The Reference Class

Author(s): Henry E. Kyburg, Jr.

Source: *Philosophy of Science*, Sep., 1983, Vol. 50, No. 3 (Sep., 1983), pp. 374-397

Published by: The University of Chicago Press on behalf of the Philosophy of Science Association

Stable URL: https://www.jstor.org/stable/187855

THE REFERENCE CLASS*

HENRY E. KYBURG, JR.

Department of Philosophy
The University of Rochester

The system presented by the author in *The Logical Foundations of Statistical Inference* (Kyburg 1974) suffered from certain technical difficulties, and from a major practical difficulty; it was hard to be sure, in discussing examples and applications, when you had got hold of the right reference class. The present paper, concerned mainly with the characterization of randomness, resolves the technical difficulties and provides a well structured framework for the choice of a reference class. The definition of randomness that leads to this framework is simplified and clarified in a number of respects. It resolves certain puzzles raised by S. Spielman and W. Harper in their contributions to *Profiles: Henry E. Kyburg, Jr. and Isaac Levi* (R. Bogdan (ed.) 1982).

1. Introduction. In Levi (1981), Isaac Levi expressed the belief that although there are a number of things wrong with the system of Kyburg (1974)—henceforth LFSI—as it stands, repairs can be effected that will overcome its technical difficulties. Levi then goes on to argue that the system is deficient, not on technical grounds, but on general philosophical grounds. My purpose here is not to argue the philosophical issues, but to provide a resolution of the technical difficulties, and to make a considerable simplification in the structure of the system. My hope is not only to demonstrate that these difficulties can be resolved, but also to make it easier to see what the philosophical issues are. There is a longer range goal as well. As both Seidenfeld (1978) and Spielman (1980) have pointed out, it is not always easy, in constructing examples and counterexamples, to know when one has got hold of the right reference class. One problem, of course, is that the examples are always very abstract, and it is more difficult than it is in real life to know what can be taken as known and what cannot. But there is also the problem that there is no systematic procedure for determining the reference class for a given statement. What one would *like* would be an algorithm that would lead from a sentence and a set of sentences representing a body of knowledge, to a term denoting an appropriate reference class for that sentence. I still have no such algorithm; but the present form of the system makes it seem more likely that such an algorithm could be developed.

The general idea of the new approach is this: given a sentence and a

*Received August 1982; revised February 1983.

Philosophy of Science, 50 (1983) pp. 374–397

body of knowledge, we define the class of *inference structures*[1] for that sentence. An inference structure consists of a quintuple which *could* serve to yield a probability. When two inference structures would yield disagreeing probabilities, at least one will be eliminated. At the end of this process, we are left with a set of inference structures that do not disagree, and that can be (partially) ordered by strength or precision. Any inference structure than which no other embodies greater statistical precision will serve to provide the probability of the sentence in question.

2. Definitions. Given a language L and a set of statements w of that language representing a body of knowledge, we begin by defining the set of *inference structures* for a given statement t of L. It is among these inference structures that we will find the statistical basis for the probability of t relative to w. We do not assume that w is deductively closed; thus, we use the notions, developed in LFSI, of being connected by a biconditional chain ($t_1 B_{w,L} t_2$), by an identify chain ($y_1 I_{w,L} y_2$), by an inclusion chain ($y_1 Incl_{w,L} y_2$), and by a membership chain ($x M_{w,L} y$). If we wish to impose the added constraint of deductive closure on w, these reduce to the presence in w (respectively) of the sentences "$t_1 \equiv t_2$", "$y_1 = y_2$", "$y_1 \subset y_2$", and "$x \in y$". To allow for unbounded reference sequences, we refer in the inference structures to the sequences themselves; the corresponding reference *set* is the union of the range of the reference sequence : $\cup \mathcal{R}y$.

For simplicity of notation in what follows, we will ignore these distinctions; they can easily be reintroduced by those to whom they seem important. Thus we will write "$t_1 \equiv t_2$" $\in w$ where we should, if w is not deductively closed, write "$t_1\ B_{w,L}\ t_2$". Similarly, we shall ignore in our notation the difference between a reference *sequence* y and the corresponding reference *set* $\cup \mathcal{R}y$, representing both simply by y. In the formal definitions we shall preserve the correct notation.

An inference structure in the language L, relative to a rational corpus w, for a sentence t of L, then, is a quintuple $\langle x,y,z,b,s \rangle$ where x denotes an individual (perhaps a complex one, such as a sample or an ordered pair consisting of a population and a sample, or a sequence), y a reference sequence, z a random quantity whose domain includes the union of the range of y, b a Borel set (which may be of any dimension; the set of Borel sets of dimension n is denoted by $\mathbf{B}^n$), s a statistical statement embodying the knowledge in w of the relative frequency with which objects in the sequence y have z-values in b, and $\ulcorner z(x) \in b \urcorner$ is connected by a biconditional chain to t. The statistical statement s is a statement of

[1]The idea of dealing with the inference structures was first suggested by William Harper in Harper (1982).

the form $\ulcorner\%(y,z,b) \in [p,q]\urcorner$ and means that the proportion of objects in y having z-values in b is in the closed interval $[p,q]$. "$PS_{w,L}(\ulcorner\%(y,z,b) \in [p,q)\urcorner,y,z,b)$" holds only if $\ulcorner\%(y,z,b) \in [p,q]\urcorner$ is a *strongest* statistical statement about y, z, and b in w, and the domain of z is known (in w) to include the union of the range of y, and b is a standard designator of a Borel set, and p and q are standard real number designators. (All this is provided for in LFSI as it stands). In addition we suppose that there is a standard set (perhaps infinite) of reference terms y and a standard set (perhaps infinite) of random quantities z. We suppose that these sets satisfy certain natural closure conditions, but we shall not characterize them in detail here.

Having entered these provisos, which are needed merely to exclude unintended cases, we will not mention them again. We may define the set of inference structures for sentence t relative to L and w thus:

D-1 $IS_{w,L}(t) = \{\langle x,y,z,b,s\rangle : tB_{w,L} \ulcorner z(x) \in b\urcorner \wedge xM_{w,L} \cup \mathfrak{R}y \wedge PS_{w,L}(s,y,z,b)\}$

Less formally, if we restrict our attention to sentences of the language L and to "strongest" statistical statements, we could write:

$$IS_w(t) = \{\langle x,y,z,b,s\rangle : \text{``}t \equiv z(x) \in b\text{''} \in w \wedge \text{``}x \in y\text{''} \in w \wedge s \in w\}$$

Any such quintuple is in principle capable of supporting a probability statement about t. But in general there may be a number of them giving rise to a variety of intervals $[p,q]$. Our problem is to single out the unique interval $[p,q]$ that represents the most appropriate constraint for rational belief in t. In view of the power of our language, we should not expect to find a *unique* reference sequence y: there will be many sequences having the same structure and differing only in logical type—something one might well exploit in the pursuit of the algorithm alluded to earlier.

There are two opposed considerations in judging among inference structures. On the one hand, we would like to use the most precise statistical knowledge we have (if $[p,q] \subset [p',q']$, we'd prefer the inference structure mentioning $[p,q]$ to the one mentioning $[p',q']$); and on the other hand, if the statistical statements mentioned in the two inference structures *differ* ($\sim[p,q] \subset [p',q']$ and $\sim[p',q'] \subset [p,q]$) we should base the probability of t on one of them only if we have good reason to ignore the other—for example, if it is based on a less determinate reference set.

Since we wish probability to be legislative for rational belief, and to serve as a guide to choice and action by entering into the calculation of expected utilities, we wish to ensure that of two conflicting inference structures no more than one will survive to impose its constraints on our degrees of belief. In general we may expect conflicting structures to cancel each other out, but there are several special cases in which one should dominate the other.

The simplest special case (inspiring Reichenbach's famous dictum that one ought to use the narrowest reference class about which one has adequate statistics) is that in which you know that one prospective reference set is included in the other.

For example, if you know the survival rate for 40-year old American males to be 0.990, and also that the survival rate for 40-year old American male white-collar workers to be 0.995, then, other things being equal, it is the latter that should constrain your beliefs and enter your utility calculations concerning the particular 40 year old male white-collar worker John Smith.

A second special case is particularly germane to statistical inference. Suppose you have observed a sample of 10,000 A's and noted that 5,236 of them are B's. Under certain circumstances we will be able to use our knowledge of this sample as the basis for a claim to the effect that the probability is high (e.g., [.99,1.0]) that the general relative frequency of B's among A's lies in a certain interval (say, .5236 ± .0100). The corresponding inference structure (which we may consider *regardless* of the circumstances) is:

$$s \quad A^{10,000} \quad z \quad [-.0100, .0100]$$

$$\%(A^{10,000}, z,[-.01, +.01]) \in [1-\epsilon,1]$$

where s is the sample at issue, and $z(x) = rf_B(x) - m(B,A)$ = the relative frequency of B's in the sample x, less the measure of B's in A in general. The statistical statement may be based on the *set theoretical* truth that, regardless of the measure of B's among A's, at least $1 - \epsilon$ of the set of all 10,000 member sequences of members of A are *representative* in the sense that $z \in [-.01,.01]$. There may even be a stronger statistical statement we can accept.

Suppose, however, that we consider a subsample $s' \subset s$, e.g., $s' \in A^{4,000}$, among which we find 3,768 B's. This gives rise to the conflicting inference structure:

$$s' \quad A^{4,000} \quad z \quad [-.4284, -.4084]$$

$$\%(A^{4,000},z,[-.4284,-.4084]) \in [0,\delta]$$

where the interval has been chosen so that both inference structures belong to the same class of inference structures: $z(s) \in [-.0100,.0100] \equiv z(s') \in [-.4284,-.4084]$.

