ERNEST W. ADAMS AND HOWARD P. LEVINE

ON THE UNCERTAINTIES TRANSMITTED FROM PREMISES TO CONCLUSIONS IN DEDUCTIVE INFERENCES

1. The problem

It is evident that where the conclusion of a deductive inference *depends* on its premises, the conclusion will not be a logical certainty unless the premises themselves are logically certain, which is rare, and uncertainties in the premises will be in some measure passed on, or transmitted, to the conclusion. Moreover, the Lottery Paradox[1] shows that there are deductively sound inferences whose premises are individually very highly probable, while their conclusions are totally improbable. The aim of this paper is to initiate systematic inquiry into the circumstances in which this sort of phenomenon can occur, and more generally into the question as to *how high* a probability is guaranteed in the conclusion of a deductively sound inference, given plausible bounds on the uncertainties of the premises. This question is important to applied logical theory, because it is not normally the case that persons making inferences are satisfied merely to know that their conclusions are entailed by premises which they accept, but they also want some assurance that their conclusions are probable. The following section will present results bearing on maximum conclusion uncertainties compatible with given premise uncertainties in certain kinds of deductively sound inferences, and the final sections make some informal remarks on the methodological significance of probability considerations in deductive logic.

Before starting we state two elementary theorems of probability theory which throw some light on our problem, and which suggest the direction of the inquiry to follow. Define the *uncertainty* of a proposition to be the probability that it is false (this uncertainty is not to be confused with the entropic uncertainty measure of Information Theory). The first theorem states that the uncertainty of the conclusion of a deductively sound inference cannot exceed the *sum* of the uncertainties of the premises.[2] Hence, the dependence of a conclusion on *many* premises, as in the Lottery Paradox example, is essential if high probabilities of the individual pre-

Synthese **30** (1975) 429–460.

mises are to be compatible with zero probability of conclusions. Conversely, persons reasoning from only two, three or four premises, which are typical in textbook applications, will not be likely to arrive at highly improbable conclusions – though the premise uncertainty sum might be 'unacceptably high'.

The second theorem, a partial converse of the first, says that if all of the premises of a deductive inference are *essential* to the inference, then for given premise uncertainties it is logically possible either that the conclusion uncertainty should *equal* the sum of the premise uncertainties, or, if that sum is greater than 1, that the probability of the conclusion should be zero. A consequence of this theorem is that if persons are to guarantee better than total improbability in the conclusions they draw from premises of *a priori* uncertainty ε, they must either restrict themselves to inferences with $1/\varepsilon$ or fewer premises or else introduce some *redundancy* into their reasoning, so that their conclusions do not depend on all of their premises. In what follows we will be largely concerned with the latter possibility, since it appears that a certain amount of redundancy is characteristic of real life reasoning from many premises, where uncertainties cannot be neglected.

Consider the following example. A telephone survey is made of the political party affiliations of 1,000 people, which results in the collection of 'data' of the form 'person 1 is a Democrat', 'person 2 is a Republican', and so on up to person 1,000. 629 of the persons interviewed report themselves to be Democrats: i.e., exactly 629 'data items' are of the form 'person i is a Democrat'. It is intuitively evident, however, that given the uncertainties typical of data collected in this way, it would be unsafe to conclude 'exactly 629 of the persons interviewed are Democrats', even though this would be a deductive consequence of the data collected, each item of which would be sufficiently probable to be accepted by itself. On the other hand, it would seem *prima facie* reasonable to conclude 'at least 600 of the persons interviewed are Democrats'. This conclusion would not depend deductively on all of the premises, or even on any one of them, and because of this 'redundancy' in the data, our second theorem does not apply. We will see in the next section that in fact the *prima facie* reasonableness of this and similar conclusions from redundant premises is partially justified theoretically.

