Journal of Experimental Psychology:
Learning, Memory, and Cognition
1998, Vol. 24, No. 3, 771–792

0278-7393/98/$3.00

Taking Into Account the Strength of an Alternative Hypothesis

Craig R. M. McKenzie
University of California, San Diego

A common phenomenon in judgment under uncertainty is that alternative hypotheses are underweighted or ignored. This article addresses when and how the strength of the alternative is taken into account when there are 2 hypotheses. A learning manipulation was used to invoke 2 representations of 2 illnesses in a medical diagnosis task. One representation tended to lead to consideration of the alternative when, for example, requesting new information, reporting confidence, and making diagnoses. The other representation tended to result in ignoring or underweighting the alternative, but a simple change in how confidence was probed increased consideration of the alternative. Costs and benefits of each representation are discussed.

Implicitly or explicitly, people are constantly assessing the likelihood of hypotheses, or unknown states of the world. Examples include estimating how likely it is that your favorite team will win its next game, that it will rain tomorrow, that a defendant is guilty, or that a new acquaintance is thoughtful. Every hypothesis has at least one alternative. For instance, when assessing your team's chances of winning, the alternative hypothesis is that the opposing team will win. If the competing hypotheses are mutually exclusive and exhaustive (i.e., exactly one of the hypotheses is true), confidence in the truth of one hypothesis should be influenced by the relative strength(s) of the other(s). Thus, when asked about the likelihood of your team winning, you should consider the quality of both teams. Ignoring the strength of the opposing team is potentially problematic because the stronger you think the opposing team is, the lower your confidence should be in your team.

This article is concerned with when and how the subjective strength of an alternative hypothesis is taken into account in judgment under uncertainty. Much evidence across several domains indicates that participants often underweight or ignore alternatives. In addition to being found in traditional hypothesis testing tasks (e.g., Beyth-Marom & Fischhoff, 1983; Doherty, Mynatt, Tweney, & Schiavo, 1979; Klayman, 1988; Schustack & Sternberg, 1981; Skov & Sherman, 1986; Snyder & Swann, 1978), underweighting alternatives has been used as an explanation of overconfidence (Arkes, Christensen, Lai, & Blumer, 1987; Koriat, Lichtenstein, & Fischhoff, 1980; Ronis & Yates, 1987; Sniezek, Paese, & Switzer, 1990), hindsight bias (Hawkins & Hastie, 1990; Slovic & Fischhoff, 1977), directional errors in belief updating (Hogarth & Einhorn, 1992; Lopes, 1987), and nonadditivity of subjective probability (Robinson & Hastie, 1985; Teigen, 1983; Van Wallendael, 1989; Van Wallendael & Hastie, 1990). In general, participants often consider the relation between data and the focal hypothesis without regard to the relation between data and the alternative or alternatives (for overviews, see Evans, 1989; Fischhoff & Beyth-Marom, 1983; Klayman & Ha, 1987).

In addition to its broad theoretical relevance, studying when people consider alternatives has implications for improving judgment. McKenzie (1994) examined through computer simulation the accuracy of intuitive strategies for covariation and Bayesian inference tasks. Although strategies that underweighted or ignored the alternative performed surprisingly well in general, their accuracy decreased dramatically under certain conditions. However, strategies that took into account the strength of the alternative performed well across conditions—even though these strategies were nonnormative. Furthermore, Koehler (1994) showed that confidence judgments were superior across several dimensions (e.g., less overconfidence and better calibration) for participants who appeared to take into account alternatives rather than ignore them (see also Koriat et al., 1980).

What, then, determines whether people consider alternatives? Successful interventions include presenting likelihoods of data under both the focal and alternative hypotheses (Bassok & Trope, 1984; Trope & Bassok, 1982, 1983; Trope, Bassok, & Alon, 1983; Trope & Mackie, 1987), asking questions that mention both hypotheses rather than just one (Baron, Beattie, & Hershey, 1988; Beyth-Marom & Fischhoff, 1983; Zuckerman, Knee, Hodgins, & Miyake, 1995), having participants generate the hypothesis set themselves (Koehler, 1994), having participants generate reasons

This article is based in part on my doctoral dissertation submitted to the University of Chicago's Department of Psychology. For their guidance, I thank my committee members, Robin Hogarth, Joshua Klayman, William Goldstein, and Lawrence Barsalou. I also gratefully acknowledge helpful comments from Michael Doherty, Kevin Dunbar, Jonathan Evans, Brian Gibbs, Gerd Gigerenzer, Jonathan Koehler, Rick Larrick, Richard Mayer, Lori Van Wallendael, John Wixted, and Mike Ziolkowski. Financial support was provided by National Science Foundation Grant SBR-9515030, by the University of Chicago's Department of Psychology Graduate Student Research Fund, and by Robin Hogarth. Portions of Experiment 1 were presented at the 1993 meeting of the Society for Judgment and Decision Making and at the 1994 meeting of the Midwestern Psychological Association. Portions of Experiments 3 and 4 were presented at the 1996 meeting of the Society for Judgment and Decision Making.

Correspondence concerning this article should be addressed to Craig R. M. McKenzie, Department of Psychology, University of California, San Diego, La Jolla, California 92093-0109. Electronic mail may be sent to cmckenzie@ucsd.edu.

why the alternative might be true (Koriat et al., 1980; Slovic & Fischhoff, 1977), and telling participants directly to consider the alternative (Lord, Lepper, & Preston, 1984). What these examples have in common is that the alternative is made salient to participants either during or immediately preceding the task of interest.

The present research takes a different approach, arguing that how confidence in competing hypotheses is cognitively represented is important for understanding both when and how alternatives are taken into account, even when participants are given the same instructions, asked the same questions, and presented with the same data. The first part of the article explains the theoretical motivation behind the research. Results from four medical decision-making experiments are subsequently reported. The final section discusses implications and limitations of the research.

Theoretical Motivation

Representation of Confidence

To illustrate the theoretical ideas (and the experimental paradigm), imagine a physician diagnosing which of two (fictitious) illnesses, puneria and zymosis, a patient suffers from. For example, it might be known that the patient has a particular class of illness, of which there are just these two types. The diagnosis is based on the presence and absence of symptoms and, because the symptoms are fallible cues, any diagnosis is uncertain; depending on the patient's symptoms, the physician will be more or less confident that the patient has puneria (or zymosis). Confidence in two hypotheses can be represented in two ways (see also Van Wallendael & Hastie, 1990). Dependent confidence is illustrated in Figure 1a, where H1 and H2 correspond to the two hypotheses. Changes in confidence in one hypothesis necessarily result in complementary changes in the other. The two poles of the scale might correspond to *The patient has puneria* and *The patient has zymosis*. Any evidence that, say, increases confidence in puneria must decrease confidence in zymosis as a result of the dependent relationship. In contrast, one might have independent confidence, as in Figure 1b, where each hypothesis can be thought of as having a separate scale. One scale might correspond to confidence in puneria, and the other to zymosis. Here, changes in confidence in one hypothesis do not necessitate changes in the other.

The distinction between dependent and independent confidence has implications for when and how people take into account the strength of the alternative. Dependent confidence implies that the alternative is always taken into account. For example, two successive changes in confidence, one increase in H1 and one (equally strong) increase in H2, would offset each other. In contrast, independent confidence implies that taking into account the alternative requires additional processing: Confidence in the alternative in addition to the focal hypothesis has to be accessed and accommodated in some fashion. For example, the above two changes in confidence would lead to increases on both independent scales, and only through a comparison of relative strengths would the changes offset each other. Ignoring the alternative would lead, in this case, to relatively high confidence in the focal hypothesis. Because the extra processing might not take place, taking into account the strength of an alternative hypothesis is less likely with independent confidence.

Exactly how the strength of the alternative is taken into account with independent confidence is not addressed in this article. The goals are to show that (a) the distinction between dependent and independent confidence in two hypotheses is psychologically real, (b) the distinction has implications for a variety of behaviors, (c) extra processing is necessary for taking into account the alternative when confidence is independent rather than dependent, and (d) an independent representation can result in behavior more similar to that of a dependent representation if consideration of the alternative is encouraged.

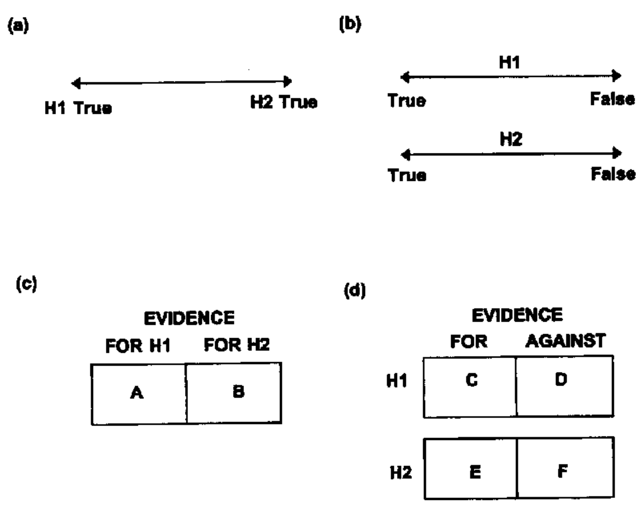

Figure 1. (a) Dependent and (b) independent confidence in two hypotheses (H1 and H2); (c) two- and (d) four-category conceptions of evidence.

Conception of Evidence

Just as there are two ways to represent confidence, there are two ways to conceive of evidence in two hypotheses. Figure 1c shows two categories of evidence: evidence for H1 (Cell A) and evidence for H2 (Cell B). In Figure 1d, however, there are four categories. Evidence for H1 in Figure 1c is broken down in Figure 1d into evidence for H1 (Cell C) and evidence against H2 (Cell F). Similarly, evidence for H2 is broken down into evidence against H1 (Cell D) and evidence for H2 (Cell E). It is important to note that the distinction in Figure 1d between evidence for one hypothesis and evidence against the other disappears in Figure 1c.

It is proposed that dependent confidence in two hypotheses results from conceiving of just two categories of evidence. Cell A evidence in Figure 1c results in movement to the left on the dependent confidence scale, and Cell B evidence results in movement to the right. Independent confidence, however, results from conceiving of four categories of evidence. Cell C evidence leads to movement to the left on the H1 confidence scale, and Cell D evidence leads to movement to the right. Similarly, Cell E evidence leads to movement to the left on the H2 scale, and Cell F evidence leads to movement to the right.

Learning

If representation of confidence affects when and how alternatives are taken into account, and if representation is affected by the conception of evidence, what, then, affects this conception? The current approach regards how people learn about the hypotheses.

Previous to diagnosing which illness the patient has, the physician might have learned to distinguish between the illnesses. Such *contrastive learning* (see also Bransford, Franks, Vye, & Sherwood, 1989; Goldstone, 1996; Klayman & Brown, 1993) is illustrated in Table 1. (The experiments reported here instantiate learning differently, but Tables 1 and 2 illustrate the present points.) The first column in Table 1 lists four symptoms; the second and third columns indicate the proportion of patients who exhibit each symptom, given each illness. Contrasting the two illnesses makes the diagnosticity of each symptom obvious. It is clear that dizziness and rash are diagnostic because dizziness is common to puneria, but not zymosis, and rash is common to zymosis, but not puneria. It is also clear that coughing and fever are nondiagnostic because the former is common to both illnesses and the latter is common to neither.

Alternatively, the physician might have learned about the two illnesses independently, called *noncontrastive learning*. The top of Table 2 illustrates learning about puneria, where patients with and without puneria are contrasted. (One can think of patients without the illness as healthy; they tend to have none of the symptoms.) Dizziness and coughing are learned to be diagnostic of puneria, but rash and fever are not. The bottom of Table 2 illustrates learning about zymosis, where patients with and without zymosis are contrasted. Rash and coughing are learned to be diagnostic of zymosis, but dizziness and fever are not.

With contrastive learning, there is no distinction between evidence for one hypothesis and evidence against the other; learning to discriminate between two hypotheses blurs this distinction and leads to conceiving of just two categories of evidence. This, in turn, leads to dependent confidence. However, noncontrastive learning leads to conceiving of four categories of evidence, resulting in independent confidence. For example, contrastive learning might lead to viewing dizziness as evidence simultaneously for puneria and against zymosis, and viewing rash as evidence against puneria and for zymosis. However, noncontrastive learning might lead to viewing dizziness as evidence for puneria but not against zymosis, and viewing rash as evidence for zymosis but not against puneria. Furthermore, noncontrastive learning might result in viewing coughing as evidence for both illnesses.