These two inference structures clearly disagree; yet it is also clear that in many cases—though perhaps not all—we would want to save the first and discard the second. The grounds are that in the second structure s' is known to be included in s, just as, in the first example, the grounds for preference are the symmetrical ones that the reference set of the desired structure is known to be included in the reference set of the one we would like to discard.

The third and most important special case is that of Bayesian inference. Consider a set of 100 balls of which 40 are black and 60 are white. The next ball chosen is one of those balls, and the obvious inference structure would yield a probability of 0.40 that it is black. But suppose we also know what the 100 balls are divided among eleven urns, in one of which there are 40 white balls and 10 black balls, and in each of the rest there are two white balls and three black balls, and that the procedure for selecting a ball consists of first selecting an urn, and second selecting a ball. The corresponding inference structure yields a probability of 0.56 that the ball is black. Intuitively and *ceteris paribus,* the second inference structure should not be undermined by the first.

This pair of inference structures has a more complicated structure than the preceding two. Let B be the set of balls, z the characteristic function of black objects, b the particular ball chosen. For the second structure, let the pair $\langle u,b\rangle$ be the urn selected paired with the ball chosen from it; let z' be the characteristic function of *pairs* of objects, of which the second is black, and let R be the set of pairs $\langle x,y\rangle$ such that x is an urn and y a ball chosen from *that* urn. The two structures we have mentioned are:

$$b \quad B \quad z \quad \{1\} \quad \%(B,z,\{1\}) \in [.40,.40]$$

$$\langle u,b\rangle \quad R \quad z' \quad \{1\} \quad \%(R,z',\{1\}) \in [.56,.56].$$

Note that while R is not a subset of B, it is a subset of a set whose structure exactly matches that of B: the set of pairs $\langle x,y\rangle$ such that x is an urn and y a ball. Where U is the set of urns,

$$\langle u,b\rangle \quad U \times B \quad z' \quad \{1\} \quad \%(U \times B,z',\{1\}) \in [.40,.40]$$

is an inference structure for "the next ball chosen is black". The grounds for preferring the second structure to the first are that there is a structure exactly matching the first whose reference set contains a proper subset entering into an inference structure that matches the second.

These three special cases seem to embody the only grounds we have for disregarding an inference structure which conflicts with another. For the moment we leave the question of greater precision to one side.

Our first task is to characterize the sort of conflict that arose in our three examples more generally. As the following example shows, conflict may also arise from our knowledge of the relative frequency of *related* values of the random quantity z. Let C be a sequence of chips drawn from an urn over a long period (with replacement), and C' be the sequence of chips drawn yesterday. We know that $\cup \mathfrak{R}\, C' \subset \cup \mathfrak{R}\, C$. Let c be a chip drawn yesterday: $c \in \cup \mathfrak{R}\, C'$. Let z be the random quantity whose value is 0 for white objects, 1 for blue objects, and 2 for red objects. Suppose our knowledge of relative frequencies is represented by the following table. (All we know about yesterday's draws is that 40% were white.)

	$z = 0$	$z = 1$	$z = 2$
C	[.50,.50]	[.25,.25]	[.25,.25]
C'	[.40,.40]	[0.0,.60]	[0.0,.60]

It seems quite clear that the probability that $z(c) = 0$, i.e., that c is white, should be [.40,.40]. But if we are interested in the probability that c is blue, the choice between C and C' is not so clear cut. The relative frequencies of blue and red chips in C and C' seem merely known *more precisely* in C than in C', and this is not the same as having different *known* values. But we do know that at least one of these frequencies is different in C' than in C. This is true in general: that if the relative frequency with which a random quantity falls in a certain Borel set differs in two reference classes R_1 and R_2, then there is also some other Borel set for which the relative frequency differs, whether we know which one or not.

We will take account of this by saying that in general two inference structures belonging to the same class *disagree*—$\langle x_1,y_1,z_1,b_1,s_1\rangle$ $DIS_{w,L}$ $\langle x_2,y_2,z_2,b_2,s_2\rangle$—not only when they are both members of $IS_{w,L}(t)$ for some t and s_1 $DIF_{w,L}$ s_2, but also when there are refinements of the quantities z_1 and z_2 which lead to conflicting inference structures for some corresponding pair of Borel sets b_1' and b_2'. As in LFSI, we say that two statistical statements differ, s_1 $DIF_{w,L}$ s_2 when the interval mentioned in neither is included in the interval mentioned in the other. We say that one random quantity is a refinement of another if it allows us to make more distinctions than the other. For example, the quantity z whose value for a blue chip is 1, and for a white chip 0, and for a red chip 2, is a refinement of the quantity z^* whose value is 1 for a blue chip and 0 for a non-blue chip. Formally, we have the following definition:

D-2 $\langle x_1,y_1,z_1,b_1,s_1\rangle\ DIS_{w,L}\ \langle x_2,y_2,z_2,b_2,s_2\rangle \equiv \bigvee t\,(\langle x_1,y_1,z_1,b_1,s_1\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle \in IS_{w,L}(t) \wedge \bigvee z_1',z_2',b_1',b_2',s_1',s_2'\,(\ulcorner\bigwedge x(x \in \cup \mathfrak{R}y_1 \wedge z_1(x) \in b_1 \supset z_1'(x) \in b_1)\urcorner \in w \wedge \ulcorner\bigwedge x(x \in \cup \mathfrak{R}y_2 \wedge z_2(x) \in b_2 \supset z_2'(x) \in b_2)\urcorner \in w \wedge \ulcorner z_1'(x_1) \in b_1'\urcorner\ B_{w,L}\ulcorner z_2'(x_2) \in b_2'\urcorner \wedge \vdash \ulcorner b_1' \cap b_1 = \emptyset\urcorner \wedge \vdash \ulcorner b_2' \cap b_2 = \emptyset\urcorner \wedge PS_{w,L}(s_1',y_1,z_1',b_1') \wedge PS_{w,L}(s_2',y_2,z_2',b_2') \wedge s_1'\ DIF_{w,L}\ s_2')$

Less formally, $\langle x_1,y_1,z_1,b_1,s_1\rangle$ $DIS_{w,L}\langle x_2,y_2,z_2,b_2,s_2\rangle$ just in case:

(1) the two inference structures belong to the same class of inference structures,

(2) there exists a random quantity z_1' such that $\ulcorner\bigwedge x(x \in y_1 \wedge z_1(x) \in b_1 \supset z_1'(x) \in b_1)\urcorner \in w$ and a random quantity z_2' such that $\ulcorner\bigwedge x(x \in y_2 \wedge z_2(x) \in b_2 \supset z_2'(x) \in b_2)\urcorner \in w$

(3) but for some b_1' disjoint from b_1 and b_2' disjoint from b_2,

(4) the strongest statistical statement about y_1, z_1', and b_1' *differs* from the strongest statistical statement about y_2, z_2', and b_2'.

In the previous example, for instance, $\langle c,C,z^*,\{1\},\ulcorner\%(C,z^*,\{1\}) \in [.25,.25]\urcorner\rangle$ $DIS_{w,L}\langle c,C',z^*,\{1\},\ulcorner\%(C',z^*,\{1\}) \in [0.0,.60]\urcorner\rangle$ even though the mentioned statistical statements do not differ, since z, whose value is 0 for a white chip, 1 for a blue chip, and 2 for a red chip is a refinement of z^* (the same random quantity happens to be involved in both inference structures), $\{0\}$ and $\{1\}$ are disjoint, and $\%(C,z,\{0\}) \in [.50,.50]$ *differs* from $\%(C',z,\{0\}) \in [.40,.40]$.

Note that every random quantity is a refinement of itself, so that D-2 encompasses the special case of "difference" with which we began our discussion, as the following Lemma and Theorem show.

L-1 $\quad t_1 B_{w,L} t_2 \supset \bigwedge I(I \in IS_{w,L}(t_1) \equiv I \in IS_{w,L}(t_2))$

This holds, since $B_{w,L}$ is a metalinguistic equivalence relation. If "$t_1 \equiv t_2$" is in w, t_1 and t_2 determine the same set of inference structures.

T-1 $\quad \langle x_1,y_1,z_1,b_1,s_1\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle \in IS_{w,L}(t) \wedge s_1\, DIF_{w,L} s_2 \supset \langle x_1,y_1,z_1,b_1,s_1\rangle\, DIS_{w,L}\langle x_2,y_2,z_2,b_2,s_2\rangle$

Proof: Take $z_1' = z_1$, $z_2' = z_2$, $b_1' = \boldsymbol{B} - b_1$, $b_2' = \boldsymbol{B} - b_2$, and s_1' and s_2' to be the obviously derivable statistical statements. Then: $\vdash\ulcorner\bigwedge x(x \in y_1 \wedge z_1(x) \in b_1 \supset z_1'(x) \in b_1)\urcorner$ and similarly for z_2 and b_2, so that the required generalizations belong to w. Since $\ulcorner z_1(x) \in b_1\urcorner\ B_{w,L}\ulcorner z_2(x_2) \in b_2\urcorner$, we also have $\ulcorner z_1(x_1) \in \boldsymbol{B} - b_1\urcorner\ B_{w,L}\ulcorner z_2(x_2) \in \boldsymbol{B} - b_2\urcorner$. And of course $\vdash\ulcorner\boldsymbol{B} - b_1 \cap b_1 = \emptyset\urcorner$ and $\vdash\ulcorner\boldsymbol{B} - b_2 \cap b_2 = \emptyset\urcorner$. Clearly s_1' and s_2' differ just in case s_1 and s_2 differ.

We note that the relation $DIS_{w,L}$ between inference structures is symmetrical:

T-2 $\quad I_1\, DIS_{w,L} I_2 \supset I_2\, DIS_{w,L} I_1$

Proof: All the clauses of D-2 are symmetrical.

