Journal of Behavioral Decision Making, Vol. 10, 293–313 (1997)

The Enhancement Effect in Probability Judgment

DEREK J. KOEHLER, LYLE A. BRENNER AND AMOS TVERSKY
Stanford University, USA

ABSTRACT

Research has shown that the judged probability of an event depends on the specificity with which the focal and alternative hypotheses are described. In particular, unpacking the components of the focal hypothesis generally increases the judged probability of the focal hypothesis, while unpacking the components of the alternative hypothesis decreases the judged probability of the focal hypothesis. As a consequence, the judged probability of the union of disjoint events is generally less than the sum of their judged probabilities. This article shows that the total judged probability of a set of mutually exclusive and exhaustive hypotheses increases with the degree to which the evidence is compatible with these hypotheses. This phenomenon, which we refer to as the enhancement effect, is consistent with a descriptive account of subjective probability called support theory. © 1997 John Wiley & Sons, Ltd.

Behav. Dec. Making, 10: 293–313, 1997.

No. of Figures: 3. No. of Tables: 6. No. of References: 19.

KEY WORDS judged probability; focal hypothesis; alternative hypothesis; enhancement effect; subjective probability; support theory

Previous research has shown that different descriptions of the same event can give rise to different judgments of the probability of that event. More specifically, the judged probability of an event can increase as its description becomes more detailed (e.g. Fischhoff *et al.*, 1978). To accommodate such findings, Tversky and Koehler (1994) developed a theory in which subjective probability is not attached to events — as in other models — but rather to descriptions of events, referred to as hypotheses. According to this account, called support theory, each hypothesis A has a support value $s(A)$, corresponding to the strength or weight of evidence for A. The judged probability that the focal hypothesis A rather than the alternative hypothesis B holds, assuming that one and only one of them is valid, is given by:

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

That is, the probability of a hypothesis is given by its support normalized relative to the support of the alternative.

CCC 0894–3257/97/040293–21$17.50

Accepted 27 November 1996

The support associated with a hypothesis is interpreted as a measure of the strength of evidence available to the judge in favor of the hypothesis. Depending on the context in which the judgment is made, this support measure may reflect an analysis of objective data (e.g. the homicide rate in the relevant population) or a subjective impression mediated by judgmental heuristics such as representativeness or availability (Kahneman *et al.*, 1982). For example, the support for the hypothesis 'Jim is an engineering major' may be based on the degree to which Jim's personality matches that of the stereotype of an engineering student.

Support theory distinguishes between explicit and implicit hypotheses. For example, 'homicide' (H) is an implicit hypothesis, whereas 'homicide by an acquaintance or homicide by a stranger' ($H_a \vee H_s$) is an explicit hypothesis describing the same event. Support theory assumes that the support of an implicit hypothesis is generally less than or equal to the support of an explicit hypothesis that refers to the same event, which in turn is less than or equal to the sum of the supports of the components. That is, if H_a and H_s form a partition of H, then

$$s(H) \leqslant s(H_a \vee H_s) \leqslant s(H_a) + s(H_s) \quad (2)$$

That is, the support function is subadditive with respect to both implicit and explicit disjunctions (Rottenstreich and Tversky, in press).

Support theory (i.e. equations (1) and (2)) implies that the judged probability of the focal hypothesis increases when the focal hypothesis is 'unpacked' into its components, and that the judged probability of the focal hypothesis decreases when the alternative hypothesis is unpacked into its components. Both memory and attention effects may contribute to this phenomenon. Unpacking a hypothesis into its components may remind the judge of possibilities that he or she might otherwise have overlooked, or it may increase the salience of possibilities that were not made explicit. The unpacking effect implies that people do not assess the probability of a hypothesis by adding the assessments of the individual components; rather, they appear to construct and evaluate a summary or composite representation of the hypothesis. This possibility is elaborated in the general discussion.

A common manifestation of the unpacking effect involves the discounting of the residual hypothesis, defined as the complement of a focal elementary hypothesis (e.g. that the patient suffers from something other than a common cold, or that the correct answer to a test question is different than one's guess, etc.). This paper investigates the factors that control the degree of discounting of the residual hypothesis. In the following experiments, subjects judged the probability of each in a set of four mutually exclusive and exhaustive hypotheses. For example, in the first experiment, subjects were asked to assess the probability that a college student has a specified major given that he or she definitely majors in a social science, which consists of Economics (A), Political Science (B), Psychology (C), or Sociology (D). Subjects were asked to evaluate the probability of each elementary hypothesis (e.g. John majors in economics, A) against its residual (e.g. John does not major in economics, $\bar{A}$). Note that the residual $\bar{A}$ (i.e not economics) is an implicit hypothesis that has the same extension as $B \vee C \vee D$ (i.e. political science or psychology or sociology).

According to support theory, the probability of the hypothesis A that John majors in economics is represented:

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

The support of the residual $s(\bar{A})$ can be expressed as $w_{\bar{A}}[s(B) + s(C) + s(D)]$, where $0 \leqslant w_{\bar{A}} \leqslant 1$ by subadditivity of $s(\bar{A})$. Hence, $w_{\bar{A}} = s(\bar{A})/[s(B) + s(C) + s(D)]$. This index measures the degree of subadditivity or discounting of the residual hypothesis $\bar{A}$. In our example, the value of $w_{\bar{A}}$ reflects the degree to which the support for the individual majors (Political Science, Psychology, Sociology) is discounted when those possibilities are included implicitly in the residual hypothesis serving as an alternative to the focal hypothesis (Economics).

The subadditivity of the support function, implied by $0 \leqslant w_{\bar{A}} \leqslant 1$, predicts that when the probability of each of the four mutually exclusive and exhaustive hypotheses is judged against its residual, their total probability (denoted T) will exceed one, contrary to standard probability theory. Thus in the example above we have $T = P(A, \bar{A}) + P(B, \bar{B}) + P(C, \bar{C}) + P(D, \bar{D}) \geqslant 1$. Considerable research supports this prediction, which has been observed even in the case of expert judgments made by physicians (Redelmeier *et al.*, 1995) and options traders (Fox *et al.*, 1996).

Our concern in this paper is to extend our understanding of the relationship between the evidence upon which a probability judgment is based and the degree to which the residual is discounted. In the most general case, each hypothesis $\bar{A}$ has a different discounting factor $w_{\bar{A}} \leqslant 1$. This form makes no prediction beyond the inequality $T \geqslant 1$. At the other extreme, the discounting factor is a constant for all hypotheses. With a constant discounting factor of 0.5, for example, the support of the residual hypothesis always equals half the sum of the support of its components. Thus evidence suggesting that John majors in Psychology would provide only half as much support for a residual hypothesis including Psychology as it would for Psychology as the focal hypothesis. This form implies that judged probabilities remain unchanged if the support values of all hypotheses under consideration are increased by a constant factor.

We introduce an intermediate form in which the discounting factor $w_{\bar{A}}$ of the residual hypothesis is inversely related to the support of the focal hypothesis $s(A)$. For simplicity, we assume that $w_{\bar{A}}$ is a linear function of $s(A)$:

$$w_{\bar{A}} = 1 - \beta s(A) \tag{3}$$

where $\beta > 0$. That is, the greater the support of the focal hypothesis A, the greater the degree of discounting of its alternative, residual hypothesis (i.e. the smaller the discounting factor $w_{\bar{A}}$).