Assume that learning builds associations between a given hypothesis and what is conceived to be evidence for and against that hypothesis. Conceiving of two categories of evidence implies that evidence for and against each hypothesis is associated with both hypotheses: Pro and con information for the two hypotheses is inextricable. Thus, the same evidence will be deemed relevant regardless of which hypothesis is focal. However, conceiving of four categories implies that different pro and con evidence can be associated with each hypothesis; different evidence might be deemed relevant depending on which hypothesis is focal.

Table 1
Illustration of Contrastive Learning of Puneria and Zymosis

Symptom	*p* (symptom| puneria)	*p* (symptom| zymosis)	Symptom type
Dizziness	.85	.15	Puneria
Rash	.15	.85	Zymosis
Coughing	.85	.85	Both
Fever	.15	.15	Neither

Table 2
Illustration of Noncontrastive Learning of Puneria and Zymosis

	Puneria	
Symptom	*p* (symptom| puneria)	*p* (symptom| no puneria)
Dizziness	.85	.15
Rash	.15	.15
Coughing	.85	.15
Fever	.15	.15
	Zymosis	
Symptom	*p* (symptom| zymosis)	*p* (symptom| no zymosis)
Dizziness	.15	.15
Rash	.85	.15
Coughing	.85	.15
Fever	.15	.15

Related Research

Others have also examined how different ways of learning concepts influence subsequent behavior. Klayman and Brown (1993) taught participants about two illnesses by presenting them with labeled profiles of each either in parallel or separately. It was hypothesized that the former group's illness concepts would consist of diagnostic symptoms, whereas the latter group's would consist of typical symptoms. One subsequent target task presented participants with unlabeled profiles and asked which illness was most likely. Klayman and Brown proposed that participants in both groups would base decisions on the same symptom-matching model (p. 107). To illustrate, consider the following example (adapted from Klayman & Brown, 1993). A patient is known to suffer from either puneria or zymosis and exhibits a cough. Coughing occurs in 60% of puneria patients and 90% of zymosis patients. Though a cough occurs more often than not in puneria patients—and is therefore typical of puneria—it is nonetheless evidence against puneria because coughing is more common in zymosis. For noncontrastive learners, the model would "score" coughing as +1 for each illness because it is typical of each. For contrastive learners, who learned that coughing was more predictive of zymosis than puneria, the model would score coughing as +1 and −1 for zymosis and puneria, respectively. A patient profile consisted of several symptoms, the total score for each illness was tallied, and the illness with the highest score was predicted to be selected by the participants. Note that, because the tallies are compared in Klayman and Brown's model, coughing should have no effect on noncontrastive learners' diagnoses, but should make zymosis more likely to be selected by contrastive learners. Thus, the model implies that both groups compare the patient profile to each illness concept; the difference between the groups lies in the different illness concepts. Though Klayman and Brown did not examine confidence, their model seems to imply that confidence would be additive; that is, high confidence in one illness will be accompanied by low confidence in the other. In contrast, the current view is concerned with, among other things, when confidence will and will not be additive. To use the coughing symptom as an example, predictions are made as to when it will increase confidence in both illnesses.

Independently, Goldstone (1996) also examined how contrastive and noncontrastive learning (among other things) affect conceptual structure, arguing that the former leads to interrelated concepts, while the latter leads to isolated concepts. Participants learned about two categories whose members consisted of line segments (Experiment 2). Some line segments were more common to one category (i.e., were diagnostic) and some were equally common to both (i.e., were nondiagnostic). He found that contrastive learners were less influenced by nondiagnostic features than were noncontrastive learners. However, the primary dependent measure was percentage of correct categorizations, whereas the present focus is on confidence and how it is represented. Perhaps more important is that Goldstone's model equates representation with consideration of the alternative, whereas the present approach distinguishes between the two. In particular, the present view allows for the possibility that the alternative can be taken into account with independent confidence, though additional processing is necessary.

The distinctions between dependent and independent confidence and between conceiving of two and four types of evidence—which lie at the heart of the present article—are not part of the theorizing of Klayman and Brown (1993) and Goldstone (1996), nor do these distinctions obviously follow from their research. Nonetheless, the previous and current research share the important trait of illustrating that how concepts are learned has a powerful influence on behavior. The current work can be seen as broadening the implications of the learning manipulation and the resulting representations.

Experiment 1

Overview

By viewing patient profiles, participants learned to recognize two illnesses, puneria and zymosis, through either contrastive or noncontrastive learning. Profiles listed 16 symptoms and whether each was present. Four symptoms were common to puneria only (puneria symptoms), four to zymosis only (zymosis symptoms), four to both (*both* symptoms), and four to neither (*neither* symptoms). The symptom types and their conditional probabilities are shown in Table 1.

Participants were then presented with three target tasks in which patients were known to have one of the illnesses, but not both. For the target tasks, puneria and zymosis symptoms were diagnostic, *both* and *neither* symptoms were not. One task involved selecting a limited number of symptoms to assess the likelihood of the focal illness. Contrastive learners were predicted to select diagnostic symptoms, independent of which illness was focal. Noncontrastive learners were predicted to select symptoms common to the focal illness, independent of diagnosticity.

Because participants sometimes fail to select relevant information but nonetheless use that information when it is presented to them (Beyth-Marom & Fischhoff, 1983), a second task displayed individual symptoms and participants responded whether their confidence in the focal illness decreased, remained the same, or increased. Contrastive learners' confidence was predicted to change only when presented with diagnostic symptoms, independent of focal illness. Noncontrastive learners' confidence was expected to change when presented with symptoms common to the focal illness, independent of diagnosticity. These responses were also timed. Noncontrastive learners' responses were predicted to take longer when they took into account the alternative rather than ignored it because the former require more processing than the latter.

The third task involved seeing two complete patient profiles, one with symptoms consistent with both illnesses, the other with symptoms consistent with neither. Contrastive learners were predicted to have medium confidence in both patients for whichever illness was focal. Noncontrastive learners were predicted to have high confidence in the focal illness for the patient with consistent symptoms, and low confidence for the patient with inconsistent symptoms. Frequency judgments were also reported to see if ignoring the alternative is limited to confidence judgments. Finally, participants made treatment decisions because consideration of alternatives might be triggered by making decisions, but not by reporting confidence.

Method

Fifty-eight participants were recruited through signs posted on the University of Chicago campus and were paid $10. The experiment was conducted entirely on computer, with up to 5 participants tested at one time in the same room. Participants played the role of physicians learning to diagnose puneria and zymosis on the basis of patient profiles listing whether 16 symptoms were present (see Table 3).

Contrastive learners were shown 20 puneria and 20 zymosis profiles in a random order. Each was labeled with the correct diagnosis and the instructions were to "learn how to distinguish patients with puneria from patients with zymosis." Participants could view each profile for up to 30 s, after which the next one appeared. They could proceed sooner if desired. They were told that they would be tested (as described below) after viewing the 40 profiles.

For each participant, names (e.g., *fever*) were randomly assigned to each symptom. The probabilities for each symptom type are shown in Table 1 and are exact for each set of 20 profiles. For example, a puneria symptom was present in 17 of 20 (85%) puneria patients, but only 3 of 20 (15%) zymosis patients. Order of symptoms on a profile was random for each participant. Symptoms were conditionally independent in that, for a given patient, the presence of one symptom was not predictive of the presence of another (i.e., symptoms did not "cluster").

After viewing 40 profiles, contrastive learners were tested. For each symptom, they were asked a pair of questions (one right after the other), namely, whether it tended to occur in patients with (a) puneria and (b) zymosis. The correct answer to each question was "yes" and "no" when the probability was .85 and .15, respectively. Symptom order was random in each testing session. If participants made any errors, a list of symptom(s) to which they responded incorrectly was presented after the testing. They were then shown 20 new cases of each illness and retested. It was only after no errors were made that participants proceeded to the next stage, consisting of viewing another 40 cases and another testing session. Contrastive learners had to complete two error-free testing sessions.

Noncontrastive participants learned about the illnesses separately. For each illness, they were shown 20 profiles of patients with the illness and 20 profiles of patients without the illness and were instructed to "learn how to distinguish patients with the illness from patients without the illness." They first learned about puneria by viewing 20 puneria profiles and 20 no puneria profiles. Probabilities are shown in the top of Table 2. After viewing the 40 profiles, participants were tested. For each symptom, they were asked whether it tended to occur in patients (a) with puneria and (b) without puneria (the correct latter answer was always "no"). Participants repeated the puneria learning until no mistakes were made and then learned about zymosis analogously. The probabilities are shown at the bottom of Table 2. Participants were then tested on zymosis and, if no errors were made, they proceeded to the next stage, which, as with contrastive learners, meant repeated learning. Noncontrastive learners had to complete two error-free testing sessions for each illness.

Table 3
Example of a Patient Profile

Symptom	Present?
Coughing	Yes
Dizziness	No
Headache	Yes
Congestion	Yes
Rash	Yes
Nausea	No
Fever	Yes
Exhaustion	Yes
Sneezing	Yes
Insomnia	No
Diarrhea	No
Sore throat	No
Itching	Yes
Earache	No
Depression	No
Swollen glands	Yes

After the first error-free testing session, participants rated how difficult it was to learn about the two illnesses on a scale from 1 (*not at all difficult*) to 9 (*extremely difficult*). When the learning was completed, participants were told the following:

> Now, imagine further that you are a specialist who deals exclusively with patients already known to have a paralymphnal illness, of which there are only two kinds: puneria or zymosis. Therefore, every patient you see has either puneria or zymosis, but not both. That is, all your patients have one—and only one—of the illnesses.

Participants were also told to consider each illness equally likely before seeing a patient's profile. They then answered the following question:

> When you see patients from now on, which of the following is true? (a) Each patient has puneria, or zymosis, or neither. That is, patients have either one of the illnesses, or neither of them. (b) Each patient has puneria, or zymosis, or both. That is, patients have either one of the illnesses, or both of them. (c) Each patient has puneria or zymosis, but not both. That is, patients have one—and only one—of the illnesses. (d) Each patient has puneria, or zymosis, or both, or neither. That is, patients may have one of the illnesses, both of the illnesses, or neither of the illnesses.

If participants did not respond with (c), they were shown the information again and retested. They were also informed that sometimes they are called upon to treat an illness. Each illness required a different treatment and, because the treatments were incompatible, treating one illness ruled out treating the other. Furthermore, failing to treat the correct illness was potentially fatal and the profile was the only available information. Participants were quizzed about the mutual exclusiveness of the treatments. If they answered incorrectly, they were shown the information again and retested.

Subsequently, participants were told they would be shown a profile for 1 min and to "form an impression of the patient." Half of the participants were shown a profile with symptoms consistent with both illnesses (H1, H2, and *both* symptoms were present and *neither* symptoms were absent), and half were shown a profile with symptoms consistent with neither illness (only *neither* symptoms were present). The profile disappeared after 1 min and participants reported their confidence that the patient had a particular illness. Half were asked about puneria, half about zymosis. The illness asked about determined the focal illness. Participants typed their responses, with 1 = *very confident that patient does NOT have* [focal illness], 5 = [focal illness] *is as likely as NOT* [focal illness], and 9 = *very confident that patient has* [focal illness]. They could change their response before proceeding. They answered two additional questions. One was, "Imagine 100 of your patients exhibited the same pattern of symptoms. How many of them do you think would have [focal illness]?" The other was, "For the patient whose profile you just saw, would you administer the [focal illness] treatment?" Half of the participants answered the frequency question first, half the treatment question.

The next task involved information search. Participants were told that a patient on the telephone wanted to know how likely it was that she had the focal illness. They selected 8 (of 16) symptoms to ask the patient about. Participants were reminded that the patient had either puneria or zymosis, but not both. They did not see whether the selected symptoms were present or absent.

Participants performed a response time (RT) task next. They were shown whether a single symptom was present or absent and responded whether it decreased, had no effect on, or increased their confidence that the patient had the focal illness. Participants were told that they would see many symptoms and that each was for a different patient. They were also told that their responses would be timed and to respond as quickly as possible while maintaining accuracy. They were again reminded that patients had either puneria or zymosis, but not both. Participants were presented with the 16 symptoms, both present and absent, in a random order, at their own pace. Responses and response times were recorded for the 32 trials.