We wish not merely to avoid inference structures that disagree, but to be directed to an inference structure that offers the maximum amount of guidance. We say that one t-inference structure $I_1 = \langle x_1,y_1,z_1,b_1,s_1\rangle$ is *stronger* than another t-inference structure $I_2 = \langle x_2, y_2, z_2, b_2, s_2\rangle$—$I_1\ ST_{w,L}\ I_2$—just in case the statistical statement s_1 is stronger than the statistical statement s_2. Formally:

D-3 $\langle x_1, y_1, z_1, b_1, s_1\rangle\ ST_{w,L}\ \langle x_2, y_2, z_2, b_2, s_2\rangle \equiv \bigvee t(\langle x_1, y_1, z_1, b_1, s_1\rangle, \langle x_2, y_2, z_2, b_2, s_2\rangle \in IS_{w,L}(t)) \wedge s_1\ STR_{w,L}\, s_2)$

As we have seen, it is possible for two inference structures for the same sentence to disagree in the sense just defined, but for one of them to be nevertheless appropriate to determine the probability of that sentence, at least as far as anything embodied in the other one is concerned. In the first example, the desired reference class is a subset of the alternative reference class. In the second example, the desired reference class consists of objects themselves larger than the objects (the smaller samples) of the alternative reference class. In the third example, the desired reference class is arrived at by a Bayesian process, while the alternative corresponds roughly to a marginalization of that process. In all such cases we shall say that the desired inference structure *dominates* the other. Often, of course, when two inference structures disagree with each other, neither will dominate the other—they simply undermine each other and require us to find yet another inference structure which will dominate those it disagrees with. The dominating inference structure may be weaker than its alternatives, and in fact may in some cases be maximally weak.

Suppose that the inference structure I_1 disagrees with the inference structure I_2 in the sense just defined. In order to preserve I_1 against I_2, we must construct an inference structure that is based on I_2, that matches I_1, and that is related to I_2 in one of the three ways already mentioned. We shall say that this constructed inference structure I *reflects* I_1 in I_2: $I\ R_{w,L}(I_1, I_2)$.

A single example may help to motivate parts of the definitions that follow; other examples will be considered later. Consider the first example of disagreement, and let M be the class of 40 year old American males, W the class of white collar workers, Y the characteristic function of those who survive for one year, and j John Smith. Then the two inference structures we considered earlier are:

$$I_2 \quad j \quad M \cap W \quad Y \quad \{1\} \quad \%(M \cap W, Y, \{1\}) \in [.995, .995]$$

$$I_3 \quad j \quad M \quad Y \quad \{1\} \quad \%(M, Y, \{1\}) \in [.990, .990]$$

In this case, I_2 itself reflects I_2 in I_3: $I_2\ R_{w,L}(I_2, I_3)$. But we may also consider the inference structure:

$$I_3' \quad \{j\} \quad [M]^1 \quad Y' \quad \{1\} \quad \%([M]^1, Y', \{1\}) \in [.990, .990]$$

where $[M]^1$ is the set of 1-membered subsets of M, and Y' is the function whose value for $\{x\}$ is exactly the value that Y has for x. We note that the reference set of I_2 is *not* a subset of $[M]^1$. But there is an inference structure I_1 which reflects I_2 in I_3':

$$I_1 \quad \{j\} \quad [M \cap W]^1 \quad Y' \quad \{1\}\ \%([M \cap W]^1, Y', \{1\}) \in [.995, .995]$$

The definitions that follow also take care of variants of the second and third examples we considered earlier.

The example shows that I need not be I_1 itself, but must merely "match" I_1. We first define what it is for one inference structure to match another belonging to the same class:

D-4 $\langle x,y,z,b,s\rangle\ MATCHES_{w,L}\ \langle x_1,y_1,z_1,b_1,s_1\rangle \equiv \bigvee t(\langle x,y,z,b,s\rangle, \langle x_1,y_1,z_1,b_1,s_1\rangle \in IS_{w,L}(t)) \wedge \bigwedge b', s'(\langle x,y,z,b',s'\rangle \in IS_{w,L}(\ulcorner z(x) \in b'\urcorner) \supset \bigvee b'', s''(\langle x_1,y_1,z_1,b'',s''\rangle \in IS_{w,L}(\ulcorner z(x) \in b'\urcorner) \wedge s'\ E\text{-}STR_{w,L}\, s''))$

Note that D-4 merely requires that every statistical distinction in y_1 be quantitatively mirrored by a corresponding distinction in y. It is possible that z makes more distinctions in y than z_1 does in y_1.

The constructed inference structure I must also be as "rich" as the structure I_2. That is, we must be able to make the same discriminations concerning the values of $z(x)$ as we can concerning the values of $z_2(x_2)$.

D-5 $\langle x,y,z,b,s\rangle\ ARA_{w,L}\langle x_2,y_2,z_2,b_2,s_2\rangle \equiv \bigvee t(\langle x,y,z,b,s\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle \in IS_{w,L}(t) \wedge \bigwedge b', s'\ (\langle x_2,y_2,z_2,b',s'\rangle \in IS_{w,L}(\ulcorner z_2(x_2) \in b'\urcorner) \supset \bigvee b'', s''\ (\langle x,y,z,b'',s''\rangle \in IS_{w,L}(\ulcorner z_2(x_2) \in b'\urcorner))$

We also require that I be statistically consistent in the sense that there be no *internal* conflict:

D-6 $SCons_{w,L}\langle x,y,z,b,s\rangle \equiv \bigwedge b', s'\ (\langle x,y,z,b',s'\rangle \in IS_{w,L}(\ulcorner z(x) \in b\urcorner)\ s'\ E\text{-}STR_{w,L}\, s)$

We must now characterize the three constructions by means of which we can generate a reflecting structure I.

The first construction is characterized by the fact that the reference class y of I is known to be a subset of the reference class y_2 of I_2; informally, "$y \subset y_2$" $\in w$. Formally, we have:

D-7 $\langle x,y,z,b,s\rangle\, SC_{w,L}\, \langle x_2,y_2,z_2,b_2,s_2\rangle \equiv \bigvee t(\langle x,y,z,b,s\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle \in IS_{w,L}(t)) \wedge \ulcorner \cup\, \mathfrak{R}y\urcorner\ Incl_{wL}\ \ulcorner \cup\, \mathfrak{R}y_2\urcorner$,

where $I\ SC_{w,L}$ means that I is obtained from I_2 by a *S*ubset *C*onstruction.

The second construction is dual to the first. It is characterized by the fact that x and x_2 are sequences and x_2 is a subsequence of the minimal sequence x; y is of essentially higher dimension than y_2—hence "*HD*":

D-8 $\langle x,y,z,b,s\rangle\ HD_{w,L}\ \langle x_2,y_2,z_2,b_2,s_2\rangle \equiv \bigvee t(\langle x,y,z,b,s\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle\ IS_{w,L}\ (t) \wedge \ulcorner\bigcup \mathcal{R}x_2\urcorner\ Incl_{w,L}\ \ulcorner\bigcup \mathcal{R}x\urcorner \wedge \bigwedge x'y'z'b's' (\langle x',y',z',b',s'\rangle\ MATCHES_{w,L}\ \langle x,y,z,b,s\rangle \supset \sim\ulcorner\bigcup \mathcal{R}x'\urcorner\ Incl_{w,L}\ \ulcorner\bigcup \mathcal{R}x\urcorner)$

The third construction is the most complex, and in fact includes the first as a special case. Here y is a subset of a product of y_2 with some other set $y_2{}^*$, where the product itself matches $I_2 = \langle x_2,y_2,z_2,b_2,s_2\rangle$. This is the *BA*yesian construction:

D-9 $\langle x,y,z,b,s\rangle\ BA_{w,L}\ \langle x_2,y_2,z_2,b_2,x_2\rangle \equiv \bigvee t(\langle x,y,z,b,s\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle \in IS_{w,L}(t)) \wedge \bigvee y_2',y_2{}^* (\ulcorner\bigcup \mathcal{R}y\urcorner\ Incl_{w,L}\ \ulcorner\bigcup \mathcal{R}y_2'\urcorner \wedge \ulcorner\bigcup \mathcal{R}y_2'\urcorner\ I_{w,L}\ \ulcorner\bigcup \mathcal{R}y_2 \times \bigcup \mathcal{R}y_2{}^*\urcorner \wedge \langle x,y,z,b,s\rangle\ MATCHES_{w,L}\ \langle x_2,y_2,z_2,b_2,s_2\rangle)$

We can now give a general definition of reflection. We say that I is a reflection of I_1 in I_2 if all these inference structures belong to a common class, I matches I_1, I is as rich as I_2, I is statistically consistent and I is obtained from I_2 in one of the three canonical ways: $I\ SC_{w,L}I_2$, or for some I_2' matching I_2, $I\ HD_{w,L}I_2'$, or $I\ BA_{w,L}I_2$.

