(Non)Complementary updating
of belief in two hypotheses

CRAIG R. M. McKENZIE
University of California, San Diego, La Jolla, California

Past research has led to the conclusion that two competing hypotheses are represented dependently, and confidence in them is updated in a complementary manner. It is argued here that confidence in two hypotheses can be represented either dependently or independently. Changes in confidence in the former case are always complementary, but changes in the latter case are complementary only under certain conditions. In three simulated medical diagnosis experiments, subjects learned about two illnesses in a manner expected to lead to either dependent or independent confidence. They were then presented with two symptoms sequentially (for each of several patients), updating confidence after each. Experiment 1 demonstrated that changes in confidence in the two illnesses were largely complementary for subjects with dependent, but not independent, confidence. Experiment 2 showed that encouraging consideration of the alternative led to more complementary changes for subjects with independent confidence. Experiment 3 succeeded in producing complementary changes from these subjects. Thus, complementarity does not imply dependent confidence, nor does independent confidence imply noncomplementarity.

Strength of belief in uncertain states of the world, or hypotheses, often changes as new information is received. Test results might lower a physician's subjective probability that a patient is suffering from a certain illness, a new experiment might lead a scientist to believe more strongly in a particular theory, or a recent encounter with a friend might lower your confidence that she is honest. When updating belief in a particular (or focal) hypothesis, there is always at least one alternative hypothesis. For present purposes, assume that there are two competing hypotheses (i.e., there is only one alternative to the focal hypothesis) and that exactly one is true. For instance, a physician might know that a patient is suffering from a class of illness, of which there are just two types, A and B, but be uncertain about which type the patient has. Any change in confidence in one type should result in a complementary change in confidence in the other.[1] Information that, for example, increases confidence in A should decrease confidence in B.

The topic of this article is when and why belief updating is (non)complementary with two mutually exclusive and exhaustive (MEE) hypotheses. Past research has found that changes in confidence are largely complementary with two hypotheses, but not with three or more. The conclusion reached was that two hypotheses are represented dependently and that three or more are represented independently. It is argued here that confidence in two hy-

potheses can be represented dependently or independently. In the former case, changes are always complementary; in the latter, changes are complementary only under certain conditions.

The first section briefly summarizes past research on complementarity and outlines theoretical notions leading to predictions regarding when changes in two MEE hypotheses will and will not be complementary. Three medical decision-making experiments are subsequently reported that support the predictions. In the final section, implications of the results are discussed.

BELIEF UPDATING AND
COMPLEMENTARITY

Though belief updating has been studied extensively (e.g., Anderson, 1981; Carlson & Dulany, 1988; Edwards, 1968; Hogarth & Einhorn, 1992; Lopes, 1985, 1987; Shanteau, 1970, 1972, 1975; Wallsten & Manley Sapp, 1977), few have examined complementarity. One exception is Van Wallendael and Hastie (1990; Robinson & Hastie, 1985; Van Wallendael, 1989), who used suspects in a murder mystery as MEE hypotheses. Subjects were presented with multiple pieces of evidence and, after each, updated belief in each suspect's guilt. The common finding was that subjects adjusted confidence in the suspect implicated by the evidence but left unchanged their confidence in the other suspects—except when there were only two suspects, in which case changes were complementary. Furthermore, subjects often failed to increase confidence in the remaining suspects after a suspect was eliminated—except when an elimination resulted in just two suspects, in which case confidence (appropriately) increased in both. Thus, there seemed to be something

Table 1
Illustration of Contrastive Learning of Puneria and Zymosis:
The Probabilities of Each Symptom Given Each Illness

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

special about two hypotheses. The conclusion drawn was that two hypotheses are represented dependently, whereas three or more are represented independently.

Dependent Versus Independent Confidence in Two Hypotheses

McKenzie (1998) has argued that confidence in two hypotheses can be represented either dependently or independently.[2] To illustrate the theoretical ideas, imagine again a physician diagnosing which of two illnesses—call them *puneria* and *zymosis*—a patient has. The physician knows that the patient has exactly one of the illnesses and bases the diagnosis on the presence and absence of various symptoms.

The role of learning. Prior to the diagnosis, the physician might have learned to discriminate between the two illnesses. Table 1 illustrates such *contrastive learning* (CL; see also Goldstone, 1996; Klayman & Brown, 1993). The left column lists four symptoms (S1–S4), which can be present or absent. The numbers correspond to the proportion of patients who have the symptom, given the illness. For example, 80% of puneria patients and 20% of zymosis patients have S1. (These numbers need not sum to one or 100%; that they do simply means that a symptom's presence and absence are equally informative.) Contrasting the two illnesses makes clear the diagnosticity of each symptom. It is obvious that S1 is strong evidence for puneria (and strong evidence against zymosis), S2 is weak evidence for puneria, S3 is weak evidence against puneria, and S4 is strong evidence against puneria.

Rather than learning to discriminate between the illnesses, the physician might have learned about them separately. For example, it is feasible that the physician could have learned to discriminate between patients with each illness and healthy patients. Table 2 illustrates such *noncontrastive learning* (NCL). The left side of the table shows learning about puneria, where patients with and without puneria are contrasted. The latter patients tend to have none of the symptoms, so one can think of them as "healthy." Here, S1–S4 are learned to be decreasingly diagnostic of puneria, with S1 strong evidence for puneria and S4 nondiagnostic. The right side of Table 2 illustrates learning about zymosis, where S1–S4 are learned to be increasingly diagnostic of zymosis, with S1 nondiagnostic and S4 strong evidence for zymosis.

Conception of evidence. There are two ways to conceive of evidence in the illnesses. One way is to conceive of just two categories of evidence: that for puneria and that for zymosis. Alternatively, there could be four cate-

gories: for puneria, against puneria, for zymosis, and against zymosis. Importantly, the four-category conception distinguishes between evidence for one illness and evidence against the other, whereas the distinction disappears in the two-category conception.

How we learn about two hypotheses affects whether we conceive of two versus four categories of evidence. CL blurs the "for one/against the other" distinction. Evidence for one illness is simultaneously evidence against the other, leading to just two categories. For example, S1 is learned to be simultaneously evidence for puneria and against zymosis. However, NCL encourages the distinction. S1 is learned to be evidence for puneria but to have no bearing on zymosis. Similarly, S2 is learned to be evidence for both illnesses. Thus, there are four categories of evidence.

Representation of confidence. Confidence in the two illnesses can be represented either dependently or independently. Dependent confidence can be thought of in terms of a single scale, with one pole corresponding to certainty in puneria and the other corresponding to certainty in zymosis. Hence, any change in confidence in one illness leads to a complementary change in the other. However, with independent confidence, each illness can be thought of as having a separate scale, the poles of each corresponding to certainty that the patient does and does not have the respective illness. (In this case, not having the illness corresponds to being healthy.) Changes in confidence in one illness do not necessarily affect confidence in the other in a complementary manner. Indeed, such a representation implies that confidence can change in one illness without any change in the other. Confidence can even increase or decrease in both simultaneously.

What determines whether we have dependent or independent confidence? How we conceive of evidence appears crucial. Conceiving of two categories, the result of CL, leads to dependent confidence. Evidence for puneria (zymosis) increases confidence in that illness and decreases confidence in zymosis (puneria). In contrast, conceiving of four categories, which results from NCL, leads to independent confidence. Evidence conceived of as for (against) puneria increases (decreases) confidence in that illness. Similarly, evidence conceived of as for (against) zymosis increases (decreases) confidence in that illness.

An important aspect of independent confidence, as depicted here, is that it does not necessitate noncomple-

Table 2
Illustration of Noncontrastive Learning of Puneria
and Zymosis: The Probabilities of Each Symptom
Given the Presence and Absence of Each Illness

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

mentarity. One could access strength of the alternative as well as the focal hypothesis and compare relative change between the two. For example, when asked about change in confidence in puneria following an increase only in zymosis, one could calculate the relative change in confidence and report a decrease in puneria. Ignoring the increase in zymosis, however, would result in a noncomplementary change. Because the extra processing involved in accessing the alternative and comparing relative change might not occur, complementary changes are less likely with independent confidence.