Participants completed the confidence, information search, and RT tasks with one illness focal, then repeated them with the other illness focal. Half of the participants began with puneria focal, half with zymosis. In the task presenting a complete profile, half of the participants first saw the consistent profile, half the inconsistent profile. They then answered the following question:

> Recall when you were seeing patients as a specialist. Which of the following best describes your understanding? (a) Each patient had puneria, or zymosis, or neither. (b) Each patient had puneria, or zymosis, or both. (c) Each patient had puneria or zymosis, but not both. (d) Each patient had puneria, or zymosis, or both, or neither.[1]

Results

Of the 58 participants, 9 (5 contrastive and 4 noncontrastive learners) failed to complete the learning phase. Data were analyzed for the remaining 24 contrastive and 25 noncontrastive learners. Mean time taken to complete the experiment was 84 min.

Learning

Contrastive learners had to complete two error-free testing sessions; they took a mean of 7.4 attempts to complete the first, and 1.3 to complete the second. Noncontrastive learners had to complete two error-free testing sessions for each of puneria and zymosis; they needed a mean of 5.8 and 3.2 attempts to complete the first session for puneria and zymosis, respectively, and 2.1 and 2.0 mean attempts to complete the second. Both groups took equally long to complete the learning phase (*M*s = 52.5 and 57.9 min, respectively, $F < 1$) and found learning equally difficult ($F < 1$).

Information Search

Participants selected 8 of the 16 symptoms to ask a patient about in order to assess the likelihood of the focal illness. For simplicity, data were collapsed across focal illness, but the pattern of results is the same for each illness separately. H1 corresponds to the focal illness, H2 to the alternative. Figure 2 shows the number of symptoms selected as a function of learning and symptom type. As predicted, contrastive learners preferred H1 and H2 symptoms, whereas noncontrastive learners preferred H1 and *both* symptoms. Because predictions regarded different preferences for H2

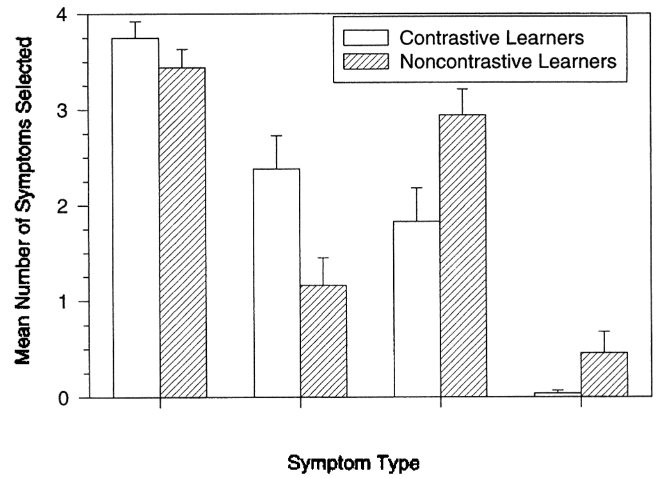

Figure 2. Experiment 1: Mean number of symptoms selected as a function of learning and symptom type. Standard error bars are shown. H1 = Hypothesis 1; H2 = Hypothesis 2.

and *both* symptoms between the groups, a 2 (learning) × 2 (symptom type: H2 and *both*) analysis of variance (ANOVA) was conducted, using symptom type as a repeated measure. The results are shown in Table 4 under "Search." The only significant effect was the predicted interaction: Relative to contrastive learners, noncontrastive learners selected fewer H2 symptoms and more *both* symptoms.

Information Use: Changes in Confidence

Participants were presented with the presence and absence of individual symptoms and responded whether their confidence in the focal illness decreased, remained the same, or increased. A "+" and "−" are used to identify a symptom's presence and absence, respectively. For example, a puneria+ symptom is common to only puneria and present. Such symptoms are evidence for puneria and against zymosis. Zymosis− symptoms, which are absent, are also evidence for puneria and against zymosis. Note the clear distinction that can be drawn between evidence for one illness (e.g., puneria+) and evidence against the other (zymosis−), a distinction to which noncontrastive learners were expected to be more sensitive.

Participants were categorized for each symptom type according to their unique modal response. The results are shown in Table 5. Data were again collapsed across focal illness; H1 corresponds to the focal illness, H2 to the alternative. Symptom types of interest were H2+, H2−, *both*+, and *both*−, each of which has one response that

[1] Only the second half of the participants were asked this question. This is also true of the two questions preceding the target tasks regarding the mutual exclusiveness and exhaustiveness of the illnesses and the mutual exclusiveness of the treatments. These were added to ensure that participants understood the task as intended. The questions did not affect any target task dependent measures.

Table 4
Analyses of Variance for Experiment 1

Source	*df*	*F* Search	*F* Confidence	*F* Frequency	*F* Decision
		Between subjects			
Learning	1	0.31	0.79	0.05	0.01
Error	47	(0.23)	(2.96)	(475.67)	(0.15)
		Within subjects			
RM	1	2.03	43.76**	58.85**	6.04*
RM × Learning	1	7.13*	7.98**	5.84*	8.83**
Error	47	(4.63)	(3.03)	(427.68)	(0.16)

Note. Values in parentheses represent mean square errors. RM = repeated measure, which is "symptom type" when the dependent variable is search and "patient" when the dependent variables are confidence, frequency, and decision.
$*p < .05$. $**p < .01$.

ignores the alternative (indicated by superscript a in Table 5), and one response that takes it into account (superscript b). The third possible response was ignored in the following analyses, and p values correspond to two-tailed Fisher's exact tests.

H2+ symptoms are evidence for the alternative. All 24 contrastive learners' modal response was "decrease," which takes into account the alternative, but this was true for only 15 noncontrastive learners; 4 participants in this condition had a modal response of "no effect," which ignores the alternative. Contrastive learners were more likely to be categorized as taking into account the alternative ($p = .031$). For H2− symptoms (evidence against the alternative), the modal response for 19 of 21 contrastive learners (who had a unique modal response) was "increase," which takes into account the alternative, whereas the majority of noncontrastive learners responded "no effect," which ignores it ($p = .0002$). When presented with *both+* symptoms (evidence for both illnesses), most contrastive learners' modal response was "no effect," which takes into account the alternative, but most noncontrastive learners responded "increase," which ignores it ($p = .025$). Finally, when presented with *both−* symptoms (evidence against both illnesses), contrastive learners were more often categorized as taking into account the alternative ("no effect") than were noncontrastive learners. Though the effect is in the predicted direction, it was not significant ($p = .20$).

Information Use: RTs

Participants' RTs were also recorded when reporting changes in confidence. A mean RT was computed for each participant, collapsing across the four symptom types (H2+, H2−, *both+*, and *both−*) for the responses that did and did not take into account the alternative (see above). Thus, each participant had a mean RT corresponding to when the alternative was taken into account and a mean RT corresponding to when it was ignored. (To eliminate outliers, the fastest 2.5% and the slowest 2.5% of all RTs were not analyzed.) Figure 3 shows that contrastive learners responded equally fast whether or not their response took into account the alternative. However, noncontrastive learners' responses were much slower when the alternative was taken into account. A 2 (learning) × 2 (take into account alternative) ANOVA indicated an effect of the latter variable ($z = 2.86$, $p = .004$), but, more important, the predicted interaction was significant ($z = 2.15$, $p = .031$). (Four contrastive and three noncontrastive learning cases were incomplete and this analysis estimated the missing data, which is why zs are

Table 5
Experiment 1: Number of Participants Categorized by Modal Response to Symptom Types Conditional on Learning

Modal response	H1+	H1−	H2+	H2−	Both+	Both−	Neither+	Neither−
				Contrastive learners				
"Decrease"	0	21	24[b]	1	0	6[a]	2	0
"No effect"	0	0	0[a]	1[a]	15[b]	17[b]	22	23
"Increase"	23	3	0	19[b]	5[a]	0	0	0
None	1	0	0	3	4	1	0	1
Total	24	24	24	24	24	24	24	24
				Noncontrastive learners				
"Decrease"	1	17	15[b]	6	1	9[a]	8	4
"No effect"	1	2	4[a]	11[a]	7[b]	10[b]	17	18
"Increase"	21	5	1	6[b]	13[a]	5	0	3
None	2	1	5	2	4	1	0	0
Total	25	25	25	25	25	25	25	25

Note. If there was no unique modal response to a particular symptom type, a participant was classified as "none." H1 = focal hypothesis; H2 = alternative hypothesis.
[a]Indicates responses that take into account only the focal hypothesis. [b]Indicates responses that take into account both the focal and alternative hypotheses.

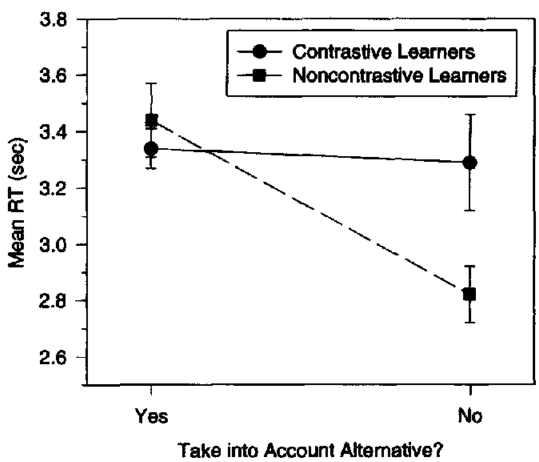

Figure 3. Experiment 1: Mean response time (RT) as a function of learning and taking into account the strength of the alternative hypothesis.

reported; the interaction is still significant if these cases are omitted.)

Confidence Judgments

Participants reported confidence in the focal illness for a profile consistent with both illnesses (the *both* patient) and for a profile consistent with neither (the *neither* patient). Figure 4 shows that though contrastive learners' confidence was higher for the *both* patient than for than the *neither* patient, the difference was greater for noncontrastive learners. The results of the 2 (learning) × 2 (patient) ANOVA are shown in Table 4. There was an effect of patient and an interaction.

Frequency Judgments

After reporting confidence, participants estimated how many of 100 patients exhibiting the same symptoms would have the focal illness. Noncontrastive learners' frequency judgments were higher than contrastive learners' for the *both* patient (53.9 vs. 44.8) and lower for the *neither* patient (11.8 vs. 22.8). A 2 × 2 ANOVA revealed an effect of patient and an interaction (see Table 4).

Decisions

Participants also decided whether they would administer the focal illness treatment for each patient. Figure 5 shows that contrastive learners were equally likely to treat either patient. However, 44% of noncontrastive learners opted to treat the *both* patient and none opted to treat the *neither* patient. There was again an effect of patient and an interaction (see Table 4). (A chi-square analysis, using learning and change in decision across the two patients as variables, led to the same conclusion.)

Participants' Understanding of the Relation Between the Illnesses

At the end of the experiment, participants were asked (in effect) whether they thought that the illnesses were mutually exclusive and exhaustive. When selecting among four responses, 83% of contrastive learners and 75% of noncontrastive learners selected the correct response (see Footnote 1). Thus, participants in both learning conditions understood the instructions equally well.

Discussion

Using several measures, it was found that contrastive but not noncontrastive learners tended to take into account the strength of the alternative hypothesis. First, when selecting symptoms to determine the likelihood of the focal illness, contrastive learners tended to select diagnostic symptoms, regardless of whether the symptoms were common to the focal illness. Noncontrastive learners, however, tended to select symptoms common to the focal illness without regard to diagnosticity. Second, when presented with individual symptoms, contrastive learners tended to change confidence when the symptoms were diagnostic. Noncontrastive learners, on the other hand, tended to change confidence when presented with symptoms common to the focal illness, without regard to diagnosticity. Put differently, noncontrastive learners appeared to distinguish between evidence for one illness and evidence against the other, whereas contrastive learners did not.

Third, when presented with a profile consistent with both illnesses (the *both* patient), contrastive learners tended to have medium confidence in whichever illness was focal, whereas noncontrastive learners had high confidence. Similarly, when the profile was consistent with neither illness (the *neither* patient), contrastive learners tended to have medium confidence in the focal illness, whereas noncontrastive learners had low confidence. Because participants did not know which illness was focal until after the profile disappeared, there is no reason to suspect that noncontras-

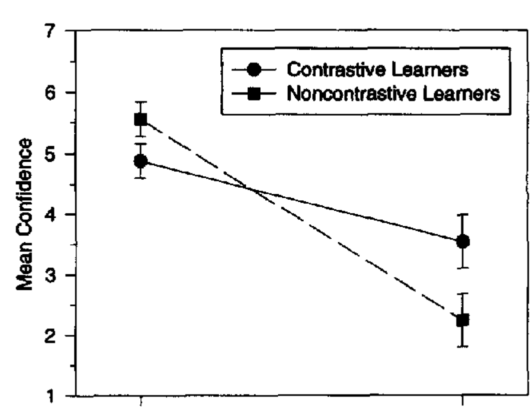

Figure 4. Experiment 1: Mean confidence as a function of learning and patient.