This linear discounting model reflects the intuition that when the focal hypothesis has a great deal of support, the corresponding residual hypothesis will be unpacked into its components to a lesser extent, and evidence supporting the individual components of the residual hypothesis will be evaluated less exhaustively, than when the focal hypothesis has less support. For example, if there is substantial evidence in favor of 'psychology', the hypothesis 'not psychology' will not be as readily unpacked into its components 'economics or political science or sociology', and the evidence suggesting each of those components will tend not be considered individually, resulting in a lower w. In contrast, if there is minimal evidence in favor of 'psychology', then 'not psychology' may be more readily unpacked into its components, and the evidence supporting each of those component hypotheses may be more likely to be considered, resulting in a higher w. Thus, greater support for the focal hypothesis is expected to be associated with smaller values of w for the corresponding residual hypothesis, as captured in equation (3).

This form implies the *enhancement effect*: If the support of each elementary hypothesis under consideration increases by a constant factor, then each judged probability should also increase. Hence, the sum of the judged probabilities of all elementary hypotheses T is predicted to increase when the support of each hypothesis is enhanced. In our college majors example, if new information about the student (e.g. that he subscribes to *Time* magazine and likes to stay up to date on current events) increases the support of each possible social science major, the total probability assigned to the four possible majors is expected to increase.

A key difference between support and probability is that support need not be compensatory: Evidence may be encountered that increases the support for each in a set of mutually exclusive and collectively exhaustive hypotheses. Classical probability theory, on the other hand, requires compensatory probability judgments: Increases in the probability of one hypothesis must be compensated for by decreases in the probability of one or more of its alternatives. According to the present account, however, the subadditivity of the support function can lead to probability judgments that are not fully compensatory. In particular, our treatment of enhancement predicts that evidence that increases the support of all the hypotheses under consideration results in probability judgments that are correspondingly less compensatory and thus yields greater total probabilities T. This prediction is tested in the following studies.

For each study, we gather independent judgments of support to test qualitative and quantitative predictions about enhancement. The qualitative prediction is that as the support for each in a set of hypotheses increases, so does the sum of the judged elementary probabilities. The quantitative prediction is that the discounting of a residual hypothesis is inversely related to the support of the focal hypothesis. Because we gather both probability judgments and independent assessments of support, we can estimate the discounting weights w and test the linear discounting model of equation (3), which extends the original formulation of support theory by providing a more precise description of how the discounting weights w are influenced by enhancement of the evidence.

The support for a hypothesis can be derived from subjects' probability judgments (see Tversky and Koehler, 1994). Alternatively, the subject can be asked to give a direct assessment of the extent to which the available evidence supports a hypothesis. The direct assessment task can be couched in generic terms applicable across a wide variety of domains in which the notion of evidential support is explicitly mentioned (e.g. to what extent does the available evidence support the hypothesis that Team X will win the game?) or it can be adapted specifically for the task at hand via a natural variable assumed to correspond to support in that domain (e.g. how strong is Team X this season?). Neither type of rating is without drawbacks: It could be argued that in the generic rating task subjects may not make the conceptual distinction between support and probability, while use of an adapted rating task depends on the assumption that the natural variable in question is in fact the basis upon which the probability judgments are made. To avoid overreliance on a single method, some of the present experiments used a generic rating task and others used an adapted rating task.

Experiment 1 demonstrates the enhancement effect and tests the linear discounting model using probability judgments of college students' majors based on courses the students had taken. The courses used as evidence varied in how strongly they supported the possible majors. We assume that the support for a hypothesis is based on representativeness; hence we use as our direct measure of support subjects' ratings of the conceptual relatedness between academic fields. These independent ratings allow estimation of the linear discounting model. Greater values of T were expected for those courses that were perceived as highly related to the possible majors. Experiment 2 tests whether the enhancement effect is observed for judgments of relative frequency as well as of probability; Experiment 3 tests whether the enhancement effect can be attributed to a confusion of (inverse) conditional probabilities.

Experiment 4 examines the case of adding new evidence rather than substituting different evidence to increase support. Subjects judged the probabilities of different college majors on the basis of an initial set of evidence and again as additional information was provided. Independent ratings of support were obtained using the more generic rating task described above; subjects assessed the extent to which the evidence supported the hypotheses under consideration. Greater values of T were expected as more evidence is encountered in favor of each of the hypotheses under consideration. Experiment 5 investigates judgments of the guilt of potential suspects in two crime stories. Judgments of the suspiciousness of the suspects are used as measures of support. In all experiments, the linear discounting model provides a good fit of the data.

EXPERIMENT 1

In Experiment 1 we test enhancement by using as evidence courses that vary in the degree to which they support the four different social science majors under consideration.

Method

Subjects were 115 Stanford University students and staff, recruited either through the introductory psychology subject pool or a newspaper advertisement run in the daily newspaper. Subjects from these two sources did not differ systematically in their responses, so this variable is not considered in the analyses which follow. Data from one additional subject were dropped as this subject failed to complete the questionnaire as instructed.

The subjects' task was to assess the probability that a college student had a specified major based on a course the student was said to have taken (cf. Mehle *et al.*, 1981). Subjects were asked to consider target students specified by first name and last initial. Each student was said to attend a large state university in the midwestern United States. Subjects were told that each student definitely had a social science major, which at this university consisted of (A) Economics, (B) Political Science, (C) Psychology, and (D) Sociology. Subjects were asked to make their judgments based on one of four different courses the target student was said to have taken in the second year: Western Civilization, French Literature, Modern Physics, or a Statistics course on data analysis. These four courses were chosen by selecting two typical and two atypical courses varying along a 'hard–soft' dimension. Statistics is a 'hard' course typical of most social science majors, while Western Civilization is a 'soft' course that is also fairly typical. French Literature is a 'soft' but atypical course for social science majors, while Physics is a 'hard' atypical course. Our treatment of enhancement suggests that the typical courses should yield a greater degree of residual discounting than should the atypical courses.

Probability judgments

On the basis of the course information, subjects were asked to judge the probability that the target student had a specific major (e.g. $P(B, \bar{B})$: the probability that Greg H. majors in Political Science). Instructions regarding this task were as follows:

> The probability you assign to a given major can vary between 0% (which means that you are certain the student does not have that major) and 100% (which means you are certain that the student does in fact have that major). Note that when you say a student has an X% (say, 35%) probability of having a given major, you are also in effect saying that the student has a $100 - X$% (in this case, 65%) probability of having some other major instead.

Subjects estimated the probability of these elementary hypotheses by circling one of the numbers 0%, 5%, ..., 100% on the scale provided.

Each subject assessed all 16 possible combinations of course with major in one of two possible orders. In other words, for each course, subjects assessed the probability that someone who had taken that course had each of the four possible majors. To avoid an explicit demand for additivity, the 16 questions were not blocked by course type, and the first names of the students involved varied from one major to the next within a course type. Male names were used with Statistics and Physics, and female names were used with Western Civilization and French Literature. The order in which the 16 probabilities were assessed had no systematic effect on the judgments and thus is disregarded in the analyses which follow.

Relatedness judgments

A separate group of 46 Stanford students judged, on a 0–10 scale, the relatedness between pairs of fields corresponding to the courses (Statistics, Western Civilization, French Literature, Physics) and the majors (Economics, Political Science, Psychology, Sociology). We treat these relatedness judgments as measures of the extent to which a particular course supports a particular major. Instructions were as follows:

> Please rate the degree of relatedness between the pairs of academic fields given below. For example, *mathematics* and *physics* would probably be considered highly related, while *chemistry* and *music* would probably be considered unrelated. For each pair of fields, make your rating by circling a number on the provided scale.

Note that no distinction was made between what served as courses and as majors in the experiment; instead both courses and majors were labeled as 'academic fields' to prevent subjects in this task from attempting to estimate p(course | major) or p(major | course). Subjects also judged the relatedness between each course and the 'social sciences as a whole'.