D-10 $I\ R_{w,L}\ (I_1,I_2) \equiv \bigvee t(I_1,I_2,I \in IS_{w,L}\ (t)) \wedge I\ MATCHES_{w,L}\ I_1 \wedge I\ ARA_{w,L}\ I_2 \wedge SCons_{w,L}\ I \wedge (I\ SC_{w,L}\ I_2 \vee \bigvee I_2' (\bigvee t(I_2,I \in IS_{w,L}\ (t)) \wedge I_2'\ MATCHES_{w,L}\ I_2 \wedge I\ HD_{w,L}\ I_2') \vee I\ BA_{w,L}I_2)$

We can now say of two disagreeing inference structures I_1 and I_2 that the first *dominates* the second—$I_1\ DOM_{w,L}I_2$—when there is an inference structure $\langle x,y,z,b,s\rangle$ which is a reflection of I_1 in I_2 and which is minimal in the sense that for any inference structure I^*, Borel set b^*, and statistical statement s^*, if I^* disagrees with $\langle x,y,z,b^*,s^*\rangle$ then there is a reflection of $\langle x,y,z,b^*,s^*\rangle$ in I^*, but no reflection of I^* in $\langle x,y,z,b^*,s^*\rangle$, and if I^* is stronger than $\langle x,y,z,b^*,s^*\rangle$, then there is an inference structure I', disagreeing with I^*, that contains no reflection of I^*. (Note that it is easy to devise inference structures I_1 and I_2 such that there exist reflections of each in the other—e.g., when $y \subset y_1 \subset y_2$ and y matches y_2, there is a reflection of y_1 in y_2 and of y_2 in y_1. In effect, we are requiring that this sort of process terminate.)

D-11 $\langle x_1,y_1,z_1,b_1,s_1\rangle\ DOM_{w,L}\ \langle x_2,y_2,z_2,b_2,s_2\rangle \equiv \langle x_1,y_1,z_1,b_1,s_1\rangle\ DIS_{w,L}\langle x_2,y_2,z_2,b_2,s_2\rangle \wedge \bigvee x,y,z,b,s [\langle x,y,z,b,s\rangle R_{w,L}(\langle x_1,y_1,z_1,b_1,s_1\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle) \wedge \bigwedge I^*, b^*, s^* [(I^*\ DIS_{w,L}\ \langle x,y,z,b^*,s^*\rangle \supset \bigvee I' (I'\ R_{w,L}\ (\langle x,y,z,b^*,s^*\rangle,I^*)) \wedge \sim\bigvee I'(I'\ R_{w,L}\ (I^*, \langle x,y,z,b^*,s^*\rangle))) \wedge (I^*\ ST_{w,L}\ \langle x,y,z,b^*,s^*\rangle \supset \bigvee I' (I'\ DIS_{w,L}\ I^* \wedge \sim\bigvee I''(I''\ R_{w,L}\ (I^*,I'))))]]$

We first show that these definitions take care of the three special cases introduced earlier. In the insurance example, the disagreeing inference structures were

I_1: j M Y $\{1\}$ $\%(M,Y,\{1\}) \in [.990,.990]$

I_2: j $M \cap W$ Y $\{1\}$ $\%(M \cap W,Y,\{1\}) \in [.995,.995]$

where for simplicity we construe M and W as sets rather than as sequences. We suppose, further, that we have no statistical knowledge other than that following from the statistical statements mentioned in I_1 and I_2. To show that I_2 dominates I_1, we must first exhibit an inference structure which is a reflection of I_2 in I_1. According to D-10, we may employ a subset construction by finding a subset of M having a structure parallel to that of I_2. In this case $M \cap W$ itself will suffice. Thus the required inference structure is I_2. Next we must argue that the last clause of *D*-11 is satisfied. Because of the paucity of our assumed statistical knowledge, this is easy: all candidates for I^* have the structure of I_1 (or its complementary structure) and I_2 contains no reflection of I_1.

The second example concerned the conflict between a sample and a subsample. The two inference structures were:

I_4 s $A^{10,000}$ z $[-.0100,.0100]$

$$\%(A^{10,000},z,[-.0100,.0100]) \in [1-\epsilon,1]$$

I_5 s' $A^{4,000}$ z $[-.4284,-.4084]$

$$\%(A^{4,000},z,[-.4284,-.4204]) \in [0,\Delta]$$

We must show that there is an inference structure matching I_5—here it is just I_5 itself—and a reflecting inference structure—here it is just I_4 itself—such that the $\cup\, \mathcal{R}s'\ Incl_{w,L} \cup \mathcal{R}s$. The other conditions of D-10 are clearly met. Continuing with D-11: that there is no reflection of I_5 in I_4 is obvious. Consider inference structures involving other Borel sets. Clearly the same considerations will apply.

The third example was the Bayesian one. The disagreeing inference structures were

I_6 b B z $\{1\}$ $\%(B,z,\{1\}) \in [.40,.40]$

I_7 $\langle u,b \rangle$ R z' $\{1\}$ $\%(R,z',\{1\}) \in [.56,.56]$

The same inessential simplifications have been made as before. Let U be the set of urns described in the example. Then the inference structures required to show that $I_7\ DOM_{w,L}\ I_6$ are:

I_8 $\langle u,b \rangle$ $U \times B$ z' $\{1\}$ $\%(U \times B,z',\{1\}) \in [.40,.40]$

and I_7. The remaining clauses of D-11 are again trivial, since we need deal only with $\{1\}$ and $\mathbf{B} - \{1\}$.

The route to randomness is now clear. Consider the set of t-inference structures each of which dominates any inference structure with which it disagrees. This set is non-empty, since if t is connected by a biconditional chain to $\ulcorner z(x) \in b \urcorner$—and there always exist such z, x, and b—then the inference structure $\langle x,\{x\},z,b,\ulcorner\%(\{x\},z,b) \in [0,1]\urcorner\rangle$ dominates any structure with which it disagrees since it disagrees with none. The inference structures $\langle x,\{x\},z,b,\ulcorner\%(\{x\},z,b) \in [1,1]\urcorner\rangle$ and $\langle x,\{x\},z,b,\ulcorner\%\{x\},z,b) \in [0,0]\urcorner\rangle$ dominate any structure with which they disagree, since we can always construct a reflection of them in the disagreeing set. These structures—those which dominate any structure with which they disagree—may be partially ordered by the strength of the statistical statements they mention. Among them, any inference structure which is as strong as or stronger than any other will serve as a basis for the probability of t, and the union of the range of the reference sequence mentioned in it is the reference class for that probability.

The term x denotes an object that is a *random member* of the reference sequence denoted by y, with respect to having a z-value in the Borel set b, relative to the body of knowledge w, just in case there is a statistical statement s such that the inference structure $\langle x,y,z,b,s\rangle$ dominates any $\ulcorner z(x) \in b\urcorner$-inference structure with which it disagrees, and is at least as strong as any $\ulcorner z(x) \in b\urcorner$-inference structure which dominates anything it disagrees with. We denote this relation by $RAN_L(x,y,z,b,w)$.

D-12 $RAN_L\langle x,y,z,b,w\rangle \equiv \bigvee s(\langle x,y,z,b,s\rangle \in IS_{w,L}\ulcorner z(x) \in b\urcorner \wedge \bigwedge I(I \in IS_{w,L}(\ulcorner z(x) \in b\urcorner) \supset (I\ DIS_{w,L}\langle x,y,z,b,s\rangle \supset \langle x,y,z,b,s\rangle\ DOM_{w,L}\ I) \wedge (I\ ST_{w,L}\langle x,y,z,b,s\rangle \supset \bigvee I^*(I^*\ DIS_{w,L}\ I \wedge {\sim}I\ DOM_{w,L}\ I^*))))$

Note that there may be—given the richness of L, there will be—a number of inference structures that are equally strong and dominating. Since it makes no difference which we use, we might as well let the randomness relation hold for any of them.

The definition of probability is clear. It is essentially the same as the definition that appears in LFSI and other places.

D-13 $Prob_L(t,w) = [p,q] \equiv \bigvee x,y,z,b(\langle x,y,z,b,\ulcorner\%(y,z,b) \in [p,q]\urcorner\rangle \in IS_{w,L}(t) \wedge RAN_L\langle x,y,z,b,w\rangle)$

3. Theorems. We can now establish some of the properties of this interpretation of probability that make it of interest, and allow us to relate the present interpretation of probability to frequency and subjectivistic interpretations. The first theorem shows that we can obtain a metalinguistic reflection of a frequency interpretation of probability.

T-3 $\bigwedge b(b \in \mathbf{B} \supset \bigvee p(\langle x,y,z,b,\ulcorner\%(y,z,b) \in [p,p)\urcorner\rangle \in IS_{w,L}(\ulcorner z(x) \in b\urcorner) \wedge RAN_L(x,y,z,b,w)) \supset \bigwedge b_1,b_2(b_1,b_2 \in \mathbf{B} \supset (\bigvee p$

$(Prob_L(\ulcorner z(x) \in b_1 \urcorner, w) = [p,p] \wedge p \in [0,1]) \wedge \bigwedge p_1,p_2$ $(\vdash \ulcorner b_1 \cap b_2 = \emptyset \urcorner \wedge Prob_L(\ulcorner z(x) \in b_1 \urcorner, w) = [p_1,p_1] \wedge Prob_L$ $(\ulcorner z(x) \in b_2 \urcorner) = [p_2,p_2] \supset Prob_L(\ulcorner z(x) \in b_1 \vee z(x) \in b_2 \urcorner,$ $w) = [p_1 + p_2, p_1 + p_2]))$

Given the antecedent, the proof is trivial. What the antecedent amounts to is (a) the stipulation that the distribution of z in the reference sequence y be completely known, and (b) that "all we know about x is that it is a member of the reference set". The latter stipulation is obscure, as conventionally stated: of course we know a lot more about x: e.g., $x \in \{x\}$, $\{x\}$ belongs to the power set of the union of the range of y, etc. The point is that we don't know anything about x that is relevant to its z-value. But this is exactly what the antecedent $RAN_L(x,y,z,b,w)$ stipulates.