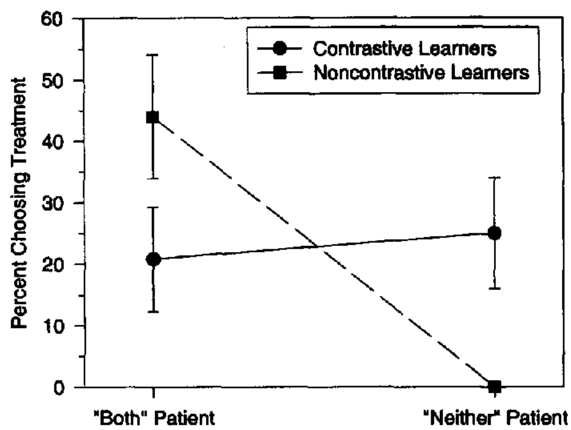

Figure 5. Experiment 1: Percentage of participants choosing to treat the focal illness as a function of learning and patient.

tive learners were attending only to symptoms regarding the focal illness. Thus, when reporting confidence after encoding symptoms regarding both illnesses, noncontrastive learners were more likely to access the strength of the focal illness and not the alternative.

Fourth, frequency judgments mirrored confidence: Relative to contrastive learners, noncontrastive learners reported higher frequencies for the *both* patient and lower frequencies for the *neither* patient. Recent studies indicate that there is an important psychological distinction between confidence in a unique case and frequency judgments for a class of cases (e.g., Gigerenzer, 1994; Gigerenzer, Hoffrage, & Kleinbölting, 1991). The frequency question was included to see if predicted differences in confidence between learning conditions would disappear with frequency questions. They did not. However, the frequency question always came after the confidence question and participants might simply have based their frequency judgments on their confidence judgments. A stronger test is needed to determine whether the present theoretical distinctions apply to frequency as well as to confidence judgments.

Fifth, the same pattern held for decisions. Contrastive learners were equally likely to administer the focal illness treatment to the *both* and *neither* patients, whereas noncontrastive learners were more likely to administer the treatment to the *both* patient. Thus, when making decisions about the focal illness, noncontrastive learners appeared prone to consult their confidence in that illness rather than compare relative strengths of the competing illnesses. The current research has direct implications for decision making as well as judgment.

An aspect of the decision results worth noting is the low overall treatment rate for both groups: On average, only 23% of contrastive learners and 22% of noncontrastive learners opted to treat a patient for the focal illness (see Figure 5). However, the results to be reported for Experiment 2 indicate that the current treatment rate is somewhat anomalous. Whatever the reason for the current low rate, it applies to both contrastive and noncontrastive learners. For present purposes, the most relevant aspect of Figure 5 is the interaction.

Finally, in the RT task, noncontrastive learners took longer to respond when they took into account the alternative than when they did not. This was not true for contrastive learners. This result is consistent with the notion that taking into account the strength of the alternative requires additional processing for noncontrastive learners. The interaction in Figure 3 cannot be explained by the fact that common responses tend to be faster than rare ones (e.g., Pike, 1968): Though noncontrastive learners' common response (ignoring the alternative) was faster than their rare one, contrastive learners' common response (taking into account the alternative) was slightly slower.

Another issue regarding the RT findings concerns the fact that contrastive learners were taking just about as long as noncontrastive learners to take into account the alternative (see Figure 3), but it is only the latter who were supposed to be doing additional processing. However, it is inappropriate to compare heights of these lines. Noncontrastive learners essentially associated eight symptoms with puneria and eight symptoms with zymosis. Ignoring the alternative for these participants implies simply responding "increase" when shown a present associated symptom, "decrease" when shown an absent associated symptom, and "no effect" otherwise. Contrastive learners, on the other hand, associated with each illness four symptoms whose presence increases confidence and whose absence decreases confidence, and four symptoms whose presence decreases confidence and whose absence increases confidence. This is clearly more complicated and it is not surprising that RTs are slower when contrastive learners take into account the alternative than when noncontrastive learners ignore it. It is the interaction in Figure 3 that is important.

A potential criticism of this experiment is that noncontrastive learners experienced three types of patients, whereas the target tasks included only two. Though participants were repeatedly reminded that patients could have only puneria or zymosis, and most participants reported believing this at the end of the experiment, perhaps noncontrastive learners' behavior could be explained by assuming that they responded as though patients could be healthy in addition to having puneria or zymosis. This is an important issue because interpretation of the results hinged on the fact that participants—noncontrastive learners in particular—viewed the illnesses as mutually exclusive and exhaustive. If noncontrastive learners believed that their new patients could be healthy in addition to having puneria or zymosis, then, for example, having low confidence in the focal illness for the *neither* patient does not necessarily indicate independent confidence or a failure to consider the strength of the alternative. One could believe strongly that the patient is healthy and thus believe only weakly in puneria and zymosis. One way around this criticism is to include a noncontrastive group who does not see a third category during learning.

Experiment 2

A third condition was added in Experiment 2. New noncontrastive participants learned about the illnesses separately, but never saw any healthy patients. Instead of distinguishing between patients with each illness and healthy patients, new noncontrastive learners saw only puneria patients when learning about puneria, and saw only zymosis patients when learning about zymosis (which is similar to the "independent" condition in Klayman & Brown, 1993). These participants should be no less likely than contrastive learners to view the illnesses as mutually exclusive and exhaustive. If noncontrastive learners' failure to view the illnesses as mutually exclusive and exhaustive explains the results of Experiment 1, then new noncontrastive learners should behave like contrastive learners. In addition, Experiment 2 attempted to replicate Experiment 1's findings using 8 symptoms rather than 16, which proved difficult (9 of 58 participants failed to complete the learning phase in Experiment 1). Another change was the participant pool. Participants in Experiment 1 were graduate and undergraduate students from the University of Chicago, whereas those in Experiment 2 were undergraduates from the University of California, San Diego (UCSD).

Method

Participants

Participants were 164 students from introductory psychology courses at UCSD who received partial course credit.

Procedure

Except for the following changes, the procedure was identical to that of Experiment 1. That there were only 8 symptoms (2 each of H1, H2, *both*, and *neither* symptoms), rather than 16, led to several small differences. First, profiles were visible up to 15 s (rather than 30 s) during learning. Second, participants selected 4 symptoms (rather than 8) in the information search task. Third, in the information use task, participants were presented twice (rather than once) with each of the symptoms both present and absent for each focal illness (resulting in 32 trials for each illness in both experiments). Finally, participants viewed the two complete target profiles, for which they reported confidence, frequencies, and decisions, for 30 s (rather than 60 s).

New noncontrastive learners first learned about puneria through viewing only puneria profiles, with instructions to "learn how to recognize patients with the illness." After viewing 20 profiles, they were tested. For each symptom, participants were asked whether it tended to occur in patients with puneria. The correct answer was "yes" and "no" when the probability was .85 and .15, respectively. If they made any mistakes, they were told which symptoms were in error and were shown another 20 puneria profiles. When no mistakes were made, participants learned about zymosis in an analogous fashion. New noncontrastive learners had to complete two error-free testing sessions for each illness. The primary difference between noncontrastive learning and new noncontrastive learners is that the latter saw no healthy patients during learning.

Results

One contrastive and 3 noncontrastive learners failed to complete the learning phase. Data were analyzed for the remaining 53 contrastive, 52 noncontrastive, and 55 new noncontrastive learners. Mean time taken to complete the experiment was 46 min.

Learning

Contrastive learners took a mean of 4.6 attempts to complete the first learning–testing session, and 1.3 attempts to complete the second. Noncontrastive learners took a mean of 3.0 and 1.9 attempts to complete the first session for puneria and zymosis, respectively, and 1.5 and 1.2 to complete the second. For new noncontrastive learners, the means were 1.8 and 1.3 for the first session for puneria and zymosis, respectively, and 1.2 and 1.1 for the second. Contrastive and noncontrastive learners took about the same time to learn (Ms = 26.6 and 29.6 min, respectively; p = .18), but new noncontrastive learners took less time (M = 14.5 min; both ts > 7). Finally, the groups did not find learning equally difficult. The means on a scale from 1 (*not at all difficult*) to 9 (*extremely difficult*) for contrastive, noncontrastive, and new noncontrastive learners were 5.4, 4.1, and 3.1, respectively. All of these means are different (ts > 3, ps < .01).

Information Search

Participants selected four of the eight symptoms in order to assess how likely it was that a patient had the focal illness. Data were again collapsed across focal illness. H1 corresponds to the focal illness, H2 to the alternative.

Replication of Experiment 1. Figure 6 shows the number of symptoms selected as a function of learning and symptom type for all three groups. For the present purposes, consider only the contrastive and noncontrastive groups. Contrastive learners were more likely than noncontrastive learners to select H2 symptoms and were less likely to select *both* symptoms. A 2 (learning: contrastive and noncontrastive) × 2 (symptom type: H2 and *both*) ANOVA revealed an

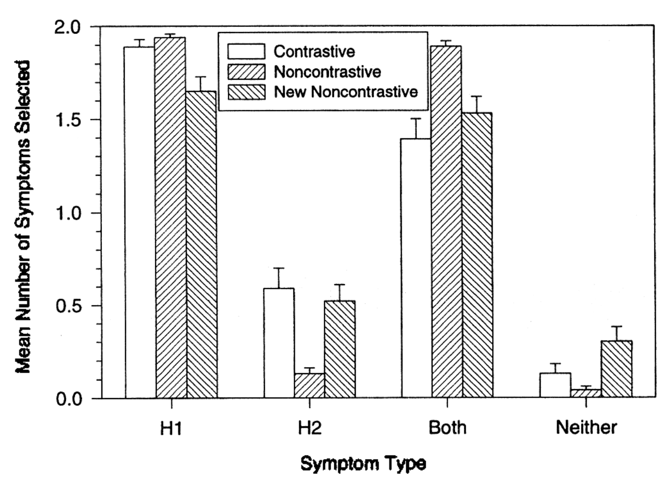

Figure 6. Experiment 2: Mean number of symptoms selected as a function of learning and symptom type. H1 = Hypothesis 1; H2 = Hypothesis 2.

effect of symptom type and an interaction (see Table 6). The interaction replicates Experiment 1.

New test. Figure 6 shows that new noncontrastive learners were more similar to contrastive than to noncontrastive learners, especially with regard to H2 and *both* symptoms. The difference between the means for those two symptom types for contrastive, noncontrastive, and new noncontrastive learners was 0.8, 1.8, and 1.0, respectively. Because predictions regard interactions, further comparisons were carried out with two-way ANOVAs. A 2 (learning: contrastive and new noncontrastive) × 2 (symptom type: H2 and *both*) ANOVA revealed only an effect of symptom type (see Table 6). Though contrastive learners selected more H2 symptoms and fewer *both* symptoms, there was no interaction. Furthermore, the other 2 (learning: noncontrastive learning and new noncontrastive) × 2 (symptom type: H2 and *both*) ANOVA revealed an effect of symptom type and an interaction (see Table 6); noncontrastive learners were less likely to select H2 symptoms and were more likely select *both* symptoms.

Taken together, the information search results regarding H2 and *both* symptoms indicate that contrastive and new noncontrastive learners were similar to each other, whereas noncontrastive learners were different from both. This is an unpredicted pattern.

Information Use: Changes in Confidence

Participants were categorized according to their unique modal response to each symptom type when presented with individual symptoms. These categorizations are shown in Table 7. H1 represents the focal illness. Responses to H2+, H2−, *both*+, and *both*− symptom types are of interest because one response takes into account the alternative (marked with superscript b in Table 7) and one ignores it (superscript a). As before, the third response was ignored in the following analyses and p values correspond to two-tailed Fisher's exact tests.

Replications. Contrastive learners were more likely than noncontrastive learners to take into account the alternative when presented with H2+ symptoms ($p = .02$), H2− symptoms ($p = .0001$), *both*+ symptoms ($p = .0005$), and *both*− symptoms ($p = .02$). These results replicate Experiment 1.

New tests. Contrastive learners were more likely than new noncontrastive learners to take into account the alternative when presented with H2+ symptoms ($p = .03$), H2− symptoms ($p = .01$), *both*+ symptoms ($p = .0004$), and *both*− symptoms ($p = .02$; see Table 7). However, noncontrastive and new noncontrastive learners did not differ on any symptom type ($ps = 1.0$, .22, 1.0, and 1.0, respectively).