Results and discussion

Relatedness judgments

Exhibit 1 displays the mean relatedness judgments for the majors and courses. A Course by Major repeated measures analysis of variance was performed on these data revealing a significant main effect of Course, $F(3, 671) = 497.7$, $p < 0.001$, but no effect of Major, $F(3, 671) < 1$. The Course by Major interaction was also significant, $F(9, 671) = 15.7$, $p < 0.001$. As expected, the two courses that would seem to be typical of a social science major (Statistics and Western Civilization) received higher ratings of relatedness to all the majors than did the 'atypical' courses (French Literature and Modern Physics), $t(671) = 38.1$, $p < 0.0001$. The judged relatedness of the courses to the 'social sciences as a whole' reinforce this finding, again indicating greater relatedness for the typical courses.

We note, incidentally, that the judged relatedness of the courses to the social sciences as a whole are generally closer to the maximum than to the mean relatedness judgment of the course to the individual majors, suggesting that when judging the course's relatedness to the social sciences as a whole, subjects may have brought to mind the specific social science major that was most compatible with the course. For example, when rating French Literature's relatedness to the social sciences as a whole ($M = 3.36$), subjects may have based their judgment on its relatedness to the major within social science they thought a French Literature course would most benefit, namely Sociology ($M = 3.50$).

Probability judgments

The probability judgments show a similar pattern to that found for the relatedness judgments, as can be seen from Exhibit 2. First, as predicted by support theory, the total probability assigned to the elementary hypotheses (T) are substantially greater than 1. Of the 115 subjects, only 7 had values of

Exhibit 1. Mean relatedness judgments for each major, listed by course, in Experiment 1

	Course			
Major	Statistics	Western Civilization	French Literature	Modern Physics
Economics	8.73	5.83	1.72	2.46
Political Science	6.22	7.26	3.22	1.70
Psychology	7.33	5.78	2.76	2.28
Sociology	6.76	7.26	3.50	1.63
All Social Science	7.31	6.91	3.36	2.44

Exhibit 2. Mean elementary probability judgments for each major and their total T, listed separately for each course in Experiment 1

	Course			
Major	Statistics	Western Civilization	French Literature	Modern Physics
Economics	0.66	0.40	0.31	0.45
Political Science	0.41	0.60	0.46	0.32
Psychology	0.59	0.37	0.39	0.35
Sociology	0.52	0.59	0.45	0.30
Elementary total (T)	2.18	1.96	1.61	1.42

T that were equal to or less than 1. Second, the predicted effect of enhancement was observed, in that the value of T increased with the typicality of the course. A Course by Major repeated measures analysis of variance revealed significant main effects of Course, $F(3, 1701) = 124.7$, $p < 0.001$, and of Major, $F(3, 1701) = 5.17$, $p < 0.01$, as well as a significant interaction, $F(9, 1701) = 52.1$, $p < 0.001$. The enhancement effect was reflected by a significant contrast between the two typical courses and the two atypical courses, $t(1701) = 9.43$, $p < 0.001$. That is, not only were the two typical courses judged as being more strongly related to the possible majors, they also yielded greater total probabilities.

The observed enhancement effect implies that, contrary to standard probability theory, people's judgments are not fully compensatory. The stronger support offered by the typical courses increased the average probability assigned to the possible majors. Observation of the enhancement effect is also not easily reconciled with nonadditive probability theories such as Shafer's theory of belief functions (Shafer, 1976), which implies superadditive judgments (i.e. $T \leqslant 1$) that become less superadditive as more evidence is revealed (see Tversky and Koehler, 1994, p. 559). Cohen's (1977) theory of Baconian probabilities, another nonadditive model, is also superadditive and therefore inconsistent with the data. It should be noted that these theories were developed as logical rather than psychological accounts of evidential evaluation. Comparison of these theories to the data is instructive, nonetheless, as some authors (Cohen, 1979; Van Wallendael and Hastie, 1990) have considered them as possible descriptive models of judgment.

Estimation of discounting model

Using the mean probability judgments and the mean relatedness judgments, w was estimated for each course–major pair using the equation:

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

where s represents the direct support rating in the form of relatedness judgments. Fitting the linear discounting model relating w to the mean relatedness judgments across the sixteen judgments yields $\beta = 0.107$. Based on this single parameter estimate, the probability judgments can be predicted from the relatedness judgments. Exhibit 3 plots the predicted probabilities based on the fitted discounting model against the mean judged probabilities. Judgments based on atypical courses are represented by filled points; judgments based on typical courses are represented by unfilled points.

The correlation between the predicted and observed mean judgments is 0.939. The mean signed and absolute differences between the predicted and observed mean judgments are -0.006 and 0.063, respectively. The probability judgments are predicted quite well from independent judgments of relatedness between fields using the simple linear model relating $s(A)$ and $w_{\bar{A}}$.

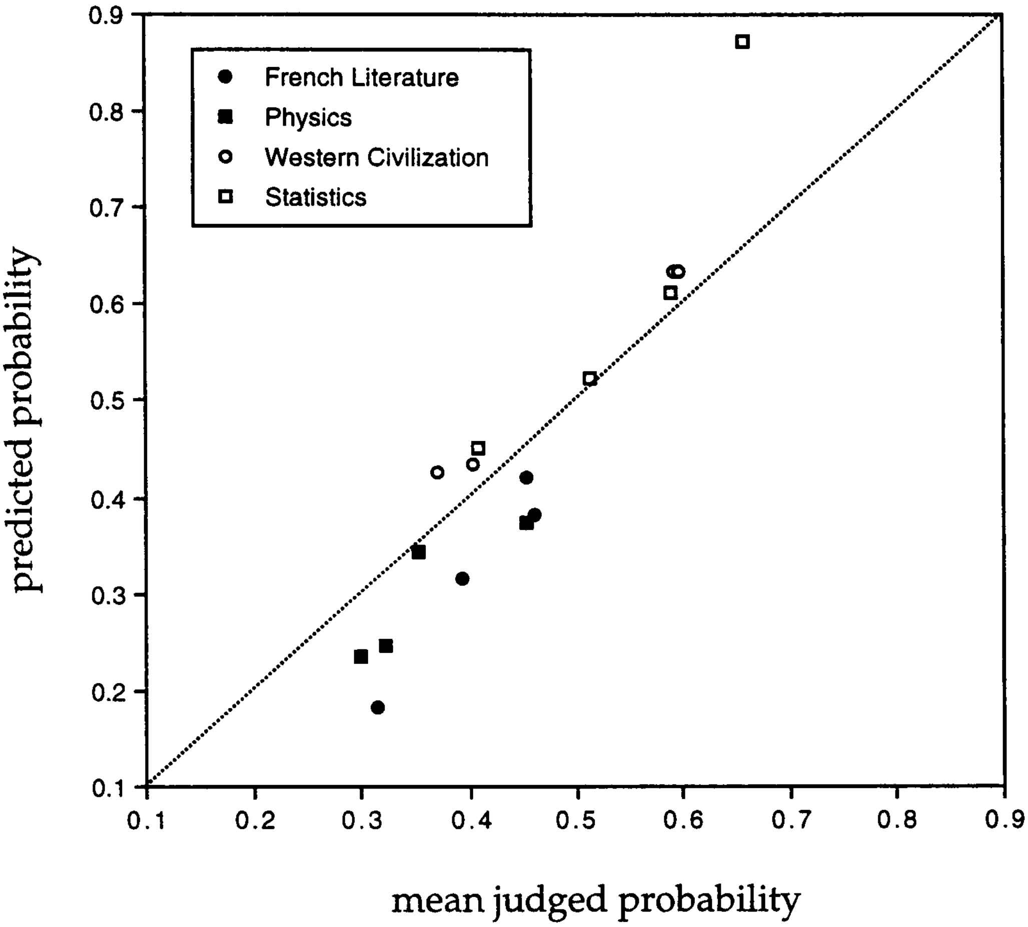

Exhibit 3. Predicted probability based on the linear discounting model plotted against mean judged probability for all course–major pairs in Experiment 1

One might ask how well the probability judgments can be predicted by assuming a constant value of w instead of the present discounting model. Exhibit 4 compares the goodness-of-fit achieved by the two models in the present experiment as well as in two subsequent experiments. The table indicates that (1) the null hypothesis $\beta = 0$ can be rejected and (2) the linear discounting model generally improves upon the fit achieved by the constant w model.

