Ampliative inference: on choosing a probability distribution

Daniel Osherson*·a, Eldar Shafirb, Edward E. Smithc

a IDIAP, CP 609, 1920 Martigny, Valais, Switzerland
b Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544-1010, USA
c Department of Psychology, University of Michigan, 330 Packard Road, Ann Arbor, MI 48104, USA

Abstract

Ampliative inference is the choice of a probability distribution on the basis of incomplete information. We consider some psychological and normative questions that arise about this kind of reasoning. The discussion is largely tutorial although a substantive hypothesis is also advanced.

1. Introduction

Suppose that three regions of unknown size are drawn on a plane surface, as in Fig. 1. A dart will be thrown at random onto the surface, and all you know about the probability of its landing in one or another region is this:

(1)

	Sentence	Symbol	Probability
i.	The dart lands in A or B (or both)	$A \vee B$	.4
ii.	The dart lands in B or C (or both)	$B \vee C$	.5
iii.	The dart lands in A or C (or both)	$A \vee C$	.6

You are invited to purchase a lottery ticket for the outcome A, that is, that the dart lands in region A. To decide whether the ticket price is fair, you need to assign a probability to this statement. What probability do you choose? Recourse
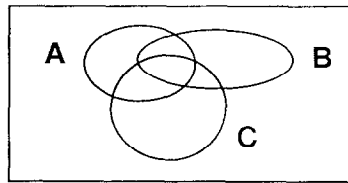
Figure 1.   *Dartboard with three regions A, B, C. The sizes shown are for illustration. The only information available is summarized in (1).*

to the probability calculus is some help, but does not resolve the problem. Probability calculations inform us only that $Pr(A) \in [.1, .4]$. Choice within $[.1, .4]$ is left open.[1] Our example illustrates in bare form the general problem of *ampliative inference*, which may be stated as follows.[2] Suppose you are given information **I** about an unknown probability distribution P. What distribution represents the most rational guess about P? Naturally, you ought to choose a distribution that does not contradict the given information **I** (since otherwise your guess is bound to be wrong). But what should you guess if it turns out that there are a multiplicity of distributions consistent with **I**? In this case the laws of probability do not impose a choice. Rather, some additional principle or leap of faith is needed to supplement **I** to cut the options down to one. Note the double uncertainty in such situations. Not only are we unsure about which events will occur; in addition, we do not know their probabilities of occurrence.

The ampliative inference problem is an idealized version of a dilemma fundamental to thought and decision. Explicitly or implicitly, we evaluate chances all the time, and we rarely have sufficient information to pin down the distribution that governs events in our environment. A descriptive, psychological question is posed thereby, bearing on the mental mechanisms underlying this kind of reasoning. Normative questions arise as well, concerning rational choice in the face of such uncertainty. The present paper attempts to clarify the descriptive issues by contrasting them with a widely accepted normative model. The discussion is largely tutorial, designed to frame the problem and set the stage for subsequent proposals. We also advance a substantive, albeit modest, hypothesis.

As a preliminary clarification, it is helpful to distinguish the topic of the present paper from usual concerns about probability judgment (as represented, for example, in Kahneman, Slovic, & Tversky, 1982). Much of the latter tradition

[1] That $Pr(A) \in [.1, .4]$ may be seen as follows. Were $Pr(A) > .4$ then $Pr(A \vee B) > .4$, contrary to (1)*i*. On the other hand, were $Pr(A) < .1$ then $Pr(C) > .5$ in view of (1)*iii*. But this implies $Pr(B \vee C) > .5$, contradicting (1)*ii*. That (1) places $Pr(A)$ in no narrower interval than $[.1, .4]$ follows from the fact that for both $Pr(A) = .1$ and $Pr(A) = .4$ there are distributions implying (1). Deductions concerning probability intervals can be carried out more efficiently via techniques due to de Finetti (1972), explained in Lad, Dickey, and Rahman (1990).

[2] The "ampliative" terminology has a long history. It appears recently in Klir and Folger (1988).

bears on the *coherence* of probability attributions, on whether such attributions are compatible with any distribution at all. For example, in situations that lead to the "conjunction fallacy" (Tversky & Kahneman, 1983) human judgment is often *in*coherent since no distribution allows the probability of a conjunctive proposition to exceed the probability of a conjunct. In contrast, it is here assumed that the reasoning agent starts with a coherent set of probability attributions that can be extended in more than one way to a complete distribution. Our question is: what kind of extension is typically preferred by humans?

The discussion proceeds as follows. First, we establish some terminology and concepts proper to ampliative inference construed normatively. Then some differences are cited between the normative picture and the mental processes that govern reasoning. Subsequently, we examine an influential proposal about ampliative inference, and try to evaluate it psychologically. Next, our own model is described. Finally, we return to the normative level, and discuss some of the issues that arise in evaluating alternative strategies of ampliative inference.

2. Ampliative inference viewed normatively

We limit attention to situations in which only a finite number of events are distinguished, and we assume that some objective distribution governs their probabilities. This distribution may be qualified as the "target" of the problem, and denoted by $\mathcal{T}$.[3] In the example above, $\mathcal{T}$ gives the probabilities of the dart's landing in the regions of Fig. 1. The smallest regions may be called "atomic" and listed as:

$$A \wedge B \wedge C \quad \neg A \wedge B \wedge C \quad A \wedge \neg B \wedge C \quad \neg A \wedge \neg B \wedge C$$
$$A \wedge B \wedge \neg C \quad \neg A \wedge B \wedge \neg C \quad A \wedge \neg B \wedge \neg C \quad \neg A \wedge \neg B \wedge \neg C$$

The eight atoms are mutually exclusive and exhaustive, and the other regions are composed of different subsets of them. For example, $A \wedge \neg B$ is the union of $A \wedge \neg B \wedge C$ and $A \wedge \neg B \wedge \neg C$, and the chance of the dart's falling into $A \wedge \neg B$ is the sum of the chances associated with the latter two regions. It is thus sufficient to conceive of $\mathcal{T}$ as assigning probabilities to the atomic events of a problem; the probabilities of complex events can be obtained by addition.

We conceive of the guess $\mathcal{G}$ to be made about $\mathcal{T}$ as a distribution over the same atoms. The choice of $\mathcal{G}$ is based on the partial information available about $\mathcal{T}$. This information often takes one of the following forms:

(2)   Three kinds of information often available about a target distribution $\mathcal{T}$:

    a. *Probability of events.* For example, in the dart problem, (1) provides the information that $\mathcal{T}(A \vee B) = .4$, etc.

    b. *Conditional probabilities.* For example, we might have been informed that $\mathcal{T}(B \mid A) = .4$, that is, that the probability is .4 of the dart's landing in B, given that it lands in A.

    c. *Expected values of random variables.* Suppose that each region in Fig. 1 is associated with a dollar payoff: \$1 for $A \wedge B \wedge C$, \$2 for $\neg A \wedge B \wedge C$, etc. Then our information might be that the average payoff for many throws of the dart is a certain sum.

