

- Zwick, R., Carlstein, E., & Budescu, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, 1, 221-242.
- Zwick, R., & Wallsten, T. S. (1989). Combining stochastic and linguistic inexactness: Theory and experimental evaluation of four fuzzy probability models. *International Journal of Man and Machine Studies*, 30, 69-111.

## COMPOSITIONAL ANOMALIES IN THE SEMANTICS OF EVIDENCE

*John M. Miyamoto, Richard Gonzalez, and Shihfen Tu*

### I. Introduction

Semantic theory plays a central role in the normative and descriptive theory of deductive inference, but its role in the study of inductive inference has been much less prominent. The disparity is both odd and understandable. It is odd because deductive and inductive reasoning both rely heavily on linguistic representations, and semantic theory is the natural tool for investigating inference within propositional structures. It is understandable because the logical approach to induction, as championed by Lambert, Keynes, and Carnap, to name only a few, was eclipsed by important developments in the theory of subjective probability, following the work of De Finetti, Savage, and others. From a semantical perspective, the basis for subjective probability theory is very elementary, namely, that there exist Boolean algebras of events that are ordered by how strongly one believes that they are true (the belief strength ordering). The axiomatic theory of subjective probability specifies further properties of belief strength that characterize so-called coherent beliefs; if strength of belief is coherent in this sense, then there exist numerical probabilities that represent the belief strength ordering and satisfy the mathematical laws of probability (Fine, 1973; Krantz, Luce, Suppes, & Tversky, 1971).<sup>1</sup>

<sup>1</sup> There is, of course, an alternative approach to formalizing subjective probability in terms of preferences among uncertain options or *acts* (Fine, 1973). This approach, due to Savage (1954), will not be discussed here because the phenomena in this paper concern strength of belief only, and not preferences for risky or uncertain options.

Historically, the study of natural language semantics has been closely linked to theories of deductive inference because deductive relations among propositions serve as clues to semantic structure (Davidson & Harman, 1975a; Quine 1960). A theory of semantics relates three aspects of language: (1) the syntactic structure of propositions, that is, a specification of how complex propositions are built from simpler parts; (2) the semantic structure of propositions, that is, a specification of the relation between propositional structure, reference, and truth values; and (3) inference rules that define inferential relations in terms of syntactic and semantic structure. Together these three aspects constitute a compositional theory of inference, a theory of the relationship between deductive inference and the compositional structure of propositions. A classical example of this line of analysis is the inference rule called *modus ponens*, according to which the truth of a proposition  $Q$  may be inferred, given that *if  $P$ , then  $Q$*  and  $P$  are true. The truth table for the material conditional (the if-then statement of the propositional calculus) is a semantic hypothesis concerning the meaning of conditional statements. It serves as part of the explanation for why inferences of the form of *modus ponens* are valid. Of course, this hypothesis is open to debate—the truth table for the material conditional is widely accepted as a semantic analysis of if-then statements in mathematical proofs, but it is quite debatable as an analysis of conditionals in ordinary (nonmathematical) discourse (Traugott, ter Meulen, Reilly, & Ferguson, 1986). Our point is simply that deductive relations between conditionals and other statements constitute evidence for the semantics of conditionals, and, more generally, deductive relations among diverse natural language statements constitute evidence for theories of the semantic structure of natural language (Davidson & Harman, 1975b; McCawley, 1993).

What we hope to show in this chapter is that there are many empirical phenomena that exhibit interesting interactions between the compositional structure of propositions and inferences drawn inductively from evidence. Just as deductive relations serve as clues to the semantic structure of propositions, relations between belief strength, propositional structure, and evidence serve as further clues to semantic structure. We begin by describing five empirical phenomena that exemplify interactions between propositional structure, evidence, and judgments of belief. We then describe the basic elements of a semantic approach to the theory of evidence. A semantic theory of evidence is a theory of how strength of belief is affected by two factors: (1) the structure of the evidence, and (2) the compositional structure of propositions. What we present is a framework for studying the interaction between propositional structure, evidence, and belief strength. Next, we consider a variety of empirical results that illustrate concretely the relationship between strength of belief, propositional structure, and the structure

of evidence. Some of these results are well known in the literature; for example, we will discuss conjunction and disjunction errors in probability judgment from the perspective of a semantic theory of evidence. Other results are less well known; for example, aspects of counterfactual reasoning can be treated as problems in the semantics of evidence. What we hope to show is that the concept of a semantic theory of evidence provides a unifying framework for seeking general answers to the question of how belief strength is affected by natural language representations and the structure of evidence.

Many of the phenomena discussed in this chapter can be characterized as compositional anomalies—they are cases in which observed relations in belief strength conflict with the semantic structure of propositions, or at least with well-established theories of this semantic structure. The term *compositional* is deserved, because the conflicts arise in the relationship between the belief strengths of complex propositions and their component propositions. One reason for thinking that semantic theories of evidence will have something new to say about language structure is that compositional anomalies demonstrate the existence of perplexing inconsistencies between inductive and deductive inference. We will return to this point after presenting examples of such inconsistencies.

## II. Five Compositional Anomalies in Probability Judgment

We present five examples of compositional anomalies in probability judgment. These examples serve as illustrations in a later section where a theoretical framework for the semantics of evidence is presented.

The first two examples are no doubt very familiar, namely, the occurrence of conjunction and disjunction errors in probability judgment. A *conjunction error* is said to occur if an individual judges  $P(A \cap B) > P(A)$  or  $P(A \cap B) > P(B)$ . A *disjunction error* is said to occur if an individual judges  $P(A \cup B) < P(A)$  or  $P(A \cup B) < P(B)$ . Many experiments have demonstrated the occurrence of conjunction and disjunction errors (Carlson & Yates, 1989; Morier & Borgida, 1984; Tversky & Kahneman, 1983; Yates & Carlson, 1986). The famous Linda problem of Tversky and Kahneman (1983) can be used to elicit both types of errors. Tversky and Kahneman (1983) presented subjects with the following description of a fictitious individual:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations

Subjects were asked to rank order a series of statements according to the likelihood that they were true. Among these were statements (2), (3), and (4) in Table I. Tversky and Kahneman (1983) found that the majority of subjects ranked the probabilities in the order:  $P(FM) > P(BT \cup FM) > P(BT)$ . Morier and Borgida (1984) replicated the Tversky and Kahneman (1983) study with various alterations, one of which was to add the disjunctive statement (5) to the set of statements (see Table I). They found that mean probability estimates were ordered as follows:

$$P(FM) > P(BT \cup FM) > P(BT \cap FM) > P(BT). \quad (1)$$

In these comparisons, judgments of  $P(BT \cap FM)$  were significantly greater than judgments of  $P(BT)$ , and judgments of  $P(FM)$  were significantly greater than judgments of  $P(BT \cup FM)$ . Although Morier and Borgida (1984) did not report a statistical test for  $P(FM)$  versus  $P(BT \cup FM)$ , one can infer that  $P(FM)$  was significantly greater than  $P(BT \cup FM)$  from the reported proportion of "union errors" and sample sizes ( $z = 2.88, p < .005$ , two-tailed sign test). These results demonstrate the occurrence of conjunction and disjunction errors with the Linda problem. Numerous other instances of conjunction and disjunction errors have been documented (Bar-Hillel & Neter, 1993; Carlson & Yates, 1989; Tversky & Kahneman, 1983; Wells, 1985; Yates & Carlson, 1986).

Conjunction and disjunction errors are anomalies of propositional composition because probability judgments for conjunctions and disjunctions are inconsistent with logical relations that are implied by the compositional

TABLE I

Statement	Probability of event	Propositional notation
2. Linda is a bank teller.	$P(BT)$	$BT$
3. Linda is active in the feminist movement.	$P(FM)$	$FM$
4. Linda is a bank teller and is active in the feminist movement.	$P(BT \cap FM)$	$BT \wedge FM$
5. Linda is a bank teller or is active in the feminist movement.	$P(BT \cup FM)$	$BT \vee FM$
6. Suppose that in addition to the other information about Linda, you are told that she is active in the feminist movement. What is the probability that she is a bank teller?	$P(BT   FM)$	$BT   FM$

structure of these propositions. Later, we will analyze these errors more carefully; but first, we present other examples of compositional anomalies.

Shafir, Smith, and Osherson (1990) discovered an interesting variant on the conjunction error. Suppose that one conducts a survey of people who are shopping in a large mall. Among the people contacted in the survey are young bankers, older bankers, and many other types of people. Now consider the two statements:

Statement	Probability of event
7. Every single banker in the group is conservative.	$P[(\forall x)(x \in B \rightarrow x \in C)]$
8. Every single young banker in the group is conservative.	$P[(\forall x)(x \in Y \cap B \rightarrow x \in C)]$

If every banker in a group is conservative, then every young banker in the group is conservative, but not vice versa. Consequently, the probability that (8) is true must be equal or greater than the probability that (7) is true, or symbolically, it must be the case that

$$P[(\forall x)(x \in Y \cap B \rightarrow x \in C)] \geq P[(\forall x)(x \in B \rightarrow x \in C)], \quad (9)$$

where  $\rightarrow$  denotes the material conditional (if-then) and  $(\forall x)$  denotes the universal quantification (for all  $x$ ). Let us say that an individual commits a *quantificational conjunction error* (QCE) if he or she judges  $P[(\forall x)(x \in Y \cap B \rightarrow x \in C)] < P[(\forall x)(x \in B \rightarrow x \in C)]$  for some choice of events  $Y, B$ , and  $C$ . Shafir et al. (1990) demonstrated empirically that QCEs occur in statements that are structurally analogous to (7) and (8). The key feature of these statements is that a prototypical  $B$  is a  $C$ , but a prototypical  $Y \cap B$  is not a  $C$ ; for example, a prototypical banker is conservative, but a prototypical young banker is not. QCEs appear to result from the fact that the more inclusive category (e.g., "bankers") is more typical of the predicate (e.g., "conservative") than the less inclusive category (e.g., "young bankers").

Miyamoto, Lundell, and Tu (1988) discovered another anomaly in the relation between conjunctive and conditional probabilities. Note that if  $1 > P(X) > 0$  and  $A$  is any other event, then  $P(A \cap X)/P(X) > P(A \cap X)$ . Applying the definition of conditional probability, we have

$$P(A | X) > P(A \cap X). \quad (10)$$

If  $A =$  "Linda is a bank teller" and  $X =$  "Linda is active in the feminist movement," then (10) shows that, normatively, Linda is more likely to be

a bank teller given that she is a feminist than she is likely to be a bank teller and a feminist. We will say that a *conditionalization error* has occurred if a subject judges  $P(A | X) < P(A \cap X)$ . Miyamoto et al. (1988) had subjects read the description of Linda and then rate the probability of various statements (4) and (6), which were intermixed with other statements that did not pertain to bank tellers or feminists. Not surprisingly, subjects gave significantly higher ratings to the conjunctive probability (4) than to the conditional probability (6) (Mann-Whitney  $U = 7,233.5$ ,  $n = 134, 167$ ,  $p < .001$ , two-tailed).

In a second problem, subjects were asked to rate the probability of various weather conditions in Seattle, among which were included statements (11) and (12):

Statement	Probability of event
11. Suppose it is an overcast day in mid-November. Rate the probability that it rains and the temperature remains below 38 degrees.	$P(RN \cap TM)$
12. Suppose it is an overcast day in mid-November and it has begun to rain. Rate the probability that the temperature remains below 38 degrees.	$P(TM   RN)$

Subjects were University of Washington undergraduates who presumably knew that rain is common in Seattle during the month of November, but the combination of near freezing weather and rain is rare but not impossible. Different samples of subjects rated statements (11) and (12) along with other filler statements. Once again, the conjunctive probability (11) received a significantly higher rating than the conditional probability (12) (Mann-Whitney  $U = 8,377.0$ ,  $n = 172, 154$ ,  $p < .001$ , two-tailed). These results demonstrate the occurrence of conditionalization errors. We will describe analogous conditionalization errors in counterfactual reasoning later in this chapter and discuss a psychological explanation for this phenomenon at that time.

Consider next a comparison of conditional and unconditional probabilities that is motivated by the following relationships.

$$P(A | X) \geq P(X | A) \text{ iff } \frac{P(A \cap X)}{P(X)} \geq \frac{P(A \cap X)}{P(A)} \quad (13)$$

$$\text{iff } P(A) \geq P(X) \quad (14)$$

Therefore, the relative ordering of the conditional probabilities,  $P(A | X)$

versus  $P(X | A)$ , should be the same as the relative ordering of the unconditional probabilities,  $P(A)$  versus  $P(X)$ . Let us say that *conditionals are inconsistent with priors* if an individual judges  $P(A | X) > P(X | A)$  and  $P(A) < P(X)$ , or vice versa. A simple example illustrates how conditionals could be inconsistent with priors. Suppose that Frank and Billy are two exconvicts who meet at a Chicago bar. Subjects were asked to rate the probability of the following statements:

Statement	Probability of event
15. What is the probability that Frank pulls a knife on Billy while in the Star Bar?	$P(K)$
16. What is the probability that Billy makes an insulting ethnic comment to Frank while they are in the Star Bar?	$P(I)$
17. Suppose Billy makes an insulting ethnic comment to Frank sometime while they are together in the Star Bar. What is the probability that Frank pulls a knife on Billy sometime while in the Star Bar?	$P(K   I)$
18. Suppose Frank pulls a knife on Billy sometime while in the Star Bar. What is the probability that Billy makes an insulting ethnic comment to Frank sometime while they are together in the Star Bar?	$P(I   K)$

The results for a sample of University of Washington undergraduates are shown in the following table ( $K$  = knife,  $I$  = insult). Subjects with tied ratings were dropped from the analysis because their inclusion complicates the picture without changing it substantially.

	$P(K) < P(I)$	$P(K) > P(I)$
$P(K   I) < P(I   K)$	42 (51%)	10 (12%)
$P(K   I) > P(I   K)$	22 (27%)	8 (10%)

The conditionals were consistent with the priors of 50 (61%) subjects, and inconsistent with the priors of 32 (39%) subjects.

Each of the phenomena described here is a compositional anomaly in the sense that probability judgments are inconsistent with logical relations deriving from the compositional structure of propositions. The phenomena are listed in Table II along with the propositions whose judged probabilities constitute the anomalies. The conjunction and disjunction errors, the QCE, and the conditionalization error have been found to occur in a significant majority of subjects, at least with particular choices of propositions. In the

TABLE II

Compositional anomaly	Probabilities to be compared
Conjunction error	$P(A), P(B), P(A \cap B)$
Disjunction error	$P(a), P(B), P(A \cup B)$
QCE	$P[(\forall x)(x \in A \rightarrow x \in C)],$ $P[(\forall x)(x \in B \rightarrow x \in C)],$ $P[(\forall x)(x \in A \cap B \rightarrow x \in C)]$
Conditionalization error	$P(A \cap B), P(A   B), P(B   A)$
Inconsistency between conditionals and priors	$P(A   B), P(B   A), P(A), P(B)$

fifth case, the inconsistency between conditionals and priors, it is possible that only a minority of individuals produce the inconsistent pattern. The purpose of the present discussion, however, is not to establish the prevalence of these anomalies, but rather to raise the question of whether the relationship between compositional structure and strength of belief can be the focus of a fruitful scientific inquiry. The phenomena listed in Table II exemplify the interaction of compositional structure with inductive inference. What we seek is a theoretical framework that promotes the systematic study of the relationship between evidence, propositional structure, and strength of belief. In the next section, we propose such a framework, and then turn to specific examples that illustrate the difficulties and rewards of this enterprise.

### III. Semantic Theories of Evidence

A semantic theory of evidence must coordinate a theory of language with a theory of how evidence affects belief strength for propositions in a language. To discuss these relationships, it will help to define some notation for formal structures. Let  $\mathcal{L}$  be a formally defined language, like a propositional calculus, predicate calculus, or intensional logic (Lewis, 1972; Montague, 1974; Quine, 1972). We assume that  $\mathcal{L}$  is defined syntactically by specifying the symbols that comprise the language, together with rules that define well-formed expressions within the language. Deductive inference in  $\mathcal{L}$  can be specified in terms of syntactic rules of inference or semantic definitions of reference and truth (or by types of rules) (van Fraassen, 1971). The former rules determine a relation of deductive consequence; the latter rules determine a relation of logical consequence. Let  $\mathcal{P}$  be the set of propositions in  $\mathcal{L}$ . Let  $\mathcal{E}$  be a set of possible bodies of evidence. The fundamental relation of a semantic theory of evidence is the ordering of proposition/evidence

pairs according to strength of belief. Stated formally, for any propositions  $p, q \in \mathcal{P}$  and evidence  $e, f \in \mathcal{E}$ , let  $(p, e) \geq (q, f)$  signify that the strength of belief for  $p$  given  $e$  is at least as great or greater than the strength of belief for  $q$  given  $f$ . For example, if  $p$  is the proposition that a patient has a cancerous tumor,  $q$  is the proposition that a patient has a noncancerous tumor,  $e$  is an x-ray picture, and  $f$  is a pathology report based on a biopsy, then  $(p, e) \geq (q, f)$  indicates that one would be more certain of a cancerous tumor given the x-ray picture (and no other evidence) than one would be of a noncancerous tumor given the pathology report (and no other evidence).

We will refer to  $\geq$  as the *belief strength order*, or more simply, the *belief order*. When we want to emphasize that the belief order is an ordering of proposition/evidence pairs, we will refer to the pairs as *p/e pairs* and the ordering as the *p/e order*. We assume that  $\geq$  has the properties of an order relation.<sup>2</sup> There are several empirical procedures that might be used to determine a belief order. One could ask individuals to make paired comparison judgments of belief strength, that is, present individuals with *p/e pairs*,  $(p, e)$  versus  $(q, f)$ , and ask, which is more believed? The notation,  $(p, e) > (q, f)$ ,  $(p, e) \sim (q, f)$ , or  $(p, e) < (q, f)$ , would indicate whether belief in  $p$  given  $e$  was judged to be strictly stronger, equally strong, or strictly weaker than belief in  $q$  given  $f$ . Alternatively, one could ask individuals to rate the degree of belief that particular propositions are true given a body of evidence. Let  $R(p, e)$  be the rated belief strength for  $p$  given  $e$ . Operationally, one could measure  $R(p, e)$  by asking the individual to place a mark on a line labeled 0 at one end and 100 at the other end, letting the position of the marked indicate the relative degree of belief. The belief order would then be inferred from the ratings under the assumption:

$$(p, e) \geq (q, f) \text{ iff } R(p, e) \geq R(q, f). \quad (19)$$

There are other methods for determining a belief order, including methods based on willingness to bet, which we will not attempt to describe here. As psychologists, we must recognize that alternative, logically equivalent methods for determining a belief order might lead to inconsistent orders (Slovic & Lichtenstein, 1983), but we will not attempt to compare different methods in this chapter.

We assume that the belief order is consistent with the existence of an underlying scale of belief strength, denoted  $S$ , that satisfies the condition:

$$(p, e) \geq 1 (q, f) \text{ iff } S(p, e) \geq S(q, f). \quad (20)$$

<sup>2</sup>Technically, by an order relation we mean a relation that is transitive and connected. Transitive: For every  $(p, e)$ ,  $(q, f)$ , and  $(r, g)$ , if  $(p, e) \geq (q, f)$  and  $(q, f) \geq (r, g)$ , then  $(p, e) \geq (r, g)$ . Connected: For every  $(p, e)$  and  $(q, f)$ ,  $(p, e) \geq (q, f)$  or  $(p, e) \leq (q, f)$  or both.