The situation characterized by the antecedent of T-3 is perhaps possible only in a general theoretical context; nevertheless it does sometimes reflect a useful idealization of an actual situation. Of more interest to us are theorems of greater generality and greater applicability. Pivotal to these theorems, and important for the soundness of the definition of probability, is the theorem that if I_1 dominates I_2, then I_2 does not dominate I_1:

T-4 $I_1 \, DOM_{w,L} \, I_2 \supset \sim I_2 \, DOM_{w,L} \, I_1$

Assume that both $I_1 = \langle x_1,y_1,z_1,b_1,s_1 \rangle \, DOM_{w,L} \, I_2 = \langle x_2,y_2,z_2,b_2,s_2 \rangle$ and $I_2 \, DOM_{w,L} \, I_1$. By D-11, this means there must exist x_1', y_1', z_1', b_1', s_1', and x_2', y_2', z_2', b_2', s_2' so that

(1) $\langle x_1',y_1',z_1',b_1',s_1' \rangle \, R_{w,L} \, (I_1,I_2)$

(2) $\langle x_2',y_2',z_2',b_2',s_2' \rangle \, R_{w,L} \, (I_2,I_1)$, and

(3) $\bigwedge I^*[I^* \, DIS_{w,L} \, \langle x_2',y_2',z_2',b_2',s_2' \rangle \supset \sim \bigvee I'(I' \, R_{w,L}(I^*, \langle x_2', y_2',z_2',b_2',s_2' \rangle))]$

Instantiating I^* to $\langle x_1',y_1',z_1',b_1',s_1' \rangle$, and recalling that since $s_1 \, DIF_{w,L}$ s_2, $s_1' \, DIF_{w,L} \, s_2'$ so that $\langle x_1',y_1',z_1',b_1',s_1' \rangle$ disagrees with $\langle x_2',y_2', z_2',b_2',s_2' \rangle$, we have

(4) $\sim \bigvee I'(I' \, R_{w,L}(\langle x_1',y_1',z_1',b_1',s_1' \rangle, \langle x_2',y_2',z_2',b_2',s_2' \rangle))$

But applying D-11 to $I_1 \, DOM_{w,L} \, I_2$, we also have:

(5) $\bigvee I'(I' \, R_{w,L}(\langle x_1',y_1',z_1',b_1',s_1' \rangle, \langle x_2',y_2',z_2',b_2',s_2' \rangle)$,

which is impossible.

We can now establish some properties of probability. We must first have two more lemmas:

Lemma 2: $PS_{w,L}(y,z,b,\ulcorner \%(y,z,b) \in [p,q] \urcorner) \equiv PS_{w,L}(y,z,\mathbf{B} - b, \ulcorner \%(y,z,\mathbf{B} - b) \in [1 - q, 1 - p] \urcorner)$

The proof depends only on the fact that if $t \in w$ and $t \vdash r$, then $r \in w$. We introduce the following notation: if $s = \ulcorner\%(y,z,b) \in [p,q]\urcorner$ we take cs to be $\ulcorner\%(y,z,\mathbf{B} - b) \in [1 - q,1 - p]\urcorner$. Then the next lemma may be stated:

Lemma 3: $\langle x,y,z,b,s\rangle R_{w,L}(\langle x_1,y_1,z_1,b_1,s_1\rangle, \langle x_2,y_2,z_2,b_2,s_2\rangle) \supset \langle x,y,z,\mathbf{B} - b,cs\rangle R_{w,L}(\langle x_1,y_1,z_1,\mathbf{B} - b_1,cs_1\rangle, \langle x_2,y_2,z_2,\mathbf{B} - b_2,cs_2\rangle)$

Proof: We need merely trace the clauses of D-10 and D7–D9, noting that if s_2' $E\text{-}STR_{w,L}$ s_2, then cs_2' $E\text{-}STR_{w,L}$ cs_2, etc.

The simplest property concerns complementary probabilities; this property, incidentally, holds for most interval measures of support, as proposed by, e.g., Dempster (1968) and Shafer (1976).

T-5 $Prob_L(t,w) = [p,q] \equiv Prob_L(\ulcorner{\sim}t\urcorner,w) = [1 - q,1 - p]$

Proof: Assume the left hand side. By D-13,

$$\bigvee x,y,z,b(\langle x,y,z,b,\ulcorner\%(y,z,b) \in [p,q]\urcorner\rangle \in IS_{w,L}(t) \wedge RAN_L(x,y,z,b,w))$$

Existentially instantiating to the same variables, we claim:

$$\langle x,y,z,\mathbf{B} - b,\ulcorner\%(y,z,\mathbf{B} - b) \in [1 - q,1 - p]\urcorner\rangle \in IS_{w,L}(\ulcorner{\sim}t\urcorner)$$

And $RAN_L(x,y,z,\mathbf{B} - b,w)$

The former claim is easy to verify, with the help of Lemma 2. To establish the latter, we need to show:

$$\bigwedge I(I \in IS_{w,L}(\ulcorner{\sim}t\urcorner) \supset (I\ DIS_{w,L}\langle x,y,z,\mathbf{B} - b,cs\rangle \supset \langle x,y,z,\mathbf{B} - b,cs\rangle\ DOM_{w,L}\ I) \wedge (I\ ST_{w,L}\ \langle x,y,z,\mathbf{B} - b,cs\rangle \supset \bigvee I^*\ (I^*\ DIS_{w,L}\ I \wedge I^* \in IS_{w,L}(\ulcorner z(x) \in \mathbf{B} - b\urcorner) \wedge {\sim}I\ DOM_{w,L}\ I^*))$$

where we have written cs for $\ulcorner\%(y,z,\mathbf{B} - b) \in [1 - q,1 - p]\urcorner$

Suppose that $I = \langle x',y',z',b',s'\rangle$, $I \in IS_{w,L}(\ulcorner{\sim}t\urcorner)$. Suppose, in addition, that $I\ DIS_{w,L}\langle x,y,z,\mathbf{B} - b,cs\rangle$. But then $\langle x',y',z',\mathbf{B} - b',cs'\rangle\ DIS_{w,L}\langle x,y,z,b,\ulcorner\%(y,z,b) \in [p,q]\urcorner\rangle$. Since $RAN_L\ (x,y,z,b,w)$, $\langle x,y,z,b,\ulcorner\%(y,z,b) \in [p,q]\urcorner\ DOM_{w,L}\ \langle x',y',z',\mathbf{B} - b',cs'\rangle$, and with the help of Lemma 3 we conclude that $\langle x,y,z,\mathbf{B} - b,cs\rangle\ DOM_{w,L}\ \langle x',y',z',b',s'\rangle$.

Suppose now that $I\ ST_{w,L}\ \langle x,y,z,\mathbf{B} - b,cs\rangle$. Then $\langle x',y',z',\mathbf{B} - b',cs'\rangle\ ST_{w,L}\ \langle x,y,z,b,s\rangle$, and therefore by $RAN_L\ (x,y,z,b,w)$, $\bigvee I^*(I^*\ DIS_{w,L}\ \langle x',y',z',\mathbf{B} - b',cs'\rangle \wedge {\sim}(x',y',z',\mathbf{B} - b',cs'\rangle\ DOM_{w,L}\ I^*)$, from which it follows that $\bigvee I^*(I^*\ DIS_{w,L}\ I \wedge {\sim}I\ DOM_{w,L}\ I^*$.

The next theorem and its corollary show, as one might expect, that if $t \in w$, then the probability of t is $[1,1]$, and if $\ulcorner{\sim}t\urcorner \in w$, then the probability of t is $[0,0]$. We require here the condition that w contain no explicit contradiction: $\sim\ulcorner 0 \neq 0 \urcorner \in w$, else all our probabilities will turn out to be [0,1] as may be seen from D-12.

T-6 $\sim\ulcorner 0 \neq 0\urcorner \in w \wedge t \in w \supset Prob_L(t,w) = [1,1]$

Proof: $\langle 0,\{0\},Id,\{0\},\ulcorner\%(\{0\},Id,\{0\}) \in [1,1]\urcorner\rangle \in IS_{w,L}(t)$, where Id is the identity function, since $\ulcorner t \equiv 0 \in \{0\}\urcorner \in w$. What we need to show is $RAN_L(0,\{0\},Id,\{0\},w)$. Suppose $\langle x,y,z,b,s\rangle\ DIS_{w,L}\langle 0,\{0\},Id,\{0\}, \ulcorner\%\{0\},Id,\{0\}) \in [1,1]\urcorner\rangle$. We show that the latter dominates the former. Since $t\ B_{w,L}\ulcorner z(x) \in b\urcorner$ (by hypothesis), $\ulcorner x \in y \wedge t\urcorner\ B_{w,L}\ulcorner x \in y \cap \{a:z(a) \in b\urcorner\}$. Since $t \vdash \ulcorner x \in y \wedge t \equiv x \in y\urcorner$, we have $\ulcorner x \in y\urcorner\ B_{w,L}\ulcorner x \in y \cap \{a:z(a) \in b\}\urcorner$. We also have $x\ M_{w,L}\ y$. MT 8.4 of LFSI states: $\ulcorner x \in y\urcorner\ B_{w,L}\ulcorner x' \in y'\urcorner \wedge x\ M_{w,L}\ y \supset x'\ M_{w,L}\ y'$. Thus we have $x\ M_{w,L}\ulcorner y \cap \{a:z(a) \in b\}\urcorner$. Clearly we have $\ulcorner\%(y \cap \{a:z(a) \in b\},z,b) \in [1,1]\urcorner$ in w. This establishes the existence of a reflection, and therefore domination. But $\langle x,y,z,b,s\rangle$ clearly cannot be stronger than $\langle 0,\{0\},Id,\{0\}, \ulcorner\%(\{0\},Id,\{0\}) \in [1,1]\urcorner$. Thus we have established our claim of randomness.

T-7 $\sim\ulcorner 0 \neq 0\urcorner \in w \wedge \ulcorner\sim t\urcorner \in w \supset Prob_L(t,w) = [0,0]$

Proof: T-6 and T-5.