Previous findings. In four experiments (McKenzie, 1998), subjects learned to diagnose two illnesses through either CL or NCL, then performed several target tasks in which it was understood that patients had exactly one of the illnesses. In one task (Experiments 1 and 2), subjects reported whether individual symptoms decreased, had no effect on, or increased their confidence in the focal illness. Subjects first performed the task with one illness serving as the focal hypothesis, then later with the other illness. In general, changes in confidence were complementary for CL, but not NCL, subjects. For example, when presented with symptoms common to both illnesses, CL subjects tended to report no change in either illness, whereas NCL subjects tended to report an increase in both.

Other related research. Klayman and Brown (1993) also taught subjects about two illnesses either in parallel or separately, hypothesizing that the former group would form illness concepts composed of diagnostic symptoms, whereas the latter would form concepts composed of typical symptoms. Klayman and Brown's model assumed that diagnoses would be based on the relative number of symptoms that matched each illness concept. Though confidence was not examined, their model seems to imply that changes in confidence in the two illnesses would be complementary. The model does predict differences between the learning conditions (and these predictions were confirmed), but the differences do not regard complementarity. Goldstone (1996) has also argued that learning affects conceptual structure. Interrelated concepts result when they are contrasted; isolated concepts result when they are not. He found that features common to both categories had a larger influence on categorizations when subjects learned about the categories independently. Like Klayman and Brown, however, Goldstone did not examine confidence (his primary dependent measure was percent correct categorizations), and it is not clear what predictions he would make regarding complementarity. Another difference, developed more fully in the General Discussion section, is that Goldstone's model essentially equates taking into account the alternative (to use my terms) with interrelated concepts, whereas the present view treats consideration of alternatives and representation as (partially) separate issues.

Despite differences, there is considerable overlap between the three articles: In each case, the learning manipulation is seen as affecting the perceived relevance of ev-idence (symptoms or features) for hypotheses (illnesses or categories). Because the current focus is on confidence in general and complementarity in particular, one way to view the present experiments is that they broaden the implications of the learning manipulation and the resulting representations.

Rationale Behind the Present Experiments

The motivation behind the present experiments stems from two general sources. The first is that McKenzie's (1998, Experiments 1 and 2) updating task was atypical. In those experiments, subjects were instructed to consider each symptom as coming from a new patient, whereas the essence of belief updating lies in receiving multiple pieces of evidence, making judgments after each for the same hypothesis. Furthermore, rather than reporting confidence per se, subjects reported qualitatively whether their confidence decreased, remained the same, or increased. A more refined measure is desirable. For example, if NCL subjects increase confidence in both hypotheses, it is of interest whether these increases are small or large. Finally, McKenzie's (1998) task was a speeded task; subjects were instructed to respond as quickly as possible while maintaining accuracy. NCL subjects' changes might be complementary without time pressure.

The second motivation regards increasing NCL subjects' complementarity. As noted, independent confidence, according to the present view, does not entail ignoring the strength of the alternative, only that it is less likely because of the additional processing required. If so, then encouraging NCL subjects to take into account the alternative (i.e., do the additional processing) should diminish differences between CL and NCL subjects' performance. McKenzie's (1998) Experiment 4 examined this, but only for confidence based on complete patient profiles, rather than updating based on sequential symptoms. Half the subjects in that experiment were asked symmetric questions (e.g., "How confident are you that the patient has puneria rather than zymosis?") rather than asymmetric questions ("How confident are you that the patient has puneria?"), which were used exclusively in Experiments 1–3. Symmetric questions were expected to encourage consideration of the alternative, thereby affecting NCL but not CL subjects. In this context, consideration of the alternative should lead to increased additivity (see note 1). That was the finding: CL subjects' confidence in the two illnesses summed relatively close to 100 for each patient independent of question type, but NCL subjects' confidence summed closer to 100 when they were asked symmetric rather than asymmetric questions. However, even with symmetric questions, NCL subjects were less additive than CL subjects, raising the question of whether differences between the groups can be erased completely. Another goal of these experiments, then, was to see whether encouragement to consider the alternative reduces—even eliminates—differences between CL and NCL subjects when updating belief.

EXPERIMENT 1

Experiment 1 attempted to replicate the finding in McKenzie's (1998) Experiments 1 and 2 that CL, but not NCL, led to largely complementary changes. In addition to the differences noted above (subjective probabilities reported for the same hypothesis for multiple sequential symptoms with no time constraint), fewer symptoms were used here (4) than in McKenzie's (1998) Experiment 1 (16) and Experiment 2 (8). This change was made because several subjects in the previous experiments failed to complete the learning phase due to its difficulty. Furthermore, the relation between the symptoms and illnesses was different. Table 1 shows that diagnosticity was graded (rather than all-or-none) in order to test more refined predictions. Finally, rather than viewing labeled profiles during learning, subjects saw unlabeled profiles, made a diagnosis, then received feedback. This allowed the monitoring of subjects' learning across trials and held constant the number of puneria and zymosis profiles seen by CL and NCL subjects.

After learning about the two illnesses, subjects were shown profiles of patients known to have either puneria or zymosis. Rather than seeing complete profiles, subjects saw just two symptoms in sequence for each profile, and confidence was updated in the focal illness after each symptom. The dependent variable of interest was change in confidence between the first and second symptoms for a given illness. Table 3 shows predicted change in confidence as a function of learning, focal illness (i.e., which illness is asked about), and second symptom presented. In the case of CL, changes in confidence were always expected to be complementary. Table 1 implies that, when presented second with S1, CL subjects should show a large increase in confidence in puneria and a large decrease in zymosis. When presented with S2, confidence should increase by a small amount in puneria and decrease by a small amount in zymosis, and so on. For NCL subjects, noncomplementary changes were expected. Table 2 implies that, when presented with S1, they should show a large increase in confidence in puneria, but no change in zymosis. When presented with S2 or S3, NCL subjects should increase confidence in both illnesses. Finally, when presented with S4, they were expected to show no change in confidence in puneria but were expected to show a large increase in zy-

Table 3
Predicted Changes in Confidence as a Function of Learning, Focal Illness, and Second Symptom Presented

Focal Illness	Second Symptom			
	S1	S2	S3	S4
	Contrastive Learning			
Puneria	↑(Large)	↑(Small)	↓(Small)	↓(Large)
Zymosis	↓(Large)	↓(Small)	↑(Small)	↑(Large)
	Noncontrastive Learning			
Puneria	↑(Large)	↑(Medium)	↑(Small)	none
Zymosis	none	↑(Small)	↑(Medium)	↑(Large)

Table 4
Example of a Patient Profile During Learning

Symptom	Present?
Coughing	No
Rash	Yes
Nausea	Yes
Fever	No

mosis. These NCL predictions are "strong" in that they assume the alternative illness is ignored, whereas the claim is only that taking it into account is less likely with NCL. To the extent that NCL subjects ignore the strength of the alternative illness, the predictions in Table 3 should hold.

Method

Subjects were 75 University of California, San Diego (UCSD) students enrolled in introductory psychology courses who received partial course credit. The experiment was conducted on computer, with up to 5 subjects tested at one time in the same room. The subjects assumed the role of physicians learning to recognize two illnesses by viewing patient profiles listing four symptoms and whether each was present or absent. Table 4 shows an example of a patient profile.

CL subjects were told they would be shown 60 puneria and 60 zymosis profiles in a random order. For each profile, the subjects made a diagnosis of either puneria or zymosis and were then told the correct diagnosis. The probability of each symptom (S1–S4) given each illness is shown in Table 1. Symptom names (e.g., "rash") were randomly assigned to the four symptoms for each subject. Order of symptoms on the profiles was also random for each subject, and symptoms were conditionally independent.