The belief function  $S$  is not directly observable, but the  $\geq$  order is observable through paired comparisons or ratings, and the existence of the scale  $S$  is guaranteed if the  $\geq$  order is transitive and connected.<sup>3</sup> In effect, we are assuming the existence of a belief scale  $S$  that is at least an ordinal scale, but a goal of research is to discover further properties of belief strength that would establish interval or ratio-scale uniqueness of  $S$  (cf. Krantz, 1991). Conditions (19) and (20) imply that  $R(p, e)$  is a strictly increasing function of the unobserved belief strength  $S(p, e)$ . In other words, ratings and belief strengths are related by the equation:

$$R(p, e) = M[S(p, e)], \quad (21)$$

for some response function  $M$  about which we assume only that it is strictly increasing.<sup>4</sup> Equation (21) is plausible because one assumes that the individual perceives her own strengths of belief, and hence, can produce larger ratings whenever the belief strengths are in fact larger.

Given these definitions, we now define an *evidence structure* to be an ordered quadruple of the form  $(\mathcal{L}, \mathcal{P}, \mathcal{E}, \geq)$ , where  $\mathcal{L}$  is a formally defined language,  $\mathcal{P}$  is the set of propositions in  $\mathcal{L}$ ,  $\mathcal{E}$  is a set of different bodies of evidence, and  $\geq$  is an ordering of  $\mathcal{P} \times \mathcal{E}$ . A semantic theory of evidence is a theory of the relationship between belief strength, semantic and syntactic properties of propositions in  $\mathcal{P}$ , and properties of the evidence in  $\mathcal{E}$ . A central goal of the semantic theory of evidence is to discover properties of evidence structures that constrain the form of a belief function  $S$  that satisfies Condition (20); specifically, can we discover empirical properties of belief strength  $\geq$  that determine how the mathematical form of  $S$  is related to the compositional structure of propositions and properties of evidence? Of course, to pose and solve problems in the semantic theory of evidence, it will be necessary to give more details about the structure of the language  $\mathcal{L}$ , its propositions  $\mathcal{P}$ , the bodies of evidence in  $\mathcal{E}$ , and the empirical belief order  $\geq$ .

To clarify the motivation for this formulation, we will sketch a concrete example. Suppose that Mr. A was found dead from a knife wound in his hotel room. Mr. X is accused of the crime. Two issues in the trial of Mr. X would be whether he had a motive for the crime and an opportunity to commit the crime. Let  $u$  be the proposition that Mr. X had a motive to murder Mr. A, let  $v$  be the proposition that Mr. X had the opportunity to

<sup>3</sup>Technically, the existence of  $S$  also requires the assumption of a countable dense subset in the belief order, but this assumption is plausible on conceptual grounds. For the definition of transitivity and connectedness, see footnote 2; for the definition of a countable dense subset, see Krantz et al. (1971).

<sup>4</sup> $M$  is strictly increasing provided that  $S(p, e) > S(q, f)$  iff  $M[S(p, e)] > M[S(q, f)]$ .

murder Mr. A in the manner in which he was killed, and let  $u \wedge v$  be the conjunctive proposition that Mr. X had both the motive and the opportunity to murder Mr. A (see Table III). Let  $f_1, f_2,$  and  $f_3$  be alternative versions of the evidence. For example,  $f_1$  might be a scenario in which witnesses testify that Mr. X had strong reasons for revenge against Mr. A, and independent witnesses saw Mr. X in the vicinity of Mr. A's room. The evidence  $f_2$  might be like  $f_1$  except that the reasons given for Mr. X's desire for revenge are less plausible. The evidence  $f_3$  might be like  $f_1$  except that the witnesses are unsure whether Mr. X was the person that they saw near Mr. A's room. One can imagine many alternative bodies of evidence, varying in the strength of the putative motives for murder, the presence and quality of an alibi, the certainty of witnesses, the existence of other suspects, and so forth.

The central problem in a semantic theory of evidence is to explain how properties of  $p$  and  $e$  affect belief strength  $S(p, e)$ . Krantz (1991) proposed an analysis of belief for conjunctions that illustrates our approach to this problem. In the murder example, the evidence decomposes naturally into evidence for  $u$  (X had a motive to kill A) and evidence for  $v$  (X had the opportunity to kill A). One can imagine evidence that goes specifically to the question of motive without pertaining in any way to opportunity. At one extreme, imagine a scenario in which Mr. X had strong reasons to want revenge against Mr. A, or a strong financial interest in Mr. A's death. At the other extreme, imagine evidence that Mr. X had a minor disagreement with Mr. A of a type that rarely produces violent consequences. Let  $\mathcal{E}_1$  denote the set of all such bodies of evidence, varying from very strong to very weak. Similarly, one can imagine different bodies of evidence pertaining to opportunity, varying also from the very strong to the very weak. We assume that  $\mathcal{E}_1$  is restricted to evidence that pertains only to motive without carrying information about opportunity, and that  $\mathcal{E}_2$  is restricted to evidence that pertains only to opportunity without carrying information about mo-

TABLE III

Rated strength of belief			State of evidence
$u$	$v$	$u \wedge v$	
$R(u, f_1)$	$R(v, f_1)$	$R(u \wedge v, f_1)$	$f_1$
$R(u, f_2)$	$R(v, f_2)$	$R(u \wedge v, f_2)$	$f_2$
$R(u, f_3)$	$R(v, f_3)$	$R(u \wedge v, f_3)$	$f_3$

tive. For any  $e_1 \in \mathcal{E}_2$  and  $e_2 \in \mathcal{E}_2$ , let  $e_1 e_2$  denote the combination of the evidence,  $e_1$  pertaining to motive and  $e_2$  to opportunity; the cartesian product,  $\mathcal{E} = \mathcal{E}_1 \times \mathcal{E}_2$ , is the set of all such evidence combinations.

How do  $S(u, e_1)$ ,  $S(v, e_2)$ , and  $S(u \wedge v, e_1 e_2)$  vary as a function of  $e_1$  and  $e_2$ ? Shafer (1976) and Krantz (1991) suggested that when  $u$  and  $v$  are supported by independent bodies of evidence, the belief for  $u \wedge v$  should be a product of the strengths of each conjunct, that is,

$$S(u \wedge v, e_1 e_2) = S(u, e_1) \cdot S(v, e_2) \quad (22)$$

for every  $e_1 \in \mathcal{E}_1$  and  $e_2 \in \mathcal{E}_2$ . Applying Eq. (21) to Eq. (22) yields

$$R(u \wedge v, e_1 e_2) = M[S(u, e_1) \cdot S(v, e_2)]. \quad (23)$$

Equations (22) and (23) are examples of compositional hypotheses in the semantic theory of evidence; they describe how belief for a conjunction  $u \wedge v$  might be related to the strength of its components. Testable assumptions that are sufficient for the validity of Eqs. (22) and (23) are described in Krantz et al. (1971). The point of this example is to show how a testable empirical hypothesis can be developed to describe belief strength for a specified propositional structure, conjunctions of propositions, and a specified class of evidence, namely, combinations of independent evidence.

A general goal of the semantic theory of evidence would be to consider a variety of propositional structures, disjunctions, negations, conditionals, statements with quantifiers, statements with complex clausal structures, counterfactual conditionals, propositional attitudes, and so forth, and to model belief strengths for this ensemble of propositions with respect to diverse evidence states. What one would seek in this investigation would be testable hypotheses, like Eqs. (22) and (23), that describe belief for complex propositions as a function of the structure of the propositions and properties of the evidence. An ultimate goal in this endeavor would be to account for belief across a wide spectrum of natural language expressions and evidence states. A more accessible, proximal goal would be to investigate belief for relatively simple propositional and evidential structures. As we hope to demonstrate, even modest goals are challenging, but also revealing of relationships that have theoretical interest.

Before continuing, we must clarify several issues that were omitted from the preceding discussion. First, we should mention that in the following discussion we will treat judgments of probability as ordinal measures of belief strength. In other words, when individuals are asked to judge the probability of events or propositions, or to rank propositions in terms of their relative probability, we interpret the data from such tasks as indicators

of the ordering of these events or propositions by degree of belief. We do not assume that degree of belief has the properties of a probability measure, even when the ordering of belief is inferred from judgments that purport to pertain to the probability of the events. We use the expression "probability judgment" as if it were synonymous with "judgment of belief strength" under the assumption that individuals typically judge strength of belief when asked to judge the probability of events. This assumption seems reasonable because it is well established that judgments of probability do not conform to the laws of probability. Hence, what subjects consult when they were asked to judge probability is not a subjective probability representation that satisfies the laws of probability, but rather a nonprobabilistic psychological representation that we are calling the belief order.

Second, we suggested earlier that belief strength is an ordering of  $p/c$  pairs, where the propositions are elements in a formalized language. But if belief strengths are determined by asking subjects to rate natural language statements, the propositions in question would appear to be statements of natural language, and not propositions in a formalized language. The response to this is that the propositions in  $\mathcal{P}$  are intended to serve as semantic representations for natural language statements. We assume that a linguistic theory specifies the mapping from natural language statements into propositions in  $\mathcal{P}$ , as well as referential relations between propositions in  $\mathcal{P}$  and states of the world. Admittedly, in our present state of knowledge, the specification of this mapping may be controversial or poorly understood. Nevertheless, for a particular research problem, it may be possible to adopt working hypotheses concerning this mapping; for example, one might assume that a natural language conjunction corresponds to a formal conjunction that satisfies the truth conditions and inference rules governing conjunctions in a propositional calculus. One can investigate consequences of this working hypothesis even if it is not embedded within a more developed linguistic theory. Of course, investigations into the semantics of evidence would advance more rapidly if there existed a widely accepted and well-established theory of the mapping from natural language statements to semantic representations, but even before such a theory is developed, it may be fruitful to adopt heuristic working hypotheses concerning this mapping and to study belief strength under these working hypotheses.

The rationale for postulating formal propositions in a theory of evidence is analogous to the rationale for postulating underlying formal structures in a theory of natural language inference. It is simpler to hypothesize a mapping from natural language statements to propositions in a formalized language and to formulate deductive relations with respect to these propositions, than it is to describe deductive relations directly with respect to natural language statements. We propose to develop a theory of belief in

which the objects of belief are propositions in a formal language rather than statements of natural language, because the syntactic and semantic relations that affect belief are more easily represented in a formal language. Of course, correspondences between natural language statements and formal propositions are empirical claims that are falsifiable, as are other aspects of a theory of belief.

Third, although our previous example ignored the existence of semantic and syntactic structure internal to propositions, for example, decompositions of propositions into subject and predicate, or quantified statements, or complex clausal structures, we recognize the likelihood that finer aspects of propositional structure affect the belief order. Indeed, it is our hope that belief strength will be sensitive to many aspects of propositional structure, for only if this is the case will the investigation of belief lead to deeper theories of the compositional structure of propositions. The examples to be discussed are confined to compositions of propositions without analysis of structure below the level of the proposition; this limitation reflects only our ignorance of the relation between belief strength and finer levels of grammatical analysis, and not some necessary limitation on the scope of a semantic theory of evidence.

Finally, we should comment briefly on the relation between the semantics of evidence and other research in philosophy and psychology. Our conceptualization of the semantics of evidence was greatly influenced by Davidson's proposal that a semantic theory of natural language should take the form of a Tarski-style truth definition (Davidson, 1967, 1970; Davidson & Harman, 1975a). In other words, Davidson proposed that a theory of natural language semantics should explain how the truth conditions for complex propositions depend on the referential relations and truth conditions of more elementary components, and on the way these components are combined. The semantics of evidence is analogous to a truth theory for natural language. Whereas a truth theory attempts to explain the relation between states of the world and the truth of propositions, a semantic theory of evidence attempts to explain the relation between states of evidence and degrees of belief for propositions. In both types of theory, propositional structure mediates the connection between language judgments and a nonlinguistic reality. In a truth theory for natural language, propositional structure is an explanatory construct for intuitions of reference and truth; and in a semantic theory of evidence, propositional structure is an explanatory construct for judgments of belief strength.

Our interest in the semantics of evidence originated in an investigation into the foundations of Shafer's (1976) theory of evidence undertaken by the first author, John Miyamoto, and David Krantz during the early 1980s. Unlike the Bayesian theory in which evidence (data) and hypotheses are

both treated as propositions in a probability space, evidence in Shafer's theory is ontologically distinct from the propositions that constitute objects of belief. The role of evidence is to induce degrees of belief for the propositions in a logical space. Shafer posits a mathematical structure for belief that is different from and more general than that of a probability measure. The precise structure of Shafer's (1976) theory need not concern us here because our formulation of the semantic theory of evidence adopts only a few general features of Shafer's (1976) theory, specifically, the fundamental assumption that belief strength is an ordering of  $p/e$  pairs and the assumption that the structure of inductive inference can be discovered by investigating how this ordering varies as a function of evidence and propositions (see also, Krantz, 1991). We depart from Shafer's theory in that our goals are descriptive and not prescriptive or normative, and for this reason, we have proposed a framework that allows for non-normative as well as normative patterns of belief. In particular, Shafer's (1976) theory of belief functions and the Dempster/Shafer rule for the combination of evidence play no role in our formulation of the semantics of evidence. Moreover, our approach emphasizes the interaction between belief strength and the compositional structure of propositions, a concern that is secondary in Shafer's research program (Shafer, 1976, 1981; Shafer & Tversky, 1985).

As will be apparent from our ensuing discussion of empirical research, the semantics of evidence is intended to serve as a framework for thinking about a number of research problems that already exist in the judgment literature. Thus, research on conjunctive probabilities (Morier & Borgida, 1984; Tversky & Kahneman, 1983; Wells, 1985; Yates & Carlson, 1986), on disjunctive probabilities (Bar-Hillel & Neter, 1993; Carlson & Yates, 1989), on verbal (non-numerical) labels for probabilities (Wallsten, Bude-scu, Rapoport, Zwick, & Forsyth, 1986), on inductive argument strength (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990), and on counterfactual inference (Miyamoto & Dibble, 1986; Miyamoto, Lundell, & Tu, 1989) can all be regarded as problems in the semantics of evidence. This is not to say that previous investigations were intended as contributions to the semantic theory of evidence, for historically, these investigations have focused on the theory of probability judgment without attempting to coordinate this theory with a theory of propositional structure. Perhaps Wallsten et al.'s investigations of verbal probability quantifiers and Osherson et al.'s studies of categorical arguments are closest to our interest in the relation between language structure and strength of belief. In general, however, research in probability judgment and inductive inference has not attempted a systematic study of how evidence and propositional structure jointly affect strength of belief.



In the next section of this chapter, we take a closer look at conjunction errors in probability judgment. Tversky and Kahneman (1983) proposed that conjunction errors result from the use of a representativeness heuristic in probability judgment. From the perspective of a semantic theory of evidence, a representativeness explanation of conjunction errors raises the issue of the compositional structure of concepts, an issue that is also raised in the theory of typicality and concept structure (Medin & Shoben, 1988; Smith & Osherson, 1984). We attempt to point out the parallels between compositional theories of typicality and inductive inference. Next, we turn to an alternative theory of conjunction errors, according to which these errors result from the application of improper mental rules for the combination of subjective probabilities (Gavanski & Roskos-Ewoldsen, 1991; Yates & Carlson, 1986). We review arguments for this theory and then present new experimental results that cannot be explained by these improper rules. In a subsequent section, we discuss compositional anomalies in counterfactual reasoning. We show that theories of counterfactual reasoning due to Braine and O'Brien (1991), Stalnaker (1968), and Lewis (1973) incorporate a strategy for conditional reasoning, called a Ramsey test, that predicts relationships between the compositional structure of counterfactuals and the belief strength of counterfactuals. We then present experimental evidence to show that the predicted relationships are systematically violated in intuitive reasoning. We propose a modification of the Ramsey test strategy that appears to account for the pattern of counterfactual reasoning observed in our experiments.

Although our discussion of probabilistic and counterfactual reasoning errors attempts to evaluate the evidence for different theories of the errors, our purpose is not to provide a definitive explanation for these phenomena. Indeed, we will argue that further research is needed before adequate explanations can be found. Rather, our purpose is to show that empirical studies of the relationship between propositional structure, evidence, and belief strength are viable and productive for theory construction. Hence, our primary goal is not to resolve the question of why reasoning errors occur, but rather to use reasoning errors as illustrations of theoretically interesting interactions between propositional structure and belief strength, that is, to use them as illustrations of interesting problems in the semantics of evidence.

#### IV. A Normative Principle Linking Deduction and Induction

Before discussing empirical examples of compositional anomalies, we must introduce a fundamental normative principle that influences this discussion. The principle may be stated as follows.

*Principle 1 (Consistency of Belief with Logical Consequence):* For any evidence  $e$ , the ordering of propositions by belief strength ought to be the same as the ordering of propositions by the relation of logical consequence.

What Principle 1 says is that if  $q$  is a logical consequence of  $p$ , then normatively, the belief strength for  $q$  should equal or exceed the belief strength for  $p$  no matter what evidence  $e$  is available. Letting  $p \models q$  symbolize the relation that  $q$  is a logical consequence of  $p$ , Principle 1 asserts that for any propositions  $p$  and  $q$ , and evidence  $e$ , if  $p \models q$ , then  $(q, e) \succeq (p, e)$ .

The rationale for Principle 1 is that if  $q$  is a logical consequence of  $p$ , then any evidence supporting  $p$  also supports  $q$ , but there could be evidence supporting  $q$  that does not support  $p$ . Because  $q$  has as much or more evidence supporting it than does  $p$ , belief strength ought to be equal or greater for  $q$  than for  $p$ . If the ordering of belief strength satisfies Principle 1, we will say that belief is consistent with logical consequence. Suppes proved that Principle 1 is satisfied in a Bayesian system of subjective probabilities (cf. Suppes, 1966, Theorem 1), and Adams (1975) proposed a similar hypothesis: "*If an inference is truth-conditionally sound then the uncertainty of its conclusion cannot exceed the sum of the uncertainties of its premises (where uncertainty is here defined as probability of falsity. . .)*" (p. 2) (italics in the original). We propose that it ought to be satisfied by any normative theory of deduction and induction, whether Bayesian or non-Bayesian. Principle 1 is closely related to the principle of extensionality as discussed in Tversky and Kahneman (1983): "If the extension of  $A$  includes the extension of  $B$  (i.e.,  $A \supset B$ ), then  $P(A) \geq P(B)$ ." (p. 294) Principle 1 differs from the extensionality principle in that Principle 1 states a relation between belief strength and logical consequence, and the extensionality principle states a relation between probability and set inclusion among events. Even if belief strength and (subjective) probability are equated as in a Bayesian system, relations of logical consequence do not necessarily reduce to inclusion relations among events (cf. our discussion of counterfactual semantics later in this chapter.)

In the following discussion, examples will be presented in which human judgment violates Principle 1. Because Principle 1 is proposed as a normative principle and not as a psychological law, empirical violations of it need not cast doubt upon its validity. Violations do bring into question, however, the relation between deductive and inductive reasoning. After examining a number of ways that Principle 1 is violated descriptively, we will discuss the implications of such violations for the semantics of evidence.