EXPERIMENT 2

Because the effect of unpacking is generally more pronounced in judgments of probability than in judgments of relative frequency (Tversky and Koehler, 1994), it is instructive to test whether the enhancement effect extends to estimates of relative frequency. To this end, both probability and (relative) frequency judgments were elicited using the same task as that of the previous experiment.

Exhibit 4. Comparison of fit to observed data (mean probability judgments) achieved by the linear discounting model and by a model assuming a constant value of w in Experiments 1, 4, and 5. Also listed for each experiment is the result of a test of the null hypothesis $\beta = 0$ implied by the constant w model

	Constant w model			Linear discounting model			Test of $\beta = 0$	
Exp.	Corr	Signed	Abs.dev.	Corr.	Signed	Abs.dev.	$t(14)$	Signif.
1	0.617	−0.015	0.072	0.939	−0.006	0.063	12.66	$p < 0.001$
4 imp	0.860	−0.009	0.047	0.925	−0.002	0.029	23.30	$p < 0.001$
4 exp	0.749	−0.011	0.046	0.816	−0.007	0.038	11.26	$p < 0.001$
5	0.587	−0.010	0.049	0.781	−0.010	0.065	0.84	n.s.

Note: Corr = correlation between predicted and observed values; Signed = average signed deviation between predicted and observed; Abs.dev. = average absolute deviation between predicted and observed. The implicit and explicit residual conditions of Experiment 4 are referred to as 4 imp and 4 exp, respectively.

Method

Subjects were Stanford University undergraduates ($N = 170$) who participated in exchange for course credit in their introductory psychology class. Subjects either estimated the probability that a student who had taken a particular course had a designated major or else estimated the proportion (i.e. relative frequency) of social science majors in the course who had that major. Only two courses were used: Statistics and French Literature. Subjects given the probability task considered two target students specified by first name and last initial. Subjects given the frequency task estimated the proportion of social science majors in a given course that had a specified social science major. Each subject gave four probability or frequency judgments (one for each of the four possible majors), two for the Statistics course and two for the French Literature course. Our goal was not to compare corresponding estimates in the two formulations directly, but rather to test whether the enhancement effect is eliminated when a frequentistic formulation is used.

Results

The mean elementary judgments of each major for the two courses appear in Exhibit 5, listed separately for the probability and frequency estimation tasks. In all cases the totals (summed over different groups of subjects) exceed 1, in accord with the predictions of support theory. The enhancement effect was replicated in this experiment: Judgments were generally higher for the Statistics course information than for the French Literature course information, paired $t(337) = 6.87$, $p < 0.001$.

As expected, the frequentistic formulation produced considerably lower values of T than the probabilistic formulation, although T was greater than 1 for both types of judgment. Most importantly, the enhancement effect was very much in evidence for the frequency judgments, $t(179) = 6.60$, $p < 0.001$, as well as for the probability judgments, $t(157) = 3.24$, $p < 0.005$.

EXPERIMENT 3

The results of the previous experiments suggest that subjects' judgments were influenced by the frequency with which social science majors take the course in question (i.e. the course's 'baserate' among social science majors). Normatively, the course baserate should not influence these judgments because they are already conditioned on the assumption that the student has taken the course.

It could be argued that the enhancement effect observed in these experiments is due in part to a confusion of the probability that someone in a given course has a specific major, p(major | course), and

Exhibit 5. Mean elementary judgments for each of the four possible majors and their total T in Experiment 2, listed by course information and type of judgment

	Type of judgment	
Course and major	Probability	Frequency
Statistics course:		
Economics	0.58	0.47
Political Science	0.40	0.38
Psychology	0.48	0.46
Sociology	0.41	0.30
Total (T)	1.87	1.61
French Literature course:		
Economics	0.24	0.19
Political Science	0.48	0.33
Psychology	0.29	0.25
Sociology	0.56	0.34
Total (T)	1.57	1.11

the probability that someone with a given major has taken a specific course, p(course | major). Experiment 3 was designed to test this possibility.

Method

Subjects ($N = 35$) were students at Stanford University who participated in exchange for credit in their introductory psychology course. Data from two additional subjects were not analyzed as these subjects failed to complete the questionnaire as instructed.

As in Experiment 1, each subject estimated all 16 possible combinations of course with major, but this time judged the probability of the course given the major. All subjects completed the same questionnaire, with instructions as follows:

> At a large, Midwestern state university, the social sciences consist of four different majors: Economics, Political Science, Psychology, or Sociology. For each major, you will be asked to estimate the percentage of students with that major who have taken a specified course sometime during their undergraduate career. Obviously, you won't know the exact figures, so you will need to base your judgments on what you know about the sorts of courses generally taken by students with different academic majors.

Judgments were blocked by major and presented alphabetically. Within each major the four courses were presented in a fixed order, again arranged alphabetically. The questions were presented in a frequentistic form (e.g. what percentage of Economics majors take a Western Civilization course as an undergraduate?). Subjects responded by circling a number between 0% and 100% on a scale, as in the previous experiment.

Results and discussion

The median judgments obtained from this study along with the probability judgments from Experiment 1 are presented in Exhibit 6, which plots separately for each course the median judgments associated with each of the four possible social science majors. There are three noteworthy features of this figure. First, focusing only on the judgments of p(course | major), we see that, for all four majors, higher probabilities were generally given for the typical courses than for the atypical courses. This