A guess $\mathcal{G}$ about $\mathcal{T}$ is said to be "consistent" with given information **I** only if $\mathcal{G}$ implies **I**. Thus, if $\mathcal{G}(A \vee B) = .5$ then $\mathcal{G}$ is *in*consistent with (1).

The three forms of information mentioned in (2) have a common property. Suppose, for example, that our information **I** is $\mathcal{T}(B \mid A) = .4$, and let $\mathcal{G}_0, \mathcal{G}_1$ be two guesses about $\mathcal{T}$ that are consistent with **I**. Then for any $\lambda \in [1, 0]$, the guess $\lambda \mathcal{G}_0 + (1 - \lambda)\mathcal{G}_1$ is also consistent with **I**. This latter distribution assigns to event X the weighted sum of the probabilities assigned to X by $\mathcal{G}_0$ and $\mathcal{G}_1$. In this sense, the class of distributions consistent with **I** yields a "convex" subset of the space of all distributions over the eight atoms of Fig. 1.[4] The class can also be shown to be "closed" in the sense of containing any distribution that is arbitrarily well approximated by its members. Closed, convex sets of distributions arise for any information of kind (2)a–c, and are central to the normative theory of ampliative inference (see below). We thus give the special name *constraint* to information **I** such that the class of guesses consistent with **I** is non-empty, closed and convex (see Shore, 1986; Shore and Johnson, 1980, for more detailed treatment).

Finally, we define an *ampliative inference method* to be any means of transforming an arbitrary constraint into a distribution consistent with it. Such a method offers a guess about the target distribution in light of available information. For example, an ampliative inference method might convert the constraint recorded in (1) into the guess $\mathcal{G}$ with:

(3) $\mathcal{G}(A \wedge B \wedge C) = .05$ $\mathcal{G}(\neg A \wedge B \wedge C) = .05$

       $\mathcal{G}(A \wedge \neg B \wedge C) = .05$ $\mathcal{G}(\neg A \wedge \neg B \wedge C) = .25$

       $\mathcal{G}(A \wedge B \wedge \neg C) = .05$ $\mathcal{G}(\neg A \wedge B \wedge \neg C) = .05$

       $\mathcal{G}(A \wedge \neg B \wedge \neg C) = .15$ $\mathcal{G}(\neg A \wedge \neg B \wedge \neg C) = .35$

---

[4]In geometric terms, a region of n-dimensional space is convex just in case it contains the line segment connecting any two points in the region (see Luenberger, 1984, Appendix B).

Observe that $\mathcal{G}$ is consistent with (1). For example: $\mathcal{G}(A \vee B) =$

$$\mathcal{G}(A \wedge B \wedge C) \;+\; \mathcal{G}(\neg A \wedge B \wedge C) \;+\; \mathcal{G}(A \wedge \neg B \wedge C) \;+$$
$$\mathcal{G}(A \wedge B \wedge \neg C) + \mathcal{G}(\neg A \wedge B \wedge \neg C) + \mathcal{G}(A \wedge \neg B \wedge \neg C)$$

$= .05 + .05 + .05 + .05 + .05 + .15 = .4$. In the same way, we compute $\mathcal{G}(A) = .3$, which provides an estimate for the chance of A, our original question about the dartboard of Fig. 1.

3. Normative ampliative inference versus human reasoning

So much for the normative picture. The descriptive side of ampliative reasoning deviates from this picture in several respects.[5] For one thing, to get on with life it is not always necessary to compute an exact probability for an uncertain event. To decide to cross the street *now*, for example, it is sufficient that the assessed probability of an unpleasant outcome fall below a certain threshold – and such a judgment need not involve comparison of the latter with any particular number.[6] Still, there are situations in which exact judgments are needed, for example, to determine the lowest price at which to offer insurance or the highest price at which to buy a lottery ticket. So we shall agree to frame the ampliative inference problem in terms of exact probability estimates. How people arrive at such estimates nonetheless diverges from our ideal picture in at least four respects.

First, it seems unlikely that estimates of needed probabilities are calculated from a distribution over atomic events. There are too many such events in situations of even modest complexity. Suppose, for example, that our problem concerns which children in a class of 20 will contract chicken-pox this week. Atomic events look like this:

Jimmy will $\wedge$ Sally won't $\wedge$ . . .

and there are more than a million of them. Instead of creating entire distributions, human ampliative reasoning should rather be seen as giving rise to selected probability judgments that need not bear on atomic events. In our dartboard example, the inference goes from (1) to the probability of non-atomic

---

[5] For a more sanguine view of the relationship between normative and descriptive considerations, see Jaynes (1988).

[6] Similarly, Jeffrey (1992) conceives (idealized) human judgment as represented not by any specific distribution but rather by a class of them, embodying a range of probability attributions. He calls such classes "probasitions".

A, without recourse to the four atoms underlying A. This inference still counts as "ampliative" inasmuch as the probability of A cannot be deduced from (1) via the probability calculus.

Second, ampliative inference methods in the normative picture return distributions that respect the laws of probability whereas the numbers produced by human reasoning are known to violate even the most elementary principles of chance. For example, since people cannot easily compute logical implication in the propositional calculus, they are likely to assign different probabilities to logically equivalent statements. As an illustration, let I and N be as follows:

Sentence	Letter
The icecaps will melt before 2050	I
New York will be submerged before 2050	N

Then $\neg(I \rightarrow N)$ and $I \wedge \neg N$ may be interpreted respectively as:

Sentence	Formula
It's false that if the icecaps melt before 2050 then New York will be submerged before 2050	$\neg(I \rightarrow N)$
The icecaps will melt before the year 2050 but New York will not be submerged before 2050	$I \wedge \neg N$

Due to the intensional character of natural language "if", the latter pair of sentences strike most people as embodying different claims, so a person who starts with the attribution of $Pr(\neg(I \rightarrow N)) = .4$ might assign a different probability to N than someone who starts from $Pr(I \wedge \neg N) = .4$. In fact, the two starting points are equivalent within propositional logic, and impose the same interval of $[.0,.6]$ on $Pr(N)$. To bring our ideal picture more in line with actual reasoning, a psychological model of deductive inference is necessary, and candidates are not lacking (e.g., Clement and Falmagne, 1986; Evans, 1982; Johnson-Laird, 1989; Rips, 1988, 1990). However, even an accurate "psycho-logic" might not suffice to predict ampliative inferences starting from equivalent but distinct statements. Experimental results reported in Shafir and Tversky (1992) and Tversky and Shafir (1992) for example, suggest that different choices sometimes result from no information at all versus "information" of the form $A \vee \neg A$. Yet even psycho-logic would be expected to recognize the vacuity of the latter statement.