NCL subjects first began learning about puneria. They were told they would be shown 40 profiles of patients with puneria and 40 profiles of patients without puneria in a random order. The probability of each symptom given puneria and no puneria is shown in Table 2. The subjects diagnosed each profile as either puneria or no puneria and received feedback. They subsequently began learning about zymosis in an analogous fashion. The relevant probabilities are shown in Table 2. NCL subjects then finished learning about puneria by being shown 20 more profiles each of puneria and no puneria, and they finished learning about zymosis similarly. NCL subjects only partially completed learning about one illness before partially learning about the other in an effort to minimize order effects. Though CL and NCL subjects saw a different number of profiles (120 and 240, respectively), both groups saw 60 puneria and 60 zymosis profiles.

All subjects then rated on a 9-point scale how difficult they found learning about the illnesses. After finishing "medical school," they were told:

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.

The subjects then answered the following on the next screen:

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 the subjects did not respond with (c), they were shown the information again and retested. They then read that they would be presented with patients and asked how confident they were in one illness or the other. They would type in 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."

They were then informed:

> Now you are going to see about a dozen patients. However, you will not see a patient's entire profile. Instead, you will first be informed about just one of the symptoms, and you will report your confidence that the patient has [focal illness], given your knowledge about that single symptom. Then, you will find out about a second symptom, and you will report your confidence that the patient has [focal illness], given your new knowledge. You will see just two symptoms for each patient. Not knowing about the other symptoms does not mean that they are absent; you just don't know about them.

They were also told to "keep in mind that each illness is equally common, so you should be about 50% confident in either illness before receiving any information about symptoms." They were reminded that each patient had either puneria or zymosis, but not both. A profile was shown to the subjects with all symptoms covered by Xs except one, which was also highlighted. For simplicity, symptoms were always present (rather than absent). The question was: "Consider the above single symptom. How confident are you that this patient has [focal illness]?" After the subjects responded, a second (highlighted) symptom was revealed, with the first symptom (no longer highlighted) and their first response still visible. The question here was: "Consider this second symptom in addition to the first. Now how confident are you that this patient has [focal illness]?". The subjects were presented, in a random order, with 12 profiles, corresponding to the 12 possible combinations of two different present symptoms. After seeing the 12 profiles with one illness serving as the focal hypothesis, the subjects repeated the task with the other illness. Half of the subjects began with puneria and half began with zymosis. There was no feedback.

Two other non–belief-updating target tasks were subsequently performed. The first involved seeing complete profiles (rather than individual symptoms) and reporting confidence. Those data are reported in McKenzie (1998, Experiment 3). The final task examined hindsight. Finally, the subjects were asked the following:

> Recall when you were a specialist and were asked about your confidence under various circumstances (i.e., AFTER you were finished with learning). Which of the following best described 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.

Results

Learning. CL subjects saw 60 puneria and 60 zymosis profiles. When learning about puneria, NCL subjects saw 60 puneria and 60 no-puneria profiles; when learning about zymosis, they saw 60 zymosis profiles and 60 no-zymosis profiles. In all three cases, performance peaked by the fifth of six blocks of 20 trials, and the two groups performed their respective tasks equally well. Not surprisingly, CL subjects completed the learning phase faster than NCL subjects (Ms = 18.2 and 29.0 min, respectively). The subjects also rated how difficult they found learning; these ratings were similar for CL and NCL subjects (Ms = 5.3 and 5.0, respectively, on a 9-point scale).

Belief updating. Change in confidence between the first and second symptoms presented was the dependent variable analyzed. Data were collapsed across the first symptom presented for each of the four symptoms presented second. For example, data regarding S1 are the mean of three values: change in confidence when presented with (1) S2 followed by S1, (2) S3 followed by S1, and (3) S4 followed by S1. Predictions were independent of the first symptom presented. In each case, confidence was expected to change in a similar manner. Furthermore, the pattern of results was the same when every combination of symptoms was analyzed separately.

Figure 1 shows mean change in confidence as a function of learning, symptom presented second, and focal illness. The top panel shows changes when puneria was focal; the bottom panel shows changes when zymosis was focal. The left group of columns corresponds to CL subjects, whose changes in confidence were qualitatively complementary. S1 and S2 increased confidence in puneria and decreased confidence in zymosis; S3 and S4 decreased confidence in puneria and increased confidence in zymosis. Not only were all eight changes in the predicted direction, but, with the exception of change following S2 when zymosis was focal, each was different from zero (ts > 2.8, ps < .01). The middle group of columns represents NCL subjects, whose changes were noncomplementary. S1 increased confidence in puneria but had no effect on confidence in zymosis; S2 and S3 increased confidence in both illnesses; and S4 increased confidence in zymosis but did not affect confidence in puneria. With the (predicted) exceptions of S4 when puneria was focal and S1 when zymosis was focal, these changes were different from zero (ts > 2.6, ps < .02).

A 2 (learning) $\times$ 2 (illness) $\times$ 4 (symptom) analysis of variance (ANOVA) was conducted on change in confidence, using illness and symptom as repeated measures. There was an effect of learning [$F(1,73) = 28.8$, $MS_e = 625.8$, $p < .001$]. An overall mean change of zero corresponds to perfect complementarity (averaged across symptoms), and this mean was smaller for CL subjects than for NCL subjects (Ms = 3.2 and 14.2, respectively). There were also two interactions: illness $\times$ symptom [$F(3,71) = 62.5$, $MS_e = 514.6$, $p < .001$] and learning $\times$ illness $\times$ symptom [$F = 3.96$, $p = .011$]. The first indicates (unsurprisingly) that change in confidence depended on which illness was focal. The second is more interesting, showing that learning affected how symptoms changed confidence in each illness (e.g., CL led S2 to increase confidence in puneria and decrease confidence in zymosis, whereas NCL led S2 to increase confidence in both illnesses).