Table 6
Analyses of Variance for Experiment 2

		F			
Source	*df*	Search	Confidence	Frequency	Decision
Contrastive vs. noncontrastive learning					
Between subjects					
Learning	1	0.79	16.33**	22.40**	2.49
Error	103	(0.02)	(4.55)	(636.36)	(0.24)
Within subjects					
RM	1	124.98**	52.10**	83.28**	28.40**
RM × Learning	1	18.17**	10.03**	9.98**	7.30**
Error	103	(0.69)	(4.96)	(533.28)	(0.17)
Contrastive vs. new noncontrastive learning					
Between subjects					
Learning	1	0.94	0.78	0.23	0.12
Error	106	(0.06)	(5.31)	(801.94)	(0.27)
Within subjects					
RM	1	42.70**	39.40**	64.59**	17.97**
RM × Learning	1	0.62	4.64*	5.31*	2.76
Error	106	(1.03)	(4.92)	(594.21)	(0.19)
Noncontrastive vs. new noncontrastive learning					
Between subjects					
Learning	1	0.17	9.14**	18.36**	1.58
Error	105	(0.06)	(4.88)	(624.68)	(0.23)
Within subjects					
RM	1	235.54**	90.08**	119.64**	54.79**
RM × Learning	1	17.63**	1.14	0.52	1.13
Error	105	(0.44)	(4.88)	(602.85)	(0.16)

Note. Values in parentheses are mean square errors. RM = repeated measure, which is "symptom type" when the dependent variable is search and "patient" when the dependent variables are confidence, frequency, and decision.
*$p < .05$. **$p < .01$.

Table 7
Experiment 2: Number of Participants Categorized by Modal Response to Symptom Types Conditional on Learning

	Symptom type							
Modal response	H1+	H1−	H2+	H2−	Both+	Both−	Neither+	Neither−
				Contrastive				
"Decrease"	0	39	41[b]	10	3	26[a]	13	6
"No effect"	1	4	9[a]	10[a]	20[b]	22[b]	38	36
"Increase"	50	8	1	28[b]	29[a]	4	0	7
None	2	2	2	5	1	1	2	4
Total	53	53	53	53	53	53	53	53
				Noncontrastive				
"Decrease"	0	35	23[b]	10	0	31[a]	13	8
"No effect"	0	5	16[a]	28[a]	5[b]	8[b]	33	38
"Increase"	51	10	5	10[b]	45[a]	9	1	6
None	1	2	8	4	2	4	5	0
Total	52	52	52	52	52	52	52	52
				New noncontrastive				
"Decrease"	1	34	23[b]	9	0	40[a]	17	2
"No effect"	1	3	15[a]	21[a]	5[b]	11[b]	31	35
"Increase"	52	7	13	15[b]	48[a]	3	3	15
None	1	11	4	10	2	1	4	3
Total	55	55	55	55	55	55	55	55

Note. If there was no unique modal response to a particular symptom type, a participant was classified as "none." H1 = focal hypothesis; H2 = alternative hypothesis; + = present; − = absent. [a]Indicates responses that take into account only the focal hypothesis. [b]Indicates responses that take into account both the focal and alternative hypotheses.

In summary, the change-in-confidence findings showed that contrastive learners were different from both noncontrastive and new noncontrastive learners, who were similar to each other. This was true for each symptom type of interest and is the predicted pattern.

Information Use: Response Times

A mean RT was computed for each participant, collapsing across responses to the four symptom types (H2+, H2−, *both*+, and *both*−) that took into account the alternative and across the responses that ignored the alternative. Each participant had a mean RT associated with taking into account the alternative and a mean RT associated with ignoring the alternative.

Replication. Figure 7 shows the results for all three groups, but consider only contrastive and noncontrastive learners for now. The difference in RT between considering and ignoring the alternative was larger for noncontrastive than for contrastive learners. A 2 (learning: contrastive and noncontrastive) × 2 (take into account alternative: yes and no) ANOVA showed only an effect of the latter variable ($z = 3.5$, $p = .0005$). Despite the predicted pattern, the interaction was not significant ($p = .13$). (Six contrastive and 5 noncontrastive learners were incomplete cases; the analyses presented here estimate missing data, thereby using all participants.)

New test. Figure 7 shows that, descriptively speaking, new noncontrastive learners were most similar to noncontrastive learners, as predicted. The difference in RT between taking into account the alternative and ignoring it was .33, .65, and .55 s for contrastive, noncontrastive, and new noncontrastive learners, respectively. A 2 (learning: contrastive and new noncontrastive) × 2 (take into account alternative) ANOVA revealed an effect of the latter variable ($z = 2.98, p = .003$), but no interaction ($p = .23$). The other 2 (learning: noncontrastive and new noncontrastive) × 2 (take into account alternative) ANOVA also revealed only an effect of the latter variable ($z = 4.6, p < .0001$).

In summary, the RT analyses showed that the predicted

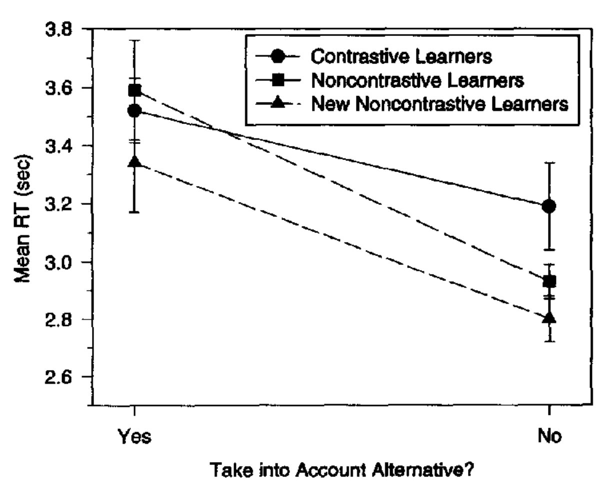

Figure 7. Experiment 2: Mean response time (RT) as a function of learning and taking into account the strength of the alternative hypothesis.

interactions were not significant. Nonetheless, new noncontrastive learners were more similar to noncontrastive than to contrastive learners, consistent with the predictions.

Confidence

Replication. Figure 8 shows that though noncontrastive learners exhibited slightly lower confidence than contrastive learners for the *both* patient, they showed a larger difference between the two patients. A 2 (learning: contrastive and noncontrastive) × 2 (patient) ANOVA indicated effects of learning and patient and an interaction (see Table 6). Except for the effect of learning, these findings replicate Experiment 1.

New test. The difference in confidence between the two patients for contrastive, noncontrastive, and new noncontrastive learners was 1.3, 3.2, and 2.6, respectively, showing that new noncontrastive learners were most similar to noncontrastive learners. A 2 (learning: contrastive and new noncontrastive) × 2 (patient) ANOVA showed an effect of patient and an interaction (see Table 6). A 2 (learning: noncontrastive and new noncontrastive) × 2 (patient) ANOVA showed only effects of learning and patient.

Thus, in terms of the Learning × Patient interactions, contrastive learners were different from both noncontrastive and new noncontrastive learners, who were similar to each other. This is the predicted pattern.

Frequency Judgments

Replication. Figure 9 shows that though noncontrastive learners' frequency judgments were slightly higher than contrastive learners' for the *both* patient (as with confidence), they showed a larger difference in judgments between the two patients. There were effects of learning and patient and an interaction (see Table 6). Except for the effect of learning, these findings replicate Experiment 1.

New test. The difference between frequency judgments for the two patients was 19.0, 39.2, and 34.3, for contrastive, noncontrastive, and new noncontrastive learners, respectively, showing that new noncontrastive learners were most similar to noncontrastive learners. A 2 (learning: contrastive and new noncontrastive) × 2 (patient) ANOVA showed an effect of patient and an interaction. The other 2 (learning: noncontrastive and new noncontrastive) × 2 (patient) ANOVA revealed effects of learning and patient, but no interaction (see Table 6).

For the frequency results in terms of Learning × Patient interactions, contrastive learners were different from both noncontrastive and new noncontrastive learners, who were similar to each other. This is the predicted pattern.

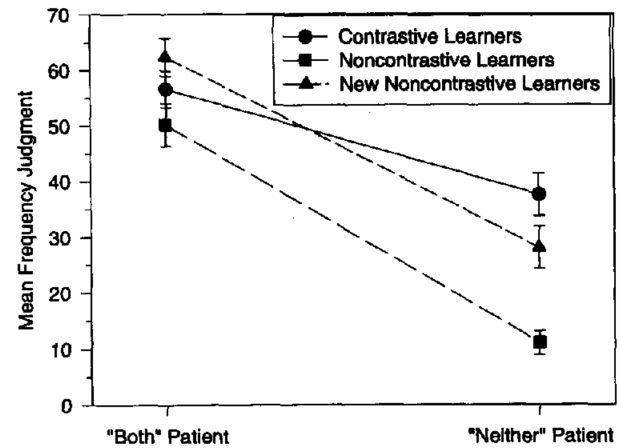

Figure 9. Experiment 2: Mean frequency judgment as a function of learning and patient.

Decisions

Replication. Figure 10 shows that noncontrastive learners were more likely than contrastive learners to treat the *both* patient and were less likely to treat the *neither* patient. There was an effect of patient and an interaction (see Table 6). These findings replicate Experiment 1. (A chi-square analysis led to the same conclusion.)

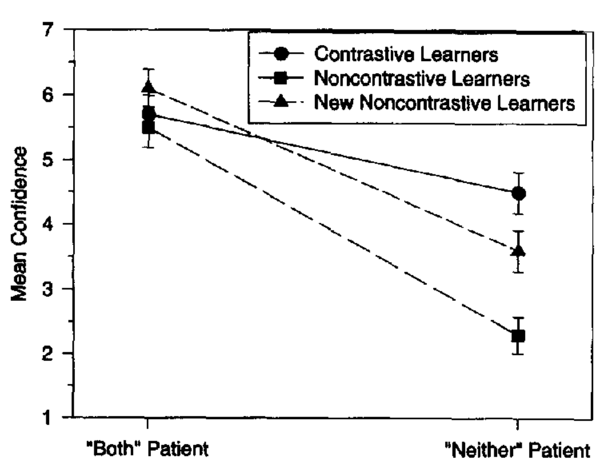

Figure 8. Experiment 2: Mean confidence as a function of learning and patient.

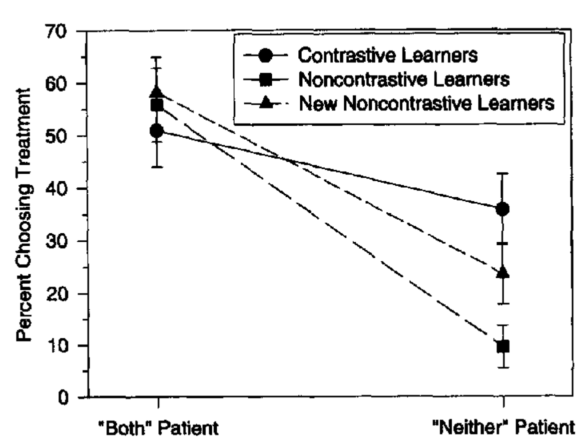

Figure 10. Experiment 2: Percentage of participants choosing to treat the focal illness as a function of learning and patient.

New test. The difference between percentages for the two patients was 15.1, 46.2, and 34.5 for contrastive, noncontrastive, and new noncontrastive learners, respectively, showing that new noncontrastive learners were most similar to noncontrastive learners. A 2 (learning: contrastive and new noncontrastive) × 2 (patient) ANOVA revealed an effect of patient. The interaction did not reach significance (p = .10). The other 2 (learning: noncontrastive and new noncontrastive) × 2 (patient) ANOVA indicated a main effect of patient and no interaction (p = .29). (Chi-square analyses led to the same conclusions.)

In summary, these results in terms of the Learning × Patient interactions show that contrastive and noncontrastive learners were different from each other, whereas new noncontrastive learners were different from neither. However, new noncontrastive learners were more similar to noncontrastive than to contrastive learners.

Participants' Understanding of the Relation Between the Illnesses

When asked at the end of the experiment about the relation between the illnesses during the target tasks, the percentage of contrastive, noncontrastive, and new noncontrastive learners choosing the mutually exclusive and exhaustive answer (from among four) was 94, 79, and 86, respectively.