Third, the probabilities people attribute to new statements may well be incompatible with those included in the given information. Attributing .4 to $A \wedge B$, for example, is incompatible with $Pr(B) = .3$ since the former probability is bounded above by the latter. For astutely chosen interpretations of A and B, just this kind of incoherency has been experimentally demonstrated, as is well known (see Shafir, Smith, & Osherson, 1990; Tversky & Kahneman, 1983; Wells,

1985). In any case, incompatibility with the starting point of an inference is virtually inevitable given the inability to compute a distribution over atoms and to calculate logical equivalence in the propositional calculus. This is another sense in which human reasoning does not constitute an ampliative inference method as defined in section 2; the conjectures produced by such methods are invariably consistent with the starting information.

Fourth, the information that triggers ampliative reasoning cannot be guaranteed to form a "constraint" in the sense discussed earlier. For example, it might be known that a certain coin has a $9:1$ bias, but unknown whether the bias is for heads or for tails. In this case, the space of distributions compatible with our information is not convex, since averaging a $9:1$ distribution for heads with one for tails leads to a fair coin.[7] Yet a person might be willing to reason ampliatively about such a situation, choosing, for example, the unbiased distribution as a representation of his uncertainty. More fundamentally, the kind of information supporting human reasoning is often neither clearly compatible nor clearly incompatible with a given distribution of probability. For example, there is evidence that probability judgment arises from operations over featural decomposition of objects and properties, as in the representation of *tiger* along the dimensions SIZE, FEROCITY, etc. (see Osherson, Smith, Meyers, Shafir, & Stob, in press; Osherson, Stern, Wilkie, Stob, & Smith, 1991). This kind of featural information does not rule in or rule out any particular distribution.

The psychology of ampliative inference is thus likely to be considerably more complicated than suggested by the abstract picture of section 2. However, inasmuch as current understanding of ampliative reasoning is close to *nil*, it may be useful to examine an influential solution to the idealized problem; perhaps a suitably degraded version of the solution can serve as a rough psychological model. Let us therefore limit attention in what follows to situations like that of Fig. 1. In such cases the given information is simple and represents a constraint in the sense of (2)a. Let us assume as well that the outcome of ampliative inference is a distribution over an appropriate set of atomic events.

4. Ampliative inference via maximum entropy

The most widely used ampliative inference method is known as *maxent*. We first describe the method, and then explain why it is so popular (for more

---

[7]In this example there are two atoms: H (for heads) and T (for not heads, or tails). There are just two distributions P_1, P_2, consistent with the stated information about bias, namely: $\{P_1(H) = .1; P_1(T) = .9\}$ and $\{P_2(H) = .9; P_2(T) = .1\}$. Now consider the distribution Q defined as $\{Q(H) = .5 \times P_1(H) + .5 \times P_2(H); Q(T) = .5 \times P_1(T) + .5 \times P_2(T)\}$. Then, $Q(H) = Q(T) = .5$, so Q is not consistent with the given information about bias. Therefore, this information does not give rise to a convex class of distributions.

complete discussion, see Klir & Folger, 1988; Rosenkrantz, 1977; Williams, 1980).

Each distribution *Pr* over a set $\mathscr{A}$ of atomic events gives rise to a number, called the "entropy" of the distribution, defined as $-\Sigma_{\alpha \in \mathscr{A}} Pr(\alpha) \log_2(Pr(\alpha))$. For example, the entropy of the distribution $\mathscr{G}$ of (3) is the negative of the sum of the following numbers:

$\mathscr{G}(A \wedge B \wedge C) \times \log_2(\mathscr{G}(A \wedge B \wedge C))$ $(=.05 \times \log_2(.05)$ or $-.216)$

$\mathscr{G}(\neg A \wedge B \wedge C) \times \log_2(\mathscr{G}(\neg A \wedge B \wedge C))$ $(=.05 \times \log_2(.05)$ or $-.216)$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$\mathscr{G}(\neg A \wedge \neg B \wedge \neg C) \times \log_2(\mathscr{G}(\neg A \wedge \neg B \wedge \neg C))$ $(=.35 \times \log_2(.35)$ or $-.530)$

yielding an entropy of 2.521 for $\mathscr{G}$. Recall that $\mathscr{G}$ is consistent with the information provided in (1). If we choose a different distribution consistent with (1), we get a different entropy. For example, the following distribution $\mathscr{G}'$ is also consistent with (1) but its entropy is 2.529, slightly higher than for $\mathscr{G}$.

(4) $\mathscr{G}'(A \wedge B \wedge C) = .0568$ $\mathscr{G}'(\neg A \wedge B \wedge C) = .0568$

$\mathscr{G}'(A \wedge \neg B \wedge C) = .0568$ $\mathscr{G}'(\neg A \wedge \neg B \wedge C) = .2364$

$\mathscr{G}'(A \wedge B \wedge \neg C) = .0568$ $\mathscr{G}'(\neg A \wedge B \wedge \neg C) = .0364$

$\mathscr{G}'(A \wedge \neg B \wedge \neg C) = .1364$ $\mathscr{G}'(\neg A \wedge \neg B \wedge \neg C) = .3636$

More generally, suppose that some constraint **c** is given, and consider the class of all distributions over the relevant atomic events that are consistent with **c**. It can be shown that there is a unique member of this class with maximum entropy.[8] It is this distribution that *maxent* chooses for purposes of ampliative inference. For example, if **c** is the constraint shown in (1), then *maxent*(**c**) turns out to be the guess $\mathscr{G}'$ of (4).[9] By adding the first column of (4) it may be seen that *maxent* assigns probability .3068 to event *A*. Such then is the counsel of *maxent* in the face of our original problem. If information about the dartboard is limited to (1), then *maxent* chooses a distribution that judges wagers on *A* to be favorable only if they offer odds better than roughly 7:3.

Let us now address the obvious question: from among the potentially infinite class of distributions that agree with a constraint **c**, why should the distribution of maximum entropy be singled out to govern our bets? In response, we should not expect a proof that *maxent* optimizes the accuracy of our judgment in some objective sense. For, if **c** represents the totality of current information, then *any*

[8]Uniqueness follows from the strict convexity of the entropy function. See Vavasis (1991, Theorem 1.3).

[9]We do not attempt to summarize here the manner in which maximum entropy distributions are calculated; see Cheeseman (1983) and Goldman and Rivest (1988).

distribution consistent with **c** might be in force. Rather, arguments in favor of *maxent* take a subtler form, illustrated by the following (see Ross, 1988, for fuller discussion).

