence recruited for the disjunction, compared with the evidence recruited when the disjunction is salient but is not the judgmental target.

References

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

Cohen, J., Dearnaley, E., & Hansel, C. (1956). The addition of subjective probabilities: The summation of estimates of success and failure. *Acta Psychologica, 12,* 371–380.

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.

Redelmeier, D. A., Koehler, D. J., Liberman, V., & Tversky, A. (1995). Probability judgment in medicine: Discounting unspecified alternatives. *Medical Decision Making, 15,* 227–231.

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

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

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

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

Received October 15, 1996
Revision received April 27, 1998
Accepted April 27, 1998 ■

Editor's Note
Diversifying the Scope of Theoretical Notes in *Psychological Review*

Traditionally, Theoretical Notes in *Psychological Review*, with rare exceptions, have consisted of critiques of prior articles and replies to such critiques. As a matter of formal policy, the *Review* is now open to Theoretical Notes of multiple types, including, but not limited to, discussions of previously published articles, comments that apply to a class of theoretical models in a given domain, critiques and discussions of alternative theoretical approaches, and metatheoretical commentary on theory testing and related topics.

This initiative represents an effort to make *Psychological Review* the home for a broad range of theoretical commentary. There will be no change, however, in the *Review*'s policy of subjecting Theoretical Notes to a rigorous review for publication, nor will there be a change in the *Review*'s policies on critiques and replies (see the January 1996 issue). Theoretical Notes will continue to be distinguished from regular articles, not only by their appearing in the Theoretical Notes section of each issue, but also by wording such as "Critique of. . .," "Reply to. . .," "Comment on. . .," "Note on. . .," and so forth, in the titles of such articles.