The big theorem that ultimately establishes the existence of an additive measure satisfying the constraints imposed by probability follows. The "proof" of this theorem offered in LFSI was invalid, as Isaac Levi noticed years ago. Worse, as stated in LFSI, the theorem wasn't even true! The latter fact stemmed from my failure to take account of the situation represented by the case discussed earlier of the red, white, and blue counters. The theorem requires the existence of a very modest degree of deductive closure: We shall say that w is *statistically closed* when the result of combining two statistical statements in w is also in w. The formal definition is in LFSI, MD 8.8.

T-8 $\ulcorner\sim(t_1 \wedge t_2)\urcorner \in w \wedge w$ *is statistically closed* $\wedge\ Prob_L(t_1,w) = [p_1,q_1] \wedge Prob_L(t_2,w) = [p_2,q_2] \supset Prob_L(\ulcorner t_1 \vee t_2\urcorner,w) \subset [p_1 + p_2,q_1 + q_2] \vee [p_1 + p_2,q_1 + q_2] \subset Prob_L(\ulcorner t_1 \vee t_2\urcorner,w)$

Proof: Let I_1 and I_2 be inference structures yielding probabilities for t_1 and t_2, and let I be that yielding the probability of $\ulcorner t_1 \vee t_2\urcorner$:

I_1: $x_1\quad y_1\quad z_1\quad b_1\quad s_1 = \ulcorner\%(y_1,z_1,b_1) \in [p_1,q_1]\urcorner$

I_2: $x_2\quad y_2\quad z_2\quad b_2\quad s_2 = \ulcorner\%(y_2,z_2,b_2) \in [p_2,q_2]\urcorner$

I: $x\quad y\quad z\quad b\quad s = \ulcorner\%(y,z,b) \in [p,q]\urcorner$

Let y' be the sequence of cross-products of the corresponding elements of y_1 and y_2—i.e., let $y'(i) = y_1(i) \times y_2(i)$. Let z' be a function on the range of y' whose value for $\langle a,b \rangle \in \cup\, \mathcal{R}y'$ is $z'\langle a,b\rangle = \langle z_1(a), z_2(b)\rangle \in \mathbf{B}^2$. We consider the following four Borel sets:

$$b_1' = b_1 \times (\mathbf{B} - b_2)$$

$$b_2' = (\mathbf{B} - b_1) \times b_2$$

$$b_3' = b_1 \times b_2$$

$$b_0' = (\mathbf{B} - b_1) \times (\mathbf{B} - b_2)$$

Note that:

$$t_1\, B_{w,L} \ulcorner z'(x_1,x_2) \in b_1' \urcorner$$

$$t_2\, B_{w,L} \ulcorner z'(x_1,x_2) \in b_2' \urcorner$$

$$\ulcorner t_1 \wedge t_2 \urcorner\, B_{w,L} \ulcorner z'(x_1,x_2) \in b_3' \urcorner$$

Since $\ulcorner \sim(t_1 \wedge t_2) \urcorner \in w$, we have by T-7 $Prob_L(\ulcorner t_1 \wedge t_2, w \urcorner) = [0,0]$, based on the inference structure (say):

$$I_3{:}\quad x^* \quad y^* \quad z^* \quad b^* \quad s^* = \ulcorner \%(y^*,z^*,b^*) \in [0,0] \urcorner$$

From statistical closure and s_1 and s_2, we can infer:

$$\ulcorner \%(y',z',b_1') \in [p_1(1-q_2), q_1(1-p_2)] \urcorner \in w$$

$$\ulcorner \%(y',z',b_2') \in [(1-q_1)p_2, (1-p_1)q_2] \urcorner \in w$$

$$\ulcorner \%(y',z',b_3') \in [p_1p_2, q_1q_2] \urcorner \in w$$

$$\ulcorner \%(y',z',b_0') \in [(1-q_1)(1-q_2), (1-p_1)(1-p_2)] \urcorner \in w$$

These statistical statements may not be the strongest statistical statements about their respective subject matters in w. Let the strongest statements be s_1', s_2', s_3', and s_0', respectively, mentioning the intervals $[p_1',q_1']$, $[p_2',q_2']$, $[p_3',q_3']$, and $[p_0',q_0']$, respectively.

Now consider the inference structures:

$$I_1'{:}\quad \langle x_1,x_2\rangle \quad y' \quad z' \quad b_1' \qquad s_1'$$

$$I_2'{:}\quad \langle x_1,x_2\rangle \quad y' \quad z' \quad b_2' \qquad s_2'$$

$$I_3'{:}\quad \langle x_1,x_2\rangle \quad y' \quad z' \quad b_3' \qquad s_3'$$

$$I_0'{:}\quad \langle x_1,x_2\rangle \quad y' \quad z' \quad b_0' \qquad s_0'$$

I_0: $\langle x_1,x_2\rangle$ y' z' $\mathbf{B} - b_0'$ cs_0'

where again we write cs_0' for the statistical statement complementary to s_0'. From statistical closure it follows that

$$[1 - q_0', 1 - p_0'] \subset [p_1' + p_2' + p_3', q_1' + q_2' + q_3']$$

Note that $I_0 \in IS_{w,L}(\ulcorner t_1 \vee t_2 \urcorner)$.

Suppose, first of all, that one of I_1', I_2', I_3' or I_0 disagrees with I_1, I_2, I_3, or I, respectively. (Note that if $0 < p_1$ and $0 < p_2$, then since s_3' is if anything stronger than $\ulcorner\%(y',z',b_3') \in [p_1p_2, q_1q_2]$, s_3' $DIF_{w,L}s''$ and I_3 $DIS_{w,L}$ I_3'.)

Case I: I_3 $DIS_{w,L}$ I_3'. By D-12, $I_3DOM_{w,L}$ I_3'. By D-11, this means that:

1. $[\bigvee x_0,y_0,z_0,b_3'',s_3''] \langle x_0,y_0,z_0,b_3'',s_3''\rangle R_{w,L} (I_3,I_3') \wedge \bigwedge I^*,b^*,s^* [(I^* \; DIS_{w,L} \; \langle x_0,y_0,z_0,b^*,s^*\rangle \supset \bigvee I'(I' \; R_{w,L}(\langle x_0,y_0,z_0,b^*,s^*\rangle,I^*)) \sim\bigvee I'(I' \; R_{w,L} \; (I^*, \langle x_0,y_0,z_0,b^*,s^*\rangle))) \wedge (I^* \; ST_{w,L} \langle x_0,y_0,z_0,b^*,s^*\rangle \supset \bigvee I' \; (I' \; DIS_{w,L} \; I^* \wedge \sim\bigvee I''(I'' \; R_{w,L} \; (I^*,I'))))]$

By D-5:

2. $\bigwedge b_2'',s_2'' \; (\langle\langle x_1,x_2\rangle, y',z',b_2'',s_2''\rangle \in IS_{w,L} \; (\ulcorner z' \; \langle x_1,x_2\rangle \in b_2''\urcorner) \supset \bigvee b_0'',s_0'' \; (\langle x_0,y_0,z_0,b_0'',s_0''\rangle \in IS_{w,L} \; (\ulcorner z' \; (x_1,x_2\rangle \in b_2''\urcorner)))$

Instantiating b_2'' and s_2'' to b_1', s_1'; then to b_2', s_2'; and then to b_0', s_0', we obtain:

3. $[\bigvee b_1'',s_1''] \langle x_0,y_0,z_0,b_1'',s_1''\rangle \in IS_{w,L} \; (t_1)$
4. $[\bigvee b_2'',s_2''] \langle x_0,y_0,z_0,b_2'',s_2''\rangle \in IS_{w,L} \; (t_2)$
5. $[\bigvee b_0'',s_0''] \langle x_0,y_0,z_0,b_0'',s_0''\rangle \in IS_{w,L} \; (\ulcorner\sim t_1 \wedge \sim t_2\urcorner)$

Since $\langle x_0,y_0,z_0,b_1'' - b_2'' - b_0'',s''\rangle \in IS_{w,L} \; (t_1)$ implies s'' $E\text{-}STR_{w,L}$ s_1'', by D-6, and similarly for b_3'', b_2'' and b_0'', we may take b_0'', b_1'', b_2'' and b_3'' to be disjoint.

Instantiate I^* in the second conjunct of line 1 to I_1, b^* and s^* to b_1'' and s_1''. Suppose I_1 $DIS_{w,L}$ $\langle x_0,y_0,z_0,b_1'',s_1''\rangle$. Then by the corresponding conditional of line 1, $\sim\bigvee I'(I' \; R_{w,L} \; (I_1, \langle x_0,y_0,z_0,b_1'',s_1''\rangle))$. But also, by D-12, I_1 $DOM_{w,L}$ $\langle x_0,y_0,z_0,b_1'',s_1''\rangle$, so that $\bigvee I'(I' \; R_{w,L}(I_1, \langle x_0,y_0,z_0,b_1'',s_1''\rangle))$. Therefore, $\sim I_1$ $DIS_{w,L}$ $\langle x_0,y_0,z_0,b_1'',s_1''\rangle$.

Suppose I_1 $ST_{w,L}$ $\langle x_0,y_0,z_0,b_1'',s_1''\rangle$. Then by the corresponding conditional of line 1, $\bigvee I'(I' \; DIS_{w,L} \; I_1 \wedge \sim \bigvee I''(I'' \; R_{w,L} \; (I_1,I')))$. But again by D-12 and D-11, this is impossible.