Discussion

Replications

The results largely replicated those of Experiment 1. For information search, contrastive learners were more likely than noncontrastive learners to select diagnostic symptoms common to the alternative and were less likely to select nondiagnostic symptoms common to the focal illness. For changes in confidence, contrastive learners were more likely than noncontrastive learners to take into account the alternative illness for each of the four symptom types of interest. For confidence judgments, the difference between *both* and *neither* patients was smaller for contrastive than noncontrastive learners. The same pattern held again for frequency judgments. And it held again for decisions regarding treatment: Noncontrastive learners were more likely to treat the *both* patient for whichever illness was focal and were less likely to treat the *neither* patient.

The sole finding that failed to replicate was that for RT in the change-in-confidence task. Though the pattern of results was as predicted—there was a larger difference in RT between considering and ignoring the alternative for noncontrastive learning versus contrastive learners—the interaction was not significant. For noncontrastive learners, the difference in RT was almost identical in each experiment: .62 and .65 s in Experiments 1 and 2, respectively. For contrastive learners, however, the differences were .05 and .33 s, respectively. Though the current difference in RT was twice as large for noncontrastive as opposed to contrastive learners, the effect is much smaller than in Experiment 1. The RT data are important because they speak to the issue of how the alternative is taken into account—specifically, whether additional processing is necessary with independent confidence—whereas the other dependent measures indicate whether it is taken into account. A more detailed examination of the processes underlying how alternatives are taken into account with independent confidence is left for future research.

New Tests

If differences between contrastive and noncontrastive learners were due solely to the latter group's failure to believe the illnesses in the target task were mutually exclusive and exhaustive, then new noncontrastive learners—who, like contrastive learners, saw no healthy patients during learning—should have behaved exactly like contrastive learners. Instead, new noncontrastive learners were more similar to noncontrastive than to contrastive learners with regard to (a) confidence judgments, (b) frequency judgments, (c) decisions, (d) changes in confidence, and (e) response latencies when reporting changes in confidence. This was true qualitatively for each dependent measure and was true statistically for all except decisions and RT. The only task that led to a different conclusion was information search, where new noncontrastive learners were most similar to contrastive learners. It is not obvious why this was the case, but it is noteworthy that the information search task is qualitatively different from the others, all of which (essentially) examine information use. Nonetheless, it does seem clear that differences between contrastive and noncontrastive learners are not satisfactorily explained by claiming that the latter failed to view the illnesses as mutually exclusive and exhaustive. Figures 7 through 10 show one line that stands out as relatively flat, and this line corresponds to contrastive learners in every case, as predicted. The predicted pattern of results is even clearer for the change-in-confidence results presented in Table 7.

Experiment 3

An important implication of the distinction between dependent and independent confidence regards additivity of subjective probability. Probability judgments are additive if they sum to one (or 100%) for a set of mutually exclusive and exhaustive hypotheses. For the confidence target task, additivity implies that if one is X% confident that a patient has puneria, then one is 100 − X% confident that the same patient has zymosis. Though Experiments 1 and 2 showed that noncontrastive learners tended to have high (low) confidence in whichever illness was focal for the *both* (*neither*) patient, the findings were between-subjects. That is, no participant reported confidence in both illnesses for the same patient. A stronger test requires a within-subjects design. Furthermore, the confidence scale should be interpretable in terms of subjective probability, rather than a 9-point scale as used in Experiments 1 and 2. These changes were made in Experiment 3.

Several other changes were made. First, only four symptoms were used. Second, the relation between the symptoms and illnesses was different. Table 8 shows that symptoms (S) differed in their degree of diagnosticity (rather than all-or-none). S1 was strong evidence for puneria (and against zymosis), S2 was weak evidence for puneria, S3 was weak evidence against puneria, and S4 was strong evidence against puneria. Tables 8 and 9 illustrate contrastive and noncontrastive learning, respectively. Note that noncontrastive learning implies that S1 through S4 are decreasingly diagnostic of puneria, with S4 nondiagnostic. For zymosis, noncontrastive learning implies that S1 through S4 are increasingly diagnostic of zymosis, with S1 nondiagnostic. Third, the learning phase was changed. Rather than being presented with labeled profiles and then tested, participants saw unlabeled profiles, made a diagnosis, then received feedback. One advantage of this change was that contrastive and noncontrastive learners saw the same number of puneria and zymosis profiles. Finally, participants were presented with all possible complete target profiles (16, or 2^4) rather than two.

Other researchers have found nonadditive probability judgments (Robinson & Hastie, 1985; Teigen, 1983; Tversky & Koehler, 1994; Van Wallendael, 1989; Van Wallendael & Hastie, 1990; Wright & Whalley, 1983). However, violations of additivity tend to occur only when there are three or more hypotheses, not two. Furthermore, when nonadditivity has been found, judgments usually sum to greater than 100%, not less (i.e., judgments are superadditive, not subadditive). The only exception to this pattern I am aware of was reported by Einhorn and Hogarth (1985), who found slight subadditivity when evidence for two hypotheses was meager and mixed. Additivity for two hypotheses is very common (e.g., Wallsten, Budescu, & Zwick, 1993) and is a fundamental implication of a recent influential theory of subjective probability (Tversky & Koehler, 1994). However, the current theoretical notions lead to predictions as to when judgments will be subadditive, additive, and superadditive—even when there are just two hypotheses.

Consider a patient exhibiting all four symptoms, similar to the *both* patient in Experiments 1 and 2, and confidence in puneria is probed. For contrastive learners, S1 and S2 increase confidence and S3 and S4 decrease confidence, resulting in medium confidence. An analogous prediction holds for confidence in zymosis. Therefore, judgments are expected to be roughly additive. However, for noncontrastive learners, S1, S2, and S3 will tend to increase confidence (S4 will tend to have no effect), leading to high confidence in puneria. It is important to note that the same profile should also lead to high confidence in zymosis: S2, S3, and S4 increase confidence in zymosis, and S1 has no effect. Thus, noncontrastive learners' judgments for the *both* patient should be superadditive. A similar argument applies to the patient with no symptoms present (the *neither* patient): Contrastive learners' judgments will be relatively additive, whereas noncontrastive learners' judgments will be subadditive.

The *both* and *neither* patients were just 2 of 16 patients, where differences in additivity between the groups were expected to be most extreme. However, greater nonadditivity was expected for noncontrastive relative to contrastive learners across all patients because the former tend to consider different evidence relevant depending on which illness is focal, whereas the latter generally consider the same evidence relevant regardless of focal illness. Additivity for each patient profile can be measured by the absolute deviation between 100 and the sum of confidence in puneria and zymosis for that profile. The mean across all profiles then provides a measure of additivity. To the extent that the strength of the alternative is ignored, the mean absolute deviation from 100 should be larger.

Table 8
Illustration of Contrastive Learning (Experiments 3 and 4)

Symptom	Puneria	Zymosis
S1	.80	.20
S2	.60	.40
S3	.40	.60
S4	.20	.80

Table 9
Illustration of Noncontrastive Learning (Experiments 3 and 4)

	Puneria	
Symptom	Puneria	No puneria
S1	.80	.20
S2	.60	.20
S3	.40	.20
S4	.20	.20
	Zymosis	
Symptom	Zymosis	No zymosis
S1	.20	.20
S2	.40	.20
S3	.60	.20
S4	.80	.20

Method

Participants were 75 UCSD students from the same population as in Experiment 2. The 38 contrastive learners were told that they would be shown 60 puneria profiles and 60 zymosis profiles in a random order. The probability of each symptom (S1 through S4) given each illness is shown in Table 8. Names (e.g., *coughing*) were randomly assigned to each symptom. Participants made a diagnosis of either puneria or zymosis, then were told the correct diagnosis.

The 37 noncontrastive learners first began learning about puneria. They were told that they would be shown 40 profiles of patients with puneria and 40 profiles of patients without puneria in a random order. The probabilities of each symptom given puneria and no puneria are shown in the top half of Table 9. Participants diagnosed each profile as either puneria or no puneria and received feedback. Subsequently, they began learning about zymosis. The relevant probabilities are shown in the bottom half of Table 9. They then finished learning about puneria by being shown 20 more profiles each of puneria and no puneria, and then finished learning about zymosis similarly. Noncontrastive learners only partially completed learning about one illness before partially learning about the other in an effort to minimize order effects. Both contrastive and noncontrastive learners saw a total of 60 puneria and 60 zymosis profiles.

After being congratulated on finishing "medical school," participants were told that they were specialists dealing exclusively with patients known to have either puneria or zymosis. They were tested to make sure that they understood the illnesses were mutually exclusive and exhaustive. Each illness was to be considered equally likely before seeing a patient's profile. Participants read that they would be presented with patients and asked how confident they were in one illness or the other. They would type numbers between 1 and 99 and these numbers

> should reflect the percentage correct that you would expect to achieve in the long run. For example, when you report confidence of 90, you should expect to be correct 90% of the time. Reporting confidence of 20 means that you expect to be correct on 20% of such occasions, and so on.

Participants then completed a belief-updating task, in which they were presented sequentially with two symptoms for each of several patients, updating subjective probability after each symptom. These data are reported in another article focusing on belief updating (McKenzie, in press). Participants received no feedback during this target task. They were then presented with the 16 (2^4) unique profiles in random order and reported subjective probabilities in each, with one illness focal. The profiles were then presented again in random order, but with the other illness focal. Thus, participants reported subjective probabilities in each illness for each profile.

Results

Learning

During learning, contrastive learners saw 60 puneria and 60 zymosis profiles. When learning about puneria, noncontrastive learners saw 60 puneria and 60 no-puneria profiles, and when learning about zymosis, they saw 60 zymosis profiles and 60 no-zymosis profiles. In all three cases, learning took place across trials and peaked by the fourth or fifth block of 20 profiles. Both groups performed their respective tasks equally well. Not surprisingly, contrastive learners completed the learning phase faster than noncontrastive learners (*M*s = 18.2 and 29.0 min). Both groups found learning about equally difficult (*M*s = 5.3 and 5.0).

Additivity

Figure 11 shows mean subjective probabilities for the *both* and *neither* patients as a function of learning and focal illness. For the *both* patient, contrastive learners' subjective probability in each illness was almost exactly 50%; thus, these judgments summed to about 100%, or were additive. However, noncontrastive learners' confidence was about 80% for each illness, summing to almost 160% and indicating superadditivity. For the *neither* patient, contrastive learners' confidence was again close to 50% for each illness, whereas noncontrastive learners' confidence was less than 15% in each, indicating subadditivity. This pattern of results is as predicted. Using summed confidence for the two

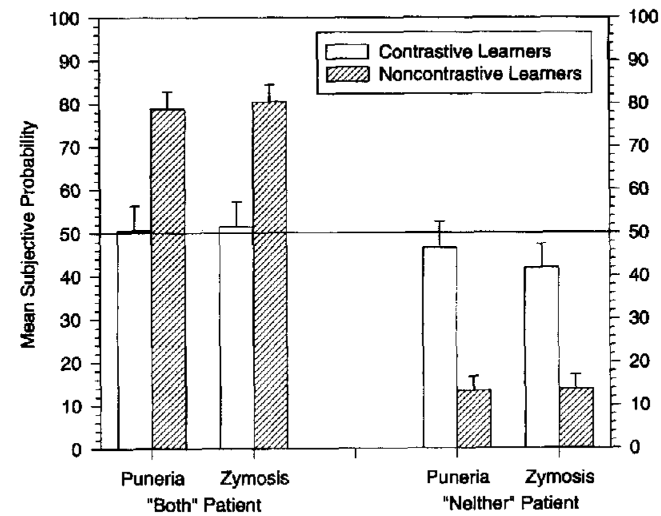

Figure 11. Experiment 3: Mean subjective probability as a function of learning and patient.

illnesses for each of the *both* and *neither* patients as the dependent measure, a 2 (learning) × 2 (patient) ANOVA revealed an effect of patient, $F(1, 73) = 117.9$, $MSE = 1{,}693.7$, $p < .0001$, and, more important, an interaction, $F(1, 73) = 77.7$, $p < .0001$. (Where *MSE*s are not reported, the last reported *MSE* applies).

Additivity across all patients was also examined. For each of the 16 target profiles, participants' judgments for the two illnesses were summed and the absolute deviation from 100 was computed. The smaller the mean absolute deviation is, the more additive the judgments are. As predicted, contrastive learners' judgments were more additive than noncontrastive learners' judgments (*M*s = 16.7 vs. 39.0), $t(73) = 8.39$, $p < .0001$.