In the face of constraint **c**, a sensible guess should be consistent with **c** but otherwise be as non-committal as possible about the chance of each atomic event. To be maximally non-committal, a distribution should leave us in the greatest doubt about which atom is the true one. The absence of doubt implies the absence of surprise, so we are led to seek a numerical measure of the surprise associated with a given atom. Within information theory, surprise is assumed to depend only on probability; more likely events are mapped to smaller numbers, and certain events are assigned 0. The measure of surprise is also assumed to be continuous and additive in the sense that the surprise associated with probability pq is the sum of the surprises associated with p and q. These assumptions lead to an insightful analysis of coding and information transmission (see, for example, McEliece, 1977). Additionally, they imply the following theorem:

(5)   Let S be any function that satisfies the preceding assumptions about surprise. Then $S(p) = -C \log_2 p$ for an arbitrary positive integer C.

For simplicity, C is set equal to 1 (nothing hinges on this choice, which simply expresses surprise in units of binary digits). Hence, the surprise of an atom α with probability $Pr(\alpha)$ is $-\log_2 Pr(\alpha)$ and the *expected* surprise associated with a distribution over a set $\mathscr{A}$ of atoms is $\Sigma_{\alpha \in \mathscr{A}} - \log_2(Pr(\alpha))Pr(\alpha)$. Expected surprise is a measure of the extent to which a distribution is non-committal about which atom is true. However, $\Sigma_{\alpha \in \mathscr{A}} - \log_2(Pr(\alpha))Pr(\alpha)$ is equivalent to $-\Sigma_{\alpha \in \mathscr{A}} Pr(\alpha) \log_2(Pr(\alpha))$, the entropy of Pr. Hence, the unique distribution that represents the information in **c** but is otherwise maximally non-committal is the one with maximum entropy that satisfies **c**. And this is the distribution proposed by *maxent*.

Other arguments have been offered in support of *maxent* (e.g., Klir & Folger, 1988; Mathai & Rathie, 1975; Rissanen, 1983; Shore & Johnson, 1980), but we agree with Hunter (1986) that the foregoing remains the most persuasive. It may also be pointed out that the entropy coefficient arises in a wide set of laws governing statistical models, information transmission, and computational complexity (see Cover & Thomas, 1991). That it should also govern ampliative reasoning is rendered more plausible thereby.

Finally, we note that entropy calculations play a central role in several expert systems for reasoning under uncertainty (e.g., Herskovits & Cooper, 1991; Wen, 1988). Maximum entropy distributions have also been claimed to fill an essential gap in Bayesian statistics by providing motivated prior distributions from which probability updating can begin (for discussion, see Earman, 1992; Jaynes, 1968; Rissanen, 1983).

5. *Maxent* as a descriptive model

On the descriptive level, *maxent* can be evaluated from both a narrow and wide point of view. Narrowly, it may be asked whether the numbers returned by *maxent* resemble those arising in spontaneous reasoning. Widely, we may ask whether the justifications offered for *maxent* provide clues to the psychology of ampliative inference.

5.1. Narrow evaluation

In view of the remarks in section 3 above, *maxent* can be evaluated narrowly only by idealizing human reasoning and considering simple cases like (1). In the case of (1) we saw that *maxent* proposes the distribution given in (4). In contrast, an informal survey of opinion suggests that (3) in section 2 is a more appealing conjecture. For one thing, the numbers in (3) are simpler, and appear less arbitrary than the highly precise probabilities in (4). For another, the probabilities in (3) are "flatter" than those in (4), inasmuch as the extremes of the former (namely, .05,.35) fit inside the extremes of the latter (.0364,.3636). Other things equal, less extreme probabilities seem preferable to more extreme ones as a representation of our uncertainty about the distribution in force.

The distribution (3) arises from an alternative to *maxent* that may be called *mindev* (for "minimal deviation"). Ampliative reasoning according to *mindev* proceeds by "anchoring and adjustment", a strategy that characterizes human reasoning in several domains (see Tversky & Kahneman, 1974; Tversky, Sattath, & Slovic, 1988). A plausible anchor in the present context is the uniform distribution, assigning the same probability to each atom, and the adjustment sought stays as close as possible to the anchor while retaining consistency with the given information. To make this idea precise, a measure of proximity to the uniform distribution is needed. One possibility is to calculate proximity using "relative entropy" (Kullback, 1968), but this leads back to *maxent*. In contrast, *mindev* relies on sums of squares: given a constraint $\mathbf{c}$ in a situation with n atoms α, *mindev* seeks the distribution Pr consistent with $\mathbf{c}$ that minimizes $\Sigma_\alpha\,[(1/n) - Pr(\alpha)]^2$. Just as with *maxent*, it can be shown that this latter distribution exists and is unique.

For simple ampliative inference problems involving random variables – as in (2)c – it has been shown that human judgment often favors the counsel of *mindev* over that of *maxent* (see Myers & Osherson, 1992). Although this result coincides with the putative preference for (3) over (4), it would be hasty to conclude that *mindev* is also descriptively superior to *maxent* for ampliative inference involving probability constraints of form (2)a. To see this, suppose that constraint (1) is altered slightly to:

(6)   *Sentence*   *Symbol*   *Probability*
         The dart lands in A or B (or both) $A \vee B$ .35
         The dart lands in B or C (or both) $B \vee C$ .5
         The dart lands in A or C (or both) $A \vee C$ .65

Applied to (6), *mindev* yields a distribution in which the atomic event $\neg A \wedge B \wedge \neg C$ has zero chance. In contrast, *maxent* assigns this atom positive probability, as it tends to do whenever permitted by the given information. In this respect, *maxent* has greater psychological fidelity than *mindev* since the information in (6) is unlikely to lead to the conclusion that some region of Fig. 1 has null area.

Another advantage for *maxent* is revealed by the following alternative to constraint (1):

(7)   *Sentence*   *Symbol*   *Probability*
         The dart lands in A A .3
         The dart lands in B B .4
         The dart lands in C C .5

Faced with this information, it is tempting to assume the independence of A, B, and C, and calculate the probability of $A \wedge B \wedge C$ to be $Pr(A)Pr(B)Pr(C)$, the probability of $\neg A \wedge B \wedge C$ to be $(1 - Pr(A))Pr(B)Pr(C)$, etc. In fact, *maxent* yields exactly this distribution, whereas *mindev* assigns, for example, .05 to $A \wedge B \wedge C$ rather than the .06 derived from independence. In this respect, *maxent*'s advice is likely to seem more natural than that of *mindev* for many people.

Is there a method descriptively superior to both *maxent* and *mindev*? Is anchoring and adjustment the key to ampliative reasoning? If so, is the uniform distribution the relevant anchor? Considerable research will no doubt be required to see more clearly into these matters.

5.2. Wide evaluation

A plausible hypothesis about ampliative reasoning is that it proceeds via the construction of "mental models" of the environment, in roughly the sense of Johnson-Laird (1983, 1989). According to this idea the reasoner enriches the given information about probabilities into a more complete account of the unknown situation, that is, into a mental model. He or she then reads off new probabilities by contemplating the model's emergent properties. In the dart example, the first step would consist in arranging the three areas A, B, C to satisfy (1). The probabilities of remaining statements could then be estimated

from the definite figure in mind. This example suggests the use of visual imagery in constructing and contemplating a mental dartboard. In general, however, the mental model strategy need have no visual component, just as probabilistic "scenarios" (Kahneman & Tversky, 1982a) need not.