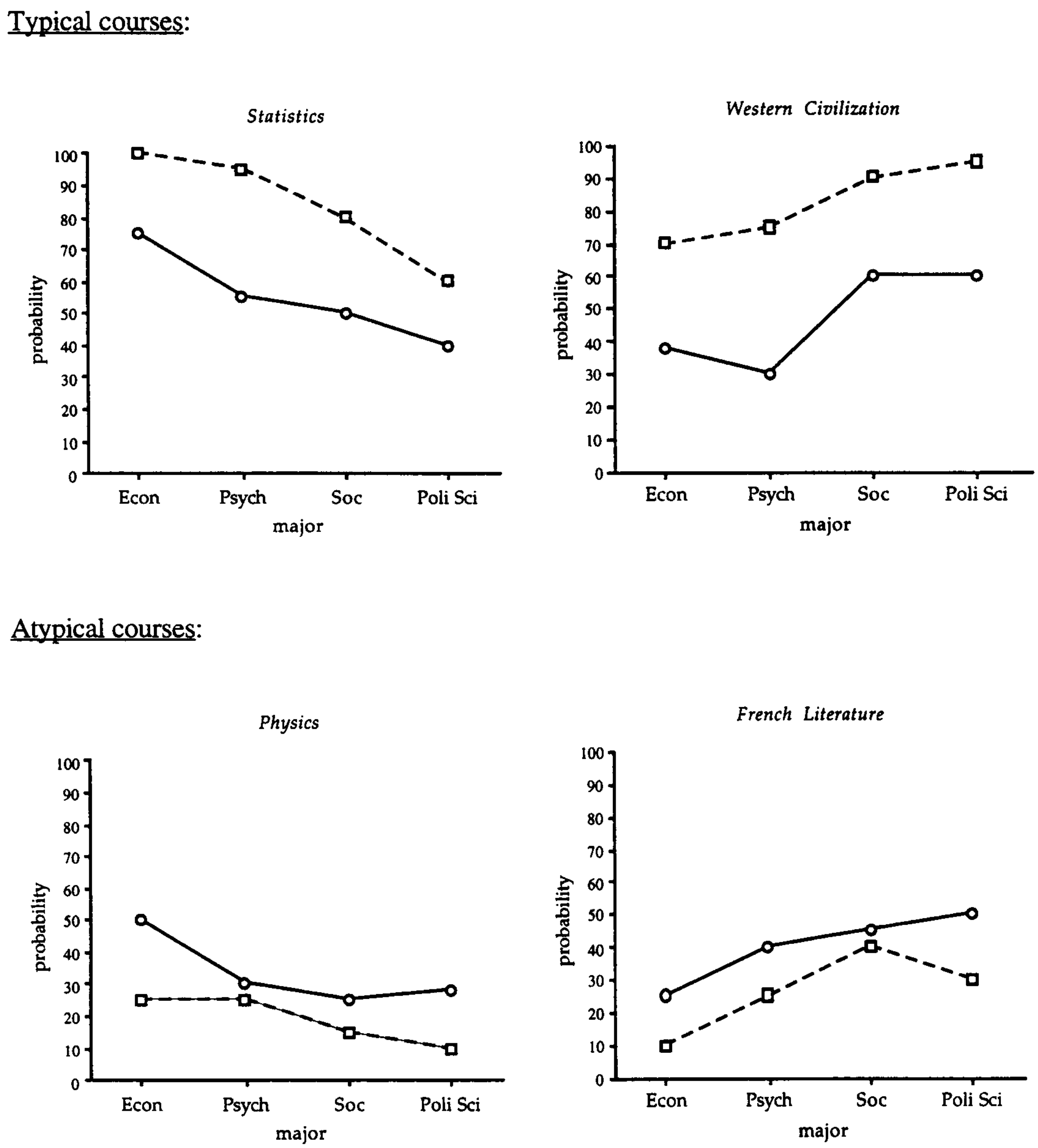

Exhibit 6. Median judgments of p(major | course) from Experiment 1, shown in solid lines, and of p(course | major) from Experiment 3, shown in dashed lines, plotted separately for each course.

observation is consistent with our assumption (which was also supported by the relatedness judgments of Experiment 1) that the Statistics and Western Civilization courses are seen as more typical of social science majors than are the French Literature and Physics courses. It also suggests that the average judgment of p(course | major) over majors can be used to predict the effect of enhancement on estimates of p(major | course). That is, in some cases at least, the inverse conditional probability may serve as a measure of 'compatibility' that can be used to compare the two bodies of evidence directly.

Second, Exhibit 6 shows that the ordering of majors is generally the same in judgments of p(major | course) and of p(course | major). Thus the support values for each major given a course obtained in the previous experiment can be predicted from the probabilities assigned to different courses given that the student has the major in question. These results suggest that the support for a major provided by a given course might be derived from the typicality or representativeness of the conjunction of the major and the course (e.g. that Greg H. majors in economics and takes a statistics course).

Third, and perhaps most importantly, Exhibit 6 clearly shows that subjects distinguish between the inverse conditional probabilities. If subjects fail to make the conceptual distinction between $p(A \mid B)$ and $p(B \mid A)$, a possibility that has been suggested by a number of investigators (e.g. Bar-Hillel, 1983; Dawes, 1988; Dawes *et al.*, 1993; Eddy, 1982), then there should be no systematic differences between the judgments obtained in Experiments 1 and 3. Instead, judgments of p(course | major) were without exception higher than judgments of p(major | course) for the typical courses (i.e. courses with high enrollment among social science majors), and lower than p(major | course) for the atypical courses.

According to probability theory, p(course | major)/p(major | course) = p(course)/p(major). If we assume, as seems reasonable in the present experiment, that the prior probabilities of the four majors are approximately equal, then the ratio p(course | major)/p(major | course) should vary as a function of course but not of major if subjects properly distinguish between the two conditional probabilities. The confusion hypothesis, in contrast, predicts that the ratio should be approximately one regardless of the prior probability of the course. We computed the ratio p(course | major)/(major | course) using the mean values from Experiments 1 and 3 for the 16 possible combinations of course and major, and then analyzed these data using a two-way analysis of variance. This analysis showed, as predicted by enhancement but not by the confusion hypothesis, that the ratio varied systematically as a function of course, $F(3, 9) = 13.8$, $p < 0.001$, but not as a function of major, $F(3, 9) < 1$. Evidently, the enhancement effect observed in Experiment 1 cannot be attributed merely to the confusion of inverse conditional probabilities.

It might be argued that this comparison does not constitute a fair test of the confusion hypothesis, however, because the judgments of p(major | course) collected in Experiment 1 were given a probabilistic formulation, while the p(course | major) ratings collected in the present experiment were given a frequentistic formulation. One way to address this problem is to compare the p(course | major) ratings with the frequentistic p(major | course) ratings collected in Experiment 2, so that the two types of judgment do not differ in how they were formulated. An analysis similar to that above of the ratio p(course | major)/p(major | course) using the mean values from Experiments 2 and 3 also revealed a main effect of course, $F(1, 3) = 24.7$, $p < 0.05$, but not of major, $F(3, 3) = 1.8$, n.s. Unless one is willing to endorse the unparsimonious assumption that the enhancement effects observed in probability and in frequency judgments arise from entirely different causes, this result renders unlikely the possibility that the enhancement effect is entirely attributable to a confusion of inverse probabilities. The present results do not show, of course, that every one of our subjects clearly distinguished between the two inverse probabilities or that such a confusion made no contribution at all to the enhancement effect observed in Experiments 1 and 2. What they do show is that the confusion hypothesis is not sufficient to account for the observed effects of enhancement.

EXPERIMENT 4

In Experiment 4 we test enhancement by adding rather than substituting evidence, that is, by examining how probability judgments change as new evidence is encountered. We compared the judgments assigned to different majors given a single course taken by the target student in his or her second year (as in Experiment 1) to the judgments given when additional information about the student was also provided. The additional information was selected to be generally compatible with all social science majors, and thus was expected to generally increase the support for each of the possible hypotheses. According to the present treatment, therefore, the introduction of this new evidence should produce the enhancement effect (i.e. greater values of T).

Method

Undergraduates at Stanford University ($N = 182$) participated in the experiment in exchange for course credit. As in the previous studies, subjects were told that all the target students attended a large state university in the midwest, and that each definitely had a social science major of Economics, Political Science, Psychology, or Sociology.

Each subject evaluated two students (Greg and Lisa) under low information and then the same two students under high information. In the low-information condition, subjects made their estimates on the basis of a single course the student had taken in his or her second year and the gender of the student in question. In the high-information condition, subjects were given additional information regarding jobs the student had held, other courses the student had taken, and issues of concern to the student. These pieces of information were combined in a way that avoided inconsistency but did not single out a specific major as especially likely. As an example, in one case subjects encountered the following description; the material in brackets constitutes the 'low-information' condition:

> [Greg H. majors in a social science. In his second year, he took a philosophy course on logic.] Greg has also taken a class on comparative government. He spent one summer working with mentally retarded children, and has also worked at an investment banking firm. He is very interested in socio-economic class differences in society.

In this example, each piece of additional evidence seems to support a different specific hypothesis (e.g. government class provides evidence for political science major, banking job provides evidence for economics major, etc.). We call this type of additional evidence 'Distinctive'.

For the other student, each piece of additional evidence supported all of the hypotheses under consideration:

> [Lisa J. majors in a social science. In her second year, she took a biology course on evolution.] Lisa has also taken a statistics class on data analysis, and a Western Civilization class. She subscribes to *Time* magazine, and likes to stay up to date on current happenings in the news.