## V. Conjunction and Disjunction Errors in Probability Judgment

In this section, we examine conjunction and disjunction errors more carefully from the perspective of the semantics of evidence. If  $p$  and  $q$  are any propositions, let  $p \wedge q$  represent the conjunction of  $p$  and  $q$ , and let  $p \vee q$  represent the inclusive disjunction of  $p$  and  $q$ . (The inclusive disjunction of  $p$  and  $q$  is the proposition that is true if  $p$  is true,  $q$  is true, or both  $p$  and  $q$  are true.) The third column of Table I shows the propositional notation for the statements in the Linda problem. In this notation, a conjunction error occurs whenever an individual judges

$$(p \wedge q, e) > (p, e) \text{ or } (p \wedge q, e) > (q, e), \quad (24)$$

and a disjunction error occurs whenever an individual judges

$$(p, e) > (p \vee q, e) \text{ or } (q, e) > (p \vee q, e). \quad (25)$$

In the Linda problem, the typical conjunction error is of the form  $(BT \wedge FM, L) > (BT, L)$ , and the typical disjunction error has the form  $(FM, L) > (BT \vee FM, L)$ , where the evidence  $L$  is the initial description of Linda.

Conjunction and disjunction errors are basic examples of violations of Principle 1. Conjunction errors violate Principle 1 because  $q$  is a logical consequence of  $p \wedge q$  but it is observed that  $(p \wedge q, e) > (q, e)$ . Disjunction errors violate Principle 1 because  $p \vee q$  is a logical consequence of  $p$  but it is observed that  $(p, e) > (p \vee q, e)$ . Because the semantic properties of conjunctions and disjunctions are central to logical theory, inconsistencies between strength of belief and the logical properties of conjunctions and disjunctions pose a fundamental difficulty for any theory of reasoning that incorporates deductive and inductive modes. Before discussing the implications of such errors, we should first examine more carefully the evidence for their existence.

### A. PRAGMATIC OBJECTIONS TO CONJUNCTION AND DISJUNCTION ERRORS

A number of authors have questioned whether conjunction errors are truly errors. Perhaps they are simply cases in which subjects interpret statements differently from the interpretation intended by the experimenters. Specifically, Marcus and Zajonc (1985) and Pennington (1984) suggested that subjects may contrast statement (2), "Linda is a bank teller," to statement

(4), "Linda is a bank teller and is active in the feminist movement," and infer that statement (2) implicitly means that

---

26. Linda is a bank teller who is not active in the feminist movement.  $(BT \wedge \neg FM)$

---

(The expression  $\neg FM$  is the negation of the  $FM$ .) Under this analysis, the apparent tendency to judge  $P(BT \wedge FM) > P(BT)$  is actually a tendency to interpret (2) as (26), together with a belief that  $P(BT \wedge FM) > P(BT \wedge \neg FM)$ . Because the ordering,  $P(BT \wedge FM) > P(BT \wedge \neg FM)$ , does not violate any laws of probability, this hypothesis would imply that the so-called conjunction error is not a true reasoning error.

Several counterarguments can be made to this suggestion. In one of their control conditions, Tversky and Kahneman (1982a, 1983) asked one sample of subjects to rank the likelihoods of a set of statements that included (2) and (3) but not (4), and asked a different sample of subjects to rank a set of statements that included (4) and excluded (2) and (3). Because the subjects who saw the constituent propositions  $BT$  and  $FM$  were different from those who saw the conjunction  $BT \wedge FM$ , there was no pragmatic reason for the subjects to interpret  $BT$  as  $BT \wedge \neg FM$ . Nevertheless, the ranking of  $BT \wedge FM$  with respect to the filler statements was significantly greater than the ranking of  $BT$  with respect to the same filler statements. In addition, Tversky and Kahneman asked a third sample of subjects to rank the likelihoods of the same statements with (4) excluded, and with (2) replaced by the following statement (2\*):

---

2\*. Linda is a bank teller whether or not she is active in the feminist movement.  $(BT^*)$

---

The mean ranking of statement (4),  $BT \wedge FM$ , was also significantly higher than the mean ranking of (2\*). Thus, clarifying the meaning of the statement that Linda is a bank teller did not eliminate the conjunction error.

Morier and Borgida (1984) asked subjects to estimate the probabilities of statements that included (2), (3), (4), and (26).<sup>5</sup> Presumably the presence of (26),  $BT \wedge \neg FM$ , among the stimuli would suppress any tendency to interpret (2),  $BT$ , as  $BT \wedge \neg FM$ . Nevertheless, a majority of subjects produced conjunction errors (77.4% errors,  $n = 30$ ,  $p < .01$  by a sign test). In addition to the Linda problem, Morier and Borgida (1984) studied

<sup>5</sup> The exact wording of the statement used by Morier and Borgida was "Linda is a bank teller who is not a feminist" rather than statement (26), but we assume that this difference in wording is unimportant.

another problem with an analogous logical structure. For this second problem, Morier and Borgida found that subjects who rated statements analogous in form to (2), (3), (4), and (26) produced significantly fewer conjunction errors than subjects who rated only the statements that were analogous to (2), (3), and (4) (48.7% errors,  $n = 39$ , vs. 76.7% errors,  $n = 30$ ;  $p < .05$  by a chi-squared test). Thus there is some evidence that clarifying the meanings of statements can reduce the frequency of conjunction errors with particular statements and stories, but the results for the Linda problem displayed a high proportion of conjunction errors even after clarifying the meaning of the statements.

Wells (1985) raised the issue whether disjunction errors occur because subjects interpret disjunctive statements as exclusive disjunctions, that is, whether subjects interpret statement (5) to mean that

---

27. Either Linda is a bank teller or she is active in the feminist movement, but not both.  $(BT \wedge \neg FM) \vee (\neg BT \wedge FM)$

---

No laws of probability are violated if someone judges  $P(FM)$  to be greater than  $P[(BT \wedge \neg FM) \vee (\neg BT \wedge FM)]$ . Under this analysis, the apparent disjunction error is actually a tendency to interpret (5) as (27), together with the belief that  $P(FM) > P[(BT \wedge \neg FM) \vee (\neg BT \wedge FM)]$ . To test this objection, Wells (1985) added the phrase "or both" to disjunctive statements that were analogous to (5). Nevertheless, a large proportion of disjunction errors was still observed. Carlson and Yates (1989) had subjects rank order the probability of statements that were analogous in form to statements (2) through (5). For some of the subjects, the phrase "or both" was appended to the disjunctive statements, whereas the phrase was omitted for the remaining subjects. No significant differences were found in the proportions of disjunction errors of subjects who saw or did not see disjunctive statements with "or both" appended to them, and the overall proportion of disjunction errors was quite high (80% disjunction errors when averaged over reasoning problems; Carlson & Yates, 1989, Study 2).

We conclude that clarifying the meaning of the statements, either by revising the wording or by eliminating contrasts between statements that suggest an unintended interpretation, does not eliminate the occurrence of conjunction and disjunction errors. Although our discussion has focused primarily on the Linda problem, conjunction and disjunction errors have been observed with many different reasoning problems (Agnoli & Krantz, 1989; Bar-Hillel & Neter, 1993; Carlson & Yates, 1989; Shafir, Smith, & Osherson, 1990; Tversky & Kahneman, 1983; Wells, 1985; Yates & Carlson, 1986). Finally, we note that the conditionalization error, the

tendency to judge  $P(A \wedge B) > P(A|B)$ , is not subject to the same pragmatic criticisms as conjunction problems, because both propositions  $A$  and  $B$  appear in the critical statements. In particular,  $A|B$  cannot be misinterpreted as  $A \wedge \neg B|B$ . Therefore, the tendency to judge  $P(A \wedge B) > P(A|B)$  is consistent with the view that conjunctions are sometimes overweighted in probability judgment. This evidence is not conclusive, however, because it may show that the conditional  $A|B$  is sometimes underweighted. One may conclude, however, that either conjunctions are sometimes overweighted, or conditionals are sometimes underweighted, or both.

## B. REPRESENTATIVENESS, TYPICALITY, AND CONJUNCTION ERRORS

In the remainder of this chapter, we focus on investigations of conjunction errors because the research on disjunction errors is less extensive. Tversky and Kahneman (1983) undertook the investigation of conjunction errors in the context of their famous studies of a general reasoning strategy, which they call the *representativeness heuristic* (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1974, 1982a, 1983). To illustrate this heuristic, suppose that Ms. A is described to have characteristics  $A_1, A_2, \dots, A_n$ . A person who employs the representativeness heuristic judges the probability that Ms. A is in category X as a function of the similarity between the given characteristics of Ms. A and the typical characteristics of category X. For example, it is hypothesized that people evaluate the probability that Linda is a bank teller as a function of the similarity between the given characteristics of Linda and bank tellers in general. By the *similarity hypothesis*, we mean the hypothesis that people judge the probability of an event by its similarity to the given information. Judging probability in terms of similarity is one of the main characteristics of the representativeness heuristic; there are other aspects that we will not discuss here because they are not relevant to conjunction errors.

Judging the probability of an event by its similarity to the given information may often be a reasonable heuristic. As Kahneman and Tversky have repeatedly pointed out, however, the intuition of similarity differs structurally from the mathematical properties of a probability measure, and where it differs, judgments of probability can systematically diverge from normative patterns of probability judgment (Kahneman & Tversky, 1973; Tversky & Kahneman, 1982a, 1983). Specifically, a conjunction of properties can be more similar to a standard than one of the properties in isolation; for example, a feminist bank teller is more like the initial description of Linda than a bank teller whose attitudes towards feminism are left unspecified. Not only is this intuitively clear, it is easily established empirically. Tversky

and Kahneman (1983) had subjects rank order the categories, feminist, bank teller, feminist bank teller, and other filler categories, in terms of "the degree to which [Linda] resembles the typical member of that class." The majority of subjects (85%,  $n = 88$ ) ranked the similarity of the categories in the order: feminist, feminist bank teller, and bank teller. Thus, the preferred rank order of the similarity between Linda and these categories coincides exactly with the preferred rank order of the probability of these categories, which is what the similarity hypothesis predicts. There is no fallacy in ordering the similarity of these categories in this fashion. The fallacy lies in equating the probability order to the similarity order.

Shafir et al. (1990) carried out a more thorough test of the similarity hypothesis. The stimuli in Shafir et al. (1990) were 14 brief personality sketches, one of which was the Linda sketch and others of which were new. For each sketch, they constructed an incompatible conjunction,  $p \wedge q$ , and a compatible conjunction,  $p \wedge r$ , such that  $p$  was the same representative outcome for both conjunctions (e.g., Linda is a feminist),  $q$  was an unrepresentative outcome (e.g., Linda is a bank teller), and  $r$  was a representative outcome (e.g., Linda is a teacher). In terms of the Linda example,  $p \wedge q$  would be "Linda is a feminist bank teller" and  $p \wedge r$  would be "Linda is a feminist teacher." One sample of subjects rated the *typicality* of the propositions  $q$ ,  $r$ ,  $p \wedge q$ , and  $p \wedge r$ , and a different sample of subjects rated the *probability* of these same propositions. There were 14 personality sketches, 4 propositions per sketch, and 2 types of ratings per proposition. Over the 56 propositions, the correlation between mean typicality rating and mean probability rating was .93 ( $p < .001$ ).<sup>6</sup> The high correlation between the typicality of a category and its probability supports the similarity hypothesis.

The high correlation between typicality and judged probability also suggests that typicality ratings should predict the locus of conjunction errors, and this is indeed the case. The incompatible conjunctions ( $p \wedge q$ ) were significantly more typical and also significantly more probable than the unrepresentative category ( $q$ ). The compatible conjunctions ( $p \wedge r$ ) were neither significantly more typical nor significantly more probable than the representative category ( $r$ ). Furthermore, typicality ratings also predicted the magnitude of conjunction errors in the following sense. Let  $p \wedge x$  stand for any of the conjunctions used in the Shafir et al. experiment, where  $x$  is either  $q$  or  $r$ . The difference in mean typicality between  $p \wedge x$  and  $x$  was moderately correlated with the difference in mean probability between  $p \wedge x$  and  $x$  (the correlation was .83,  $n = 28$ ,  $p < .01$ ). Thus, differences in typicality predicted the magnitude of differences in probability between

<sup>6</sup> This correlation was not reported by Shafir et al. (1990), but it can be computed from the statistics that they did report.

$p \wedge x$  and  $x$ . One might object that a request to judge the typicality of an outcome has the same meaning as a request to judge the probability of that outcome, and thus the results are unsurprising. To do so, however, is to concede the validity of the similarity hypothesis in the guise of objecting to it. The objection would have teeth, so to speak, if the intuition of typicality conformed to the mathematical laws of probability but the finding of Shafir et al. was rather to the contrary—estimates of probability conformed to the psychological laws of typicality (cf. Osherson, 1990, and Tversky & Kahneman, 1982a, for elaboration on this point).

In the cognitive theory of typicality, the problem of determining the typicality of conjunctive categories is referred to as the problem of conceptual combination (Medin & Shoben, 1988; Osherson & Smith, 1982; Smith & Osherson, 1984, 1989; Smith, Osherson, Rips, & Keane, 1988). Perhaps the most complete model of conceptual combination is the prototype combination model (Smith & Osherson, 1984, 1989; Smith et al., 1988); in this model, categories are represented as lists of attribute-value combinations, with each attribute value weighted by its diagnosticity and salience. The attribute values for a category are not regarded as necessary and sufficient features; rather, they define a prototype or most typical member of the category. The typicality of instance  $X$  in category  $A$  is determined by the similarity of  $X$  to the category  $A$  prototype, where the measure of similarity is a variant of Tversky's (1977) feature contrast rule.

To determine the prototype for a combined category like "feminist bank teller," the Smith/Osherson theory proposes that the category prototype of "bank teller" serves as a default representation. Attributes and values of the noun that are unspecified by the adjective remain unchanged in the combined category, but attributes and values of the noun that differ from those of the adjective are modified to conform more closely to the adjective. We will not describe the precise mathematical rule that governs conceptual combination in the Smith/Osherson theory other than to note that its effect is to increase the diagnosticity of attributes and the salience of values that are specified by the adjective. Smith et al. (1988) have shown that their model is able to predict conditions under which an instance is more typical of a combined category than of a constituent category (e.g., Linda is more typical of "feminist bank teller" than of "bank teller"), as well as cases in which it is not (e.g., Linda is less typical of "uninformed teacher" than of "teacher").<sup>7</sup>

<sup>7</sup> The example of "uninformed teacher" versus "teachers" does not appear in Smith and Osherson (1989) or Smith et al. (1988). Their experiments concern combinations of fruit and vegetable categories with color and shape adjectives. The claim that Linda is more typical of "teacher" than of "uninformed teacher" has not been verified empirically, although it is plausible based on intuition and also on the Smith et al. (1988) results.

A thorough discussion of typicality is beyond the scope of this chapter, but we wish to point out the analogy between issues in the theory of typicality and the concern in the present chapter with compositional models of belief. The central problem in the semantics of evidence is to predict relations in belief strength from the compositional structure of propositions and from properties of evidence. If belief strength is determined by a representativeness heuristic, and if this heuristic reduces to judgments of typicality in questions of category membership, then for propositions asserting category membership, a compositional theory of belief must be based on a compositional theory of typicality. The theory of typicality is obviously not a finished product, but rather a theory under construction with many viewpoints and empirical phenomena to take into account (Hampton, 1987a, 1987b; Huttenlocher & Hedges, 1994; Medin & Shoben, 1988; Murphy, 1988; Osherson & Smith, 1982; Smith & Medin, 1981; Smith & Osherson, 1984; Smith et al., 1988; Zadeh, 1982). What we can look for in the near future are not ultimate answers to questions concerning the compositional structure of concepts, but rather informative ways to investigate this structure and its relation to strength of belief.

### C. PROBABILITY COMBINATION MODELS FOR CONJUNCTION ERRORS

The main competitors to the representativeness explanation of conjunction errors are various probability combination models. For example, let  $R(A, e)$  denote a rating of the probability of  $A$  given evidence  $e$ . A weighted averaging model asserts that

$$R(A \wedge B, e) = \frac{v_1 R(A, e) + v_2 R(B, e)}{v_1 + v_2}, \quad (28)$$

where  $v_1$  and  $v_2$  are positive weights (Wyer, 1976). We are interested in more general models in which the combination process occurs at the level of belief strength rather than at the level of ratings. For example, an averaging model for belief strength asserts that

$$S(A \wedge B, e) = \frac{w_1 S(A, e) + w_2 S(B, e)}{w_1 + w_2}, \quad (29)$$

where  $w_1$  and  $w_2$  are positive weights. The difference between models (28) and (29) is that (28) hypothesizes a relation between ratings of probability, whereas (29) hypothesizes a similar relation between the unobserved belief strengths. Weighted averaging models like (28) and (29) are natural models for conjunctive probability judgments because they predict that  $R(A,$

$e) > R(A \wedge B, e) > R(B, e)$ , if  $A$  is the more probable and  $B$  is the less probable outcome given  $e$ .

Our discussion will focus on combination models for belief strength, like model (29), rather than on combination models for ratings, like model (28), but our analysis will be equally relevant to both types of models. To see how this could be, recall that ratings and belief strengths are related by the condition  $R(p, e) = M[S(p, e)]$ . Thus, model (29) is equivalent to the hypothesis that

$$R(A \wedge B, e) = M \left[ \frac{w_1 S(A, e) + w_2 S(B, e)}{w_1 + w_2} \right]. \quad (30)$$

If one assumes that  $M$  is linear, then model (30) reduces to model (28). Model (30) allows for the possibility that  $M$  is nonlinear. We will present several probability combination models of belief strength, as in (29) and (30), and derive ordinal predictions from these models. These predictions are also implied by corresponding models that assume a linear response function, like (28). If empirical findings contradict an ordinal prediction derived from these models, all versions of the model, those that assume a linear response function and those that do not, are equally rejected.

Probability combination models are models, like (28), (29), and (30), that treat the belief strengths of conjunctions as functions of the belief strengths of the components. Their essential characteristic is that the component propositions contribute to the belief strength of the conjunction only through their belief strengths and not through semantic or conceptual relations between the components. A formal definition of the class of probability combination models is given with Eqs. (42) and (43). Although probability combination models for conjunctions had been proposed as early as Wyer (1976), interest in these models was stimulated by the finding of Yates and Carlson (1986) that conjunction errors occur even among conjunctions of conceptually unrelated propositions. For example, consider the propositions:

- 
- |  |          |
|--|----------|
| 31. Governor Blanchard will succeed in raising the Michigan state income tax.      | <i>A</i> |
| 32. Bo Derek will win an Academy Award for the movie that she is currently making. | <i>B</i> |
- 

At the time of the Yates and Carlson (1986) study, Governor Blanchard was lobbying to raise the Michigan state income tax, and his chances were generally regarded as good. Bo Derek was a well known, sexy but untalented

attress whose chances of winning an Academy Award were widely recognized to be slight. Subjects were asked to rank the probability of a series of propositions, among which were included  $A$ ,  $B$ , and  $A \wedge B$ . Of 78 subjects, 44 (56.4%) ranked the probability of  $A \wedge B$  higher than the probability of  $B$  alone. Because the component propositions,  $A$  and  $B$ , are obviously unrelated semantically, it is difficult to see how the judged probability of  $A \wedge B$  could be based on the representativeness of the conjunction, that is, the typicality of the conceptual combination of Blanchard's success and Bo Derek's winning. This finding suggests that conjunction errors may result simply from the combination of particular belief strengths, without regard to semantic relations between the component propositions.