For such an account of ampliative reasoning to be predictive, more needs to be said about which model people tend to choose as a representation of initial information; there are, after all, many ways to draw the eight regions of Fig. 1 consistently with (1), and they lead to divergent inferences. The justification for *maxent* discussed in section 4 is suggestive in this connection. Just as the maximum entropy distribution is as non-committal as possible while agreeing with the given constraint, so the chosen mental model would be expected to incorporate the weakest possible assumptions necessary to represent the information available at the outset.[10] Within this perspective, the problem is how to measure the psychological strength or weakness of assumptions. It is tempting to follow information theory by defining an assumption to be strong if the person in question finds it to be *a priori* unlikely. Such a definition, however, threatens to bring us full circle, since the estimation of such likeliness can be counted on to involve some form of ampliative reasoning.

Whatever progress is made on which model people choose to represent available information, there remains the question of how new probabilities are extracted from it. Again, *maxent* might provide an initial clue to this process. Proponents of *maxent* conceive it as generalizing the "principle of insufficient reason" usually credited to Laplace. The principle enjoins us to attribute equal probabilities to alternative states of nature that cannot be distinguished in terms of their chance of occurring (see Howson & Urbach, 1989, Ch. 3). Jaynes (1968) presents an argument according to which the *maxent* distribution can be mathematically realized in more ways than any other distribution satisfying a given constraint. Assuming all these distributions to be of equal likelihood, the *maxent* one is favored by insufficient reason (see Brillouin, 1962; Jaynes, 1979; Rosenkrantz, 1977, for fuller treatment of this argument; see Dias & Shimony, 1981; Neapolitan, 1990, Section 2.6, for a critique).

People might well reason in this way when extracting probabilities from a mental model. In the case of Fig. 1, for example, it might be assumed that each point has the same chance of receiving the dart. Probabilities can then be computed from areas. In the general case, application of insufficient reason requires that the model be analyzed as a set of mutually exclusive and exhaustive possibilities among which the person can find no grounds for favoring one over another. The difficulty thus arises that there are competing analyses possible for any given model, as illustrated by the numerous examples to which Laplace's principle gave rise.[11] A simple example is as follows (taken from Howson &

---

[10] A similar idea underlies many artificial systems of reasoning (e.g., Lifschitz, 1985; McCarthy, 1986; Reiter, 1978).

[11] A famous case is "Bertrand's paradox", described in Ross (1988, pp. 161–162).

Urbach, 1989, Section 3.b.3). Suppose you are drawing balls from an urn that contains white, red, and blue balls in unknown proportions. What is the probability of drawing white (W)? If we distinguish only the cases of colored versus uncolored balls, then insufficient reason yields $Pr(W) = 1/2$, since nothing in the problem gives us reason to attribute different probabilities to the two cases. However, if we distinguish among red, blue, and uncolored balls, then insufficient reason yields $Pr(W) = 1/3$. Of course, it is non-trivial to formulate principles that predict which analysis a person will choose from among a class of competing possibilities.

6. A descriptive hypothesis

The present section offers a hypothesis about the psychology of ampliative inference. Attention is limited to situations in which there is no ambiguity about the set of atoms likely to be recognized by the subject (as in the dartboard of Fig. 1). These atoms will be denoted by $\mathscr{A} = \{\alpha_1 \ldots \alpha_n\}$. Let $\mathscr{T}$ be the distribution over $\mathscr{A}$ that is the target of ampliative inference. Let $\mathbf{D}$ be the class of all distributions over these atoms, and let U be the uniform distribution assigning $1/n$ to each. We suppose that subjects reason with the help of an implicit measure of the separation between two distributions. This measure is embodied by a function that maps pairs of distributions drawn from $\mathbf{D}$ into the set $\mathfrak{R}$ of real numbers, where greater numbers represent greater disparity. We further assume that faced with given probabilistic information, subjects employ an "anchor and adjustment" heuristic, attempting to find the distribution that respects the constraint but is otherwise as close as possible to U. The form of our model is thus as follows:

(8)   *Model form*: For every person S there is a separation measure $m: \mathbf{D} \times \mathbf{D} \to \mathfrak{R}$ such that given starting information $\mathbf{I}$ and distributions $P,Q \in \mathbf{D}$ both of which are consistent with $\mathbf{I}$, S prefers P to Q as a guess about $\mathscr{T}$ if and only if $m(P, U) < m(Q, U)$. Moreover, m has the special properties . . .

Now it is clear that *any* method of ampliative inference can be attributed to subjects via a hypothesis of form (8) unless its ellipsis can be filled with non-trivial properties of m. So we focus on these properties by stating a genuine model of ampliative reasoning:

(9)   *Model*: For every person S, the separation measure m that satisfies (8) has the special properties:

a. For all $P,Q \in \mathbf{D}$, $m(P,Q) \geq 0$. Moreover, $m(P,Q) = 0$ if and only if $P = Q$ (i.e., if and only if P and Q are the same distribution).

    b. Suppose that $P' \in \mathbf{D}$ results from $P \in \mathbf{D}$ by permuting the probabilities that P assigns to two atoms, and suppose that $Q' \in \mathbf{D}$ results from $Q \in \mathbf{D}$ by permuting the probabilities that Q assigns to the same two atoms.[12] Then $m(P',Q') = m(P,Q)$.

    c. Let $P,Q,P',Q' \in \mathbf{D}$ be such that for all atoms α_i, $|P(\alpha_i) - Q(\alpha_i)| \leq |P'(\alpha_i) - Q'(\alpha_i)|$. Then $m(P,Q) \leq m(P',Q')$. If, in addition, $|P(\alpha_j) - Q(\alpha_j)| < |P'(\alpha_j) - Q'(\alpha_j)|$ for some atom α_j, then $m(P,Q) < m(P',Q')$.

Condition (b) claims that m is insensitive to which atoms get which probabilities; all that matters are the numbers assigned. Condition (c) may be paraphrased as follows. Suppose that for every atom, P and Q are as close to each other as are P' and Q'. Then m assigns no greater separation to the pair P,Q than to the pair P',Q'. Such would seem to be true of people's intuitive judgment about the difference between two distributions.

A variety of other conditions may be added to the list, for example, guaranteeing that for every constraint $\mathbf{c}$ there is a unique distribution consistent with $\mathbf{c}$ that minimizes $m(P,U)$. Each condition yields predictions about subjects' intuitions of distribution-separation, and thus indirectly about ampliative reasoning. The list provided by (9) is minimal, but nonetheless open to challenge. We note that it is consistent with both *maxent* and *mindev* as descriptive models. This is easy to see in the case of *mindev*. Suppose that $m(P,Q)$ is defined as $\Sigma_{i \in \{1 \ldots n\}} (P(\alpha_i) - Q(\alpha_i))^2$. Then according to (8), if P,Q are both consistent with the given information $\mathbf{I}$, P is preferred to Q just in case P is closer than Q to U in the squared-deviation sense.