The top of Table 5 shows the percent of subjects in each condition exhibiting complementarity for at least two of the four symptoms. (Changes had to be nonzero and in the correct direction.) Twice as many CL subjects as NCL subjects exhibited such complementarity (87% vs. 43%; $p < .001$, two-tailed Fisher's exact test), showing that differences between the groups in terms of mean
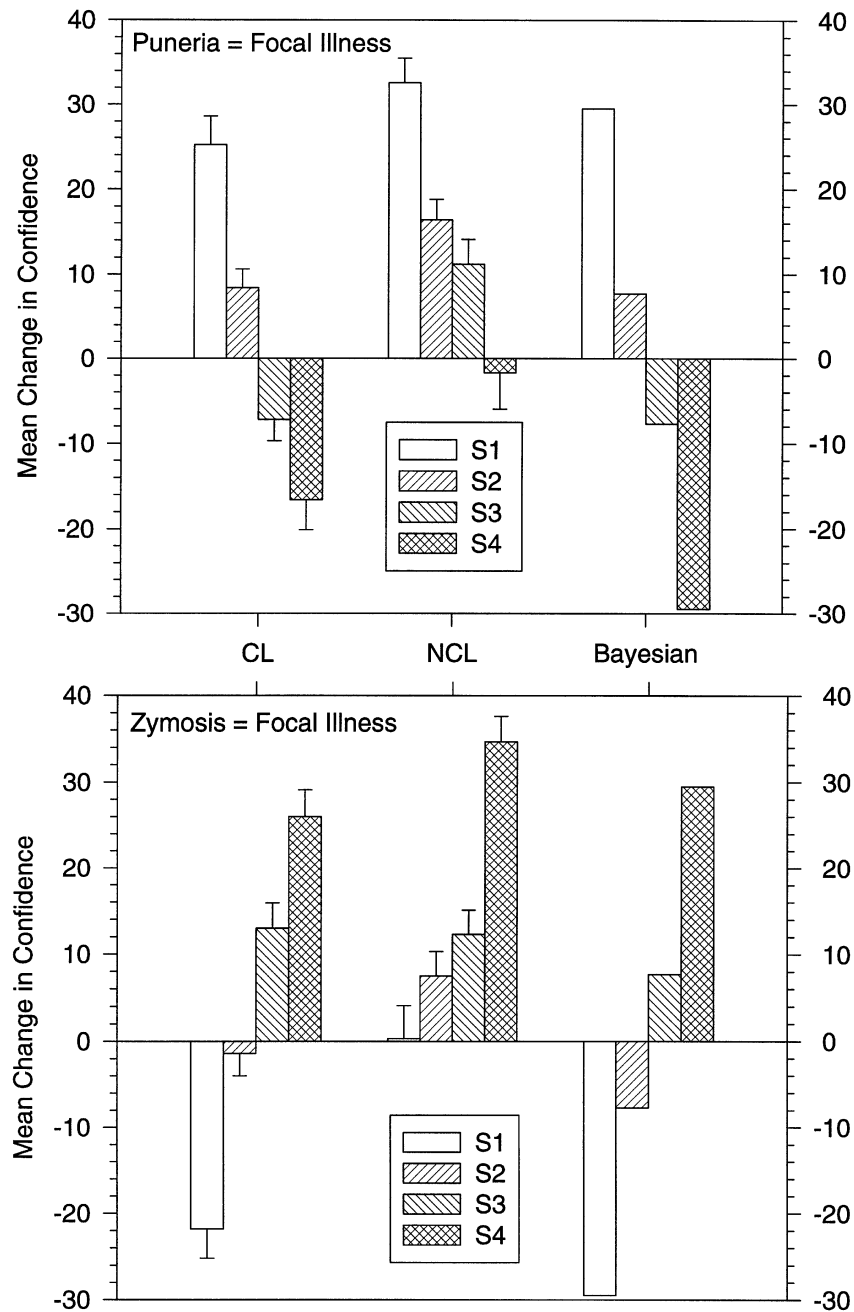
Figure 1. Experiment 1: Mean change in confidence for each of the four symptoms (S1–S4) as a function of learning (CL vs. NCL) and focal illness (the top panel shows changes in confidence when puneria was focal, the bottom when zymosis was focal). The right group of columns represents the Bayesian changes. Standard error bars are shown.

change in confidence were not due to a small number of subjects.

Though it is obvious from Figure 1 that NCL subjects' changes were quantitatively noncomplementary in every case, one might wonder about CL subjects' changes. For example, were the absolute changes in confidence following S1 different when puneria versus zymosis was focal?

They were not ($p > .2$). However, the other three comparisons revealed noncomplementarity (ts > 2.3, ps $< .03$). In each case, the decrease in confidence was smaller than the corresponding increase.

Finally, the right group of columns in Figure 1 shows how confidence should change, according to Bayes's theorem, in puneria (top panel) and zymosis (bottom panel)

Table 5
Percent of Subjects Exhibiting Complementarity
for Two or More of the Four Symptoms

Condition	% Exhibiting Complementarity
Experiment 1	
CL	87
NCL	43
Experiment 2	
CL/asymmetric	86
CL/symmetric	79
NCL/asymmetric	30
NCL/symmetric	55
Experiment 3	
CL/asymmetric/noncontiguous	85
CL/symmetric/noncontiguous	81
CL/asymmetric/contiguous	84
CL/symmetric/contiguous	94
NCL/asymmetric/noncontiguous	49
NCL/symmetric/noncontiguous	60
NCL/asymmetric/contiguous	53
NCL/symmetric/contiguous	80

Note—CL, contrastive learning; NCL, noncontrastive learning.

following each symptom.[3] Note the similarity between the Bayesian changes and those made by CL subjects, and the dissimilarity between the Bayesian changes and those made by NCL subjects.

Subjects' understanding of the relation between the illnesses. The subjects were asked at the end of the experiment about the relation between the illnesses in the target tasks. When selecting among four statements, 36 of 38 CL subjects and 36 of 37 NCL subjects chose the statement indicating that they believed the illnesses were MEE. Eliminating the 3 subjects who had not selected the MEE response had virtually no effect on the results.

Discussion

The results were consistent with the predictions in Table 3. CL led symptoms to be evaluated in terms of both illnesses, resulting in largely complementary changes in confidence. NCL, however, led symptoms to be evaluated in terms of their relation to the focal illness without regard to the alternative, resulting in noncomplementary changes. These findings lend further credence to the psychological reality and importance of the distinction between dependent and independent confidence. Experiment 1 was a particularly strong test because even researchers who conclude that confidence in alternative hypotheses has little (or no) impact on focal hypotheses state that there is one exception to this rule—namely, when there are just two competing hypotheses (Robinson & Hastie, 1985; Van Wallendael, 1989; Van Wallendael & Hastie, 1990; see also Teigen, 1983; Tversky & Koehler, 1994). Thus, NCL subjects' lack of complementarity in the two illnesses is noteworthy.

Can the fact that CL and NCL subjects saw, respectively, 120 and 240 patients during learning account for the re-

sults? One might argue that because puneria and zymosis patients comprised only half of the learning trials, NCL subjects attended less well to them. However, NCL subjects' responses were not simply "noisier" than those of CL subjects, they were consonant with predictions based on ignoring the strength of the alternative illness as opposed to learning less well.

Another potential alternative explanation of the present results is that, because NCL subjects were shown patients who had neither illness during the learning phase, they did not believe the illnesses were exhaustive during the target task. Though 36 of the 37 NCL subjects reported that they believed the illnesses during the updating task were MEE, can the results be explained by assuming that NCL subjects responded normatively given the third possibility that patients had neither illness? No. With this hypothesis set (the patient has puneria, zymosis, or neither illness), S1 should (normatively) decrease confidence in zymosis, but NCL subjects did not change confidence; S2 should not affect confidence in zymosis, but NCL subjects increased confidence; S3 should have no impact on confidence in puneria, but NCL subjects increased confidence; finally, S4 should reduce confidence in puneria, but NCL subjects did not change confidence.[4]

Though CL subjects' changes were qualitatively complementary, they were not quantitatively complementary for three of the four symptoms. The asymmetry was systematic in that the decrease in confidence in one illness was always smaller than the increase in the other. On the face of it, this noncomplementarity is inconsistent with the predictions. However, such asymmetry is common. For example, Robinson and Hastie (1985) found that clues implying guilt were more likely to increase confidence in guilt than clues implying innocence were to decrease confidence. Koriat, Lichtenstein, and Fischhoff (1980) found that self-generated reasons for each of two competing hypotheses had a greater impact on confidence than did reasons against. This general asymmetry says little about the distinctions drawn here regarding dependent versus independent confidence and conceiving of two versus four categories of evidence.

Figure 1 shows a high similarity between CL subjects' changes in confidence and the Bayesian changes (and a low similarity between each of these and NCL subjects' changes). The claim is not that CL subjects are using Bayes's theorem, only that they are taking into account the alternative—an important variable in judgment, generally speaking. McKenzie (1994) showed through computer simulation that judgment strategies taking into account the alternative were highly accurate across a variety of conditions. Considering alternatives also leads to better calibration of subjective probability (Koehler, 1994; Koriat et al., 1980). In none of these cases is the argument made that subjects are using a normative strategy—in fact, in McKenzie (1994), the strategies are explicitly nonnormative—only that a qualitative strategy that takes into account the alternative leads to improved performance. Indeed, the theoretical notions outlined above lead to the

prediction that encouraging consideration of the alternative will lead NCL subjects to perform more similarly to CL subjects, who should not be affected.