We call this type of additional evidence 'Common'. Each subject saw Common evidence for one student and Distinctive evidence for the other.

Probability judgments

Each subject evaluated two different target majors for each of the two students considered in a given information condition. Different groups of subjects evaluated different majors within each case so that, over subjects, we obtained estimates of the probability of all four majors for each case. Subjects circled their judgments on the same scale used in the previous experiments.

The explicitness of the residual hypothesis was also varied. Subjects in the implicit residual condition ($n = 151$) rated, for instance, 'the probability that Greg majors in Economics rather than *another social science*'. Subjects in the explicit residual condition ($n = 31$) rated 'the probability that Greg majors in Economics rather than *Political Science, Psychology, or Sociology*'.

Support judgments

A separate group of 48 paid Stanford subjects and 34 San Jose State introductory psychology students judged the extent to which each student description 'supported' the conclusion that the student had a particular major. Subjects made judgments of the support for the four majors for Greg under low and high-common information and for Lisa under low and high-distinctive information (a total of 16 judgments).

Results and discussion

Probability judgments

The mean judged probabilities for the different majors, residual descriptions, students, and levels of information are displayed in Exhibit 7. Analysis of variance revealed main effects of Major, $F(3, 1397) = 8.57$, $p < 0.001$, Student (Greg versus Lisa), $F(1, 1397) = 3.77$, $p < 0.06$, and Information condition, $F(2, 1397) = 9.92$, $p < 0.001$, as well as a significant Major by Information interaction, $F(6, 1397) = 12.70$, $p < 0.001$. The results can be summarized as follows: (1) the totals (T) of the mean judged probabilities across the four majors are uniformly greater than 1; (2) T is greater for high information than for low information, $t(1397) = -4.3$, $p < 0.001$; (3) enhancement is observed under both Common and Distinctive added evidence (the difference between the two conditions is non-significant, $t(1397) = -1.1$); and (4) enhancement is observed for both implicitly and explicitly framed residual hypotheses, and is more pronounced in the former than in the latter case, $F(1, 1397) = 10.41$, $p < 0.001$.

Estimation of discounting model

The support ratings, combined with the mean probability judgments from Exhibit 7, were used to relate $w_{\bar{A}}$ to $s(A)$, separately for the implicit and explicit residual conditions. The estimated values of β

Exhibit 7. Mean elementary probability judgments for each major and their total T assigned for each target student under low, high-common, and high-distinctive information conditions, listed separately for the implicit and explicit residual conditions of Experiment 4

Target/information	Econ.	Poli. Sci.	Psych.	Soc.	total (T)
Implicit residual					
Greg: low	0.32	0.38	0.40	0.38	1.48
Greg: high-common	0.43	0.55	0.25	0.45	1.68
Greg: high-distinctive	0.52	0.51	0.35	0.50	1.88
Lisa: low	0.22	0.25	0.50	0.41	1.38
Lisa: high-common	0.45	0.49	0.30	0.39	1.63
Lisa: high-distinctive	0.43	0.41	0.36	0.50	1.70
Explicit residual					
Greg: low	0.31	0.35	0.36	0.32	1.34
Greg: high-common	0.48	0.47	0.20	0.44	1.59
Greg: high-distinctive	0.32	0.34	0.31	0.56	1.53
Lisa: low	0.23	0.24	0.41	0.39	1.27
Lisa: high-common	0.38	0.34	0.23	0.40	1.35
Lisa: high-distinctive	0.41	0.40	0.24	0.43	1.48

Note: Econ. = economics; Poli.Sci. = political science; Psych. = psychology; Soc. = sociology.

were 0.096 and 0.077 for the implicit and explicit residual conditions, respectively. Exhibit 8 displays the predicted and observed probabilities based on the fitted linear discounting model, and shows that the model provides a very good fit to the data. Exhibit 4 indicates that, as in the first experiment, the linear discounting model outperformed a model assuming a constant value of w.

Experiment 3 showed that enhancement is not attributable to a confusion of inverse conditional probabilities. The current results show further that the ordering of the inverse conditional probabilities, that is, the probability of the information about the student given a specific major, does not always predict the effect of enhancement. In contrast to Experiments 1 and 2, in Experiment 4 the evidence actually became (objectively) less likely given any specific major under the high-information condition, because the evidence in this condition was a conjunction of the low-information evidence with further information about the student's preferences and values. Even though the inverse conditional probabilities were necessarily lower under the high-information condition than under the low-information condition, the additional information nonetheless induced enhancement (i.e. greater values of T). Such results suggest that the psychological concept of support cannot be equated with the probability of the evidence given the hypothesis.

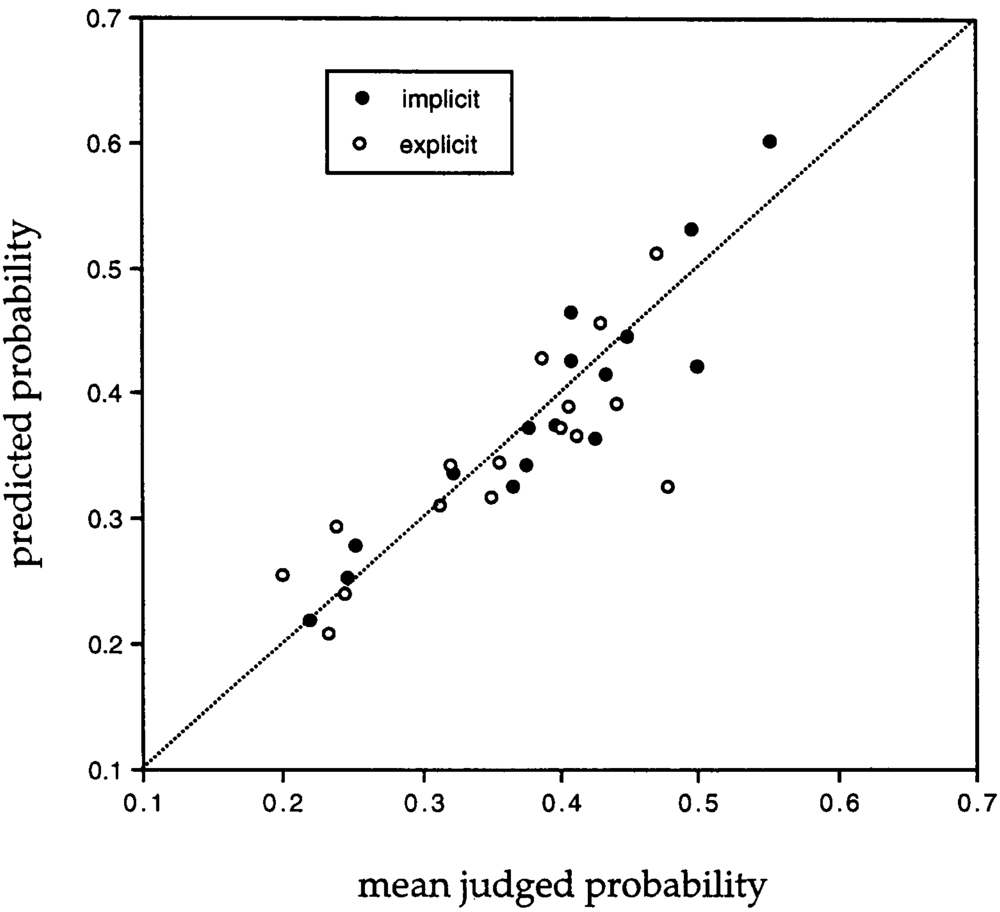

Exhibit 8. Predicted probability based on the linear discounting model plotted against mean judged probability for the implicit and explicit residual conditions of Experiment 4.