To illustrate a prediction of the model, suppose that our dartboard is limited to two regions, A,B. If the starting information is $Pr(B) = .6$ then (8) and (9)c imply a preference for the distribution

$$\mathcal{G}(A \wedge B) = .3 \quad \mathcal{G}(\neg A \wedge B) = .3 \quad \mathcal{G}(A \wedge \neg B) = .2 \quad \mathcal{G}(\neg A \wedge \neg B) = .2$$

compared to

$$\mathcal{G}(A \wedge B) = .1 \quad \mathcal{G}(\neg A \wedge B) = .5 \quad \mathcal{G}(A \wedge \neg B) = .2 \quad \mathcal{G}(\neg A \wedge \neg B) = .2$$

We suspect that this prediction is correct for almost everyone.[13]

---

[12]For example, in the case of four atoms, P might assign .2, .3, .25, .25 to α_1, α_2, α_3, α_4, whereas P' assigns them .3, .2, .25, .25.

[13]It has been suggested to us that the prediction might fail if the propositions A and B are highly meaningful for the subject. In this case, the uniform distribution may not turn out to be a natural anchor and property (9)b might also prove too strong. To stimulate discussion, we stick with our prediction.

7. Some normative questions about ampliative inference

The preceding discussion has attempted to clarify the descriptive problem of ampliative inference by comparing it to a widely held normative theory based on maximizing entropy. We must not leave the impression, however, that the normative theory of ampliative inference is free of controversy and paradox. In this section we describe two difficulties that arise at the normative level. The first is relevant to *maxent* and other methods of a similar cast. The second is more general in nature. Both are inspired by discussion in Seidenfeld (1986).

7.1. The equivocal character of maxent (and of similar methods, like mindev)

Not all the arguments advanced in favor of *maxent* are convincing (as observed by Hunter, 1986; Skyrms, 1985), and the fundamental idea of being maximally non-committal raises unanswered questions. Recall from section 4 that this latter idea was linked to the surprise connected with an event, and that surprise was measured by $-\log_2 p$, where p is the probability of the event in question. To justify this measure it was assumed, among other things, that the surprise associated with probability pq is the sum of the surprises associated with p and q. The additivity of surprise, however, is not an evident truth (like the postulates of geometry), and to our knowledge no serious argument has been offered in its favor.[14]

In addition to doubt about the arguments supporting *maxent*, questions have arisen about the inferences it endorses (see Dias & Shimony, 1981; Hunter, 1989; Seidenfeld, 1986, for extensive discussion). One troublesome feature of *maxent* is illustrated by the dartboard of Fig. 2.[15] Suppose that the information in hand is just:

(10) | Sentence | Symbol | Probability |
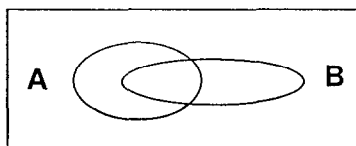|---|---|---|
| The dart lands in A | A | .3 |
| The dart lands in B | B | .8 |

Figure 2.   *Two regions A and B drawn on a plane surface. The regions intersect in $A \wedge B$.*

[14]On the descriptive, psychological level, one may even question whether surprise is a function of probability; see Kahneman and Tversky (1982b).

[15]The present example is a simpler version of a case discussed by Seidenfeld (1986).

Then, *maxent* applied to (10) yields the following guess $\mathscr{G}$ about the target distribution over the four atoms:

(11) $\mathscr{G}(A \wedge B) = .24$ $\mathscr{G}(\neg A \wedge B) = .56$

$\mathscr{G}(A \wedge \neg B) = .06$ $\mathscr{G}(\neg A \wedge \neg B) = .14$

Now suppose that the very same dartboard is conceived slightly differently. In addition to the regions A and B, we imagine a new region R within their intersection. A may thus be represented as the union of R and $A - R$, and B as the union of R and $B - R$. Labeling $A - R$ and $B - R$ by P and Q, respectively, we arrive at the dartboard of Fig. 3. Since A is just $P \cup R$ and B is just $Q \cup R$, the information in (10) is equivalent to:

(12) | *Sentence* | *Symbol* | *Probability* |
|---|---|---|
| The dart lands in P or R | $P \vee R$ | .3 |
| The dart lands in Q or R | $Q \vee R$ | .8 |

It is clear that (10) and (12) express identical information about the board. Indeed, the construal of A and B as unions of other events is an arbitrary decision, and R might well have no area (which would equate A with P and B with Q). Consequently, a rational estimate of the probability of the dart landing in $A \wedge B$ should match the estimate for $(P \vee R) \wedge (Q \vee R)$. Otherwise, our ampliative inference method is sensitive to inessential features of the given information (namely, to the manner of its expression), and is ambiguous about which probabilities it proposes. In fact, *maxent* applied to (12) yields the following guess $\mathscr{G}'$ about the target distribution over the eight atoms generated by P,Q,R.

(13) $\mathscr{G}'(P \wedge Q \wedge R) = .0564$ $\mathscr{G}'(\neg P \wedge Q \wedge R) = .0564$

$\mathscr{G}'(P \wedge \neg Q \wedge R) = .0564$ $\mathscr{G}'(\neg P \wedge \neg Q \wedge R) = .0564$

$\mathscr{G}'(P \wedge Q \wedge \neg R) = .0564$ $\mathscr{G}'(\neg P \wedge Q \wedge \neg R) = .5190$

$\mathscr{G}'(P \wedge \neg Q \wedge \neg R) = .0190$ $\mathscr{G}'(\neg P \wedge \neg Q \wedge \neg R) = .1800$

Adding the probabilities of the five relevant atoms shows that $\mathscr{G}'((P \vee R) \wedge (Q \vee R)) = .282$ which is, unfortunately, different from $\mathscr{G}(A \wedge B)$. Exactly the
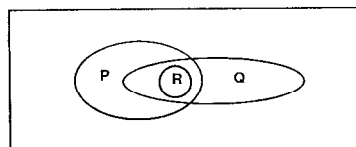

Figure 3.   *The same two regions as in Fig. 2, but with A conceived as $P \vee R$ and B as $Q \vee R$.*

same kind of difficulty arises for *mindev* and for many alternative doctrines of ampliative inference.[16]

The *maxent* doctrine can be revised to avoid undue sensitivity to the manner in which information is expressed. The trick is to effect a prior partition of atoms on the basis of their compatibility with different subsets of the given information. For details of the construction we refer the reader to Nilsson (1986, Section 4), which discusses essentially this new version of *maxent*. The new version yields the same distribution as before when applied to (10). When applied to (12), however, it assigns .24 to $(P \lor R) \land (Q \lor R)$, corresponding to the value given $A \land B$.