EXPERIMENT 2

The subjects in Experiment 1 were asked asymmetric questions (e.g., "How confident are you that the patient has puneria?"). Considerable evidence indicates that symmetric questions ("How confident are you that the patient has puneria *rather than zymosis*?") encourage subjects to take into account the alternative (e.g., Baron, Beattie, & Hershey, 1988; Beyth-Marom & Fischhoff, 1983; Zuckerman, Knee, Hodgins, & Miyake, 1995). Because only NCL subjects ignore the alternative, symmetric questions should affect NCL subjects but not CL subjects. The more that NCL subjects take into account the alternative, the more similar their changes in confidence should be to those of CL subjects—and to the Bayesian changes. In terms of Figure 1, when puneria is focal, NCL subjects' changes for S2, S3, and S4 should decrease. Indeed, if NCL subjects were to take the alternative completely into account, the latter two changes should become negative. Similarly, when zymosis is focal, changes when presented with S1, S2, and S3 should decrease, with full consideration of the alternative leading the former two changes to become negative.

Method

The subjects were 172 UCSD students from the same population as in Experiment 1. The procedure was identical to that of Experiment 1, except that during the target task, half of the CL and NCL subjects were asked asymmetric questions (ns = 44) and half were asked symmetric questions (ns = 42). Results from the task in which the subjects reported confidence in complete profiles are reported in McKenzie (1998, Experiment 4) and were discussed above.

Results

Learning. As in Experiment 1, both learning groups' number of correct diagnoses appeared to have reached asymptote by the end of the learning phase, and both groups found learning about equally difficult (Ms = 5.2 and 5.0).

Belief updating. Figure 2 shows mean change in confidence for the CL asymmetric, CL symmetric, NCL asymmetric, and NCL symmetric conditions as a function of focal illness and symptom. The top panel shows change in confidence when puneria was focal; the bottom panel shows change in confidence when zymosis was focal. Note first that the pattern of results for the CL asymmetric condition (first set of columns) replicates Experiment 1: S1 and S2 increased confidence in puneria and decreased confidence in zymosis, and S3 and S4 decreased confidence in puneria and increased confidence in zymosis. Furthermore, asking CL subjects symmetric questions had virtually no effect (cf. first and second set of columns), though perhaps changes after S2 and S3 moved toward zero. Regardless of question type, CL

subjects' changes were qualitatively similar to those prescribed by Bayes's theorem (Figure 1, right side).

The pattern of results for the NCL asymmetric condition (third set of columns) was also similar to Experiment 1 in that changes in confidence were clearly non-complementary. Importantly, NCL subjects were affected by question type. As predicted, symmetric questions (last set of columns) "pushed down" changes for each of S2, S3, and S4 when puneria was focal (top panel) and each of S1, S2, and S3 when zymosis was focal (bottom panel). Symmetric questions moved NCL subjects' changes in the direction of CL subjects' and the Bayesian changes. Furthermore, the effect was strong enough to reverse the qualitative impact of two symptoms: S4 (S1) tended to increase confidence in puneria (zymosis) with asymmetric questions but decreased confidence with symmetric questions.

A 2 (learning) $\times$ 2 (question type) $\times$ 2 (illness) $\times$ 4 (symptom) ANOVA revealed an effect of learning: CL subjects' changes in confidence were again more complementary than those for NCL subjects, evidenced by an overall mean closer to zero (Ms = 3.7 and 11.9, respectively) [$F(1,168) = 39.22, MS_e = 581.6, p < .001$]. There was also an effect of question type ($F = 8.83, p = .003$), but it was due to its interaction with learning ($F = 10.12, p = .002$): Question type had no effect on CL subjects (asymmetric and symmetric Ms = 3.5 and 3.8, respectively) but had a large effect on NCL subjects (Ms = 15.8 and 7.8, respectively). Thus, symmetric questions led to more complementary changes for NCL subjects but not CL subjects. That is the most important finding. In addition, there was a main effect of symptom [$F(3,166) = 2.84, MS_e = 141.4, p = .04$]; increases in confidence were larger than decreases for S1 and S4 to a greater extent than for S2 and S3. This effect was small and not meaningful in the present context. There were also illness $\times$ symptom and learning $\times$ illness $\times$ symptom interactions (ps < .001), replicating Experiment 1.

At the individual-subject level, the middle of Table 5 shows that most CL subjects reported complementary changes for at least two symptoms regardless of question type ($p = .40$, two-tailed Fisher's exact test). In contrast, whereas only 30% of the NCL subjects who were asked asymmetric questions reported such changes, symmetric questions nearly doubled that number to 55% ($p = .028$). Finally, across the four conditions, 86%–90% of the subjects indicated at the end of the experiment that they believed the illnesses were MEE. Eliminating those not choosing the MEE response had virtually no impact on the results.

Discussion

Symmetric questions, which encourage consideration of the alternative, led to increased complementarity for NCL subjects but not CL subjects. The effect was strong enough such that the qualitative impact of some symptoms was reversed: These symptoms tended to increase NCL
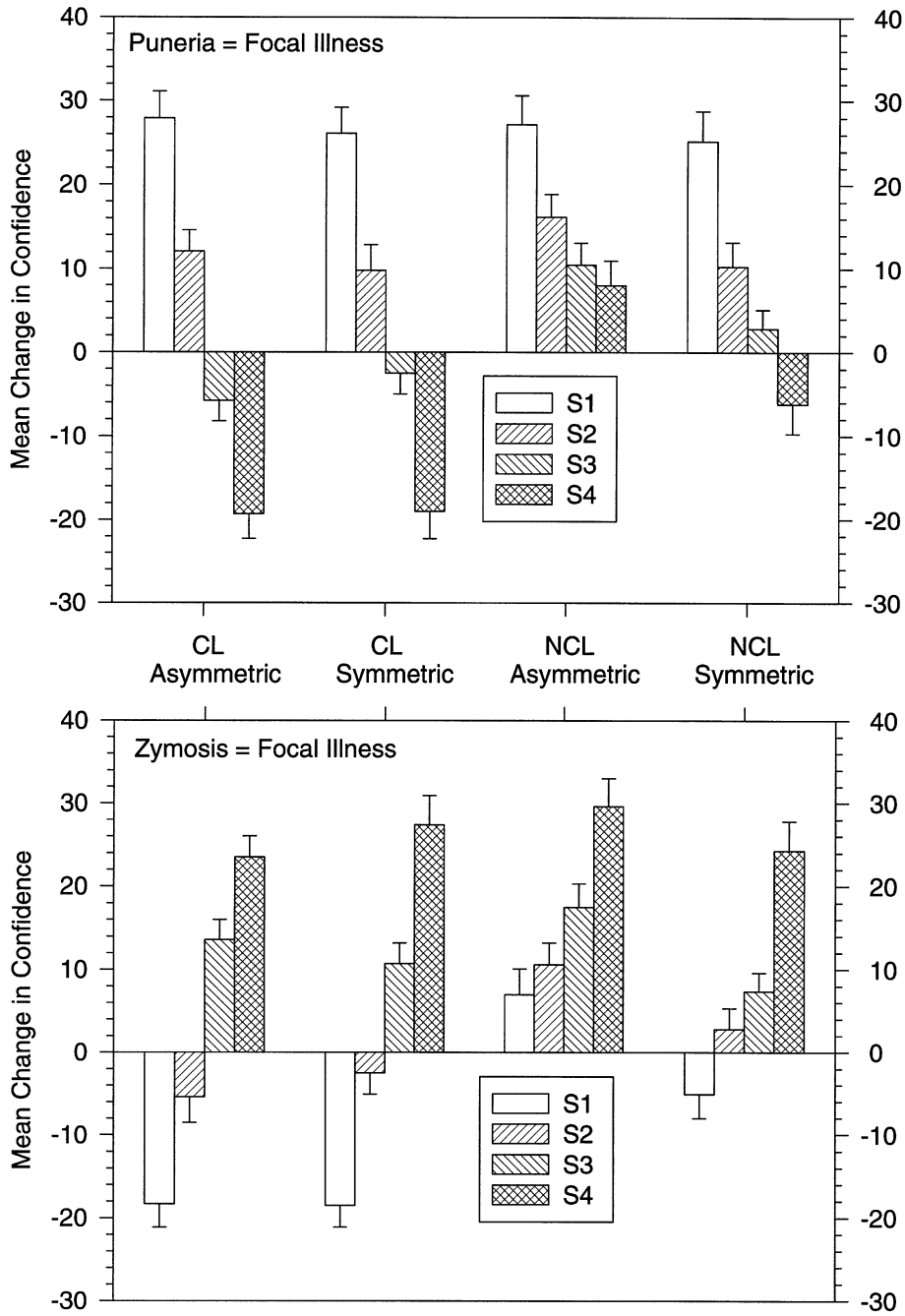
Figure 2. Experiment 2: Mean change in confidence for each symptom (S1–S4) as a function of learning (CL vs. NCL), question type (asymmetric vs. symmetric), and focal illness (the top panel corresponds to puneria, the bottom to zymosis).