We conclude,

6. s_1'' $E\text{-}STR_{w,L}$ $s_1 \vee s_1''$ $STR_{w,L}$ s_1

7. $s_2'' \textit{ E-STR}_{w,L}\ s_2 \vee s_2'' \textit{ STR}_{w,L}\ s_2$

8. $s_0'' \textit{ E-STR}_{w,L}\ s_0 \vee s_0'' \textit{ STR}_{w,L}\ s_0$

Therefore,

10. $\ulcorner\%(y_0,z_0,b_1'') \in [p_1,q_1]\urcorner \in w$

11. $\ulcorner\%(y_0,z_0,b_2'') \in [p_2,q_2]\urcorner \in w$

12. $\ulcorner\%(y_0,z_0,b_3'') \in [0,0]\urcorner \in w$

13. $\ulcorner\%(y_0,z_0,b_0'') \in [p_0,q_0]\urcorner \in w,$

where $p_0 = 1 - q$ and $q_0 = 1 - p$. By statistical closure, therefore,

14. $q_0 \leq 1 - p_1 - p_2$ and $1 - q_1$ and $q_2 \leq p_0$,

or $p_1 + p_2 \leq p$ and $q \leq q_1 + q_2$, as advertised.

Case II, Case III, and Case IV, where $I_1\ DIS_{w,L}\ I_1'$, $I_2\ DIS_{w,L}\ I_2'$, and $I_0\ DIS_{w,L}\ I_0'$, are similar, and lead to the same conclusion.

In the remaining case we have no disagreements; in particular, I_0 does not disagree with I, from which it follows immediately that $[p,q] \subset [p_1 + p_2, q_1 + q_2]$ or $[p_1 + p_2, q_1 + q_2] \subset [p,q]$.

It is worth noting that the circumstances that yield the case in which $[p_1 + p_2, q_1 + q_2] \subset [p,q]$ are rather odd. We must have $p_1 = 0$ or $p_2 = 0$, for otherwise we would have $I_3\ DIS_{w,L}\ I_3'$ automatically. We must therefore have an inference structure yielding a probability for t_1 (for which the lower bound is 0), and an inference structure yielding a probability for t_2. Furthermore, if we form the cross product of the reference sets, we don't get anything different, with respect to either t_1 or t_2. But when we contemplate the probability of the *disjunction* of t_1 and t_2—or, alternatively, the probability of the joint denial of t_1 and t_2—we obtain an interval which is *broader* than that we would get from the cross product structure. By D-12, the only way in which this can happen is that there is an inference structure I^* which disagrees with the cross product inference structure, but which is such that neither dominates the other. Thus we are thrust back to the broader interval.

A somewhat strained example: Experiment E consists of drawing a chip from urn 1, and then drawing a chip from urn 2. Between 0% and 20% of the chips from urn 1 are black; between 10% and 30% of the chips from urn 2 are black. On a given draw, D, performed by Susan on Tuesday, we know that two black chips were not drawn. But in the long run of actual trials of the experiment in which two black chips are not drawn, between 5% and 60% have yielded one black chip. Let t_1 be: the draw D from urn 1 yields a black chip; let t_2 be: the draw D from urn 2 yields a black chip.

The frequencies in the cross product set (corresponding to the notation of theorems 8 and 9) are:

$$p_1' = p_1(1 - q_2) = 0 \qquad q_1' = q_1(1 - p_2) = .18$$

$$p_2 = p_2(1 - q_1) = .08 \qquad q_2' = q_2(1 - p_1) = 0.30$$

$$p_3' = p_1 p_2 = 0 \qquad q_3' = q_1 q_2 = .06$$

$$p_0' = (1 - q_1)(1 - q_2) = .56 \quad q_0' = (1 - p_1)(1 - p_2) = .90$$

Of course we know that the draw D comes from a certain subset of the cross product—that in which there are no pairs of black draws—but the frequencies there aren't very different:

$$p_1'' = 0 \qquad q_1'' = .18/.98$$

$$p_2'' = .08/.98 \quad q_2'' = .30/1.0$$

$$p_3'' = 0.0 \qquad q_3'' = 0.0$$

$$p_0'' = .56/.94 \quad q_0'' = .90/1.0$$

In addition, suppose we know that between 5% and 10% of the draws performed by Susan yield a black chip followed by a non-black chip; and that the frequency with which pairs of non-black chips occur on Tuesday lies between .50 and .75.

So much for frequencies; how about probabilities?

(1) The probability that D yields a black chip on the first draw is not [0,.18] nor [0,.18/.90] (simply the proportion of pairs of chips of which the first is white), since we know that the draw D was performed by Susan, and that the frequency with which she draws a black chip followed by a non-black chip lies between .05 and .10, and that the first draw of D yields a black chip if and only if it yields a black chip followed by a non-black chip. The latter reference class yields a more precise frequency.

(2) The probability that D yields a black chip on the second draw is not [.08,.30] nor even [.08/.98,.30], since, although we have no conflicting information, we have stronger statistical knowledge about the frequency of black balls from urn 2: namely that it lies in [.10,.30].

(3) Finally, although we know that the frequency of pairs of non-black chips lies between .56 and .90 and that, among those pairs in which two black chips do not occur, the frequency of pairs of non-black chips lies between .56/.94 and .90, we also know that on Tuesday-draws between .50 and .75 of the draws result in pairs of non-black chips. These two bits of statistical knowledge

conflict; neither dominates the other; and so we fall back on our vague general knowledge that among performances of E, between 40% and 95% yield pairs of non-black chips.

To summarize: the probability of t_1—that the first chip is black—is $[.05,.10]$; the probability of t_2—that the second chip is black—is $[.10,.30]$; the probability of $\ulcorner t_1 \vee t_2 \urcorner$ that either the first or the second is black—is $[.05,.60]$ which includes and is not included in $[.05 + .10, .10 + .30] = [.15,.40]$.

The following corollary is therefore of some intrinsic interest:

C-1 $\ulcorner\sim(t_1 \wedge t_2)\urcorner \in w \wedge w$ *is statistically closed* $\wedge\ Prob_L(t_1,w) = [p_1,q_1] \wedge Prob_L(t_2,w) = [p_2,q_2] \wedge \vdash\ulcorner p_1p_2 > 0\urcorner \supset Prob_L(\ulcorner t_1 \vee t_2\urcorner,w) \subset [p_1 + p_2, q_1 + q_2]$.

We are now ready for the final theorem: the coherence theorem. The theorem states that if w is statistically closed, and T is any finite set of sentences in L, then there exists a function B (one might call it a belief function) whose domain is T, whose maximum value is less than or equal to 1, whose minimum value is greater than or equal to 0, whose value for the disjunction of t_i and t_j is the sum of its values for t_i and t_j when the denial of their conjunction is in w, and which is such that $B(t_i) \in Prob_L(t_i,w)$. In short: there is a coherent belief function satisfying the constraints of probability.

T-9 $T = \{t_0,\ldots,t_{n-1}\} \wedge T \subset ST_L \wedge w$ *is statistically closed* $\supset \bigvee B(B$ *is fun* $\wedge\ \mathfrak{D}B = T \wedge \mathfrak{R}B \subset [0,1] \wedge \bigwedge i, j < n\ (\ulcorner t_i \vee t_j\urcorner \in T \wedge \ulcorner\sim(t_i \wedge t_j)\urcorner \in w \supset B(\ulcorner t_i \vee t_j\urcorner) = B(t_i) + B(t_j)) \wedge \bigwedge i(B(t_i) \in Prob_L(t_i,w)))$

Proof:

Let $S = \{s_1,\ldots,s_m\}$ be any smallest set of sentences of L such that every member of T is connected in w to some truth functional combination of elements of S by a biconditional chain. (Such a set always exists, and has at most n members.) Construct the algebra whose basis is S—i.e., whose atoms are sentences of the form $\ulcorner Ns_1 \wedge Ns_2 \wedge \ldots \wedge Ns_m\urcorner$, where each "$N$" represents an occurrence of "$\sim$" or nothing. We show that a function B satisfying the conditions of the theorem can be constructed for any such algebra.

There are 2^m atoms in the algebra. For every atom a_i, let $I_i = \langle x_i,y_i,z_i,b_i,s_i\rangle$ be an inference structure yielding the probability $[p_i,q_i]$ of that atom. As in the proof of T-8, we construct the cross product sequence y of the sequences y_i and a function z such that for every atom a_i there is a (2^m)-dimensional Borel set b_i such that $\ulcorner z(x) \in b_i \equiv a_i\urcorner \in w$, where $x = \langle x_1,x_2,\ldots,x_{2m}\rangle$. If d is a disjunction of atoms, then where b is the union of the corresponding b_i, $\ulcorner z(x) \in b \equiv d \urcorner \in w$. Every statement in T (except a contradiction, about which there is no problem) can be expressed as a disjunction of atoms. Finally, for every Borel set b, there is in w a strongest statistical statement concerning the frequency with which objects in y have z-values in b.

We consider two cases, corresponding to the two main cases considered in the proof of T-8.

Case I. For some disjunction d of one or more atoms, the probability of d, $Prob_L(d,w)$, based on the inference structure $I_d = \langle x_d, y_d, z_d, b_d, s_d \rangle$, is neither included in, nor includes, the interval mentioned in the corresponding statistical statement for the cross product reference set: $\ulcorner \%(y,z,b_d) \in [p_{d'}, q_{d'}] \urcorner$. Thus I_d disagrees with $\langle x,y,z,b_d, \ulcorner \%(y,z,b_d) \in [p_{d'}, q_{d'}] \urcorner \rangle$. From D-6, therefore, we know that there is an inference structure $I = \langle x', y', z', z', b_d', s_d' \rangle$ which is a reflection of I_d in the inference structure based on the cross product reference set. Furthermore, by D-5 we know that for every b_i and s_i, if $\langle x,y,z,b_i,s_i \rangle$ is an inference structure, there is a corresponding inference structure $\langle x', y', z', b', s' \rangle$ such that if $I'' = \langle x'', y'', z'', b'', s'' \rangle$ is an inference structure yielding the probability of $\ulcorner z'(x') \in b' \urcorner$ then s' is stronger than s'' or s' and s'' are equally strong. Finally, by the argument of T-8, the Borel sets b_i' corresponding to the original Borel sets b_i may be taken to be disjoint.