7.2. Ampliative inference versus conditionalization

We turn now to an argument against the acceptibility of *any* form of ampliative reasoning. Consider again the dartboard of Fig. 2 and the initial information in (10). We apply our favorite method M of ampliative inference to (10) and arrive at some particular distribution $\mathcal{G}$. To be a sensible guess, $\mathcal{G}$ must be consistent with the information in (10), so we have the following requirement on M:

(14)   *First condition of adequacy on M:* The result of applying M to (10) must be a distribution $\mathcal{G}$ such that $\mathcal{G}(A) = .3$ and $\mathcal{G}(B) = .8$.

Now suppose that we receive new information, compatible with (10) but not implied by it, namely:

(15)   *Sentence*   *Symbol*   *Probability*
          The dart lands in A or in B $A \lor B$ 1.0

How do we revise our initial guess $\mathcal{G}$ in light of (15)? There are two possibilities:

- We start over again, and apply M to the total information now available, namely, to (15) conjoined with (10). The result will be some new distribution $\mathcal{G}_1$.
- We *conditionalize* $\mathcal{G}$ on the information in (15). The result will be the new distribution $\mathcal{G}_{A \lor B}$ defined by the stipulation that for all events, $X, \mathcal{G}_{A \lor B}(X) = \mathcal{G}(X \mid A \lor B)$.

The second strategy is justified by theorems that underline the risks of updating a distribution by any other means than conditionalization, when new information

---

[16]The root of the problem already appears in the attempt to represent total ignorance via the uniform distribution. Equivocation over the atoms in play leads to difficulties similar to those outlined above. See Shafer (1976, Ch. 1) for discussion.

like (15) becomes available (for discussion, see Jeffrey, 1992, Ch. 5; Skyrms, 1987). If we are to have confidence in M, we should therefore expect that $\mathcal{G}_1 = \mathcal{G}_{A \vee B}$. Otherwise, M is irrational either in selecting $\mathcal{G}$ in the face of (10), or in selecting $\mathcal{G}_1$ in the face of (10) and (15); for, if both guesses were rational but $\mathcal{G}_1 \neq \mathcal{G}_{A \vee B}$, then it would be rational to update $\mathcal{G}$ by some other means than conditionalization. Of course, it is also expected that $\mathcal{G}_1$ remain consistent with (10), since (10) is still part of the available information. We thus have the following requirement on M.

(16)   *Second condition of adequacy on M:* The result of applying M to the conjunction of (10) and (15) must be a distribution $\mathcal{G}_1$ with the following properties:

   a. $\mathcal{G}_1$ is the conditionalization of $\mathcal{G}$ with respect to $A \vee B$, where $\mathcal{G}$ is the result of applying M to (10).

   b. $\mathcal{G}_1(A) = .3$ and $\mathcal{G}_1(B) = .8$.

Now suppose that instead of receiving (15) as a supplement to (10), we received:

(17)   *Sentence*   *Symbol*   *Probability*
       The dart lands outside A or inside B $\neg A \vee B$ 1.0

In such a case we must also revise our initial guess $\mathcal{G}$, this time to take account of (17). The same alternatives arise as before, namely, applying M to (17) plus (10) or conditionalizing $\mathcal{G}$ on $\neg A \vee B$. And, as before, the correctness of our method requires that the two strategies yield the same answer. We thus have the following additional requirement on M:

(18)   *Third condition of adequacy on M:* The result of applying M to the conjunction of (10) and (17) must be a distribution $\mathcal{G}_2$ with the following properties:

   a. $\mathcal{G}_2$ is the conditionalization of $\mathcal{G}$ with respect to $\neg A \vee B$, where $\mathcal{G}$ is the result of applying M to (10).

   b. $\mathcal{G}_2(A) = .3$ and $\mathcal{G}_2(B) = .8$.

The difficulty is that *no* method whatsoever of ampliative inference respects all three conditions (14), (16), and (18). The Appendix provides a simple proof of this fact. It seems therefore to follow that ampliative reasoning must either contradict probabilistic information already in hand or else respond to new

information by offering successive distributions that are not related by conditioning; and both alternatives are unappealing.[17]

8. Concluding remarks

The goal of the present paper has been to clarify some of the issues surrounding ampliative reasoning. We have seen that complex questions arise on both the descriptive and normative levels. As in other areas of cognitive science involving reasoning, judgment, and choice, we suspect that joint investigation from both perspectives has the greatest chance of yielding insights of value to psychologists and epistemologists alike.

References

Brillouin, L. (1962). *Science and information theory*. New York: Academic Press.

Cheeseman, P. (1983). A method of computing generalized bayesian probability values for expert systems. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe. San Mateo, CA: William Kauffman.

Clement, C., & Falmagne, R. (1986). Logical reasoning, world knowledge, and mental imagery: interconnections in cognitive processes. *Memory & Cognition, 4*, 299–307.

Cover, T.M., & Thomas, J.A. (1991). *Elements of information theory*. New York: Wiley–Interscience.

de Finetti, B. (1972). *Probability, induction and statistics*. New York: Wiley.

Dias, P., & Shimony, A. (1981). A critique of Jaynes' maximum entropy principle. *Advances in Applied Mathematics, 2*, 172–211.

Earman, J. (1992). *Bayes or bust?* Cambridge MA: MIT Press.

Evans, J.St.B.T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.

Goldman, S., & Rivest, R. (1988). A non-iterative maximum entropy algorithm. In Lemmer, J.F. & Kanal, L.N. (Eds.), *Uncertainty in artificial intelligence* (pp. 133–148). Amsterdam: Elsevier.

Herskovits, E., & Cooper, G. (1991). Kutató: an entropy-driven system for construction of probabilistic expert systems from databases. In P.P. Bonissone, M. Henrion, L.N. Kanal & J.F. Lemmer (Eds.), *Proceedings of the sixth workshop on uncertainty in artificial intelligence* (pp. 117–125). Amsterdam: Elsevier.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: the Bayesian approach*. La Salle, IL: Open Court.

Hunter, D. (1986). Uncertain reasoning using maximum entropy inference. In L.N. Kanal & J.F. Lemmer (Eds.), *Uncertainty in artificial intelligence*. Amsterdam: Elsevier.

Hunter, D. (1989). Causality and maximum entropy updating. *Journal of Approximate Reasoning, 3*, 87–114.

Jaynes, E. (1979) Where do we stand on maximum entropy? In R. Levine & M. Tribus (Eds.), *The maximum entropy formalism* (pp. 15–117). Cambridge, MA: MIT Press.