Let us define zero, single, and double conjunction errors by the conditions:

Zero:  $(A, e) \geq (A \wedge B, e)$  and  $(B, e) \geq (A \wedge B, e)$   
 Single:  $(A, e) \geq (A \wedge B, e) > (B, e)$  or  $(A, e) < (A \wedge B, e) \leq (B, e)$   
 Double:  $(A \wedge B, e) > (A, e)$  and  $(A \wedge B, e) > (B, e)$

Yates and Carlson (1986) factorially combined events that were either low or high in probability according to public opinion of the time with events that were either conceptually related or unrelated. Conceptually related events pertained to interrelated political events or interrelated sports events, whereas conceptually unrelated problems were like the preceding Governor Blanchard/Bo Derek example. Subjects rank ordered the probabilities of conjunctions and their component propositions, and the frequencies of zero, single, and double conjunction errors were counted.

The results in Table IV are a reanalysis of the results in Table III of Yates and Carlson (1986). The top half of Table IV shows the percentage of zero, single, and double conjunction errors in conjunctions of related and unrelated events. High/high, high/low, and low/low refer to the combination of high and low probabilities in the particular conjunction. The bottom half of Table IV shows the number of problems and sample sizes that contributed to the percentages in the top half of the table. The most striking feature of Table IV is the systematic relation between the probabilities of the component events and the types of conjunction errors that were most common in those problems. For both related and unrelated events, zero conjunction errors occurred most frequently in the low/low problems, single conjunction errors occurred most frequently in the high/low problems, and double conjunction errors occurred most frequently in the high/high problems. The similarity in the pattern of conjunction errors produced by problems with related and unrelated events lends credibility to the

TABLE IV  
 PERCENTAGE OF ZERO, SINGLE, AND DOUBLE CONJUNCTION ERRORS

	Related events			Unrelated events		
	0	1	2	0	1	2
High/High	38.7	15.2	46.1	40.7	29.4	30.0
High/Low	31.0	67.7	1.3	42.7	54.8	4.5
Low/Low	58.0	27.3	14.8	58.5	35.7	5.8
High/High	2 problems in study 4, $N = 105$			4 problems in study 4, $N = 105$		
High/Low	3 problems in study 2, $N = 46$			3 problems in study 2, $N = 46$		
	3 problems in study 3, $N = 33$			3 problems in study 3, $N = 33$		
Low/Low	2 problems in study 4, $N = 105$			3 problems in study 4, $N = 105$		

Note. Percentages are averaged over problems in Yates and Carlson's (1986) Studies 2, 3, and 4. Within any study, the problems were administered within subject. (From Yates & Carlson, 1986, Table 3.)

conjecture that the causes of conjunction errors are similar in the two types of problems.

We will shortly present evidence that is inconsistent with all probability combination models, but we first note that the data from Yates and Carlson (1986) are sufficient to rule out the weighted averaging model as a model of conjunctive probabilities. An averaging model like (28) or (29) predicts that single conjunction errors should always occur, because an average of two numbers is always between the numbers in magnitude (Yates & Carlson, 1986). Clearly the data in Table IV rejects this prediction, as do the data reported in other studies (Tversky & Kahneman, 1983; Wells, 1985). A modified version of the averaging model allows for the occurrence of all three patterns—zero, single, and double conjunction errors—for different conjunctions.

$$S(A \wedge B, e) = \frac{w_0 s_0 + w_1 S(A, e) + w_2 S(B, e)}{w_0 + w_1 + w_2} \quad (33)$$

In this model,  $s_0$  is an initial impression or default level of belief for an arbitrarily chosen conjunction. Model (33) will be referred to as the *initial impression averaging model*. Model (33) is equivalent to the model

$$S(A \wedge B, e) = p_0 s_0 + (1 - p_0)[p_1 S(A, e) + (1 - p_1)S(B, e)], \quad (34)$$

where  $p_0 = w_0/(w_0 + w_1 + w_2)$ , and  $p_1 = w_1/(w_1 + w_2)$ . Equation (34) makes it clear that zero or single conjunction errors should occur whenever  $s_0$  is

less than the smaller of  $S(A, e)$  and  $S(B, e)$ , single conjunction errors should occur whenever  $s_0$  is between  $S(A, e)$  and  $S(B, e)$ , and single or double conjunction errors should occur whenever  $s_0$  is greater than the larger of  $S(A, e)$  and  $S(B, e)$ . Thus, the initial impression averaging model predicts that zero conjunction errors should be most prevalent when  $S(A, e)$  and  $S(B, e)$  are both large, and double conjunction errors should be most prevalent when  $S(A, e)$  and  $S(B, e)$  are both small. This is precisely opposite to the pattern of conjunction errors displayed in Table IV. To test the statistical reliability of this pattern, we treated the percentage of double conjunction errors for each problem in the high/high and low/low conditions as the data in a two-factor analysis of variance. The factors were relatedness (related events vs. unrelated events) and probability levels (high/high vs. low/low). The percentages from the high/low condition were excluded because they came from Yates and Carlson's Studies 2 and 3, whereas the high/high and low/low percentages were from a single study, Study 4. The main effect of probability level was highly significant ( $F(1, 7) = 41.9$ ,  $M_e = 46.4$ ,  $p < .001$ ). On the average, the double conjunction errors were 27.7 percentage points more common in the high/high problems. We conclude that an initial impression averaging model is inconsistent with the data in Yates and Carlson (1986), and must be rejected.

Yates and Carlson (1986) proposed a *signed summation* model that correctly predicts the qualitative pattern of conjunction errors in Table IV. According to this model, the belief strength of a conjunction is a sum of the strengths of its components:

$$S(A \wedge B, e) = S(A, e) + S(B, e), \quad (35)$$

where the belief strength scale  $S$  is allowed to take on either positive or negative values. Let  $\neg A$  denote the negation of  $A$ , and let us say that  $A$  is "likely" if  $S(A, e) > S(\neg A, e)$ , "unlikely" if  $S(A, e) < S(\neg A, e)$ , and "neutral" if  $S(A, e) = S(\neg A, e)$ . The essence of the signed summation rule is the assumption that  $S(A, e) > 0$  if  $A$  is likely,  $S(A, e) = 0$  if  $A$  is neutral, and  $S(A, e) < 0$  if  $A$  is unlikely. Given these assumptions, the signed summation model predicts that double conjunction errors will occur when  $A$  and  $B$  are both likely, single conjunction errors will occur when one proposition is likely and the other is unlikely, and zero conjunction errors will occur when  $A$  or  $B$  are both unlikely. This is precisely the qualitative pattern in Table IV.

One defect of the signed summation model is that it predicts the occurrence of single conjunction errors for any self-contradictory conjunction.  $A \wedge \neg A$ . Thus, "Jones will win the race and Jones will lose the race" should be judged more probable than the less likely outcome. This criticism

can be avoided if one assumes that individuals detect self-contradictions and apply some other reasoning rule to them. One might also object that it is counterintuitive to have negative values of belief strength. This criticism is also easily avoided. From the standpoint of ordinal predictions of belief strength, the signed summation model (35) is equivalent to a multiplicative model,

$$S(A \wedge B, e) = S(A, e) \cdot S(B, e), \quad (36)$$

where the belief scale  $S$  ranges from 0 to  $+\infty$ . The qualitative equivalence of signed summation (35) and the multiplicative model (36) is readily seen if one takes logarithms of both sides of Eq. (36), yielding Eq. (35) after resetting  $\log(S) \rightarrow S$ .

It could be objected that Yates and Carlson's (1986) experiment confounded the distinction between related/unrelated events with the specific propositions used to instantiate this distinction. In other words, the problems with related events were constructed from different positions than those used in the problems with unrelated events. Perhaps the similarity of the results for related and unrelated events is merely a coincidence resulting from a particular choice of propositions. Gavanski and Roskos-Ewoldsen (1991) provided evidence against this objection, as well as a sharper test of the generality of probability combination methods. They selected propositions  $A_1, A_2, B_1$ , and  $B_2$  such that  $A_1$  and  $A_2$  pertained to the same issue, and  $B_1$  and  $B_2$  pertained to some other issue. Let us call  $A_1 \wedge A_2$  and  $B_1 \wedge B_2$  the conceptually related conjunctions, and let us call  $A_1 \wedge B_1$  and  $A_2 \wedge B_2$  the conceptually unrelated conjunctions, or more simply, the related and unrelated conjunctions, respectively.<sup>8</sup> For example,

---

37. Linda is both a bank teller [ $A_1$ ] and extremely fashion conscious [ $A_2$ ].	$A_1 \wedge A_2$
38. Jason is both a computer programmer [ $B_1$ ] and is very shy [ $B_2$ ].	$B_1 \wedge B_2$
39. Linda is a bank teller [ $A_1$ ] and Jason is a computer programmer [ $B_1$ ].	$A_1 \wedge B_1$
40. Linda is extremely fashion conscious [ $A_2$ ] and Jason is very shy [ $B_2$ ].	$A_2 \wedge B_2$

---

For conjunctions of the form (37) through (40), probability combination models predict the following equivalence, which we prove in Appendix 2:

$$S(A_1 \wedge A_2, e) > S(A_1 \wedge B_1, e) \text{ iff } S(B_1 \wedge B_2, e) < S(A_2 \wedge B_2, e), \quad (41)$$

that is,

<sup>8</sup> Our notation differs from that of Gavanski and Roskos-Ewoldsen (1991) in order to be consistent with the notation used later in this chapter.

*Related* > *Unrelated* iff *Related* < *Unrelated*.

Thus, a probability combination model predicts that unrelated conjunctions will be neither systematically stronger nor weaker in strength of belief than the corresponding related conjunctions. Gavanski and Roskos-Ewoldsen (1991) did not test (41) directly, although it is a testable condition, but rather they tested a consequence of (41), namely, that the average number of conjunction errors should be about the same for related and unrelated conjunctions. Gavanski and Roskos-Ewoldsen (1991) found no significant differences in the frequency of conjunction errors produced by related and unrelated conjunctions. The statistics reported in their study do not permit one to evaluate the power of their tests, but the fact that the null result was found in experiments with samples of 180 and 153 subjects, respectively, suggests that statistical power was sufficient to detect differences of moderate size. Thus, conjunctions of related and unrelated events appear to produce conjunction errors equally often when the components of the conjunctions are drawn from the same set of propositions. This result supports the hypothesis that the same processes underly conjunction errors in related and unrelated conjunctions.

Representativeness does not provide a straightforward explanation for conjunction errors among unrelated events. Therefore, if conjunction errors are produced by the same process in conjunctions of related and unrelated events, the most plausible explanation would seem to be that they result from improper probability combination rules. Reasoning along these lines, Gavanski and Roskos-Ewoldsen (1991) boldly advanced the hypothesis that probability combination models could account for all types of conjunction errors, including those involving conceptually related events.

To what extent does judgment by representativeness contribute to the conjunction fallacy? Our results support a surprising answer: The only contribution of representativeness stems from its effects on a conjunction's component events. Conjunction fallacies, even in exemplar representativeness problems, stem primarily from the incorrect rules people use to combine probabilities. These rules are likely to yield the fallacy with certain combinations of component probabilities, regardless of whether these probabilities are arrived at by means of assessments of representativeness. (p. 190)

We construe this remark as implying the following hypothesis: For any propositions  $p$  and  $q$  and evidence  $e$ ,

$$R(p \wedge q, e) = F[R(p, e), R(q, e)], \quad (42)$$

where  $F$  is a function specified in a particular probability combination model, for example, averaging in an averaging model or addition in the

signed summation model. Model (42) formalizes the intuition expressed in the remark from Gavanski and Roskos-Ewoldsen (1991) that the conjunctive probability  $R(p \wedge q, e)$  is simply a function of component probability ratings  $R(p, e)$  and  $R(q, e)$ . Representativeness influences the conjunctive probability  $R(p \wedge q, e)$  only by affecting the component ratings. In particular, the conjunctive probability does not depend on any semantic or conceptual relationship between the components. Assuming that  $R(p, e) = M[S(p, e)]$ , and that  $M$  and  $F$  are continuous and strictly increasing, Equation (42) is equivalent to the hypothesis that there exists a continuous, strictly increasing function  $G$  such that

$$S(p \wedge q, e) = G[S(p, e), S(q, e)] \quad (43)$$

for every proposition  $p$  and  $q$ , and every body of evidence  $e$ .<sup>9</sup> Equation (43) implies that belief strength of a conjunction depends only on the belief strengths of the components, and not on any relationship between the components. Thus, the essential feature of the probability combination models, as characterized by either Eq. (42) or (43), is that they deny that semantic or conceptual relations *between* the components can affect the belief strengths of conjunctions.

Equation (43) implies four simple predictions that every probability combination model must satisfy. First, for any propositions  $A$ ,  $C$ ,  $D$ , and  $X$ , and any evidence  $e$ ,

$$(C, e) > (D, e) \text{ iff } (A \wedge C, e) > (A \wedge D, e) \quad (44)$$

and

$$(C, e) > (D, e) \text{ iff } (C \wedge X, e) > (D \wedge X, e) \quad (45)$$

The intuition behind (44) and (45) is that if the belief strength of a conjunction is unaffected by conceptual relations between the components, then the belief order between  $C$  and  $D$  should predict the belief order after they are conjoined with any other proposition,  $A$  or  $X$ . For example, if "Linda is a teacher" seems more probable than "Linda is a bank teller," then "Linda is a teacher and wears bifocals" should seem more probable than "Linda is a bank teller and wears bifocals." Equation (43) also predicts that for any propositions  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $X$ , and  $Y$ , and any evidence  $e$ ,

$$(A \wedge C, e) > (A \wedge D, e) \text{ iff } (B \wedge C, e) > (B \wedge D, e) \quad (46)$$

and

<sup>9</sup> Proof: Eq. (42) is true iff  $S(p \wedge q, e) = M^{-1}\{F[M(S(p, e)), M(S(q, e))]\}$  iff Eq. (43) is true.



$$(C \wedge X, e) > (D \wedge X, e) \text{ iff } (C \wedge Y, e) > (D \wedge Y, e). \quad (47)$$

The intuition behind (46) and (47) is that if the belief strength of a conjunction is unaffected by conceptual relations between the components, then the belief order between conjunctions that share a common component should not change if that common component is replaced by some other proposition. For example, if "Linda is a teacher and wears bifocals" seems more probable than "Linda is a bank teller and wears bifocals," then "Linda is a teacher and walks to work" should seem more probable than "Linda is a bank teller and walks to work." The proof that conditions (44) through (47) follow from Eq. (43) is given in Appendix 2.

In the terminology of the analysis of variance, conditions (44) through (47) assert that crossover interactions should not occur if one changes the component shared by a pair of conjunctions. Crossover interactions are excluded by probability combination models because these models deny that relations between the components of a conjunction can have interactive effects on belief strength. The analysis of variance will not be used to test Eqs. (44) through (47) because ordinal tests have greater generality of interpretation, but the intuition of crossover versus no crossover is appropriate. Conditions (44) through (47) are independence assumptions in the sense of conjoint measurement theory (Krantz et al., 1971). As shown in conjoint measurement, ordinal tests of independence assumptions are preferable to the analysis of variance because one is not forced to assume interval scale responses and the linearity of the response function (the function  $M$  in our notation).

Our strategy for testing (44) through (47) is straightforward. To test (45), choose  $C$ ,  $D$ , and  $X$  such that  $C$  and  $X$  are conceptually compatible and  $C$  and  $X$  are conceptually incompatible, and  $C$  is less believed than  $D$  given evidence  $e$ , that is,  $(C, e) < (D, e)$ . If conceptual relations actually affect the judged probability of conjunctions, then the ordering  $(C \wedge X, e) > (D \wedge X, e)$  might be observed, in violation of condition (45). To test (47), choose  $C$ ,  $D$ , and  $X$  as before, and in addition, choose some other  $Y$  such that  $C$  and  $Y$  are conceptually incompatible and  $D$  and  $Y$  are conceptually compatible. With appropriately chosen propositions, it should be possible to find  $(C \wedge X, e) > (D \wedge X, e)$  and  $(C \wedge Y, e) < (D \wedge Y, e)$  in violation of condition (47). Violations of (44) and (46) might be obtained by analogous constructions. The following experiment tests the validity of conditions (44) through (47).

### Experiment 1

**Subjects:** Subjects were 422 undergraduates at the University of Washington who received credit in a psychology course for participation in the experiment. Subjects were tested in large groups.

**Stimulus materials:** The experimenters wrote 10 vignettes and 10 statements that pertained to the events in each vignette.<sup>10</sup> The vignettes and statements are contained in Appendix 1. For any vignette, the statements were of the following types: two filler statements, four simple propositions ( $A_1, A_2, B_1, B_2$ ), and four conjunctions ( $A_1 \wedge A_2, A_1 \wedge B_2, B_1 \wedge A_2, B_1 \wedge B_2$ ). The statements were written such that  $A_1$  and  $A_2$  were conceptually compatible, as were  $B_1$  and  $B_2$ , whereas  $A_1$  and  $B_2$  were conceptually incompatible, as were  $B_1$  and  $A_2$ . (See the following examples and Appendix 1).

**Procedure:** The 10 vignettes were arbitrarily divided into five pairs of vignettes. Subjects were randomly assigned to one of these pairs. The vignettes and statements were presented to subjects in the form of a questionnaire. Subjects were taught to rate the probability of a statement by placing a mark on a response line labeled "Absolutely Impossible" at one end and "Absolutely Certain" at the other end. They then read a first vignette, rated the 10 statements for that vignette, read a second vignette, and rated the 10 statements for that vignette. The statements for each vignette were presented in one of two randomly selected orders. The data were converted to a 30-point scale by measuring the position of the marks on the response line; a rating of 1 indicated the lowest probability, and 30 the highest probability.

**Results:** Because of the relations of compatibility and incompatibility among the stimulus statements, we predicted that

$$(A_1 \wedge A_2, e) > (B_1 \wedge A_2, e) \text{ and } (A_1 \wedge B_2, e) < (B_1 \wedge B_2, e) \quad (48)$$

in contradiction to condition (46), and

$$(A_1 \wedge A_2, e) > (A_1 \wedge B_2, e) \text{ and } (B_1 \wedge A_2, e) < (B_1 \wedge B_2, e) \quad (49)$$

in contradiction to condition (47). For example, in Problem 4-1 (Appendix 1), a high school senior, Joe B, had applied for admission to Harvard, Princeton, Oklahoma, and Texas. The statements were: ( $A_1$ ) Joe is accepted at Harvard, ( $A_2$ ) Joe is accepted at Princeton, ( $B_1$ ) Joe is rejected at Oklahoma, and ( $B_2$ ) Joe is rejected at Texas. Acceptance at Harvard and Princeton ( $A_1 \wedge A_2$ ) should appear more probable than rejection at Oklahoma and acceptance at Princeton ( $B_1 \wedge A_2$ ), because rejection at Oklahoma is diagnostic of characteristics that lower chances of acceptance at Princeton, and acceptance at Princeton is diagnostic of characteristics that reduce the chance of rejection at Oklahoma. Therefore, one predicts that  $A_1 \wedge A_2 > B_1 \wedge A_2$ . However, acceptance at Harvard and rejection at Texas ( $A_1 \wedge B_2$ ) should seem less probable than rejection at Oklahoma and Texas

<sup>10</sup> We would like to thank Rob Flaherty for his able assistance in constructing stimulus materials and running subjects.