When selecting among four responses at the end of the experiment regarding the relation between the illnesses, 36 of 38 contrastive learners and 36 of 37 noncontrastive learners reported that they thought that the illnesses were mutually exclusive and exhaustive.

Discussion

As predicted, noncontrastive learners' judgments were less additive than contrastive learners' judgments. The former group's massive nonadditivity is particularly noteworthy because (a) only two hypotheses were used, (b) subadditivity as well as superadditivity was predicted and found, and (c) the response scale was well defined and interpretable in terms of subjective probability. The results also speak to an issue addressed in Experiment 2, namely, whether noncontrastive learners might respond as though patients could have neither illness in addition to puneria or zymosis (despite denying this when asked directly). This explanation could account for why noncontrastive learners' judgments were subadditive for the *neither* patient (because they might have allocated about 70% confidence to the third "neither illness" hypothesis, which was not asked about), but it cannot explain their superadditive judgments for the *both* patient. Confidence in the two illnesses sums to almost 160%, so assuming a third hypothesis will not help confidence sum to 100%.

As mentioned, a common finding is that subjective probabilities are additive for two hypotheses (even when there are many intervening judgments, as there were in Experiment 3). However, the typical measure used to determine additivity is the mean sum of probabilities for each pair of hypotheses across all items. This mean is virtually always near 100. Indeed, contrastive and noncontrastive learners' mean sums were equally close to 100, but the claim here is that (only) the latter group's judgments were grossly nonadditive. There are two reasons for this apparent contradiction. First, current predictions led to examining specific items: Judgments were examined for the *both* and *neither* patients, where super- and subadditivity for noncontrastive learners was predicted (and found). Second, a different measure was used here than in past studies: Mean absolute deviation from 100 across all items (profiles) was used rather than mean sum. The current measure appears more appropriate for assessing nonadditivity. Note, for example, that noncontrastive learners' judgments are nonadditive for each of the *both* and *neither* patients (confidence sums to 159% and 27%, respectively), but are largely additive if averaged across the 2 patients (93%). Thus, probabilities summing, on average, to 100% for two hypotheses is not inconsistent with the present view. Finally, the systematic nonadditivity revealed here for noncontrastive learners is problematic for any account of subjective probability that assumes or implies additivity for two mutually exclusive and exhaustive hypotheses (e.g., Tversky & Koehler, 1994). How confidence in outcomes is represented is an important variable.

Experiment 4

As depicted here, independent confidence does not necessitate ignoring the alternative; it is only more likely to occur because of the additional processing required. This implies that differences between the two groups should be decreased by encouraging noncontrastive learners to take into account the strength of the alternative—that is, to do the additional processing. In Experiments 1–3, participants were asked asymmetric questions, such as "How confident are you that the patient has puneria?" In Experiment 4, participants were asked symmetric questions, such as "How confident are you that the patient has puneria rather than zymosis?" Symmetric questions encourage consideration of the alternative (e.g., Beyth-Marom & Fischhoff, 1983) and should therefore affect noncontrastive but not contrastive learners.

Method

Participants were 172 UCSD students from the same population as in Experiments 2 and 3. The procedure was identical to that of Experiment 3 except that during the target tasks, half of the contrastive and noncontrastive learners were asked asymmetric questions ($ns = 44$) and half were asked symmetric questions ($ns = 42$).

Results

Learning

As in Experiment 3, each group's number of correct diagnoses increased during the learning phase and appeared to have reached asymptote before the end. Also as before, noncontrastive learners took longer to complete the learning phase (Ms = 28.7 and 18.0 min) and both groups found learning equally difficult (Ms = 5.2 and 5.0).

Additivity

Figure 12 shows mean subjective probabilities for the *both* and *neither* patients as a function of learning and question type. As shown in the top panel, contrastive learners were about 50% confident in each illness for each patient, regardless of question type. However, when noncontrastive learners (bottom panel) were asked asymmetric questions (clear columns), their judgments were again superadditive for the *both* patient (left side); judgments for each illness were nearly 90% (summing to almost 180%). Also as before, their judgments were subadditive for the *neither* patient (right side): Judgments were only about 15%

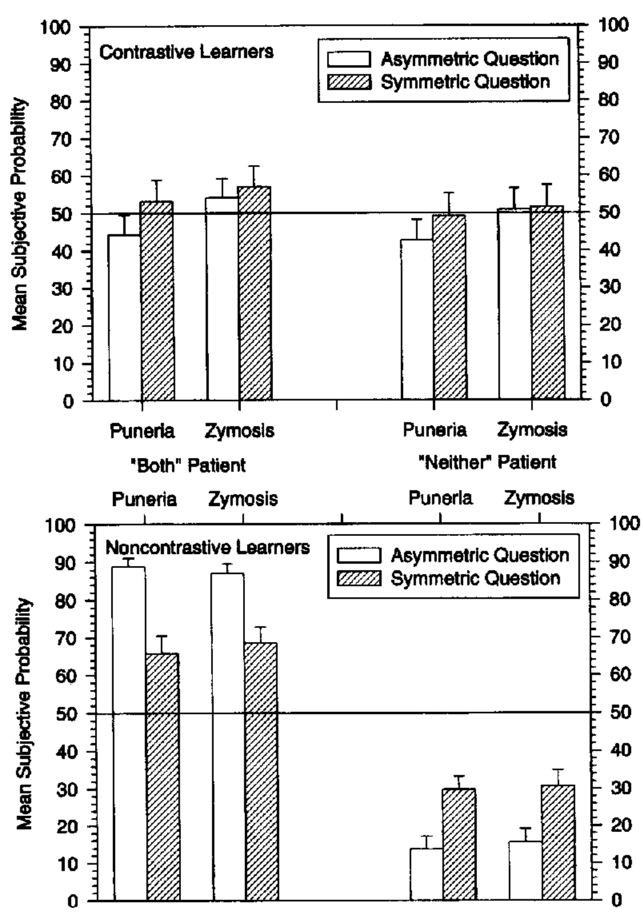

Figure 12. Experiment 4: Mean subjective probability as a function of learning, patient, and question type.

in each illness. Symmetric questions (hatched columns) aided noncontrastive learners' additivity considerably, evidenced by the fact that each judgment moved toward 50%—the normative response for these two patients. Thus, symmetric questions led not only to more additive judgments for noncontrastive learners, but also to more normative judgments.

Using summed confidence in puneria and zymosis for each of the two patients as the (repeated) dependent measure, a 2 (learning) × 2 (question type) × 2 (patient) ANOVA revealed a Learning × Question Type interaction, $F(1, 168) = 4.5$, $MSE = 1{,}031.5$, $p = .036$: Symmetric questions led to an increase in summed confidence for contrastive learners and a decrease for noncontrastive learners. This effect was not predicted, but is small relative to the others. There was also an effect of patient, $F(1, 168) = 171.5$, $MSE = 1{,}730.5$, $p < .0001$; confidence was highest in the *both* patient. Patient interacted with learning, $F(1, 168) = 132.4$, $p < .0001$, and question type, $F(1, 168) = 14.5$, $p = .0002$. The former interaction is due to noncontrastive learners' larger difference in summed confidence between the two patients; the latter is due to symmetric questions decreasing confidence in the *both* patient and increasing confidence in the *neither* patient. Most important was the three-way interaction, $F(1, 168) = 18.6$, $p < .0001$: For noncontrastive learners, symmetric questions led to a large decrease in summed confidence for the *both* patient and a large increase for the *neither* patient, whereas for contrastive learners these questions led to slightly higher summed confidence in each patient.

Figure 13 shows mean absolute deviation from 100 averaged across all 16 profiles as a function of learning and question type. It is evident that (a) contrastive learners' judgments were more additive than noncontrastive learners' judgments, and (b) question type had virtually no effect on contrastive learners, but had a large effect on noncontrastive learners. A 2 (learning) × 2 (question type) ANOVA indicated effects of both learning, $F(1, 168) = 86.9$, $MSE = 146.5$, $p < .0001$, and question type, $F(1, 168) = 22.2$, $p < .0001$, and, most important, an interaction, $F(1, 168) = 13.2$, $p = .0004$. Contrasts showed that question type had no effect on contrastive learners, $t(84) = 0.9$, $p = .37$, but symmetric questions increased additivity for noncontrastive learners, $t(84) = 5.2$, $p < .0001$.

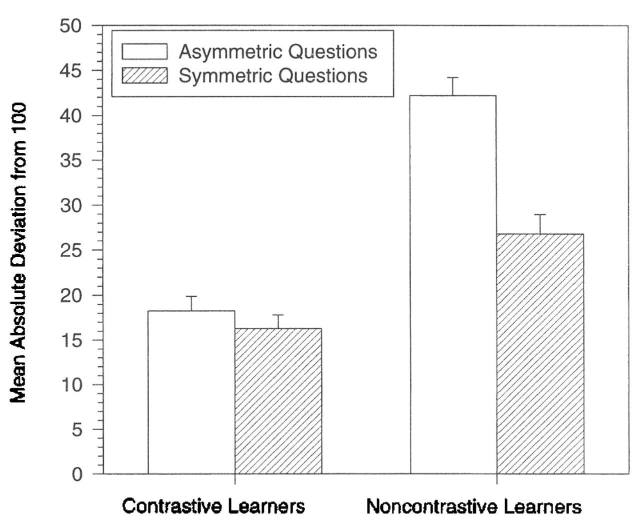

Figure 13. Experiment 4: Mean absolute deviation from 100 for subjective probabilities summed across both illnesses for all 16 patient profiles as a function of learning and question type.

When asked at the end of the experiment about the relationship between the illnesses during the target tasks, 86% to 90% of the participants in the four conditions selected the answer (from among four) indicating that the illnesses were mutually exclusive and exhaustive.

Discussion

The extreme nonadditivity found for noncontrastive learners in Experiment 3 was replicated, and it was found that asking symmetric questions, which encourages taking into account the alternative, had a large impact on noncontrastive but not contrastive learners. Given that others have found that mentioning the alternative leads to changes in behavior (e.g., Baron et al., 1988; Beyth-Marom & Fischhoff, 1983; Zuckerman et al., 1995), perhaps most informative from the current perspective was not so much that question type affected noncontrastive learners, but that it had no effect on contrastive learners.

Symmetric questions did not eliminate differences between contrastive and noncontrastive learners, but the difference between symmetric and asymmetric conditions was only three words ("How confident are you that the patient has [focal illness] *rather than* [*alternative*]?"). Nonetheless, the finding does highlight the question of how, exactly, noncontrastive learners take into account the alternative. The process seems akin to normalization because summed confidence moves toward 100%. For the *both* patient, confidence was high in each illness separately (>50%), and taking into account the alternative resulted in lower values. Similarly, confidence was low in each illness for the *neither* patient (<50%), and taking into account the alternative resulted in higher values. Modeling such changes in confidence is the subject of ongoing research.

Because of the relatively small number of target profiles and lack of feedback during the target tasks, it seems safe to assume that noncontrastive learners' underlying representations did not change as a function of question type. Instead, it seems that symmetric questions changed how noncontrastive learners processed information. That differences in certain behaviors (e.g., reported probability judgments) between contrastive and noncontrastive learners can be reduced (possibly eliminated), even though underlying representations remain unchanged, touches on an important point: According to the current perspective, strength of the alternative can be taken into account without implying a dependent representation (but additional processing is necessary in this case). Goldstone (1996) found that nondiagnostic features had less of an impact on noncontrastive learners as they saw more category instances. His interpretation of this finding was that noncontrastive learners' concepts of the two categories were becoming more interrelated. This is a reasonable interpretation, because participants were continually receiving feedback over a large number of trials. However, Goldstone's model essentially equates taking into account the alternative with interrelated concepts. Experiment 4 indicates that the two need not be equated.

General Discussion

The distinction between dependent and independent confidence appears psychologically real and has important implications for when and how an alternative hypothesis is taken into account. Predictions were largely confirmed across four experiments examining information search, confidence judgments, frequency judgments, decisions, and changes in confidence and their accompanying response times. These results, together with those of Klayman and Brown (1993) and Goldstone (1996), highlight the importance of studying how experience with hypotheses affects representation, which in turn affects subsequent inferences. The results are also consistent with the conclusion by Doherty, Chadwick, Garavan, Barr, and Mynatt (1996) that participants are more likely to be sensitive to the relation between data and alternative hypotheses when the task is highly structured for them. (A similar point has been made by Markman & Medin, 1995, in the context of choice.) One can think of contrastive learning as structuring the target task for the participants, whereas the structure must be imposed by noncontrastive learners.