---

[17]Seidenfeld (1986) develops a related argument for the specific method of *maxent*. The present objection might be put this way: faced with any constraint c, an ampliative inference method should produce a probability distribution all of whose conditionalizations (including itself) satisfy c. This is impossible except in degenerate cases. (The present formulation is due to an anonymous referee.)

Jaynes, E. (1988). How does the brain do plausible reasoning? In G. Erickson & C. Smith (Eds.), *Maximum entropy and Bayesian methods in science and engineering* (Vol. 1, pp. 1–24). Dordrecht: Kluwer.

Jaynes, E.T. (1968) Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics, SSC-4* 227–241.

Jeffrey, R. (1983). *The logic of decision*, 2nd edn. Chicago: University of Chicago Press.

Jeffrey, R. (1992). Probability and the art of judgment. In *Probability and the art of judgment* (pp. 44–76). Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. (1983). *Mental models: towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P. (1989) Mental models. In M. Posner (Ed.), *Foundations of cognitive science*, pp. 469–499. Cambridge, MA: MIT Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: heuristics and biases.* Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982a) The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases.* Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982b) Variants of uncertainty. *Cognition, 11*, 143–158.

Klir, G. (1988). Methodological principles of uncertainty in inductive modeling: a new perspective. In G. Erickson & C. Smith (Eds.), *Maximum-entropy and Bayesian methods in science and engineering* (Vol. 1, pp. 295–304). Dordrecht: Kluwer.

Klir, G., & Folger, T. (1988). *Fuzzy sets, uncertainty, and information.* Englewood Cliffs, NJ: Prentice-Hall.

Kullback, S. (1968). *Information theory and statistics.* New York: Dover.

Lad, F., Dickey, J.M., & Rahman, M.A. (1990). The fundamental theorem of prevision. *Statistica, 50*, 19–38.

Lifschitz, V. (1985). Closed-world databases and circumscription. *Artificial Intelligence, 27,* 229–235.

Luenberger, D. (1984). *Linear and nonlinear programming.* Reading, MA: Addison-Wesley.

Mathai, A., & Rathie, P. (1975). *Basic concepts of information theory and statistics.* New York: Wiley.

McCarthy, J. (1986). Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence, 28,* 89–116.

McEliece, R. (1977). *Theory of information and coding.* Reading, MA: Addison-Wesley.

Myers, T., & Osherson, D. (1992). On the psychology of ampliative inference. *Psychological Science, 3,* 131–135.

Neapolitan, R. (1990) *Probabilistic reasoning in expert systems.* New York: Wiley.

Nilsson, N.J. (1986). Probabilistic logic. *Artificial Intelligence, 28,* 71–87.

Osherson, D., Smith, E.E., Meyers, T.S., Shafir, E., & Stob, M. (in press). Extrapolating human probability judgment. *Theory and Decision.*

Osherson, D., Stern, J., Wilkie, O., Stob, M., & Smith, E. (1991). Default probability. *Cognitive Science, 15,* 251–70.

Reiter, R. (1978). On closed world data bases. In H. Gallaire & J. Minker (Eds.), *Logic and data bases* (pp. 55–76). New York: Plenum Press.

Rips, L. (1988). Deduction. In R. Sternberg & E.E. Smith (Eds.), *The psychology of human thought.* Cambridge, UK: Cambridge University Press.

Rips, L. (1990). Reasoning. *Annual Review of Psychology, 41,* 321–353.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Annals of Statistics, 17,* 416–431.

Rosenkrantz, R. (1977). *Inference, method, and decision.* Boston, MA: Reidel.

Ross, S. (1988). *A first course in probability.* New York: Macmillan.

Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science, 53,* 467–491.

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

Shafir, E., Smith, E., & Osherson, D. (1990). Typicality and reasoning fallacies. *Memory & Cognition, 18,* 229–239.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: nonconsequential reasoning and choice. *Cognitive Psychology, 24,* 449–474.

Shore, J., (1986). Relative entropy, probabilistic inference, and ai. In L.N. Kanal & J.F. Lemmer (Eds.), *Uncertainty in artificial intelligence* (pp. 211–215). Amsterdam: Elsevier.

Shore, J., & Johnson, R. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions of Information Theory, IT-C 26,* 26–37.

Skyrms, B. (1985). Maximum entropy inference as a special case of conditionalization. *Synthese, 63,* 55–74.

Skyrms, B. (1987). Dynamic coherence and probability kinematics. *Philosophy of Science, 54,* 1–20.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 5,* 305–309.

Vavasis, S. (1991). *Nonlinear optimization.* New York: Oxford University Press.

Wells, G. (1985). The conjunction error and the representativeness heuristic. *Social Cognition, 3,* 266–279.

Wen, W.X. (1988). Minimum cross entropy reasoning in recursive causal networks. In R.D. Schachter, T.S. Levitt, L.N. Kanal & J.F. Lemmer (Eds.), *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence,* pp. 360–367. Amsterdam: North-Holland.

Williams, P. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science, 31,* 131–144.

Appendix

Let M be any ampliative inference method, and suppose that M satisfies (14) and (16). We will show that M does not satisfy (18). Faced with the information in (10), let x_{AB}, $x_{\bar{A}B}$, $x_{A\bar{B}}$, $x_{\bar{A}\bar{B}}$ be the probabilities that M assigns to $A \wedge B$, $\neg A \wedge B$, $A \wedge \neg B$ and $\neg A \wedge \neg B$, respectively. By (14), we have:

(19)   a. $x_{AB} + x_{A\bar{B}} = .3$

b. $x_{AB} + x_{\bar{A}B} = .8$

By (16)a, M applied to (10) and (15) is the conditionalization of the previous distribution on the fact that $x_{\bar{A}\bar{B}} = 0.0$. By (16)b, it thus follows that:

(20)   a. $\dfrac{x_{AB} + x_{A\bar{B}}}{x_{AB} + x_{\bar{A}B} + x_{A\bar{B}}} = .3$

b. $\dfrac{x_{AB} + x_{\bar{A}B}}{x_{AB} + x_{\bar{A}B} + x_{A\bar{B}}} = .8$

It is easy to verify that (19) and (20) imply:

(21)   a. $x_{AB} = .1$

b. $x_{\bar{A}B} = .7$

c. $x_{A\bar{B}} = .2$

d. $x_{\bar{A}\bar{B}} = .0$

We now produce a contradiction from the assumption that M also satisfies (18). By (18)a, M applied to (10) and (17) is the conditionalization of the initial distribution on the fact that $x_{A\bar{B}} = 0.0$. By (18)b, it thus follows that:

(22)   a. $\dfrac{x_{AB} + x_{A\bar{B}}}{x_{AB} + x_{\bar{A}B} + x_{\bar{A}\bar{B}}} = .3$

    b. $\dfrac{x_{AB} + x_{\bar{A}B}}{x_{AB} + x_{\bar{A}B} + x_{\bar{A}\bar{B}}} = .8$

However, both equalities in (22) are contradicted by (21).