$(B_1 \wedge B_2)$ . Thus,  $A_1 \wedge B_2 < B_1 \wedge B_2$ . Hence, condition (48) should be observed. Similarly, acceptance at Harvard and Princeton ( $A_1 \wedge A_2$ ) should appear more probable than acceptance at Harvard and rejection at Texas ( $A_1 \wedge B_2$ ), but rejection at Oklahoma and acceptance at Princeton ( $B_1 \wedge A_2$ ) should seem less probable than rejection at Oklahoma and Texas ( $B_1 \wedge B_2$ ). Hence condition (49) should be observed.<sup>11</sup>

Tables V and VI show the results of tests of (44) through (47). Problem  $k$ - $j$  refers to the  $j$ th vignette ( $j = 1$  or  $2$ ) in condition  $k$  ( $k = 1$ - $5$ ). The same subjects rated the statements for Problems  $k$ -1 and  $k$ -2. We will explain the format of Table V in detail; the format of Table VI is analogous. Each cell displays the proportion of times that the first proposition was rated higher than the second proposition in the column heading (ties were dropped). For example, in Problem 1-1,  $A_2$  received a higher rating than  $B_2$  in 68% of the untied responses, and  $A_1 \wedge A_2$  was rated higher than  $A_1 \wedge B_2$  in 78% of the untied responses. Asterisks and daggers indicate the  $p$ -values of the proportions as calculated by two-tailed sign tests. Conditions

TABLE V  
TESTS OF CONDITIONS (44) AND (46)

	$A_2$ vs. $B_2$	$A_1 \wedge A_2$ vs. $A_1 \wedge B_2$	$B_1 \wedge A_2$ vs. $B_1 \wedge B_2$	Pattern
Problem 1-1	.68*	.78	.67*	◆◆◆
Problem 1-2	.43	.80††	.03††	◊ ◆◆
Problem 2-1	.92††	.89††	.74††	◆◆◆
Problem 2-2	.84††	.91††	.60	◆◆◆
Problem 3-1	.45	.56	.35*	○○○
Problem 3-2	.51	.71†	.28††	○○○
Problem 4-1	.73††	.71†	.39	◆◆○
Problem 4-2	.87††	.93††	.38*	◆◆◆
Problem 5-1	.12††	.42	.11††	◆◆○
Problem 5-2	.71†	.69†	.65*	◆◆◆

Note. All proportions refer to the proportion of times the first proposition was rated higher than the second proposition in the column heading (ties were dropped).  $P$ -values are for two-tailed sign tests: \*  $p < .05$ ; \*\*  $p < .01$ ; †  $p < .005$ ; ††  $p < .001$ . Pattern column: The triples of arrows indicate the direction of greater belief strength in the three comparisons; solid black arrows indicate significant comparisons ( $p < .05$ ) and open white arrows indicate nonsignificant comparisons.

<sup>11</sup> Experiment 1 and the example of Joe B were inspired by the following remark of Tversky and Kahneman (1983, p. 305): "it is more representative (as well as more probable) for a student to be in the upper half of the class in both mathematics and physics or to be in the lower half of the class in both fields than to be in the upper half in one field and in the lower half in the other. Such observations imply that the judged probability (or representativeness) of a conjunction cannot be computed as a function (e.g., product, sum, minimum, weighted average) of the scale values of its constituents. This conclusion excludes a large class of formal models that ignore the relation between the constituents of a conjunction."

TABLE VI  
TESTS OF CONDITIONS (45) AND (47)

	$A_1$ vs. $B_1$	$A_1 \wedge A_2$ vs. $B_1 \wedge A_2$	$A_1 \wedge B_2$ vs. $B_1 \wedge B_2$	Pattern
Problem 1-1	.68**	.57	.47	→⇔⇐
Problem 1-2	.55	.99††	.18††	⇔→⇐
Problem 2-1	.82††	.76††	.61	→→⇐
Problem 2-2	.63*	.63*	.20††	→→⇐
Problem 3-1	.44	.66**	.44	⇔⇐⇐
Problem 3-2	.38	.58	.13††	⇔⇐⇐
Problem 4-1	.70	.67†	.33†	→→⇐
Problem 4-2	.64*	.89††	.12††	→→⇐
Problem 5-1	.37*	.76††	.24††	⇐→⇐
Problem 5-2	.36*	.48	.28†	⇐⇐⇐

Note. The notation is the same as for Table V.

(44) and (46) predict that in every row of Table V the proportions will all be greater than .5, or all less than .5. We will say that condition (44) or (46) is significantly violated if a row of Table V contains at least one proportion that is significantly greater than .5 and at least one proportion that is significantly less than .5 ( $p < .05$  in both cases).

The arrows in the far right column indicate whether the proportions in a row are consistent with or in violation of conditions (44) and (46). The arrows point in the direction from the higher to the lower rated probability; bold arrows indicate proportions that are significantly different from .5 ( $p < .05$ ), and non-bold arrows indicate proportions that are not significantly different from .5. For example, Problem 1-1 of Table V displays the pattern →→→ indicating that every proportion in the row was significantly greater than .5, and that they were consistent with each other as predicted by (44) and (46). Problem 4-2 displays the pattern →→⇐ indicating significant violations of conditions (44) and (46); the proportions in the first and third data columns violate (44), and the proportions in the second and third data columns violate (46). Problem 4-1 displays the pattern →→⇐ indicating that the last proportion in the row was inconsistent with the first two proportions, but that the proportion .39 was not significant. The format and notation of Table VI is exactly the same as that of Table V, except that the results test conditions (45) and (47) rather than (44) and (46).

The probability combination models predict that conditions (44) through (47) should be satisfied by every row in Tables V and VI. Even one violation of one of the conditions, (44) through (47), is sufficient to reject this class of models, provided that one can establish that the apparent violation is not due to sampling error. An examination of the arrow patterns in Tables

V and VI indicates that there were five significant violations of conditions (44) and (45) (Problem 4-2 of Table V; Problems 2-2, 4-1, 4-2, and 5-1 of Table VI), and eight significant violations of conditions (46) and (47) (Problems 1-2, 3-2, and 4-2 of Table V; Problems 1-2, 2-2, 4-1, 4-2, and 5-1 of Table VI). All of the violations were in the predicted direction, that is, an incompatible conjunction was rated below a corresponding compatible conjunction where condition (44), (45), (46), or (47) would require the opposite ordering.

We also tested whether conjunction errors occurred in the probability ratings of these reasoning problems. We will not describe this analysis in detail because it simply replicates the findings of other experiments, but the point to be made here is that conjunction errors did occur among the same problems that produced significant violations of conditions (44) through (47). Specifically, conjunction errors occurred in the comparisons of  $A_1$  to  $A_1 \wedge A_2$ , and  $B_1$  to  $B_1 \wedge B_2$  in Problem 1-2; in the comparison of  $B_2$  to  $B_1 \wedge B_2$  in Problem 2-2; and in the comparisons of  $A_2$  to  $A_1 \wedge A_2$ , and  $B_2$  to  $B_1 \wedge B_2$  in Problems 4-2 and 5-1. All of these comparisons were statistically significant by two-tailed sign tests ( $p < .05, .01, .005, .001, .001, .05$ , respectively). Note that these problems also produced violations of conditions (44) through (47) (see Tables V and VI). If violations of probability combination models occurred in a different class of problems from those that produce conjunction errors, one could retain probability combination models as explanations of conjunction errors, leaving the violations of (44) through (47) to be explained by a different theory pertaining to this other class of problems. This defense of probability combination models is not viable, however, for violations of (44) through (47) occurred among the same problems that produced conjunction errors, and hence, they must be explained by any theory of conjunction errors.

#### D. CONCLUSIONS FROM CONJUNCTION AND DISJUNCTION ERRORS IN PROBABILITY JUDGMENT

Both the representativeness hypothesis and probability combination models have difficulty accounting for all results on conjunction errors. For the representativeness theory, the main difficulty is to explain conjunction errors with conceptually unrelated propositions (Gavanski & Roskos-Ewoldsen, 1991; Yates & Carlson, 1986). According to this theory, conjunction errors occur because the conjunction evokes a composite representation that is more similar to the given information than the less representative component. It is not obvious what mechanisms would lead to a composite representation of conceptually unrelated components, especially one that is more similar to the given information than the less probable component.

Another potential difficulty for the representativeness theory are the disjunction errors (Carlson & Yates, 1989; Morier & Borgida, 1984; Wells, 1985). The problem for the representativeness theory lies in explaining how a disjunction of propositions is represented cognitively, and how its similarity to the given information is evaluated.

The results of Experiment 1 demonstrate that conceptual relations between the components of a conjunction affect belief for the conjunction in ways not predicted by the strengths of the components. Conceptual compatibility enhances the belief strength of a conjunction and conceptual incompatibility reduces belief strength. The violations of conditions (44) through (47) establish that probability combination models cannot account for judgments of conjunctive probabilities in general. Thus, weighted averaging, weighted averaging with initial impression, signed summation, multiplicative combination, min, and max can all be rejected as models for conjunctive probabilities in general. The results of Experiment 1 do not eliminate the possibility that conjunctions of conceptually unrelated components are evaluated by a probability combination model, because all of the problems in Experiment 1 involved conjunctions of conceptually related events. Indeed, the construction of counterexamples to the probability combination models was based on conceptually related events that varied in compatibility. Therefore, we cannot exclude the possibility that a probability combination model governs judgments of likelihood for conjunctions of conceptually unrelated components, and representativeness governs judgments of likelihood for conjunctions of conceptually related components.

We should mention several lines of inquiry that deserve further attention. First, can probability combination models explain judgments of probability for disjunctions? The finding that the probability of disjunctions correlates with the representativeness of disjunctions suggests that probability combination models will also fail as a general account of the belief strength of disjunctions. Perhaps an experiment analogous to Experiment 1 can be devised to test whether conceptual relations between the components affect the belief strengths of disjunctions. Second, we need to explore how models of representativeness can be extended to conjunctions of unrelated events and to disjunctions of events, whether related or unrelated. Third, we need to see how the belief strength of conjunctions varies as a function of variation in the evidence. Oddly, most studies that we are aware of (our own included) confound propositions with evidence, in the sense that a given proposition is always presented with the same, possibly implicit body of evidence. Experiments that hold propositions constant while varying the evidence can be contrasted with experiments that hold evidence constant while varying the semantic properties of the propositions. Such contrasts

might permit us to identify the independent contributions of propositional structure and evidence structure. Fourth, we believe that the relationship between compositional theories of typicality and compositional theories of belief deserves continued investigation. The theory of typicality is arguably the best developed theory of the representations and mechanisms underlying judgments of representativeness. Accordingly, the attempt to establish the relationship between belief strength and typicality constitutes an important test of representativeness explanations of belief.

## VI. Anomalous Compositions in Counterfactual Reasoning

In this section, we extend the discussion to the semantics of evidence for counterfactual conditionals. Counterfactual conditionals are statements of the form.

50. If  $A$  were the case, then  $X$  would be the case.

where  $A$  and  $X$  are propositions. Typically, the antecedent of a counterfactual,  $A$ , is a proposition that is known to be false, and the consequent,  $X$ , is a proposition whose truth is related to the truth of the antecedent. For example, the following is a counterfactual.

51. If Richard Nixon had not resigned from the presidency, he would have been impeached.

The semantic properties of counterfactuals have been the focus of philosophical investigations because important, yet problematic, questions appear to depend on the analysis of counterfactual inference (Chisholm, 1946; Gardenfors, 1988; Goodman, 1947; Harper, Stalnaker, & Pearce, 1981; Lewis, 1973; Sosa, 1975; Stalnaker, 1984).

Let  $A \mapsto X$  stand for a counterfactual like statement (50). A theory of evidence for counterfactuals must account for the belief strengths of  $p/e$  combinations,  $(A \mapsto X, e)$ , as a function of the semantic content of  $A$  and  $X$ , properties of the evidence  $e$ , and other beliefs that are not explicitly represented in the propositions or evidence. In the Nixon example, implicit beliefs would include beliefs concerning Richard Nixon, his political situation, and the conditions that would have influenced events had Nixon not resigned. Although we will not attempt to discuss philosophical issues pertaining to counterfactuals in this chapter, we would like to point out that the theory of evidence provides a new perspective on counterfactuals. Whereas previous studies have tended to ask the questions, what are the truth conditions for counterfactuals, and what are the logical entailments of counterfactuals, the theory of evidence focuses on the question of how

belief strength for  $(A \mapsto X, e)$  varies as a function of  $A$ ,  $X$ , and  $e$ . This is a natural question psychologically because many counterfactuals produce degrees of belief that are intermediate between certainty of truth or falsity. The theory of evidence attempts to model these intermediate degrees of belief as well as the special cases in which belief is at the extremes of certainty.<sup>12</sup>

What we examine in this section are compositional anomalies in counterfactual reasoning, in other words, inconsistencies between the propositional structure of counterfactuals and the belief strengths of counterfactuals. We first discuss the Ramsey test, a schema for conditional inference that was proposed by the English philosopher Frank Ramsey (1931). Next, we consider the Theory of IF due to Braine and O'Brien (1991). The Theory of IF includes a theory of counterfactual reasoning as part of a larger theory of conditional reasoning. We show that the Theory of IF incorporates a Ramsey test as part of a psychological model of conditional reasoning, and from the properties of the Ramsey test, one can derive implications among counterfactuals. In combination with Principle 1, these implications predict relations in belief strength among counterfactuals. We also consider alternative theories of counterfactual inference that embody versions of the Ramsey test. These theories also predict relations in belief strength among counterfactuals. We then show empirically that belief strengths are inconsistent with the predictions that were derived from the hypothesis of a Ramsey test. We conclude from this that the Ramsey test, which has influenced many modern theories of conditional reasoning, is incorrect as a descriptive model of counterfactual inference. We will attempt to pinpoint how actual reasoning with counterfactuals differs from the pattern of the Ramsey test.

### A. THE RAMSEY TEST HYPOTHESIS

The semantic properties of counterfactual conditionals are quite different from the properties of the material conditional, the if-then construction of the propositional calculus (Quine, 1972). A material conditional, "if  $A$ , then  $X$ ," is true if  $A$  is false or if  $X$  is true, or both. Interpreted as material conditionals, the following conditionals are both true because their antecedents are both false:

52. If  $2 + 2 = 5$ , then the moon is made of cheese.

53. If  $2 + 2 = 5$ , then the moon is not made of cheese.

<sup>12</sup> The observation that counterfactuals vary through intermediate levels of belief is not new: "Between subjunctive conditionals in a reasonably dispositional spirit and subjunctive conditionals at their wildest there is no boundary, but only a gradation of better and worse" (Quine, 1960, pp. 225). What is new is the proposal to use these gradations of belief as the data for a semantic theory.

Because a material conditional is true when its antecedent is false, material conditionals with self-contradictory antecedents are always true. The same is not the case for counterfactual conditionals. Consider the following example (from Goodman, 1947). Suppose that a match is sitting dry and unused in a box, and one says of this match:

54. If that match had been scratched, it would have lighted.

Clearly, this statement is not intended as a material conditional for if it were, the following would also be true (because (54) and (55) both have false antecedents):

55. If that match had been scratched, it would not have lighted.

Statements (54) and (55) illustrate the fact that the truth of a counterfactual depends on semantic relations between the antecedent and consequent beyond whether they are true or false. Rather, a counterfactual asserts or presupposes that some sort of relation obtains between antecedent and consequent, and the problem for philosophical analysis is to elucidate the nature of this relation (Chisholm, 1946; Goodman, 1947). To interpret (54) and (55), one assumes the validity of causal laws (flammable materials ignite when heated; scratching heats materials) and background conditions (oxygen was present; temperatures were not too cold; the surroundings were dry). In the context of these laws and background conditions, the scratching of a match results in the lighting of the match. The basic schema of counterfactual inference would thus seem to be that in the context of implicit background conditions and natural laws, the antecedent of a counterfactual implies the consequent of the counterfactual. Essentially, this schema was proposed by Ramsey (1931), in his influential analysis of conditional statements:

In general we can say with Mill that "if  $p$ , then  $q$ " means that  $q$  is inferrible from  $p$ , that is, of course, from  $p$  together with certain facts and laws not stated but in some way indicated by the context.

As Goodman (1947) pointed out, one cannot evaluate the truth of a counterfactual simply by inferring the consequences of the antecedent in the context of one's current beliefs because the antecedent typically contradicts other propositions that are currently believed. For example, the antecedent of (51) contradicts the belief that Nixon did resign. If the premises of a counterfactual inference consisted of current beliefs supplemented by the antecedent, then, typically, the antecedent would contradict other current beliefs, and from these contradictory premises, any conclusion whatsoever could be inferred.

To avoid these difficulties, let us amend the proposal by supposing that after adding the antecedent of a counterfactual to one's current beliefs, one temporarily suspends belief in propositions that contradict the antecedent. Stalnaker (1968) expressed this strategy as follows:

First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true. (p. 102)

The adjustments in beliefs made in the second step have the effect of eliminating beliefs that are inconsistent with the antecedent, and adding beliefs that are consequences of the antecedent and other consistent beliefs. A counterfactual is true, according to this theory, if the consequent is among the propositions that are believed after existing beliefs are modified to accommodate the hypothetical truth of the antecedent.

We will say the truth of a counterfactual is evaluated by a *Ramsey test* if the evaluation proceeds through the three steps described in the Stalnaker quotation (Harper, 1981). By the *Ramsey test hypothesis* we mean the hypothesis that counterfactuals are evaluated by a Ramsey test in actual reasoning. Neither Ramsey (1931) nor Stalnaker (1968) proposed their theories as descriptive psychological theories, for their primary interests were normative. We will see, however, that psychologists and researchers in artificial intelligence have proposed Ramsey tests as part of descriptive theories of conditional inference.

## B. BRAINE AND O'BRIEN'S THEORY OF IF

The Theory of IF is part of a descriptive theory of deductive reasoning developed by Braine and his colleagues in a series of publications (Braine, 1978, 1990; Braine, Reiser, & Rumin, 1984; Braine & Rumin, 1981). The general theory describes a set of inference schemata, and a reasoning program that characterizes how people select and apply the schemata in the construction of deductive sequences. The Theory of IF is the subtheory of the general theory that pertains to natural inferences with conditionals. In the Theory of IF, the truth of a conditional is evaluated by means of a Ramsey test:

To derive or evaluate *if  $p$  then . . .*, first suppose  $p$ ; for any proposition  $q$  that follows from the supposition of  $p$  taken together with other information assumed, one may assert *if  $p$  then  $q$* . (Braine & O'Brien, 1991, p. 183)

Braine and O'Brien call this the *Schema for Conditional Proof*. An important difference between the Theory of IF and standard propositional logic

lies in the treatment of contradictory premises. Whereas by standard logic, inconsistent premises imply that every proposition is true, in the Theory of IF, nothing follows from inconsistent premises other than the conclusion that at least one premise is false. This restriction on deduction is called the *Constraint on Conditional Proof*:

A supposition can be the antecedent of a conditional conclusion reached via Schema 2 [Schema for Conditional Proof] only if it is consistent with prior assumptions (i.e., premise assumptions plus any previously made suppositions). [Furthermore] an assumption reiterated into a conditional argument cannot contradict the supposition that is to be the antecedent of the conditional. (Braine & O'Brien, 1991, p. 185)

The Constraint on Conditional Proof implies that one cannot infer "if  $p$ , then  $q$ " when  $p$  is inconsistent with other premises.