Assumptions of Mutual Exclusiveness and Exhaustiveness

An assumption throughout the article has been that exactly one of the two competing hypotheses is true. However, this does not always hold in real-world contexts. When diagnosing a patient, even if there are two leading candidates, the patient might have both illnesses (i.e., the hypothesis set is not mutually exclusive) or neither illness (i.e., the set is not exhaustive). However, from the current standpoint, the two assumptions collapse into one, namely, exhaustiveness, because the possibility of both hypotheses (or neither) being true is simply a third hypothesis. If the illnesses were neither mutually exclusive nor exhaustive, this would be a four-hypothesis testing situation: The patient can have A, B, both, or neither. The hypothesis set, theoretically speaking, is back to being mutually exclusive and exhaustive. The issue, then, is not one of whether the set is mutually exclusive and exhaustive, but one of clearly explicating what the mutually exclusive and exhaustive set is. One could, for example, learn what distinguishes each of the four cases, or learn about each separately. With the larger hypothesis set, the same analysis holds, and the same questions are relevant: When and how are alternatives taken into account? For example, if a symptom indicates that the patient has neither A nor B, does confidence in the other three hypotheses decrease, remain the same, or some combination? Are dependent representations of three or more hypotheses feasible? The theoretical distinctions and framework are easily generalized beyond two mutually exclusive and exhaustive hypotheses.

The Role of Environment

Can one make a priori predictions regarding when people will have dependent versus independent confidence? Relatedly, when would it make sense from an adaptive point of view to have one versus the other? The key may lie in whether, when assessing a particular hypothesis, the alternative tends to be the same each time. An example is found in the present experiments, where, for the "paralymphnal specialist," the alternative to puneria is always zymosis. Under such conditions, one will undoubtedly eventually learn to distinguish between the illnesses, resulting in dependent confidence. Mundane examples might include assessing guilt versus innocence, rain versus no rain, or male versus female—hypotheses that are constant competitors. Note that in these examples there appears to be no distinction between evidence for one hypothesis and evidence against the other. For instance, having a motive is likely to be seen as evidence for guilt and, simultaneously, as evidence against innocence.

Though one might learn to distinguish between hypotheses that consistently compete with each other, alternatives to a particular hypothesis often vary. Rather than a specialist, consider a general practitioner who has to determine which of many possible maladies a patient might have. The competing hypotheses might be A versus B in one instance, A versus C in another, and A versus either B or C in yet another. Here, one might expect independent confidence, allowing for the possibility of comparing relative strengths of any hypotheses. Another example is a sports fan who assesses her favorite team's chances of winning each game throughout the season; the alternative (i.e., the opposing team) often changes. (Note that there appears to be a distinction between evidence for one team winning—for example, your team signed a top player—and evidence against the other team winning—for example, their team's star player is injured.) Having independent structures, each composed of the strengths and weaknesses of a particular team, allows for a comparison of relative strengths for each game. This seems more efficient than having a dependent structure corresponding to each pairwise comparison in the league. Furthermore, independent confidence is more efficient when hypotheses are being added to or eliminated from the set. With dependent confidence, each datum is evaluated in terms of all of the hypotheses in the set, and adding or eliminating a hypothesis would mean reevaluating all the evidence. This is not so with independent confidence because each datum is evaluated in terms of individual hypotheses.

Independent confidence allows for still more flexibility. Consider a physician entertaining four hypotheses: The patient has either A, B, both, or neither. Rather than four independent cognitive scales, the physician could conceivably have just two, one corresponding to confidence in A and one to confidence in B. Low confidence on both scales could imply high confidence in neither illness, and high confidence on both scales could imply high confidence that the patient has both. Thus, a separate cognitive scale need not correspond to each hypothesis when confidence is independent.

Another way to view the benefits of independent confidence can be found in a study by Morris and Larrick (1995), who proposed a normative model of causal discounting. A typical discounting scenario involves first estimating the probability that a particular cause (A) produced an effect (E), or $p(A|E)$. It then becomes known that another cause (B) was present, and the question is how confident one should be that A was present, given B is known to have been present, or $p(A|E\&B)$. The decrease between the first and second conditional probability is the amount of discounting. Though Morris and Larrick were primarily concerned with situations in which the second cause is known to be present, their model can be extended to cases in which the presence of the second cause is uncertain (p. 348), which corresponds more closely to the current experiments. One variable in their model is the correlation between the causes. The only situation in which dependent confidence would be functional is when the two causes are perfectly negatively correlated, or mutually exclusive. As the correlation increases, the normative amount of discounting decreases. Dependent confidence cannot accommodate varying degrees of discounting, but independent confidence could. Given that a perfect negative correlation between causes represents a special case, independent confidence appears to have wider applicability.

Though independent confidence has potential advantages, exploiting them requires accessing and comparing relative strengths of the competing hypotheses—additional processing that might not occur, at least under some conditions, as the present experiments demonstrate. Nonetheless, any shortcoming lies in ignoring the strength of alternatives, not in independent confidence per se.

Nonadditive Views of Probability

Traditional normative theories view probabilities as degrees of likelihood, indicating how likely it is that a hypothesis is true. Because only one hypothesis in a mutually exclusive and exhaustive set can be true, additivity holds. Psychologists interested in probability judgment have largely subscribed to such theories and, therefore, to additivity. However, there are normative theories that view probability not as indicating likelihood, but justification (Cohen, 1977; Shafer, 1976). Here, assigning a probability indicates the amount of support for that hypothesis. Such views have potential implications for the present research because they do not assume additivity: Support may increase or decrease for one hypothesis independent of support for the alternative or alternatives.

When assessing the implications of such views of probability for the current research, it is important to distinguish between descriptive and normative issues. The descriptive issue is the following: Can support theories of probability explain the current data? Indeed, nonadditive judgments motivated such theories. One might argue that although participants were asked about likelihood, considerations of support might have influenced responses. However, it is not obvious why only noncontrastive learners' responses would be influenced by consideration of support, or why symmetric questions would lead to less consideration of support for these participants. Furthermore, responses were consistent across confidence judgments, frequency judgments, and decisions. The latter two measures are difficult to interpret in terms of support.

The second issue raised by support theories is a normative one, namely, whether it is an error to ignore the strength of alternatives when reporting subjective probability. A common way to determine error is through appeal to a normative theory. However, different theories claiming normative status can disagree (Gigerenzer et al., 1991; Gigerenzer & Murray, 1987). Such is the case here, where traditional probability theories espouse additivity, but support theories do not. A second way to determine errors is to ask participants what they, upon reflection, think they should do (Kahneman & Tversky, 1982; Slovic & Tversky, 1974; but see Baron, 1994). Participants were asked at the end of Experiments 1 and 2 whether they thought that their confidence in one illness should be affected by symptoms for or against the other. Most participants in each condition thought so. However, judgment errors were not the main focus of the present research. It is important not only that noncontrastive learners often do not take into account the alternative, but also that contrastive learners do. To use just two examples, Figure 5 shows that the two groups made different decisions given the same information, and Figure 11 illustrates the large differences in confidence between groups. Such findings are interesting independent of any normative theory.

Dependent–Independent Confidence as a Continuum Rather Than Dichotomy

The distinction between dependent and independent confidence has been discussed as a dichotomy. But the distinction might best be thought of as a continuum, with complete dependence as one pole and complete independence as the other. The dichotomy implies that dependent confidence entails consideration of the alternative, but empirically this was not always so. For example, 50% or more of contrastive learners were categorized as ignoring the alternative when responding to certain symptoms in Experiment 2 (see Table 7). In addition, though noncontrastive learners exhibited much less confidence in the *neither* patient than the *both* patient, there was also a considerable difference in confidence for contrastive learners (at least in Experiments 1 and 2), indicating some degree of ignoring the strength of the alternative (see Figures 4 and 8). Frequency judgments showed a similar pattern. Though there were clear and important differences between contrastive and noncontrastive learners, these differences were not as extreme as the dichotomy predicts.

Indeed, Goldstone (1996), in his work on isolated and interrelated concepts, proposed a continuum between the two. His instantiation of contrastive and noncontrastive learning is particularly relevant. Both groups saw a series of members from both categories, but the categories alternated more frequently for contrastive learners. That is, noncontrastive learners tended to see long successions of members of the same category, and contrastive learners tended to see a member from one category, then the other. Manipulating the probability of alternation results in learning that is more or less contrastive, and might result in confidence that is more or less dependent. However, Goldstone used only two probabilities of alternation, one high and one low. How might intermediate-contrast participants behave? In terms of the present experiments, as contrast is increased, symptoms common to both illnesses might simply become less likely to affect confidence in either. Similarly, symptoms common to only one illness might become more likely to affect confidence in the illness that, under low contrast learning, is largely unaffected. Such graded effects might occur at the individual level, or perhaps only show up at the group level. How varying degrees of contrast during learning affect information search and use is an open question.

Whether best described as a dichotomy or a continuum, the cognitive distinction between dependent and independent confidence appears useful because of its implications for taking into account alternative hypotheses. The theoretical ideas presented here are simple, yet they led to predictions—some surprising, and all largely confirmed—across a variety of tasks. Because every judgment under uncertainty involves at least two competing hypotheses, understanding when and how alternatives are taken into account has far-reaching theoretical and practical implications. The current perspective is an attempt to help understand these important issues.

References

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes, 39,* 133–144.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences, 17,* 1–42.

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes, 42,* 88–110.

Bassok, M., & Trope, Y. (1984). People's strategies for testing hypotheses about another's personality: Confirmatory or diagnostic? *Social Cognition, 2,* 199–216.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology, 45,* 1185–1195.

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 470–497). Cambridge, England: Cambridge University Press.

Cohen, L. J. (1977). *The probable and the provable.* Oxford, England: Clarendon Press.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory & Cognition, 24,* 644–654.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43,* 111–121.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review, 92,* 433–461.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Hillsdale, NJ: Erlbaum.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90,* 239–260.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129–162). New York: Wiley.

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

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24,* 608–628.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311–327.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24,* 1–55.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11,* 123–141.

Klayman, J. (1988). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 317–330.

Klayman, J., & Brown, K. (1993). Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment. *Cognition, 49,* 97–122.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

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

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica, 64,* 167–185.

Lord, C. G., Lepper, M., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47,* 1231–1243.

Markman, A. B., & Medin, D. L. (1995). Similarity and alignment in choice. *Organizational Behavior and Human Decision Processes, 63,* 117–130.

McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology, 26,* 209–239.

McKenzie, C. R. M. (in press). (Non)complementary updating of belief in two hypotheses. *Memory & Cognition.*

Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review, 102,* 331–355.

Pike, A. R. (1968). Latency and relative frequency of response in psychophysical discrimination. *British Journal of Mathematical and Statistical Psychology, 21,* 161–182.

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 443–456.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40,* 193–218.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120.

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

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 544–551.

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science, 19,* 368–373.

Sniezek, J. A., Paese, P. W., & Switzer, F. S., III. (1990). The effect of choosing on confidence in choice. *Organizational Behavior and Human Decision Processes, 46,* 264–282.

Snyder, M., & Swann, W. B., Jr. (1978). Hypothesis-testing in social interaction. *Journal of Personality and Social Psychology, 36,* 1202–1212.

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

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology, 43,* 22–34.

Trope, Y., & Bassok, M. (1983). Information-gathering strategies in hypothesis testing. *Journal of Experimental Social Psychology, 19,* 560–576.

Trope, Y., Bassok, M., & Alon, E. (1983). The questions lay interviewers ask. *Journal of Personality, 52,* 90–106.

Trope, Y., & Mackie, D. (1987). Sensitivity to alternatives in social hypothesis testing. *Journal of Experimental Social Psychology, 23,* 445–459.

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

Van Wallendael, L. R. (1989). The quest for limits on noncomplementarity in opinion revision. *Organizational Behavior and Human Decision Processes, 43,* 385–405.

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

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probabilistic judgments. *Management Science, 39,* 176–190.

Wright, G., & Whalley, P. (1983). The supra-additivity of subjective probability. In B. P. Stigum & F. Wenstøp (Eds.), *Foundations of utility and risk theory with applications* (pp. 233–244). Dordrecht, The Netherlands: Reidel.

Zuckerman, M., Knee, C. R., Hodgins, H. S., & Miyake, K. (1995). Hypothesis confirmation: The joint effect of positive test strategy and acquiescence response set. *Journal of Personality and Social Psychology, 68,* 52–60.

Received June 24, 1996
Revision received July 24, 1997
Accepted August 27, 1997 ■