subjects' confidence when questions were asymmetric but reduced confidence when questions were symmetric. NCL subjects behaved more similarly to CL subjects and to the prescriptions of Bayes's theorem. Consistent with the theoretical ideas presented earlier, ignoring the alternative is not a necessary outcome for NCL subjects.

Though differences in performance between the two learning groups were reduced, they were not eliminated. However, the difference between the two types of question was rather subtle; only three words ("rather than [alternative illness]") distinguished the conditions. Can differences in behavior be reduced further?

EXPERIMENT 3

As mentioned, Van Wallendael and Hastie (1990; Robinson & Hastie, 1985; Van Wallendael, 1989) found that subjects updated confidence in two hypotheses in a complementary fashion. An important difference between those studies and the present Experiments 1 and 2 concerns how confidence ratings were assessed. In the other studies, confidence was reported in each of the hypotheses after each datum. In the present Experiments 1 and 2, the subjects rated confidence in both hypotheses, but at different times. The former method is conceivably more likely to lead to complementarity. In Experiment 3, an attempt was made to further reduce differences between CL and NCL subjects by asking about confidence in each illness contiguously. After being presented with a symptom and reporting confidence in one illness, some subjects were immediately asked about their confidence in the other illness.

Method

The subjects were 281 UCSD undergraduates from the same population as in Experiments 1 and 2. The procedure was identical to that of Experiment 2, but whether or not the questions were contiguous was manipulated in addition to learning and question type, resulting in eight groups.

Results

Data from 7 subjects were not included in the analysis because they did not perform better than chance during the last half of the learning trials. (NCL subjects' performance was averaged across the last half of trials for both illnesses.) An additional subject was eliminated for responding "50" to every question. Data from the remaining 273 subjects were analyzed. There were between 31 and 36 subjects in each condition.

Learning. Both CL and NCL subjects again learned across trials and had essentially reached asymptote by the end of the learning session, and both groups again found learning about equally difficult ($Ms = 5.5$ and 5.3).

Belief updating. Figure 3 shows the results for each of the eight conditions. The left side (the first four sets of columns) corresponds to the CL subjects. The top panel shows change when puneria was focal; the bottom panel shows change when zymosis was focal. There is no obvious trend for CL subjects when moving from the asymmetric/noncontiguous condition (far left) to the symmetric/contiguous condition (fourth set of columns). There are some apparent anomalies, however. In particular, the asymmetric/noncontiguous CL subjects, on average, tended not to change confidence in puneria when presented with S3, nor did they change confidence in zymosis when presented with S2. Furthermore, for the asymmetric/contiguous CL group, the increase in confidence in zymosis when presented with S4 was unusually small. For each CL group, though, changes in confidence were largely complementary.

Changes in confidence for NCL subjects are shown on the right of Figure 3 (last four sets of columns). There is

a trend moving from left to right: S2–S4 were "pushed down" when puneria was focal, as were S1–S3 when zymosis was focal. Symmetric questions (third set from right) and contiguous questions (second from right) independently resulted in S4 decreasing confidence in puneria and S1 decreasing confidence in zymosis. Symmetric, contiguous questions in combination (far right) resulted not only in larger effects on S4 (S1) when puneria (zymosis) was focal but led S3 to decrease confidence in puneria and S2 to decrease confidence in zymosis. The end result was changes in confidence that were complementary and qualitatively similar to both CL subjects' and the Bayesian changes (Figure 1).

The bottom of Table 5 shows that these mean between-group differences were not due to just a few extreme subjects. Note that 80% of the NCL subjects exhibited complementarity for at least two symptoms when asked symmetric, contiguous questions. This number was not only larger than that of the other NCL groups combined ($p = .009$, two-tailed Fisher's exact test), it was not different from that of the CL subjects ($p = .43$).

A 2 (learning) $\times$ 2 (question type) $\times$ 2 (contiguity) $\times$ 2 (illness) $\times$ 4 (symptom) ANOVA revealed an effect of learning, with CL subjects exhibiting greater overall complementarity than NCL subjects ($Ms = 2.7$ and 7.9, respectively) [$F(1,265) = 30.56$, $MS_e = 241.3$, $p < .001$]. There was also an effect of question type ($F = 11.31$, $p < .001$) due to its interaction with learning ($F = 9.43$, $p = .002$). Symmetric rather than asymmetric questions led to more complementarity for NCL subjects [$Ms = 4.9$ and 11.0, respectively; $t(137) = 3.52$, $p < .001$] but had no effect on CL subjects ($Ms = 2.6$ and 2.9, respectively; $p = .67$). All these findings replicate Experiment 2. In addition, there was an effect of contiguity, with contiguous questions leading to more complementarity than noncontiguous questions ($Ms = 3.5$ and 7.1, respectively) ($F = 15.54$, $p < .001$). The learning $\times$ contiguity interaction was not significant ($p = .14$), although contiguous rather than noncontiguous questions led to a larger increase in complementarity for NCL subjects ($Ms = 5.3$ and 10.4, respectively) than for CL subjects ($Ms = 1.6$ and 3.8, respectively). Both these differences were significant ($ts = 2.9$, $ps = .004$). There were also the usual illness $\times$ symptom and learning $\times$ illness $\times$ symptom interactions ($ps < .001$).

Finally, 94% of both CL and NCL subjects selected the MEE option at the end of the experiment. Across the eight conditions, the range was 88%–100%. There was no systematic preference for any one of the three remaining responses, and eliminating the subjects not choosing the MEE response had virtually no impact on the results.