EXPERIMENT 5

Subjects read two crime stories and were asked to make judgments about the suspects involved. As in the previous experiment, subjects gave their judgments twice: once under low-information conditions and again after further evidence was revealed. One group of subjects was asked to assign probabilities that various suspects were guilty of the crime in question. With these data we are able to test enhancement in a domain quite different from that of the college majors. This task is similar to that used previously by Teigen (1983) and Robinson and Hastie (1985; see also Van Wallendael and Hastie, 1990), except that in the present experiment a given subject evaluated only a subset of the suspects to avoid any explicit prompt to give additive judgments. Another group of subjects was asked to rate the suspiciousness of the suspects rather than their probability of guilt. We used these ratings as a direct measure of support.

Method

The subjects from Experiment 1 ($N = 115$) also participated in Experiment 5. Again, there were no systematic differences among subjects obtained from the two different recruitment procedures, so this variable is not considered further. The data from one additional subject were dropped as that subject failed to complete the questionnaire appropriately.

Two crime stories were constructed, one involving embezzlement (Case 1) and the other a murder (Case 2). All subjects were presented first with the embezzlement case and then with the murder case. Each case involved four suspects, exactly one of which was guilty. In the low-information condition the four suspects were introduced with a short description of their relationship to the case and possible motive. Tentative alibis were also provided in the murder case but not in the embezzlement case, where alibis seem less relevant. In the high-information condition the motive of each suspect was enhanced, and in the murder case the alibis were sometimes brought into question. In general we constructed the cases so that certain suspects appeared somewhat more guilty than others, and that all suspects generally seemed more suspicious as more evidence was revealed. This is characteristic of mystery novels, in which all suspects become increasingly suspicious as the story unfolds. In fact, the murder case was adapted from P. D. James's *Devices and Desires*.

Probability judgments

Subjects were assigned either to the probability judgment task or to the suspiciousness rating task, and gave ratings after reading the low-information material and again after reading the high-information material. Subjects in the probability judgment task ($n = 60$) made judgments for two of the four suspects. The same suspects were assessed by a given subject under low- and high-information conditions.

Suspiciousness ratings

Subjects in the suspiciousness rating task ($n = 55$) rated two suspects for each case, both under low- and high-information conditions. Target suspects were paired in the same way as in the probability judgments. Subjects were instructed to make their ratings by giving a number between 0 (indicating that the suspect is 'not at all suspicious') and 100 (indicating that the suspect is 'maximally suspicious'). Subjects were told:

> Your ratings of a given suspect should reflect his or her relative suspiciousness in proportion to the other suspects. So, for instance, if you feel that one suspect is twice as suspicious as the other, that

suspect should be given a suspiciousness rating that is twice that of the other suspect. As another example, assigning a suspiciousness rating of 60 to Suspect A and a rating of 20 to Suspect B indicates your belief that Suspect A is 3 times as suspicious as Suspect B.

These instructions were intended to encourage subjects to give their ratings using a ratio scale.

Although subjects in both groups made evaluations for only two of the four suspects, the information regarding all four suspects was presented to all subjects.

Results and discussion

Suspiciousness ratings

The mean suspiciousness ratings and probability judgments are listed in Exhibit 9 by case and information level. As intended, the additional information provided under the high-information condition made all of the suspects seem more suspicious, paired $t(239) = 5.64$, $p < 0.001$. (The one exception is the artist in Case 2, who was given an alibi in the high-information condition and, as would be expected, was consequently rated as less suspicious and less likely to be guilty.)

Probability judgments

While it is perfectly acceptable for all suspects to be rated as more suspicious under the high-information condition, probability theory requires that the total probability T equal 1 even if all of the suspects appear more suspicious. Nonetheless, as predicted by the present treatment of enhancement, judged probability did in fact increase for all suspects (except the artist) under the high-information conditions, paired $t(211) = 8.53$, $p < 0.001$. As shown in Exhibit 9, the probability judgments in both cases added to more than 1, and the sum increased considerably as more evidence was provided.

Estimation of discounting model

As in the previous experiments, the discounting model (3) was used to relate the support estimates to the rate of discounting of the residual w. The estimated value of the single parameter β was 1.20. The fit

Exhibit 9. Mean suspiciousness rating and judged probability of each suspect under low- and high-information conditions for the two criminal cases of Experiment 5

	Suspiciousness		Probability	
Suspect	Low info	High info	Low info	High info
Case 1: Embezzlement				
Accountant	0.41	0.53	0.40	0.45
Buyer	0.50	0.58	0.42	0.48
Manager	0.47	0.51	0.48	0.59
Seller	0.32	0.48	0.37	0.42
Total	1.70	2.10	1.67	1.94
Case 2: Murder				
Activist	0.32	0.57	0.39	0.57
Artist	0.27	0.23	0.37	0.30
Scientist	0.24	0.43	0.34	0.40
Writer	0.38	0.60	0.33	0.54
Total	1.22	1.84	1.43	1.81

of the model is indicated in Exhibit 4. As in Experiments 1 and 4, the correlation between predicted and observed probability judgments was higher for the linear discounting model than for the constant w model. It should be noted, however, that in terms of absolute and signed deviation the linear discounting model did not perform better than the constant w model in this case.

GENERAL DISCUSSION

This paper reports several studies of the enhancement effect, in which the sum of judged probability for a set of mutually exclusive and exhaustive hypotheses is increased by evidence that is compatible with these hypotheses. We have demonstrated the effect using different manipulations (substituting versus adding evidence) and different domains (college majors and crime stories). In these studies the effect was of substantial magnitude, increasing T on average by approximately 20%. In this final section we review some related findings from other studies and offer some speculations regarding the causes of enhancement.

As mentioned above, Robinson and Hastie (1985; also Van Wallendael and Hastie, 1990) investigated probability judgments using crime stories, as we did in Experiment 5 (see also Teigen, 1983). Subjects were presented with some background information regarding the case and then received a series of clues, giving revised probability judgments after each. In contrast to our experiment, each subject gave probability judgments for all the suspects. Despite the possibility that this method would prompt subjects to make additive judgments, the sum of the judgments were consistently greater than one. Consistent with our treatment of enhancement, Robinson and Hastie (1985) found that subjects tended only to adjust the probability assigned to the suspect directly implicated by the clue and not any of the remaining suspects, with the result that the presentation of clues suggesting the guilt of a suspect increased the total probability assigned to the set of suspects. In their Experiment 3, for example, the number of clues implying guilt was varied between subjects, and the resulting total probabilities (estimated from Robinson and Hastie, 1985, Figure 5) were approximately 1.5, 2.5, and 4.3 given 0, 2, and 4 guilty clues, respectively. As would be expected, the introduction of clues implying innocence (which presumably decreased the compatibility of the evidence with the set of hypotheses) had the opposite effect.

Peterson and Pitz (1988, Experiment 3) asked subjects to assess the probability that the number of games won in a season by a particular baseball team fell in a given interval, based on either one, two, or three cues (batting average, earned-run average, and total home runs). Subjects considered a number of different teams and, without their knowledge, assigned probabilities to all three components of a partition (e.g. less than 80 wins, between 80 and 88 wins, more than 88 wins) for each team. Consistent with enhancement, subjects assigned a greater total probability to the components of the partition as the number of cues increased, with totals of 1.26, 1.61, and 1.86 for one, two, and three cues, respectively. Increasing the number of cues only increased the total probability, however, when the different cues had conflicting or mixed implications for the outcome variable (e.g. high batting average, suggesting a winning season, combined with a high earned-run average, suggesting a losing season). As implied by the current treatment, adding evidence yields enhancement only when the additional information increases the compatibility of the evidence with each in the set of possibilities.