It might seem that the Constraint on Conditional Proof would prevent Braine and O'Brien from giving an analysis of counterfactuals, for the antecedent of a counterfactual is typically inconsistent with other beliefs of the discourse participants. This is not the case, however, for Braine and O'Brien (1991) draw a distinction between the totality of beliefs that the reasoner holds and the premises that the reasoner introduces into a deduction that originates in a counterfactual supposition:

In this case of a deliberate counterfactual supposition, the premise assumptions can never be a record of an actual state of affairs. For example, if we wished to argue from the supposition *If Dukakis had won the 1988 election*, our premise assumptions could not be a record of the actual events of 1988; for example, they could not include the fact the Bush had won. (p. 184)

Thus, to infer the consequences of a counterfactual supposition,  $A$ , the reasoner cannot take the totality of propositions believed to be true as premises. Rather, the reasoner adopts as premises only the propositions that are consistent with  $A$  that would have been true had  $A$  been true. We note in passing that Braine and O'Brien (1991) are vague concerning the criteria or process by which these auxiliary premises are selected, but our critique of the Theory of IF can be given without explicitly spelling out these criteria.

In summary, Braine and O'Brien (1991) propose that a counterfactual,  $A \mapsto X$ , is evaluated by supposing that  $A$  is true and attempting to derive  $X$  from this supposition. Other propositions can be introduced as premises in this derivation provided that these propositions are consistent with  $A$ . Every inference in the derivation must be an instance of a basic inferential schema in Braine's general theory of inference.

We now state two implications of the Theory of IF. If an individual reasons in accordance with the Theory of IF, then he or she will satisfy the following hypotheses:

*Hypothesis 1:* Let  $p$ ,  $q$ , and  $r$  be any three propositions. If the reasoner believes that "if  $p$  were the case, then  $q$  and  $r$  would be the case," then she must also believe that "if  $p$  were the case, then  $q$  would be the case" and "if  $p$  were the case, then  $r$  would be the case." In other words,  $p \mapsto q \wedge r$  implies  $p \mapsto q$  and  $p \mapsto r$ .

*Hypothesis 2:* Let  $p$ ,  $q$ , and  $r$  be any three propositions. If the reasoner believes that "if  $p$  were the case, then  $q$  and  $r$  would be the case," then she must also believe that "if  $p$  and  $q$  were the case, then  $r$  would be the case" and "if  $p$  and  $r$  were the case, then  $q$  would be the case." In other words,  $p \mapsto q \wedge r$  implies  $p \wedge q \mapsto r$  and  $p \wedge r \mapsto q$ .

The proof that these hypotheses are implied by the Theory of IF is given in Appendix 2.

The Theory of IF was not proposed as a theory of belief strength; thus, it would be improper to claim that it predicts relations in belief strength on its own. In combination with Principle 1, however, Hypothesis 1 implies that

$$(p \mapsto q \wedge r, e) \leq (p \mapsto q, e) \text{ and } (p \mapsto q \wedge r, e) \leq (p \mapsto r, e), \quad (56)$$

and Hypothesis 2 implies that

$$(p \mapsto q \wedge r, e) \leq (p \wedge q \mapsto r, e) \text{ and } (p \mapsto q \wedge r, e) \leq (p \wedge r \mapsto q, e) \quad (57)$$

for all propositions  $p$ ,  $q$ , and  $r$ , and evidence  $e$ . Condition (56) is analogous to the probability relation  $P(B \cap C | A) \leq P(B | A)$ ; violations of (56) are analogous to conjunction errors in probability judgment. Condition (57) is analogous to the probability relations  $P(B \cap C | A) \leq P(C | A \cap B)$ ; violations of (57) are analogous to conditionalization errors in probability judgment. If judgments of belief strength violate (56), we call this a *counterfactual conjunction error*, and if they violate (57), we call this a *counterfactual conditionalization error*.

### C. ALTERNATIVE APPROACHES TO COUNTERFACTUAL INFERENCE

Perhaps the most influential among current theories of counterfactual inference is a model-theoretic analysis due to Stalnaker (1968) and Lewis (1973).

We will refer to this theory as the Stalnaker/Lewis theory, because Lewis's (1973) theory is a generalization of the basic principles proposed by Stalnaker (1968). The Stalnaker/Lewis theory is based on the concept of a *possible world*. Possible worlds are abstract entities relative to which propositions have truth values. The actual world is a possible world, but there are infinitely many possible worlds that differ from the actual world in a multitude of ways. According to the Stalnaker/Lewis theory, to evaluate whether  $A \mapsto X$  is true, one considers alternative possible worlds where  $A$  is true, and tests whether  $X$  is true in these worlds (a Ramsey test). In carrying out the Ramsey test, it is proposed that one considers only those worlds that are as similar as possible to the actual world, subject to the constraint that  $A$  is true in those worlds. If  $X$  is true in these maximally similar alternative worlds where  $A$  is true, then  $A \mapsto X$  is true. For example, to decide whether statement (51) is true, consider alternative possible worlds in which Nixon did not resign, subject to the constraint that these worlds should otherwise be as similar to the actual world as possible. If Nixon was impeached in all of these worlds, then (51) is true; if Nixon was not impeached in at least some of these worlds, then (51) is false.

Our description of the Stalnaker/Lewis theory attempts to convey the intuition behind the theory without delving into the formal structure of the theory. A more precise, mathematical formulation was given in Miyamoto et al. (1989) and, of course, in Stalnaker (1968) and Lewis (1973). The point we want to make here is that the Stalnaker/Lewis theory implies Hypotheses 1 and 2 (Miyamoto et al., 1989); hence, in combination with Principle 1, it predicts conditions (56) and (57). We should hasten to add that the Stalnaker/Lewis theory was proposed as a normative theory, and therefore empirical tests of (56) and (57) cannot refute it. Nevertheless empirical tests of these predictions may shed light on the relationship between reasoning norms and actual reasoning processes.

The Stalnaker/Lewis theory also implies another logical relation among counterfactuals:

*Hypothesis 3:* Let  $p$ ,  $q$ , and  $r$  be any three propositions. If the reasoner believes that "if  $p$  were the case, then  $q$  would be the case" or "if  $p$  were the case, then  $r$  would be the case," then he or she must also believe that "if  $p$  were the case, then  $q$  or  $r$  would be the case." In other words,  $p \mapsto q$  implies  $p \mapsto q \vee r$  and  $p \mapsto r$  implies  $p \mapsto q \vee r$ .

In combination with Principle 1, Hypothesis 3 predicts that

$$(p \mapsto q, e) \leq (p \mapsto q \vee r, e) \text{ and } (p \mapsto r, e) \leq (p \mapsto q \vee r, e). \quad (58)$$

Formal derivations of Hypotheses 1, 2, and 3 from the Stalnaker/Lewis theory are given in Miyamoto et al. (1989). The reason we did not attempt to derive Hypothesis 3 from the Theory of IF is that the inference schema,  $p$  implies  $p \vee q$ , is not a basic inference schema in the Theory of IF. Thus, it could be argued that in the Theory of IF, the inference of  $p \vee q$  from  $p$  involves greater cognitive complexity than the inferences by which Hypotheses 1 and 2 were derived.

Beginning with Ginsberg (1986), researchers in artificial intelligence (AI) have investigated the relation between the intuitions underlying the Stalnaker/Lewis theory and knowledge representations in large databases (see Eiter & Gottlob, 1992, for a review). In the AI literature, counterfactual reasoning is treated as a problem of resolving inconsistencies in a database when new, contradictory information is added to it (so-called non-monotonic reasoning). A counterfactual is evaluated as true if the consequent is a logical consequences of the database after adding the antecedent and updating the database. The updating process is constrained to maximize the similarity between the initial and updated database (Ginsberg, 1986) or to minimize the changes from the initial database (Eiter & Gottlob, 1992). Without attempting to explore the formal details of the AI approach, we note that these models incorporate Ramsey tests in the evaluation of counterfactuals (Eiter & Gottlob, 1992; see also Gardenfors, 1988, for similar ideas in epistemic logic); consequently, the AI analyses of counterfactuals imply Hypotheses 1, 2, and 3.

Rips and Marcus (1979) proposed a theory of conditional reasoning that combined the intuitive structure of the Stalnaker/Lewis theory with explicit assumptions concerning the representation and processing of propositions in working memory. The Rips and Marcus (1979) theory can be viewed as a precursor of the AI database update models discussed in Eiter and Gottlob (1992). Once again, we will not describe the details of Rips and Marcus (1979) other than to note that their theory incorporates a Ramsey test in the evaluation of conditionals, and, as such, it implies Hypotheses 1, 2, and 3. The simulation heuristic proposed by Tversky and Kahneman (1982b) is essentially a Ramsey test, although it was not presented as a logical analysis of counterfactual inference, but rather as a study of the situational factors that inspire counterfactual reasoning.

Before turning to empirical studies of belief strength for counterfactuals, we should mention that the mental modes theory of conditional reasoning due to Johnson-Laird (1986) has some of the flavor of a Ramsey test theory, but it differs from it in an important way. We will discuss this theory and

its relation to the Ramsey test hypothesis after presenting empirical findings for counterfactuals.

#### D. EMPIRICAL TESTS OF HYPOTHESES 1, 2, 3

Miyamoto and Dibble (1986) tested conditions (56) and (58), and Miyamoto et al. (1989) tested (56), (57), and (58). We will describe the latter study. Miyamoto et al. constructed counterfactual statements that exemplified the propositional structures in (56), (57), and (58). The statements pertained to the vacation plans of the Conley's, as described in the following story.

##### Vacation Plans

When Bill and Lucy Conley were planning their vacation last year, they had a friendly disagreement over what to do.

Bill wanted to go hiking and camping in the Canadian Rockies. He collected brochures showing the beauty of the lakes and mountains, spectacular ice fields, and magnificent forests. Lucy wanted to visit New York City where she had never been. She argued for the cultural advantages of New York City—the great art museums, theaters on and off Broadway, elegant and exotic restaurants, and famous stores.

Bill had to admit that New York had a lot to offer. He shared Lucy's taste for culture. His main concern was that New York would cost a lot more than camping and hiking in the Canadian Rockies. Food and lodging were much more expensive in New York. Theater tickets were very expensive and hard to get. The opera was also extremely expensive, but this didn't matter because they didn't like opera.

Of course, some things were not expensive even in New York. For example, the price of admission to art museums was very reasonable, and they both loved to see great art. It wouldn't cost anything to browse in the stores, as long as they didn't buy anything. Some friends had told them where they could hear good, live jazz for under \$5 apiece.

When the Conley's gave careful consideration to the cost of visiting New York, however, they decided that they really couldn't afford it this year. They briefly considered a compromise solution, a vacation in Denver, where they could combine city amusements with outdoor activities. It was quickly obvious, however, that neither Bill nor Lucy would be happy in Denver, so they scrapped that idea. In the end, they went camping in the Canadian Rockies and had a lot of fun for relatively little money. They promised themselves that they would save money for a vacation in New York someday.

Table VII lists the critical statements for tests of (56), (57), and (58). In addition to these statements, there were 15 other statements that were irrelevant to our present discussion. These 25 statements were presented to subjects in one of four random orders.

Subjects were 70 University of Washington undergraduates. None of the subjects had had a course in logic and all were native English speakers. Subjects read the story and rated the statements for "how true or false they seem based on the information in the preceding story and whatever else you know about the world." Ratings were made by placing a mark on a horizontal line labeled "Absolutely True" at one end and "Absolutely

TABLE VII

Statements: Topic 1	Form	Median
59. If the Conley's had vacationed in New York, they would have visited art museums.	$A \mapsto X$	26.0
60. If the Conley's had vacationed in New York, they would have attended the opera.	$A \mapsto Y$	3.0
61. If the Conley's had vacationed in New York, they would have visited art museums, and they would have attended the opera.	$A \mapsto X \wedge Y$	11.0
62. If the Conley's had vacationed in New York, they would have visited art museums, or they would have attended the opera, or both.	$A \mapsto X \vee Y$	18.0
63. If the Conley's had vacationed in New York and visited art museums, they would also have attended the opera.	$A \wedge X \mapsto Y$	5.0
Statements: Topic 2		
64. If the Conley's had vacationed in New York, they would have heard outstanding live jazz.	$A \mapsto X$	24.0
65. If the Conley's had vacationed in New York, they would have gone for late evening walks in Central Park.	$A \mapsto Y$	15.0
66. If the Conley's had vacationed in New York, they would have heard outstanding live jazz, and gone for late evening walks in Central Park.	$A \mapsto X \wedge Y$	19.0
67. If the Conley's had vacationed in New York, they would have heard outstanding live jazz, or gone for late evening walks in Central Park, or both.	$A \mapsto X \vee Y$	21.0
68. If the Conley's had vacationed in New York and had heard outstanding live jazz, they would have gone for late evening walks in Central Park.	$A \wedge X \mapsto Y$	15.0

False" at the other end. Numerical ratings were derived by measuring the position of the mark on a scale from 1 (= maximum falsity) to 30 (= maximum truth). Median ratings for the 10 statements are listed next to the statements. Note that the ordering of the medians violated statements (56), (57), and (58) for both topics:  $A \mapsto X \wedge Y$  received a higher median rating than  $A \mapsto Y$  in violation of (56);  $A \mapsto X \vee Y$  received a lower median rating than  $A \mapsto X$ , in violation of (58); and  $A \mapsto X \wedge Y$  received a higher median rating than  $A \wedge X \mapsto Y$ , in violation of (57). Table VIII shows the results of two-tailed sign tests for the critical comparisons. Each cell shows the percentage of times the first counterfactual in the column heading was rated higher than the second counterfactual (ties were dropped). As one can see, the violations of (56), (57), and (58) in Topics 1 and 2 were all significant.

As with probabilistic conjunction errors, it can be objected that counterfactual conjunction errors result from subjects and experimenters having



TABLE VIII

## COUNTERFACTUALS IN THE COMPARISON

	$A \mapsto X \wedge Y$ $A \mapsto Y$	$A \mapsto X \vee Y$ $A \mapsto X$	$A \mapsto X \wedge Y$ $A \wedge X \mapsto Y$
Topic 1	87.9, $p < .001$	9.7, $p < .001$	91.1, $p < .001$
Topic 2	70.0, $p < .01$	27.8, $p < .005$	80.0, $p < .001$

different interpretations of the stimulus statements. Specifically, subjects might contrast  $A \mapsto Y$  to  $A \mapsto X \wedge Y$ , leading them to interpret  $A \mapsto Y$  as  $A \mapsto Y \wedge \neg X$  (cf. Marcus & Zajonc, 1985, and Pennington, 1984). One way to counter this objection is to have different subjects rate  $A \mapsto Y$  and  $A \mapsto X \wedge Y$ . We conducted such an experiment using the Vacation Plans story. Approximately half the subjects rated  $A \mapsto X$  and  $A \mapsto Y$  for Topic 1, and  $A \mapsto X \wedge Y$  for Topic 2. The remaining subjects rated  $A \mapsto X \wedge Y$  for Topic 1 and  $A \mapsto X$  and  $A \mapsto Y$  for Topic 2. For both topics,  $A \mapsto X \wedge Y$  received significantly higher ratings than  $A \mapsto Y$  ( $U = 6,642.5$ ,  $p < .001$  for Topic 1;  $U = 8,008.5$ ,  $p < .001$  for Topic 2;  $n = 158, 157$ ). Thus, counterfactual conjunction errors occurred even when the subject did not have an opportunity to contrast  $A \mapsto Y$  and  $A \mapsto X \wedge Y$  for the same topic, which was the rationale for interpreting  $A \mapsto Y$  as  $A \mapsto Y \wedge \neg X$ . It appears that counterfactual conjunction errors do not result from this kind of miscommunication.

We conclude that belief strengths for counterfactuals violate Principle 1. The violations have the form:

Counterfactual conjunction error:	$(A \mapsto X \wedge Y, e) > (A \mapsto Y, e)$
Counterfactual disjunction error:	$(A \mapsto X, e) > (A \mapsto X \vee Y, e)$
Counterfactual conditionalization error:	$(A \mapsto X \wedge Y, e) > (A \wedge X \mapsto Y, e)$

#### E. THE RAMSEY TEST HYPOTHESIS AND ANOMALIES OF COUNTERFACTUAL BELIEF

The experimental results demonstrate the occurrence of counterfactual conjunction, disjunction, and conditionalization errors. We will argue that these results suggest a modification of the Ramsey test in which judgments of similarity replace judgments of truth in a model. The argument applies to many versions of the Ramsey test, not only to the Theory of IF and the Stalnaker/Lewis semantics, but also to the AI theories reviewed in Eiter and Gottlob (1992) and to Rips and Marcus's (1979) theory of conditionals.

The following argument attempts to identify the features of the Ramsey test that are questionable from a descriptive standpoint.

First, let us introduce some notation. Suppose that  $A \mapsto X$  is a counterfactual, and that  $\mathcal{B}(e)$  represents the current state of belief. Depending on the theoretical framework,  $\mathcal{B}(e)$  might be a propositional network, possible world, or a knowledge base. We treat current belief as a function of the evidence  $e$  because it is natural to expect belief to be influenced by evidence. It is likely that belief is also affected by other factors, such as learning history, but such factors remain implicit in our notation because we are not attempting to model them at this time. Let  $\mathcal{W}[A, \mathcal{B}(e)]$  denote the revision of  $\mathcal{B}(e)$  to accommodate the hypothetical truth of  $A$ . For purposes of our critique, we need not specify the criteria that  $\mathcal{W}[A, \mathcal{B}(e)]$  must satisfy other than that  $\mathcal{W}[A, \mathcal{B}(e)]$  is a set of one or more representations that are a function of  $\mathcal{B}(e)$  and  $A$ .

In any theory that embodies a Ramsey test, the truth of  $A \mapsto X$  is evaluated by testing whether  $X$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$ . Hypothesis 1 is derived from the fact that in any such theory,  $X \wedge Y$  implies  $Y$  and, hence,  $Y$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$  if  $X \wedge Y$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$ . In most such theories,  $X$  implies  $X \vee Y$ , and Hypothesis 3 results from the fact that  $X \vee Y$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$  if  $X$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$ . Hypothesis 2 was derived from the fact that in the Theory of IF and in the Stalnaker/Lewis semantics, if  $X$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$ , then  $\mathcal{W}[A \wedge X, \mathcal{B}(e)] = \mathcal{W}[A, \mathcal{B}(e)]$  (see Appendix 2 and Miyamoto et al., 1989). Although we do not have a general proof that this must hold in every version of a Ramsey test, this equation is plausible because if  $X$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$ , then a situation in which  $\mathcal{B}(e)$  represents one's beliefs and  $A$  is posited to be true must be very similar to a situation in which  $\mathcal{B}(e)$  represents one's beliefs and  $A \wedge X$  is posited to be true. If  $X \wedge Y$  is true in  $\mathcal{W}[A, \mathcal{B}(e)]$ , and  $\mathcal{W}[A \wedge X, \mathcal{B}(e)] = \mathcal{W}[A, \mathcal{B}(e)]$ , then  $Y$  is true in  $\mathcal{W}[A \wedge X, \mathcal{B}(e)]$ , implying Hypothesis 2. The purpose of these arguments is to show that Hypotheses 1, 2, and 3 follow from the general structure of the Ramsey test. Therefore, most versions of the Ramsey test will combine with Principle 1 to yield the predictions, (56), (57), and (58) that were found to be violated empirically.