Discussion

A combination of symmetric and contiguous questions succeeded in getting NCL subjects to update confidence in a qualitatively complementary fashion. Under these conditions, NCL subjects' changes in confidence were qualitatively similar to both CL subjects' responses
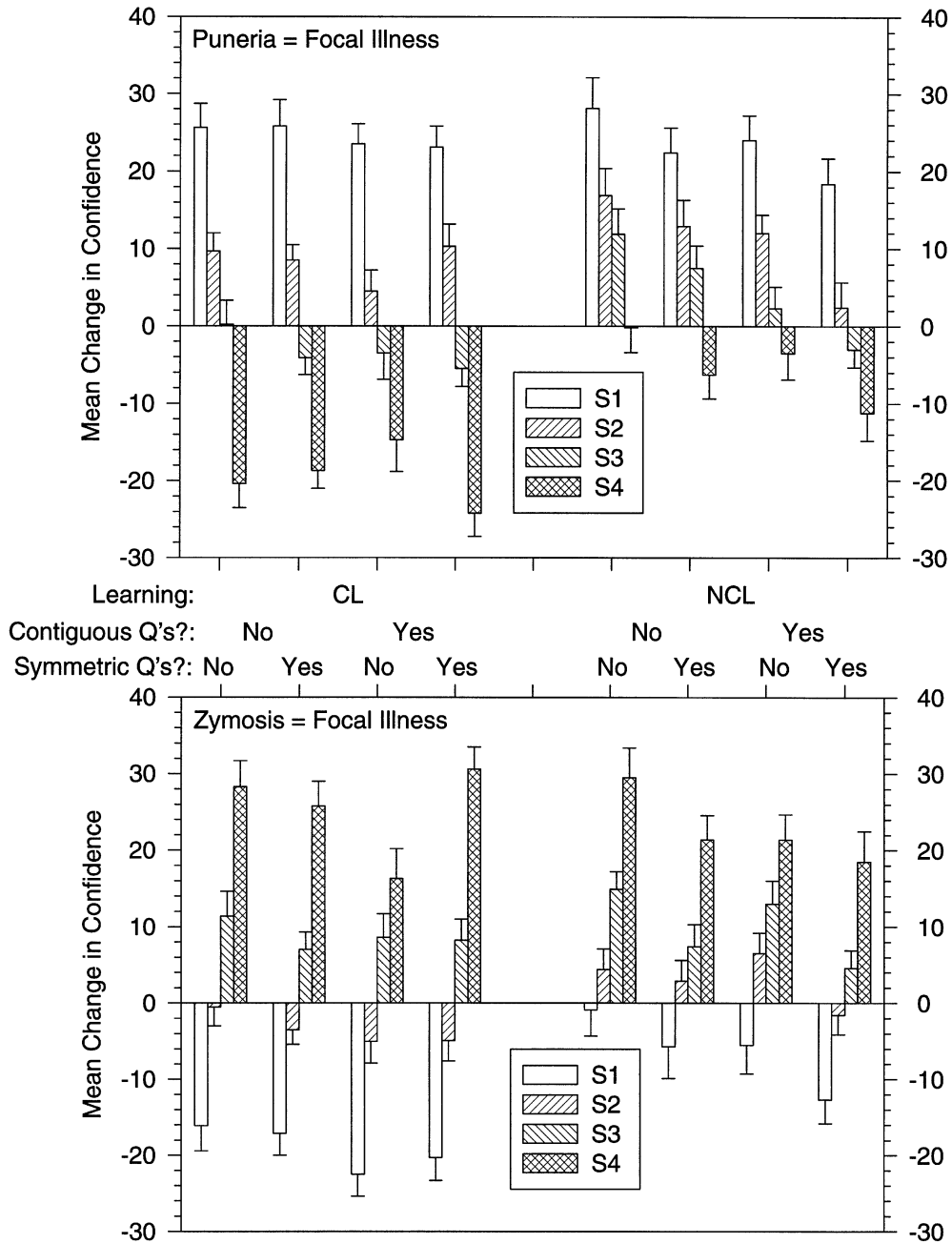
Figure 3. Experiment 3: Mean change in confidence for each symptom (S1–S4) as a function of learning (CL vs. NCL), question type (asymmetric vs. symmetric), contiguity (contiguous vs. noncontiguous questions), and focal illness (puneria is focal in the top panel, zymosis in the bottom).

and the Bayesian responses. Noncomplementarity is more likely with independent confidence, but it is not necessary. Task variables can overcome differences in representation.

These findings speak against an alternative view of the earlier results—namely, that CL subjects outperformed NCL subjects in the target task because it was more similar to the learning conditions for CL subjects. Not only are the (categorization) learning task and the

(belief-updating) target task different, the fact that NCL subjects performed similarly to CL subjects in Experiment 3 (and Experiment 2, to some extent) shows that different learning histories are not the whole story. Furthermore, task variables affected NCL subjects—who had the same learning history—in Experiments 2 and 3.

There are at least two interpretations of the effect of contiguity. One is that contiguous questions, like symmet-

ric questions, encouraged consideration of the alternative. The other is that they simply led the subjects to respond with complementary changes. To see the difference, consider NCL subjects who are asked contiguous questions when S1 (strong evidence for puneria) is presented and puneria is asked about first. If contiguous questions encourage consideration of the alternative, then confidence will increase in puneria and decrease in zymosis. Furthermore, it should not matter which illness is asked about first; confidence should increase in puneria and decrease in zymosis. Indeed, illness order should not effect any symptom. In contrast, contiguous questions might simply lead subjects to make their second judgment complementary to their first. If puneria were asked about first, S1 would increase confidence in puneria and, in order to be complementary, would decrease confidence in zymosis. However, if zymosis were asked about first, there would be no change in confidence, and, in order to be complementary, there would be no change reported in puneria. Thus, if contiguous questions led NCL subjects to report complementary judgments without consideration of the alternative, there should be an effect of illness order. This is true for all four symptoms. Using only NCL subjects who were asked contiguous questions, a 2 (illness order) $\times$ 2 (question type) $\times$ 2 (illness) $\times$ 4 (symptom) ANOVA was performed on change in confidence. There were no effects involving illness order, lending credence to the hypothesis that the effect of contiguity for NCL subjects was due to increased consideration of the alternative.

Unfortunately, it is impossible to determine which interpretation of the contiguity effect led to more complementarity for CL subjects because both make the same predictions. It is worth noting, though, that the apparently anomalous changes in confidence for the CL subjects might have led to the effect. The lack of change for S3 when puneria was focal and S2 when zymosis was focal in the asymmetric/noncontiguous condition, as well as the relatively small increase following S4 when zymosis was focal in the asymmetric/contiguous condition, all favored complementarity for contiguous questions.

GENERAL DISCUSSION

Three experiments showed that learning to distinguish between two illnesses (CL) led to dependent confidence in the illnesses, whereas learning about the illnesses separately (NCL) led to independent confidence. With dependent confidence, changes in confidence were always (qualitatively) complementary. With independent confidence, changes were complementary to varying degrees, depending on task variables. When questions were symmetric, which encouraged consideration of the strength of the alternative, or when questions about the competing hypotheses were asked contiguously, NCL subjects' complementarity increased. When symmetric and contiguous questions were asked in combination, changes were com-

plementary. Furthermore, under these conditions, NCL subjects' responses were similar to both CL subjects' responses and the Bayesian responses.

Representation and Consideration of Alternatives

Because NCL subjects saw relatively few patients during the updating task and did not receive feedback, it is unlikely that how questions were asked affected their representation of confidence. Instead, symmetric and contiguous questions apparently changed how NCL subjects processed information—that is, these subjects did (to various degrees) the additional processing necessary for taking into account the alternative with independent confidence. One cannot conclude that if the alternative is taken into account, confidence is dependent (though it does appear reasonable to conclude that if the alternative is *not* taken into account, confidence is independent). Goldstone (1996) found that nondiagnostic features had less impact on NCL subjects as they saw more category instances, and he concluded that their concepts of the two categories were becoming more interrelated. This is a reasonable interpretation because his subjects continually received feedback over a large number of trials. However, Goldstone's model equates taking into account the alternative with interrelated concepts. The present results indicate that consideration of the alternative is also possible with isolated concepts.

Similarly, Van Wallendael and Hastie (1990) equated complementarity with dependent representations of hypotheses. When their subjects were presented with clues regarding which of two suspects committed a crime, changes in confidence were complementary (see also Robinson & Hastie, 1985; Van Wallendael, 1989). Was that the result of dependent confidence in the two hypotheses (either Suspect A or Suspect B was guilty) or the contiguous questions? According to the present view, it was most likely the latter, because, if there is a distinction between evidence for one hypothesis and evidence against the other, confidence is independent. Note that one can distinguish between evidence for A's guilt (e.g., A has a motive) and evidence against B's guilt (e.g., B has an alibi). However, consider a case with a single suspect. This situation is logically identical to the two-suspect case in that there are still two MEE hypotheses: The suspect is either innocent or guilty. Here, though, confidence in the two hypotheses is expected to be dependent because there is no distinction between, for example, evidence for guilt and evidence against innocence. If the single suspect had a motive to commit the crime, is that evidence for guilt or against innocence? It is, of course, both simultaneously. Thus, with a single suspect, confidence in the two hypotheses is expected to be dependent and changes in confidence complementary. With two suspects, however, confidence in the two hypotheses is expected to be independent and changes in confidence sensitive to task variables, such as how confidence is probed.