A recent study (Koehler, 1995) has tested enhancement using a classification learning task, in which subjects were presented with fictitious 'patients' and asked to diagnose which of three influenza (flu) strains each is suffering from on the basis of a set of four symptoms. After a series of 240 training trials during which subjects learned how the symptoms and flu strains are related, they were asked to judge the probability that a patient with a given pattern of symptoms had a designated flu strain. Each subject assessed the probability of all three flu strains given each of 16 possible symptom patterns. As expected, these judgments added to more than 100%. As predicted by our discussion of enhancement, and consistent with the results of Peterson and Pitz (1988), T increased with the conflict among the symptoms in the symptom pattern: Subjects gave greater total probabilities for those patterns implying two or more different flu strains ($M = 1.28$) than for those implying only a single flu strain or no particular strain ($M = 1.07$).

Taken together, the findings from all these experiments suggest that a wide variety of manipulations and experimental contexts give rise to enhancement. The manipulations of evidence producing enhancement appear to fall into one of two general kinds: increasing the prevalence or baserate of the evidence given the set of hypotheses under consideration, and increasing the number of hypotheses implicated by or consistent with the evidence. What both manipulations do, we suggest, is increase the extent to which the evidence supports each of the hypotheses under consideration. Other manipulations with this consequence would also be expected to produce enhancement.

When a judge evaluates the probability of a focal hypothesis against its alternatives, it is likely that the evidence will be assessed in terms of its implications for the focal hypothesis relative to a composite representation of the residual. Thus, rather than first evaluating the implications of the evidence for each elementary hypothesis separately and then aggregating the support of those hypotheses included in the residual, it appears that people first aggregate the hypotheses in the residual and then evaluate how well the resulting composite is supported by the evidence. Put this way, it seems that the support offered by a body of evidence is used or absorbed more efficiently by a hypothesis when it is represented explicitly than when it is included implicitly in the residual. From this view, the task of identifying the psychological mechanisms underlying enhancement becomes one of identifying how the composite residual hypothesis is created and evaluated.

We conclude by offering a few speculations regarding this process. In creating a composite representation of the residual, certain common features (e.g. those shared by all elementary hypotheses included in the residual) will presumably become features of the composite. Other features of the component elementary hypotheses will probably not be incorporated into the residual, which could contribute to the discounting of the residual hypothesis. In a criminal case, for example, the feature 'male' might be incorporated into the residual representation if all or most of the alternative suspects to the focal suspect are male. A feature such as 'left-handed', in contrast, is unlikely to be incorporated into the residual if it characterizes only a single member of the residual.

The evaluation of evidence is likely to be influenced by the formation of a composite representation of the residual. As an illustration, suppose that a designated piece of evidence supports any hypothesis with a specific feature. In the case of a store burglary, for example, evidence that the burglar had a key to the building would support the conjecture that the crime was committed by an employee rather than an outsider. If the feature (in this case, 'employee') is shared by the focal hypothesis and the residual, the evidence may be counted as equally supportive of both. If the feature in question is actually a characteristic of each specific hypothesis included in the residual (i.e. all the suspects are employees), however, then the evidence should normatively favor the residual over the focal hypothesis by a factor corresponding to the number of elementary hypotheses (suspects) included implicitly in the residual. Thus the focal hypothesis will be favored by a support assessment process that compares features of the focal hypothesis to those of the composite rather than to those of the component elementary hypotheses included implicitly in the residual.

Finally, the focal hypothesis may determine which aspects of the evidence are used to assess support in the first place. Evidence that supports an elementary hypothesis included in the residual may not be considered relevant (i.e. may be treated as nondiagnostic) if the critical feature of the elementary hypothesis has not been incorporated into the residual's composite representation. Evidence that the criminal must have been capable of moving some heavy items, for example, may not be treated as diagnostic if the relevant feature (e.g. that one of the alternative suspects has a bad back) is not included in the residual representation. Such biases in the evaluation of evidence may contribute to the enhancement effect, in which the addition of compatible evidence adds disproportionately to the support of the focal hypothesis.

ACKNOWLEDGEMENTS

Support for this research was provided by Grant MH 53046 from the National Institutes of Health to Amos Tversky. During the course of this research, Derek Koehler was supported by a National Defense Science and Engineering Graduate Fellowship and by the National Science Foundation's Program for Long- and Medium-term Research at Foreign Centers of Excellence; Lyle Brenner was supported by a National Science Foundation Graduate Fellowship. Correspondence concerning this article should be addressed to Derek J. Koehler, who is now at the Department of Psychology, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

REFERENCES

Bar-Hillel, M., 'The base-rate fallacy controversy', in Scholz, R. W. (ed.), *Decision Making Under Uncertainty* (pp. 39–61), Amsterdam: North-Holland, 1984.

Cohen, L. J., *The Probable and the Provable*, Oxford: Oxford University Press, 1977.

Cohen, L. J., 'On the psychology of prediction: Whose is the fallacy?' *Cognition*, **7** (1979), 385–407.

Dawes, R. M., *Rational Choice in an Uncertain World*. San Diego: Harcourt Brace Jovanovich, 1988.

Dawes, R. M., Mirels, H. L., Gold, E. and Donahue, E., 'Equating inverse probabilities in implicit personality judgments', *Psychological Science*, **4** (1993), 396–400.

Eddy, D. M., 'Probabilistic reasoning in clinical medicine: Problems and opportunities', in Kahneman, D., Slovic, P. and Tversky, A. (eds), *Judgment under Uncertainty: Heuristics and biases* (pp. 249–267), Cambridge: Cambridge University Press, 1982.

Fischhoff, B., Slovic, P. and Lichtenstein, S. 'Fault trees: Sensitivity of estimated failure probabilities to problem representation', *Journal of Experimental Psychology: Human Perception and Performance*, **4** (1978), 330–44.

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

Kahneman, D., Slovic, P. and Tversky, A. (eds), *Judgment under Uncertainty: Heuristics and biases*, Cambridge: Cambridge University Press, 1982.

Koehler, D. J., 'Probability judgment in three-category classification learning', unpublished manuscript, University of Waterloo, 1995.

Mehle, T., Gettys, C. F., Manning. C., Baca, S. and Fisher, S. 'The availability explanation of excessive plausibility assessment', *Acta Psychologica*, **49** (1981), 127–40.

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

Redelmeier, D. A., Koehler, D. J., Liberman, V. and Tversky, A. 'Probability judgment in medicine: Discounting unspecified causes', *Medical Decision Making*, **15** (1995), 227–30.

Robinson, L. B. and Hastie, R., 'Revision of beliefs when a hypothesis is eliminated from consideration', *Journal of Experimental Psychology: Human Perception and Performance*, **4** (1985), 443–56.

Rottenstreich, Y. S. and Tversky, A., 'Unpacking, repacking, and anchoring: Advances in support theory', *Psychological Review*, in press.

Shafer, G., *A Mathematical Theory of Evidence*, Princeton, NJ: Princeton University Press, 1976.

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

Tversky, A. and Koehler, D. J., 'Support theory: A nonextensional representation of subjective probability', *Psychological Review*, **101** (1994), 547–67.

Van Wallendael, L. R. and Hastie, R., 'Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing', *Memory & Cognition*, **18** (1990), 240–50.

Authors' biographies:
Derek Koehler received his PhD in psychology from Stanford University in 1993, and is now assistant professor in the Department of Psychology at the University of Waterloo.

Lyle Brenner received his PhD in psychology from Stanford University in 1995, and is now assistant professor at The Anderson School of Management at the University of California, Los Angeles.

Amos Tversky received his PhD in psychology from the University of Michigan in 1965, and until his death in June 1996 held the Davis Brack Chair of Behavioral Sciences at Stanford University.

Authors' addresses:
Derek Koehler, Department of Psychology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

Lyle Brenner, The Anderson School of Management, Box 951481, University of California, Los Angeles, CA, 90095-1481, USA.