To see what psychological mechanisms would produce the observed pattern of belief strengths, consider first the representativeness explanation of conjunction errors. According to the representativeness hypothesis, the similarity of a conjunction to a standard of comparison can be greater than the similarity of a single component. Applying this idea to the Ramsey test, one could argue that the procedure for testing the truth of a conjunctive consequent,  $X \wedge Y$ , against the revised beliefs,  $\mathcal{W}[A, \mathcal{B}(e)]$ , is also a judgment of similarity. From this, we infer that the conjunction  $X \wedge Y$  could be more similar to  $\mathcal{W}[A, \mathcal{B}(e)]$  than a single component,  $Y$  alone, and that

these differences in similarity could account for counterfactual conjunction errors. The essence of our proposal is to alter the "test" in the Ramsey test from a test of truth of the consequent in a revision of current beliefs to a judgment of similarity between the consequent and these revised beliefs.

Let us say that counterfactuals are evaluated by a *modified Ramsey test* if the evaluation of belief strength for  $(A \mapsto X, e)$  follows these four steps: first, add the antecedent  $A$  to one's current beliefs  $\mathcal{B}(e)$ ; second, construct a revised belief representation  $\mathcal{W}[A, \mathcal{B}(e)]$  that modifies  $\mathcal{B}(e)$  to accommodate the hypothetical truth of  $A$ ; third, evaluate the similarity of  $X$  to  $\mathcal{W}[A, \mathcal{B}(e)]$ ; fourth, let the belief strength for  $(A \mapsto X, e)$  be an increasing function of this similarity. The revised representation constructed at the second stage may satisfy some criterion of maximal similarity to current belief, as in the Stalnaker/Lewis theory, or a criterion of minimal change from current belief, as in the AI theories reviewed in Eiter and Gottlob (1992). The modified Ramsey test differs from the original Ramsey test at step 3, where a judgment of similarity replaces a judgment of truth, and at step 4, where belief strength is evaluated. Step 4 is omitted from the original Ramsey test because the original Ramsey test only allows for counterfactuals to be true or false without qualification.

To explain how  $A \mapsto X \wedge Y$  could have greater belief strength than  $A \mapsto Y$ , one must postulate a cognitive representation in which  $X \wedge Y$  is more similar to  $\mathcal{W}[A, \mathcal{B}(e)]$  than is  $Y$  alone. Given such a representation, it follows that  $(A \mapsto X \wedge Y, e) > (A \mapsto Y, e)$  because belief strength is an increasing function of similarity. In effect, the representativeness explanation for conjunction errors in probability judgment has a straightforward generalization to counterfactual conjunction errors. It is not clear how a modified Ramsey test could explain counterfactual disjunction errors, but the difficulty here is much like the problem of providing a representativeness explanation for disjunction errors in probabilistic reasoning. As yet, it is not clear how to represent the typically or similarity of a disjunction. A representativeness account of counterfactual conditionalization errors seems to be quite promising. If the belief strength of  $A \mapsto X$  is high, then  $X$  and  $\mathcal{W}[A, \mathcal{B}(e)]$  must be highly similar. But if  $X$  and  $\mathcal{W}[A, \mathcal{B}(e)]$  are highly similar, then  $\mathcal{W}[A, \mathcal{B}(e)]$  must be highly similar to  $\mathcal{W}[A \wedge X, \mathcal{B}(e)]$ . Therefore, it is possible for  $X \wedge Y$  to be more similar to  $\mathcal{W}[A, \mathcal{B}(e)]$  than  $Y$  alone is to  $\mathcal{W}[A \wedge X, \mathcal{B}(e)]$ . Hence, the belief strength for  $A \mapsto X \wedge Y$ , which is a function of the similarity of  $X \wedge Y$  to  $\mathcal{W}[A, \mathcal{B}(e)]$ , could be higher than the belief strength for  $A \wedge X \mapsto Y$ , which is a function of the similarity of  $Y$  to  $\mathcal{W}[A \wedge X, \mathcal{B}(e)]$ .

Elsewhere, we attempted to explain compositional anomalies in counterfactual reasoning by means of a mechanism that is structurally different from the modified Ramsey test (Miyamoto et al., 1989). The essence of

this alternative theory is to propose that people compare two different representations,  $\mathcal{W}[A, X, \mathcal{B}(e)]$  and  $\mathcal{W}[A, \neg X, \mathcal{B}(e)]$ , when evaluating the truth of  $A \mapsto X$ . The former,  $\mathcal{W}[A, X, \mathcal{B}(e)]$ , is the most similar world in which  $A$  and  $X$  are both true, and the latter,  $\mathcal{W}[A, \neg X, \mathcal{B}(e)]$ , is the most similar world in which  $A$  is true and  $X$  is false. For example, in deciding whether statement (51) is true, one compares the most plausible scenario in which Nixon does not resign and Nixon is impeached to the most plausible scenario in which Nixon does not resign and Nixon is not impeached. Statement (51) is believed to be true to the degree that the former scenario is more plausible than the latter. In effect,  $\mathcal{W}[A, X, \mathcal{B}(e)]$  is the most plausible example, and  $\mathcal{W}[A, \neg X, \mathcal{B}(e)]$  is the most plausible counterexample to  $A \mapsto X$ .  $A \mapsto X$  is judged to be true to the extent that the most plausible example is more plausible than the most plausible counterexample to the counterfactual hypothesis.

This approach to counterfactual judgments is similar to the role of examples and counterexamples in Johnson-Laird's theory of deductive reasoning (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991), and especially to his proposal that for certain classes of counterfactuals, the truth of a counterfactual is judged in terms of "[whether] the consequent is true with respect to the model based on the antecedent and any relevant beliefs (*including those triggered by the consequent*)." (p. 72) (Johnson-Laird, 1986; the italics are ours). It is this last point that separates Johnson-Laird's (1986) theory of counterfactuals from the various Ramsey test theories and our own modified Ramsey test theory. In a Ramsey test, the consequent of  $A \mapsto X$  has no influence over the construction of a revised belief representation,  $\mathcal{W}[A, \mathcal{B}(e)]$ . In contrast, Johnson-Laird's (1986) theory proposes that the consequent can influence the construction of a mental model of the antecedent. Similarly, the relative plausibility theory of Miyamoto et al. (1989) proposes that the consequent could influence the construction of a model,  $\mathcal{W}[A, X, \mathcal{B}(e)]$ , and countermodel,  $\mathcal{W}[A, \neg X, \mathcal{B}(e)]$ , upon which the judgment of relative plausibility is based. We will not attempt to compare the modified Ramsey test theory presented here to the relative plausibility theory presented in Miyamoto et al. (1989), for our discussion is already lengthy and at present there are no data that distinguish between these theories.

## VII. Conclusions

The semantic theory of evidence attempts to integrate two domains of research, the semantic theory of natural language and the study of inductive inference. To be sure, theorists have long envisaged a marriage of these

disciplines (Keynes, 1921; Carnap, 1952), but their conceptualizations focused on normative theories of deductive and inductive inference. Our proposal is to integrate these theories at the descriptive level of the psychology of reasoning. Since the linguistic revolution of the late 1950s and 1960s, there has been an explosion of studies of natural language inference and the semantic structure of natural language. At the same time, but without strong interconnections to linguistic research, the study of the Bayesian theory and expected utility theory stimulated enormous progress in the psychology of inductive reasoning. We believe there exists the potential for a productive interchange between these lines of research.

The theory of natural language semantics emphasizes the idea that the semantic structure of propositions is a compositional structure—complex expressions are built from more elementary parts by means of syntactic rules, and the semantic properties of expressions are determined by this compositional structure. In a sense, what the theory of inductive inference brings to this enterprise is a new dependent variable, namely, the study of relations in belief strength. Whereas the principal objects of classical semantic studies are referential relations, truth conditions, and relations of logical consequence, we suggest that the study of belief strength will provide additional insights into language structure and the role of this structure in reasoning. The theory of inductive inference also contributes a new independent variable to the study of language, namely, variations in the evidence with respect to which the belief strengths of propositions are evaluated. Table IX summarizes the propositional structures that have been discussed in this chapter along with the p/e pairs that enter into a particular structure. Many of these structures were only briefly discussed in this chapter; we included them here to emphasize the variety of research problems

that will emerge in a semantic theory of evidence. As these examples show, it is not so much the belief strengths of individual p/e pairs that require explanation when taken in isolation, rather, it is the relations in belief strength among different p/e pairs that are the main objects of study.

The interaction between compositional structure and belief strength is richly illustrated in the study of conjunctive propositions. Conjunction is one of the most basic forms of semantic composition. An interesting point that emerged from the investigation of conjunctions was the finding that the belief strength of a conjunction is not simply a function of the strengths of its components. Contrary to the predictions of probability combination models, the belief strength of a conjunction is a function of conceptual relations between the component propositions as well as their respective belief strengths. For the sake of brevity, we did not explore the details of a representativeness theory of conjunctive probabilities, but it, too, can be viewed as a compositional theory (Medin & Shoben, 1988; Osherson & Smith, 1982; Smith & Osherson, 1984, 1989; Smith et al., 1988). In a representativeness theory, what are composed are the property structures and prototype representations of categories, not the belief strengths of propositions. The belief strength of a conjunction derives from the conceptual combination of its components, and the psychological similarity of this combination to a target event or population. It is still too early to evaluate whether this program will succeed in explaining the belief strengths of conjunctions in general, but the point we wish to make is that in the case of conjunctive propositions, the attempt to explain belief strength is leading towards a more refined analysis of the internal structure of propositions and concepts.

We conjecture that for many, perhaps all, propositional structures, the attempt to explain relations in belief strength will lead to a more refined theory of the cognitive representation of semantic structure, and the processes that access or transform this structure. We did not attempt to survey the work on belief for disjunctive propositions, but it should be clear even from our superficial discussion that the modeling of belief strength for disjunctions presents an interesting challenge for both the representativeness theory and probability combination models. We proposed the study of counterfactual conditionals as another domain in which compositional structure is likely to interact in interesting ways with evidence and strength of belief. Experimental results suggest that the Ramsey test, which has served as a paradigm for many theories of counterfactual reasoning, is descriptively inadequate. Where the Ramsey test tests the truth of a consequent in a model of the antecedent, or the proof of a consequent under the supposition of the antecedent, we propose that intuitive reasoning tests the representativeness of the consequent in a model of the antecedent.

TABLE IX

Compositional structure	Comparison of P/E combinations
Conjunctions	$(A \wedge B, e), (A, e), (B, e)$
Disjunctions	$(A \vee B, e), (A, e), (B, e)$
Quantificational conjunctions	$[(\forall x)(x \in A \rightarrow x \in C), e], [(\forall x)(x \in B \rightarrow x \in C), e],$ $[(\forall x)(x \in A \cap B \rightarrow x \in C), e]$
Conditionalization	$(A \wedge B, e), (A   B, e), (B   A, e), (A, e), (B, e)$
Conditionals and priors	$(A, e), (B, e), (A   B, e), (B   A, e)$
Counterfactual conjunctions	$(A \mapsto X \wedge Y, e), (A \mapsto X, e), (A \mapsto Y, e)$
Counterfactual disjunctions	$(A \mapsto X \vee Y, e), (A \mapsto X, e), (A \mapsto Y, e)$
Counterfactual conditionalization	$(A \mapsto X \wedge Y, e), (A \wedge X \mapsto Y, e), (A \wedge Y \mapsto X, e),$ $(A \mapsto X, e), (A \mapsto Y, e)$

Kahneman and Tversky have often emphasized that the properties of similarity are structurally different from the properties of a probability measure, and, we would add, they are also structurally different from the properties of logical consequence. It remains to be shown, however, that this intuitively appealing approach to counterfactual beliefs can develop into a well-articulated theory. The point we make here is merely that the semantic theory of evidence is likely to be productive because there are a number of research problems in this area that have interesting structure.

Another feature of the semantic theory of evidence that was implicit in our treatment of conjunctive propositions is the central role of conjoint measurement theory in theory construction. Conjoint measurement theory is a theory of qualitative or ordinal assumptions that are implied by quantitative models of judgment like the probability combination models for conjunctions. In many cases, it has been possible to axiomatize psychological models in the conjoint measurement framework, in other words, to discover qualitative assumptions that are jointly sufficient for the validity of a particular model. The axiomatic analysis of subjective probability and utility are among the most famous examples of the conjoint measurement methodology (Luce & Suppes, 1965; Wallsten, 1977), but the methodology can now be applied quite generally to the analysis of psychological models (Falmagne, 1985; Krantz et al., 1971; Luce, Krantz, Suppes, & Tversky, 1990; Suppes, Krantz, Luce, & Tversky, 1989). The value of an axiomatic analysis is that it identifies critical ordinal predictions of a psychological model. For example, the analysis of probability combination models presented in this chapter was essentially an axiomatic analysis. The critical predictions derived in this analysis, conditions (46) and (47), are independence assumptions, a very common form of axiom in conjoint measurement analyses (Krantz et al., 1971; Falmagne, 1986). We used a standard line of measurement theoretic reasoning to discover predictions whose empirical testing lead to the rejection of the probability combination models for conjunctions. The rejection of probability combination models does not lead away from an axiomatic analysis, however, for the similarity theory that underlies the representativeness account is equally amenable to an axiomatic analysis (Osherson, 1987; Tversky, 1977).

Conjoint measurement theory should be an effective tool in the semantic theory of evidence because it is well adapted to attacking the kinds of research problems that arise in this area. Whereas physical measurement systems are largely (but not exclusively) based on concatenations of objects, psychological measurement of the rigorous axiomatic variety has been based more often on trade-offs between different orderings that arise within structured sets of objects (Krantz et al., 1971; Luce et al., 1990). If we are correct in supposing that belief strength is systematically related to the

compositional structure of propositions, as suggested by the examples in Table IX, then the belief orderings between related propositions are likely candidates for conjoint measurement analysis. We have not attempted a full-blown measurement analysis of belief in this chapter, but the examples that we have discussed are suggestive of the kinds of relationships that can be investigated effectively in the conjoint measurement framework (see also Krantz, 1991).

Finally, we must ask what the anomalies discussed in this chapter tell us about the semantics of evidence. Belief strength was found to be inconsistent with relations of logical consequence among conjunctive and disjunctive propositions, and among counterfactuals with conjunctive or disjunctive consequents. We also noted that conditionalization, the shifting of a proposition from a conjunctive consequent into the antecedent, reduced the belief strength of counterfactuals, whereas a logical analysis would require that belief should stay constant or increase. Evidently, Principle 1 is descriptively invalid, that is, relations in belief strength are inconsistent with logical consequence. This would seem to show that deductive reasoning plays a limited role in the formation of belief strength, although this may be an overgeneralization. Perhaps it plays a role in specific contexts that are different from the contexts that have been studied in the experiments discussed here. In view of the clear evidence demonstrating that naive reasoning violates norms of probability theory (Kahneman, Slovic, & Tversky, 1982) and norms of deductive inference (Evans, 1982; Johnson-Laird & Byrne, 1991), it is not surprising that inconsistencies should be found between deductive and inductive reasoning.

What may be less obvious, however, is that these findings cast doubt upon the proposal that the study of belief strength will contribute to a compositional theory of natural language semantics. This pessimistic view can be argued as follows. Much of what we know about semantic structure derives from studying intuitions of reference, truth and logical consequence, and inferring propositional structures as explanatory hypotheses for these intuitions. At least, such is the explicit strategy in the Davidson program for natural language semantics (cf. Davidson & Harman, 1975b, and other references therein), and something very like it motivates the work of theorists as diverse as McCawley (1993), Montague (1974), and Quine (1972).<sup>13</sup> If belief strength is inconsistent with logical consequence, then it is unlikely that the study of belief strength can contribute to a theory of compositional structure, for this structure has been inferred from intuitions of logical consequence and the postulated structure is intended to explain these intuitions.

<sup>13</sup> Semantic theory in the Chomskian tradition is less closely tied to referential semantics, although it is by no means independent from it (cf. Chomsky, 1986; Jackendoff, 1990).

An optimist might reply that the inconsistency between belief strength and logical consequence gives us reason to hope that belief strength is sensitive to aspects of semantic structure to which logical consequence is insensitive. Specifically, belief strength appears to be sensitive to psychological representations of similarity and typicality that run counter to truth functional operations like conjunction and disjunction (Tversky & Kahneman, 1982a, 1983), and to the test of truth in a model that is at the heart of the Ramsey test for conditionals. To develop a representativeness theory of belief strength, it will be necessary to determine the conceptual representations of composite categories and propositions, and to describe the computations of similarity or typicality that apply to these representations. We hope that this endeavor will discover aspects of semantic structure that could never be revealed solely by the study of reference, truth, and logical consequence.

#### ACKNOWLEDGMENTS

Many of the ideas concerning the semantic theory of evidence were developed in discussions between John Miyamoto and David Krantz while Miyamoto was a graduate student at the University of Michigan. We acknowledge Krantz's invaluable influence on this work. We thank Jay Lundell for his contributions to the work on conjunctions and counterfactuals reported here.

### Appendix 1: Vignettes and Statements for Experiment 1

#### PROBLEM 1-1

*Background Information.* Tracy is 35 years old. She works in a major hospital. She has always been a very happy person, who enjoys life a great deal and has lots of friends. She likes to travel; her most recent trips were to Tahiti and Bali.

##### Statements

A1	Tracy has never been married.
A2	Tracy's boyfriend is a devout Catholic.
B1	Tracy is divorced.
B2	Tracy has two children.
A1 $\wedge$ A2	Tracy has never been married and her boyfriend is a devout Catholic.
A1 $\wedge$ B2	Tracy has never been married and has two children.
B1 $\wedge$ A2	Tracy is divorced and her boyfriend is a devout Catholic.
B1 $\wedge$ B2	Tracy is divorced and has two children.
filler	Tracy is afraid to fly in airplanes.
filler	Tracy earns over \$20,000 per year.

#### PROBLEM 1-2

*Background Information.* The choices presented below represent different possible forms that a seven-letter word might take. For example, "----i-" represents a seven-letter word with "i" as the sixth letter. "Acclaim" and "abstain" are examples of words that fit this form. Consider each of the forms described below. For each form, rate the probability that a seven-letter word randomly chosen from the dictionary would have that form.

##### Statements

A1	----n-
A2	-----g
B1	----l-
B2	-----y
A1 $\wedge$ A2	----n g
A1 $\wedge$ B2	----n y
B1 $\wedge$ A2	----l g
B1 $\wedge$ B2	----l y
filler	-----s
filler	-----x

#### PROBLEM 2-1

*Background Information.* Joan W. is a highly ambitious woman whose professional career is the single most important interest in her life. Eight years ago, Joan graduated in the top 3% of her class at a prestigious law school. She works many long hours and most of her friends are associated in some way with her professional life. Her principal indulgence is that she enjoys buying expensive, fashionable clothing.

##### Statements

A1	Joan is a highly respected corporate lawyer.
A2	Joan's husband is the chairman of a powerful banking conglomerate.
B1	Joan is a well-known criminal lawyer.
B2	Joan's husband is a rather macho police captain.
A1 $\wedge$ A2	Joan is a highly respected corporate lawyer and is married to the chairman of a powerful banking conglomerate.
A1 $\wedge$ B2	Joan is a highly respected corporate lawyer and is married to a rather macho police captain.
B1 $\wedge$ A2	Joan is a well-known criminal lawyer and is married to the chairman of a powerful banking conglomerate.
B1 $\wedge$ B2	Joan is a well-known criminal lawyer and is married to a rather macho police captain.
filler	Joan knows how to type.
filler	Joan watches soap operas for at least 2 hours each day.