In virtue of statistical closure, which ensures statistical consistency, we may be sure that there exists a function B' defined on the disjoint Borel sets $b_i' (1 \leq i \leq 2^m)$ such that $B'(b_i') \in [p_i', q_i']$, where $\ulcorner \%(y', z', b_i') \in [p_i', q_i'] \urcorner$ represents our statistical knowledge about y', z', and b_i', and also such that $\Sigma_{1 \leq i \leq 2^m} B'(b_i) = 1$. Extend B' to unions of the b_i by taking $B'(b_i \cup b_j) = B'(b_i) + B'(b_j)$ for $i \neq j$, and in general $B'(c \cup d) = B'(c) + B'(d)$ for $c \cap d = \emptyset$. This function B' just represents a possible distribution of z'-values in y' that is consistent with our statistical knowledge of y'. For every union b_d of the b_i, therefore, $B'(b_d) \in [p,q]$, where $\ulcorner \%(y', z', b_d) \in [p,q] \urcorner$ represents our strongest knowledge in w about y', z', and b_d. Since $\ulcorner \%(y', z', b_d) \in [p,q] \urcorner$ is, if anything, stronger than the statistical statement on which the probability of the corresponding disjunction of statements d is based, $B'(b_d) \in Prob_L(d,w)$.

The function called for by the theorem is just $B(d) = B'(b_d)$, with its domain restricted to T.

Case II. For every disjunction d of one or more atoms, corresponding to the Borel set b_d, if the probability of d is based on the inference structure $I_d = \langle x_d, y_d, z_d, b_d^*, s_d \rangle$, then s_d does not differ from $\ulcorner \%(y,z,b_d) \in [p_d, q_d] \urcorner$. There is no disagreement that throws us into a reflection.

Two subcases may arise:

Subcase II A. For every disjunction d of atoms $\{a_i : i \in D\}$, $[\Sigma_{i \in D} p(a_i), \Sigma_{i \in D} q(a_i)] \subset Prob_L(d,w)$, where the probability of a_i is represented by

$[p(a_i),q(a_i)]$. In this case no problems arise and the function B as defined for case I will satisfy the theorem.

Subcase II B. For some disjunction d of atoms $\{a_i : i \in D\}$, $Prob_L(d,w) \subset [\Sigma_{i\in D}p(a_i),\Sigma_{i\in D}q(a_i)]$ and not $[\Sigma_{i\in D}p(a_i),\Sigma_{i\in D}q(a_i)] \subset Prob_L(d,w)$.

Let $\mathcal{B}$ be the set of functions B' such that for every i, $B'(b_i) \in [p(a_i),q(a_i)]$, and $B'(c \cup d) = B'(c) + B'(d)$, for $c \cap d = \emptyset$. Suppose there is no $B' \in \mathcal{B}$ that satisfies the condition of T-9 that for any $t \in T$, $B(t) \in Prob_L(t,w)$, where $B(t)$ is given by the value of B' for the corresponding Borel set. Since for every $B' \in \mathcal{B}$, $B'(a_i) \in Prob_L(a_i,w)$, the failure of T-9 has to concern disjunctions of the a_i.

Let $\{d_1,\ldots,d_k\}$ be a minimal set of disjunctions such that for no $B' \in \mathcal{B}$ will $B(d_i) \in Prob_L(d_i,w)$ hold for all i from 1 to k. Construct a new algebra, as before, where the basis is now $\{s_1,\ldots,s_m,d_1,\ldots,d_k\}$. Let the corresponding product sequence be y'', and construct z'' as before. Let the inference structures making use of y'' and z'' have the general form:

$$\langle x'',y'',z'',b,s\rangle$$

By construction, for each j from 1 to k, there is a Borel set b_j such that:

$$(1j) \qquad \langle x'',y'',z'',b_j,\ulcorner\%(y'',z'',b_j) \in [p(d_j),q(d_j)]\urcorner\rangle \in IS_{w,L}(d_j)$$

But also, where $\bigcup_{i\in D_j} b_i$ is the union of the Borel sets corresponding to the disjuncts a_i in d_j, we have

$$(2j) \quad \left\langle x'',y'',z'',\bigcup_{i\in D_j} b_i,\ulcorner\%(y'',z'',\bigcup_{i\in D_j} b_i) \in \left[\sum_{i\in D_j} p(a_i),\sum_{i\in D_j} q(a_i)\right]\urcorner\right\rangle \in IS_{w,L}(d_j)$$

We also have:

$$(3j) \quad \left\langle x'',y'',z'',b_j \cup \bigcup_{i\in D_j} b_i,\ulcorner\%(y'',z'',b_j \cup \bigcup_{i\in D_j} b_i) \in \left[p(d_j) + \sum_{i\in D_j} p(a_i),q(d_j) + \sum_{i\in D_j} q(a_i)\right]\urcorner\right\rangle \in IS_{w,L}(d_j)$$

If, for any j, $(1j)$ *DIS* $(3j)$, then we are forced to seek a reflection of $(1j)$, and an argument analogous to that of Case I will go through, showing that our supposition is false. But we can avoid disagreement only if all the p's—the lower probabilities—are 0, or if all the q's—the upper probabilities—are 1.

Suppose the p's are all 0. If it is possible for a measure B'' on the Borel sets to assign 0 to each b_j and each b_i involved, then it was already possible for B' to make that assignment, again contradicting the case II B hypothesis. If it is *not* possible for a measure B'' on the Borel sets to assign 0 to each b_j and each b_i involved, then statistical closure implies that some of the p's cannot be 0.

Similar considerations apply to the supposition that all of the relevant q's are 1.

In either case, we obtain the function called for by the theorem.

4. Illustrations and Comments. An illustration of the case in which the probability of an exclusive disjunction is properly included in the interval formed by the bounds on the probabilities of the disjuncts is easy to come by. Consider a die known to land 1, 2, 3, and 4, 1/6 of the time, but which is either biased to some degree in favor of 5 and the cost of a corresponding bias against 6, or vice versa. The probability of a 5 may well be (say) [.1,.2], and the probability of a 6 correspondingly [.1,.2], but the probability of the disjunction 5 or 6 will be exactly $[1/3,1/3] \subset [.2,.4]$.

A general illustration of the procedure outlined in theorem 8 may be of interest. Suppose we consider experiments consisting of tossing a bent coin (frequency of heads between .4 and .6), and throwing a biased die (frequency of 1 between .1 and .2). We consider a *specific* performance of this experiment. Assuming that the conditions of randomness are met, we can calculate that the probability of heads is [.4,.6]; the probability of 1 is [.1,.2]; and using the product class as reference class, the probability of heads and 1 is [.04,.12]—supposing that all the statistical knowledge we have is that stated.

Now suppose that we know of this trial that it is one that does not yield both heads and a 1. It might be thought that we would still have a probability of [.4,.6] for heads, and of [.1,.2] for a 1, and therefore of [.5,.8] for the disjunction of heads and 1. But this is not so, since the trial belongs to a special subset of the set of trials—those in which both heads and a 1 do not occur—in which the relative frequencies are

H and ~ 1: [.35,.57]:	this differs from [.4,.6] and dominates it, and yields the probability of heads.
$\sim H$ and 1: [.04,.13]:	this differs from [.1,.2], dominates it, and yields the probability of 1.
$\sim H$ and ~ 1: [.36,.56]:	this yields the probability of neither heads nor 1, and therefore the complementary interval yields the probability of the disjunction: [.44,.64].

Note that $[.44,.64] \subset [.35 + .04, .57 + .13] = [.39,.70]$ as required by T-8.

There are two projects which would contribute to the usefulness of this system. First, as observed in LFSI, we cannot allow any arbitrary terms to serve as reference terms, nor can we allow any arbitrary terms to serve as random variables. Both must be constrained. Certain closure conditions were suggested in LFSI, but it is not clear that these are the best. Furthermore, we must then consider the question of the initial stock of reference terms and of random quantities to which the closure conditions are to be applied. These initial stocks must be considered simply characteristic of the language. This should not be considered too much of a shortcoming, though. First, if we have a way of choosing reference classes for a particular piece of science, using the reference terms and random variables appropriate to that science, that is no small boon. Second, if (as I believe) we can develop criteria for choosing between scientific languages for a particular domain of inquiry, then in a given domain we will have determinate probabilities that are legislative for rational belief.

The second project, and a rather more important one, to my mind, is to develop an explicit algorithm for determining a reference class relative to a body of knowledge. The present system lends itself better to that than did the system of LFSI. It may be that both projects must go hand in hand; but in any event the development I have offered here provides a more convenient framework for pursuing those projects than did LFSI, as well as providing grounds for some unwarranted claims made there.

REFERENCES

Dempster, A. (1968), "A Generalization of Bayesian Inference", *Journal of the Royal Statistical Society,* series B, 30: 205–247.

Harper, W. (1982), "Kyburg on Direct Inference", in R. Bogdan (ed.), *Henry E. Kyburg, Jr. and Isaac Levi*. Dordrecht: D. Reidel.

Kyburg, H. (1974), *The Logical Foundations of Statistical Inference*. Dordrecht: D. Reidel.

Levi, I. (1981), "Direct Inference and Confirmational Conditionalization", *Philosophy of Science 48*: 532–552.

Seidenfeld, T. (1978), "Direct Inference and Inverse Inference", *The Journal of Philosophy 75*: 709–797.

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

Spielman, S. (1980), "Seidenfeld's Critique of Kyburgian Statistics", *The Journal of Philosophy 77*: 791–797.