Did NCL Subjects Ignore the Alternative or Consider the Wrong Alternative?

When NCL subjects were asked asymmetric and/or noncontiguous questions, they tended to respond as though the alternative to the focal illness was "healthy" rather than the other illness. Is their behavior, then, better described as considering the wrong alternative rather than ignoring the alternative? The answer is that these two descriptions are identical. Whenever an uncertain judgment is made regarding a hypothesis, there is always at least one alternative. What the alternative is (or alternatives are) depends on the context. One might assess the likelihood of rain versus no rain one day, and the likelihood of rain versus snow another. A physician might assess the probability of Illness A versus not-A in one case, and the probability of A versus B in another. When $X\%$ confidence is reported, the judge is presumably $100 - X\%$ confident in some alternative. The question is whether it is the "correct" alternative. When it is agreed on as to what the correct alternative is, claims about ignoring the strength of the (correct) alternative can be made. Such is the case for the experiments reported here. The subjects were told repeatedly that their patients had exactly one of the two illnesses, they had to show an understanding of this before performing the target tasks, and they reported such an understanding at the end of the experiment. At issue was whether NCL subjects would take into account the strength of the alternative illness. They tended not to, except when task variables encouraged them to do so.

Is CL Necessarily Superior to NCL?

CL and NCL are defined relative to a target task. For example, if the present NCL subjects' target task consisted of assessing the likelihood of puneria versus no puneria, they would then be CL subjects. As such, they would be expected to have dependent confidence in the competing hypotheses and update belief in a complementary manner. For a given target task, having dependent rather than independent confidence will generally lead to better performance because the alternative is more likely to be taken into account.

However, there might be advantages to independent confidence because of its flexibility (McKenzie, 1998). Consider again the fact that alternatives to a given hypothesis change with context. If a physician were continually evaluating A versus B, dependent confidence in the two illnesses would be ideal. But B might not always be A's alternative; different illnesses might compete with A under different circumstances. Independent confidence in individual illnesses allows for comparing the relative strengths of any competing illnesses. Koehler (1996) has made a similar point in the context of applying a strength model of probability judgments to tournaments. Note, though, that exploiting the flexibility of independent confidence requires accessing the strength of the alternative and comparing it with the strength of the focal hypothesis— processing that might not occur, as the present experiments demonstrated. Part of becoming an expert diagnostician (or oddsmaker) might not be forming dependent confidence so much as accessing relevant alternatives.

Noting costs and benefits of the different representations leads to speculation about when each might be expected. If the competing hypotheses are often the same (as in the case of a specialist), one might learn to distinguish between them, resulting in dependent confidence. If the alternatives to a particular hypothesis tend to differ depending on the circumstances (perhaps in the case of a general practitioner), then independent confidence might result.

Conclusion

When updating belief in a particular hypothesis, there is always at least one alternative. Whether (or to what extent) the strength of the alternative is taken into account affects complementarity in belief updating. The distinction between dependent and independent confidence has implications for consideration of alternatives and, therefore, for complementarity. It was found that learning to discriminate between two hypotheses led to dependent confidence and to complementary updating. In contrast, learning separately about the hypotheses led to independent confidence and to noncomplementary updating. However, differences in behavior were reduced by task variables. Asking about confidence in a manner that encouraged consideration of the alternative increased complementarity for subjects with independent representations. The final experiment succeeded in getting these subjects to update beliefs in a complementary manner. Thus, independent confidence does not necessarily imply noncomplementarity, nor does complementarity necessarily imply dependent confidence.

REFERENCES

Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.

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

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

Carlson, R. A., & Dulany, D. E. (1988). Diagnostic reasoning with circumstantial evidence. *Cognitive Psychology*, **20**, 463-492.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmutz (Ed.), *Formal representation of human judgment* (pp. 17-52). New York: Wiley.

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

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

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.

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

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

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

Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society*, **23**, 509-512.

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

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. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 771-792.

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

Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, **85**, 181-191.

Shanteau, J. (1972). Descriptive versus normative models of sequential inference judgment. *Journal of Experimental Psychology*, **93**, 63-68.

Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica*, **39**, 83-89.

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

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 & Human Decision Processes*, **43**, 385-405.

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

Wallsten, T. S., & Manley Sapp, M. (1977). Strong ordinal properties of an additive model for the sequential processing of probabilistic information. *Acta Psychologica*, **41**, 225-253.

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 & Social Psychology*, **68**, 52-60.

NOTES

1. This normative claim is based on the assumption—met in the experiments reported here—that *confidence* is measured in terms of subjective probability. It is also worth distinguishing between what I refer to as *complementarity* and *additivity*. With two mutually exclusive and exhaustive hypotheses, I use the former term to mean that a change of $X\%$ in one hypothesis results in a change of $-X\%$ in the other. I use the latter term to mean that the probabilities for the hypotheses sum to 100%. Thus, complementarity is necessary, but not sufficient, for additivity.

2. Van Wallendael and Hastie (1990) refer to *hypotheses* as being (in)dependent, whereas I refer to *confidence*. The distinction in this case appears semantic. We both largely rely on patterns of reported changes in confidence to make inferences about representation.

3. Bayesian changes were calculated in the following manner. Assume that puneria is focal and S1 is presented first. According to Bayes's theorem, the posterior probability, $p(\text{puneria} \mid \text{S1})$, is equal to $[p(\text{puneria}) \times p(\text{S1} \mid \text{puneria})] / p(\text{S1})$. The prior probability, $p(\text{puneria})$, is the probability that the patient has puneria before S1 is observed. The likelihood, $p(\text{S1} \mid \text{puneria})$, is the probability of observing S1, given the patient has puneria. The likelihoods for each symptom and illness are shown in Table 1 (e.g., $p[\text{S1} \mid \text{puneria}] = .8$). Furthermore, the unconditional probability of observing the symptom, $p(\text{S1})$, is equal to $p(\text{puneria}) \times p(\text{S1} \mid \text{puneria}) + p(\text{zymosis}) \times p(\text{S1} \mid \text{zymosis})$. Before the first symptom was presented, it was assumed (as the subjects were instructed) that $p(\text{puneria}) = p(\text{zymosis}) = .5$. Substituting the appropriate quantities into the above equations, $p(\text{puneria} \mid \text{S1}) = (.5)(.8)/[(.5)(.8) + (.5)(.2)] = .8$. The measure of interest was change in confidence between the first and second symptoms presented. For the second symptom, the former posterior probabilities become the new prior probabilities; that is, $p(\text{puneria}) = .8$, and $p(\text{zymosis}) = .2$. Assume that S2 is presented after S1. Bayes's theorem is applied again using the new priors and the likelihoods for S2 (Table 1): $p(\text{puneria} \mid \text{S2}) = (.8)(.6)/[(.8)(.6) + (.2)(.4)] = .857$. Thus, the Bayesian change in confidence when S2 follows S1 is $.857 - .8 = .057$. Like the empirical results, the Bayesian changes for each symptom presented second were averaged across the other three symptoms presented first. The normative change when S2 was presented second is the mean change when it followed S1 (.057, calculated above), S3 (.1), and S4 (.073), which equals .077. This is the value ($\times 100$) shown on the right side of Figure 1 for the Bayesian change for S2 when puneria was focal. Such calculations were done for each symptom and illness.

4. I am assuming equal priors for the three hypotheses, but NCL subjects saw twice as many patients without the illnesses as patients with puneria or zymosis. With these different priors—$p(\text{puneria}) = p(\text{zymosis}) = .25$, $p(\text{no illness}) = .5$—S1 and S4 should still change confidence in zymosis and puneria, respectively (a decrease of .213, or 21.3%, in both cases).