#### PROBLEM 2-2

*Background Information.* Jane is a sophomore in the University of Washington and her cumulative GPA is 3.4. She is taking Psych 101 this quarter. Please predict her performance in this course.

## Statements

A1	Jane gets a 3.7 on her first test.
A2	Jane gets a 3.9 for the class at the end of the quarter.
B1	Jane gets a 3.1 on her first test.
B2	Jane gets a 2.9 for the class at the end of the quarter.
A1 $\wedge$ A2	Jane gets a 3.7 on her first test and gets a 3.9 for the class at the end of the quarter.
A1 $\wedge$ B2	Jane gets a 3.7 on her first test and gets a 2.9 for the class at the end of the quarter.
B1 $\wedge$ A2	Jane gets a 3.1 on her first test and gets a 2.9 for the class at the end of the quarter.
B1 $\wedge$ B2	Jane gets a 3.1 on her first test and gets a 3.9 for the class at the end of the quarter.
filler	Jane drops out of Psych 101.
filler	Jane passes Psych 101.

## PROBLEM 3-1

*Background Information.* Susan is a quiet girl and spends a great deal of her time in the school library. She is shy and does not like parties. Most of her energies are directed toward her school work, and her grades are very good. She has only a few friends, most of whom are students as well.

## Statements

A1	Susan is a music student.
A2	Susan plays the saxophone for relaxation.
B1	Susan is a business administration student.
B2	Susan reads the Wall Street Journal every morning at breakfast.
A1 $\wedge$ A2	Susan is a music student and plays the saxophone for relaxation.
A1 $\wedge$ B2	Susan is a music student and reads the Wall Street Journal every morning at breakfast.
B1 $\wedge$ A2	Susan is a business administration student and plays the saxophone for relation.
B1 $\wedge$ B2	Susan is a business administration student and reads the Wall Street Journal every morning at breakfast.
filler	Susan always cheats on her exams.
filler	Susan's GPA is above 3.0.

## PROBLEM 3-2

*Background Information.* Edward L. is married and has four children. He has a good income. Due to the nature of his job, he often needs to work nights and weekends. He dislikes this aspect of his job because it keeps him from spending time with his family.

## Statements

A1	Edward works in a noisy factory.
A2	Edward is becoming prematurely deaf.
B1	Edward is a policeman who patrols a rough neighborhood.
B2	Edward was stabbed last year by a drug addict.

A1 $\wedge$ A2	Edward works in a noisy factory and is becoming prematurely deaf.
A1 $\wedge$ B2	Edward works in a noisy factory and was stabbed last year by a drug addict.
B1 $\wedge$ A2	Edward is a policeman who patrols a rough neighborhood and is becoming prematurely deaf.
B1 $\wedge$ B2	Edward is a policeman who patrols a rough neighborhood and was stabbed last year by a drug addict.
filler	Edward is under 20 years old.
filler	Edward is over 20 years old.

## PROBLEM 4-1

*Background Information.* Joe B. is a high school senior. He is a good student who is extremely active in extracurricular activities, especially football, rugby, school theatricals, and school politics. His high school GPA is 3.4. Joe has applied to four universities: Harvard, Princeton, Oklahoma and Texas at Austin.

## Statements

A1	Joe is accepted at Harvard.
A2	Joe is accepted at Princeton.
B1	Joe is rejected at Oklahoma.
B2	Joe is rejected at Texas.
A1 $\wedge$ A2	Joe is accepted at Harvard and accepted at Princeton.
A1 $\wedge$ B2	Joe is accepted at Harvard and rejected at Texas.
B1 $\wedge$ A2	Joe is rejected at Oklahoma and accepted at Princeton.
B1 $\wedge$ B2	Joe is rejected at Oklahoma and rejected at Texas.
filler	Joe will go to college.
filler	Joe will take an extra year to finish high school.

## PROBLEM 4-2

*Background Information.* The following problems ask you to evaluate the likelihood of various conditions that might affect the price of oil and gasoline in 1991<sup>14</sup>.

## Statements

A1	Oil exports from Saudi Arabia are severely restricted during 1991, due to military conflicts between Iran and its neighbors.
A2	The price of gasoline rises over \$2.50 per gallon in 1991.
B1	In 1991, Dow Chemicals discovers an inexpensive way to manufacture synthetic gasoline.
B2	The price of gasoline drops below \$.50 per gallon in 1991.
A1 $\wedge$ A2	Oil exports from Saudi Arabia are severely restricted during 1991, due to military conflicts between Iran and its neighbors, and the price of gasoline rises over \$2.50 per gallon.
A1 $\wedge$ B2	Oil exports from Saudi Arabia are severely restricted during 1991, due to military conflicts between Iran and its neighbors; nevertheless the price of gasoline drops below \$.50 per gallon.

<sup>14</sup>The experiment was conducted during January to March of 1989. Iraq invaded Kuwait on August 2, 1990.

$B1 \wedge A2$	In 1991, Dow Chemicals discovers an inexpensive way to manufacture synthetic gasoline; nevertheless, the price of gasoline rises over \$2.50 per gallon.
$B1 \wedge B2$	In 1991, Dow Chemicals discovers an inexpensive way to manufacture synthetic gasoline, and the price of gasoline drops below \$.50 per gallon.
filler	Saudi Arabia is a major oil producer in 1991.
filler	Japan is a major oil producer in 1991.

## PROBLEM 5-1

*Background Information.* The following problems ask you to evaluate the likelihood of various events that might affect the space shuttle program in 1990<sup>15</sup>. Read each statement carefully. Place a vertical mark on the rating line indicating how sure you are that the statement is true or false.

Statements	
A1	During 1990, another space shuttle blows up during lift-off.
A2	During 1990, Congress votes to eliminate the space shuttle program.
B1	During 1990, the Soviet Union launches a successful manned spaceflight to Mars.
B2	During 1990, Congress votes to divert \$4 billion from the defense budget to the space shuttle program.
$A1 \wedge A2$	During 1990, another space shuttle blows up during lift-off, and Congress votes to eliminate the space shuttle program.
$A1 \wedge B2$	During 1990, another space shuttle blows up during lift-off, and Congress votes to divert \$4 billion from the defense budget to the space shuttle program.
$B1 \wedge A2$	During 1990, the Soviet Union launches a successful manned spaceflight to Mars, and Congress votes to eliminate the space shuttle program.
$B1 \wedge B2$	During 1990, the Soviet Union launches a successful manned spaceflight to Mars, and Congress votes to divert \$4 billion from the defense budget to the space shuttle program.
filler	During, 1990, a space shuttle flight sets a record for the most number of orbits by a manned space vehicle.
filler	During 1990, a robot is sent into outer space along with human crew members.

## PROBLEM 5-2

*Background Information.* Tom is currently retired. He was born in a small town in the Midwest, and has returned to his hometown after retirement. Tom has made enough money in his working days to retire comfortably. He spends his time fishing and playing card games with his friends.

Statements	
A1	Tom was a professional football player.
A2	Tom has knee problems.
B1	Tom is a cigarette smoker.

<sup>15</sup> The experiment was conducted during January to March of 1989. The space shuttle Challenger exploded during take off on January 28, 1986.

B2	Tom has had lung cancer, which is presently in remission.
$A1 \wedge A2$	Tom was a professional football player, and has knee problems.
$A1 \wedge B2$	Tom was a professional football player, and he has had lung cancer, which is presently in remission.
$B1 \wedge A2$	Tom is a cigarette smoker, and has knee problems.
$B1 \wedge B2$	Tom is a cigarette smoker, and he has had lung cancer, which is presently in remission.
filler	Tom is under 40 years old.
filler	Tom is at least 50 years old.

## Appendix 2

*Proof that Eq. (43) implies conditions (44)–(47):* Choose any propositions  $C$ ,  $D$ , and  $X$ , and evidence  $e$ . Then,

$$(C, e) > (D, e) \text{ iff } S(C, e) > S(D, e) \quad (69)$$

$$\text{iff } G[S(C, e), S(X, e)] > G[S(D, e), S(X, e)] \quad (70)$$

$$\text{iff } S(C \wedge X, e) > S(D \wedge X, e) \quad (71)$$

$$\text{iff } (C \wedge X, e) > (D \wedge X, e), \quad (72)$$

where (70) follows from (69) because  $G$  is strictly increasing in both of its arguments, and (71) follows from (70) by Eq. (43). Therefore (45) follows from (43). But this shows that  $(C, e) > (D, e)$  iff  $(C \wedge Y, e) > (D \wedge Y, e)$  for any other  $Y$ . Hence  $(C \wedge X, e) > (D \wedge X, e)$  iff  $(C \wedge Y, e) > (D \wedge Y, e)$  as required by (47). Therefore (47) follows from (43). The proof that Eq. (43) implies conditions (44) and (46) is analogous and is therefore omitted.

*Proof that probability combination models predict condition (41):*

$$S(A_1 \wedge A_2, e) > S(A_1 \wedge B_1, e) \text{ iff } S(A_2, e) > S(B_1, e) \quad (73)$$

$$\text{iff } S(A_2 \wedge B_2, e) > S(B_1 \wedge B_2, e) \quad (74)$$

where (73) follows by (44), and (74) follows by (45).

*Proof that the Theory of If implies Hypotheses 1 and 2:* The derivation of Hypothesis 1 is straightforward. Let  $\Sigma(p)$  denote the set of propositions that can be introduced as premises in an inference based on the supposition

of  $p$ . According to the Theory of IF, if  $p \mapsto q \wedge r$  is true, then  $q \wedge r$  can be derived from  $\Sigma(p)$  and  $p$ . But then  $q$  can be derived from  $q \wedge r$  and  $r$  can be derived from  $q \wedge r$ . Therefore  $q$  can be derived from  $p$  and  $\Sigma(p)$ , and similarly for  $r$ . Hence,  $p \mapsto q$  and  $p \mapsto r$  are true, so the Theory of IF implies Hypothesis 1. To derive Hypothesis 2, if  $p \mapsto q \wedge r$  is true, then  $q \wedge r$  can be derived from  $p$  and  $\Sigma(p)$ . Therefore  $p$ ,  $\Sigma(p)$ ,  $q$ , and  $r$  must be consistent, for if not, the Constraint on Conditional Proof would be violated. But then  $\Sigma(p)$  and  $p \wedge q$  must be consistent, so  $\Sigma(p) = \Sigma(p \wedge q)$ . Hence,  $q \wedge r$  can be derived from  $p \wedge q$  and  $\Sigma(p \wedge q)$ , so  $r$  can be derived from  $p \wedge q$  and  $\Sigma(p \wedge q)$ ; hence,  $p \wedge q \mapsto r$  is true. This establishes that  $p \mapsto q \wedge r$  implies  $p \wedge q \mapsto r$ . A similar argument shows that  $p \mapsto q \wedge r$  implies  $p \wedge r \mapsto q$ . Thus, the Theory of IF implies Hypothesis 2.

## REFERENCES

- Adams, E. W. (1975). *The logic of conditionals*. Dordrecht, The Netherlands: Reidel.
- Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, 21, 515–550.
- Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology*, 65, 1119–1131.
- Braine, M. D. S. (1978). On the relation between the natural logic of reasoning and standard logic. *Psychological Review*, 85, 1–21.
- Braine, M. D. S. (1990). The "natural logic" approach to reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives* (pp. 133–157). Hillsdale, NJ: Erlbaum.
- Braine, M. D. S., & O'Brien, D. P. (1991). A theory of IF: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203.
- Braine, M. D. S., Reiser, B. J., & Rumin, B. (1984). Some empirical justification for a theory of natural propositional logic. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and thinking* (Vol. 18, pp. 317–371). New York: Academic Press.
- Braine, M. D. S., & Rumin, B. (1983). Logical reasoning. In J. H. Flavell & E. M. Markman (Eds.), *Carmichael's handbook of child psychology: Vol. 3. Cognitive development*. San Diego: Academic Press.
- Carlson, B. W., & Yates, J. F. (1989). Disjunction errors in qualitative likelihood judgment. *Organizational Behavior and Human Decision Processes*, 44, 368–379.
- Carnap, R. (1952). *The continuum of inductive methods*. Chicago: University of Chicago Press.
- Chisholm, R. M. (1946). The contrary-to-fact conditional. *Mind*, 55, 289–307.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Davidson, D. (1967). Truth and meaning. *Synthese*, 17, 304–323.
- Davidson, D. (1970). Semantics for natural languages. *Linguaggi nella Società e nella Tecnica*, 117–188. (Reprinted in D. Davidson & G. Harman (Eds.), (1975b). *The logic of grammar*. Encino, CA: Dickenson.)
- Davidson, D., & Harman, G. (1975a). Introduction. In D. Davidson & G. Harman (Eds.), *The logic of grammar* (pp. 1–14). Encino, CA: Dickenson.
- Davidson, D., & Harman, G. (Eds.). (1975b). *The logic of grammar*. Encino, CA: Dickenson.
- Eiter, T., & Gottlob, G. (1992). On the complexity of propositional knowledge base revision, updates, and counterfactuals. *Artificial Intelligence*, 57, 227–270.
- Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge and Kegan Paul.
- Falmagne, J. C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.
- Fine, T. L. (1973). *Theories of probability: An examination of foundations*. New York: Academic Press.
- Gärdenfors, P. (1988). *Knowledge in flux*. Cambridge, MA: MIT Press.
- Gavanski, L., & Roskos-Ewoldsen, D. R. (1991). Representativeness and conjoint probability. *Journal of Personality and Social Psychology*, 61, 181–194.
- Ginsberg, M. L. (1986). Counterfactuals. *Artificial Intelligence*, 30, 35–79.
- Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy*, 44, 113–128.
- Hampton, J. A. (1987a). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition*, 15, 55–71.
- Hampton, J. A. (1987b). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 12–32.
- Harper, W. L. (1981). A sketch of some recent developments in the theory of conditionals. In W. L. Harper, R. Stalnaker, and G. Pearce (Eds.), *IFS: conditionals, belief, decision, chance, and time*. Dordrecht, The Netherlands: Reidel.
- Harper, W. L., Stalnaker, R., & Pearce, G. (Eds.). (1981). *IFS: conditionals, belief, decision, chance, and time*. Dordrecht, The Netherlands: Reidel.
- Huttenlocher, J., & Hedges, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review*, 101, 157–165.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
- Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P. N. (1986). Conditionals and mental models. In E. C. Traugott, A. ter Meulen, J. S. Reilly, & C. A. Ferguson (Eds.), *On conditionals*. Cambridge: Cambridge University Press.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 3, 430–454.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan.
- Krantz, D. H. (1991). From indices to mappings: The representational approach to measurement. In D. R. Brown & J. E. K. Smith (Eds.), *Frontiers of mathematical psychology: Essays in honor of Clyde Coombs*. New York: Springer-Verlag.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.
- Lewis, D. K. (1972). General semantics. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht, The Netherlands: Reidel.
- Lewis, D. K. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement* (Vol. 3). San Diego: Academic Press.



- Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3). New York: Wiley.
- Marcus, H., & Zajonc, R. (1985). The cognitive perspective in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed.). Reading, MA: Addison-Wesley.
- McCawley, J. D. (1993). *Everything that linguists have always wanted to know about logic but were ashamed to ask* (2nd ed.). Chicago: University of Chicago Press.
- Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, 20, 158–190.
- Miyamoto, J. M., & Dibble, E. (1986). Counterfactual conditionals and the conjunction fallacy. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*.
- Miyamoto, J. M., Lundell, J. W., & Tu, S. (1988). Conditional fallacies in probability judgment. *Bulletin of the Psychosomatic Society*, 26, 516.
- Miyamoto, J. M., Lundell, J. W., & Tu, S. (1989). Anomalous conditional judgments and Ramsey's thought experiment. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 212–220). Hillsdale, NJ: Erlbaum, 1989.
- Montague, R. (1974). *Formal philosophy: Selected papers of Richard Montague*. Edited and with an introduction by R. H. Thomason. New Haven, CT: Yale University Press.
- Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task-specific phenomenon? *Personality and Social Psychology Bulletin*, 10, 243–252.
- Murphy, G. L. (1988). Comprehending complex concepts. *Cognitive Science*, 12, 529–562.
- Osherson, D. N. (1987). New axioms for the contrast model of similarity. *Journal of Mathematical Psychology*, 31, 93–103.
- Osherson, D. N. (1990). Judgment. In D. N. Osherson & E. E. Smith (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3). Cambridge, MA: MIT Press.
- Osherson, D. N., & Smith, E. E. (1982). Gradedness and conceptual combination. *Cognitive Science*, 12, 299–318.
- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 100, 254–278.
- Pennington, N. (1984). *Technical note on conjunctive explanations*. Center for Decision Research, Graduate School of Business, University of Chicago, February, 1984.
- Quine, W. V. (1960). *Word and object*. Cambridge, MA: M.I.T. Press.
- Quine, W. V. (1972). *Methods of logic*. New York: Holt, Rinehart, and Winston.
- Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays*. London: Kegan Paul, Trench, Trubner.
- Rips, L. J., & Marcus, S. L. (1979). Suppositions and the analysis of conditional sentences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 185–220). Hillsdale, NJ: Erlbaum.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
- Shafer, G. (1981). Constructive probability. *Synthese*, 48, 1–60.
- Shafer, G., & Tversky, A. (1985). Languages and designs for probability judgment. *Cognitive Science*, 9, 309–339.
- Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory & Cognition*, 18, 229–239.
- Slovic, P., & Lichtenstein, S. (1983). Preference reversals: A broader perspective. *American Economic Review*, 73, 596–605.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science*, 8, 337–361.
- Smith, E. E., & Osherson, D. N. (1989). Similarity and decision making. In S. Vosniadou & A. Ortony (Eds.), *Similarity, analogy, and thought*. New York: Cambridge University Press.
- Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, 12, 485–527.
- Sosa, E. (Ed.). (1975). *Causation and conditionals*. New York: Oxford University Press.
- Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory*. Oxford: Blackwell. (ApQ Monograph No. 2. Reprinted in E. Sosa, *Causation and conditionals*. Oxford: Oxford University Press, 1975.)
- Stalnaker, R. C. (1984). *Inquiry*. Cambridge, MA: MIT Press.
- Suppes, P. (1966). Probabilistic inference and the concept of total evidence. In J. Hintikka & P. Suppes (Eds.), *Aspects of inductive logic* (pp. 49–65). Amsterdam: North-Holland.
- Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement* (Vol. 2). San Diego: Academic Press.
- Traugott, E. C., ter Meulen, A., Reilly, J. S., & Ferguson, C. A. (Eds.). (1986). *On conditionals*. Cambridge: Cambridge University Press.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 2, 105–110.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Tversky, A., & Kahneman, D. (1982a). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1982b). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- van Fraassen, B. C. (1971). *Formal semantics and logic*. New York: Macmillan.
- Wallsten, T. S. (1977). Measurement and interpretation of beliefs. In H. Jungermann & G. de Zeeuw (Eds.), *Decision making and changes in human affairs* (pp. 369–393).
- Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, 115, 348–365.
- Wells, G. L. (1985). The conjunction error and the representativeness heuristic. *Social Cognition*, 3, 266–279.
- Wyer, R. S. (1976). An investigation of the relations among probability estimates. *Organizational Behavior and Human Performance*, 15, 1–18.
- Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation." *Organizational Behavior and Human Decision Processes*, 37, 230–253.
- Zadeh, L. A. (1982). A note on prototype theory and fuzzy sets. *Cognition*, 12, 291–297.