The problem of calculating the probability that at least 600 out of

1,000 people are Democrats, with 'data' like that just described with given individual datum 'error probabilities', looks at first sight like a very elementary one of probability theory. However, our approach to it will differ from the usual approach in statistics, in that we will not assume *a priori* that errors are independent (e.g., that the chance that both of two reports are incorrect is equal to the product of their individual uncertainties). Independence assumptions are plausibly regarded as *inductive* in character and we want to ascertain how high a conclusion uncertainty is *logically* compatible with given premise uncertainties (we will assume in fact that any probability function satisfying the Kolmogorov Axioms is a logically possible one). Comparing the *logical* bounds on conclusion uncertainties with those which follow when independence assumptions are invoked will in fact afford us some indication of the degree to which certain kinds of deductively sound inferences in reality rest on inductive assumptions for their justification. We will even find some which are deductively sound, are totally unjustified when logically possible conclusion uncertainties are considered, but which are inductively justified by independence assumptions! Such inferences are plausibly termed *deceptively deductive.*

Our formulation of the problem of determining logically maximum conclusion uncertainties compatible with given premise uncertainties reduces it to one of maximizing a linear function representing the conclusion uncertainty, subject to linear constraints representing *a priori* bounds on the premise uncertainties. This is simply a problem of *linear programming.* Most of the results we shall state are in fact fairly straightforward applications of basic theorems of that theory (principally the so-called *duality theorems*), and for this reason we shall state them rather informally, and refer to the relevant literature for the proofs. Our primary concern will be with the usefulness and significance of these results in application, and not with the development of yet another unwanted 'system' of non-standard logic.

A significant limitation on the present study must be noted. This is that it is restricted to inferences involving only what might be called 'factual' propositions, to which probabilities satisfying the Kolmogorov Axioms properly apply. In particular, we shall not consider inferences involving *conditional* propositions, whose probabilities are plausibly measured as conditional probabilities.[3] It turns out that when conditionals are introduced it can happen that conclusion probabilities are bounded by *products* *of premise probabilities* rather than by *sums of premise uncertainties*, and therefore conclusion uncertainties can sometimes be greatly reduced in the conditional case. However, our investigation of this case is still at an early stage, and we hope to present results on it in a later paper.

2. Uncertainty Maxima

An *inference* will be a system $I = \langle\langle\phi_1, \ldots, \phi_n\rangle, \psi\rangle$, where $\phi_1, \ldots, \phi_n$ and ψ are sentences of an unspecified first-order language; $\phi_1, \ldots, \phi_n$ are the *premises* of I, $\{\phi_1, \ldots, \phi_n\}$ is its *total premise set* (subsets of which are simply *premise sets*), ψ is its *conclusion*, and $\{\phi_1, \ldots, \phi_n, -\psi\}$ is its *refutation set*. Occasionally we will use the alternative notation $(\phi_1, \ldots, \phi_n)\, I(\psi)$ for I. The usual concepts of logical consequence, contradiction, and so on are presupposed (when explicitly specified the language may have consistent axioms which are treated as logical truths), as well as that of a probability function for the language. It is not presupposed that the inferences we deal with are deductively sound, or that their total premise sets are consistent – though inconsistent premise sets introduce some surprising uncertainty phenomena whose interpretations involve problems, and which will not be entered into in detail.

Arbitrary first-order inferences are trivially reducible to sentential inferences for the purpose of uncertainty maximum determination. Two inferences $(\phi_1, \ldots, \phi_n)\, I(\psi)$ and $(\phi'_1, \ldots, \phi'_n)\, I(\psi')$ are *equivalent with respect to uncertainty maximization* if any given subset of the first refutation set $\{\phi_1, \ldots, \phi_n, -\psi\}$ is consistent if and only if the corresponding subset of $\{\phi'_1, \ldots, \phi'_n, -\psi'\}$ is consistent. It is easy to show that two inferences which are equivalent in this sense have the same conclusion uncertainty maxima. Obviously any first-order inference is equivalent to a sentential inference in the defined sense. We will show next that the foregoing reduction can be carried considerably farther.

Fixing attention on the inference $(\phi_1, \ldots, \phi_n)\, I(\psi)$, two kinds of premise sets will prove important where the total premise set $P = \{\phi_1, \ldots, \phi_n\}$ is consistent, and a third kind must be considered when it is inconsistent. A premise set $P' \subseteq P$ is *sufficient for I* if ψ is a logical consequence of P', and is *essential for I* if ψ is not a logical consequence of $P \sim P'$. Sufficient and essential premise sets are dual in certain respects. Every sufficient premise set intersects (has a non-empty intersection with) every essential one, and the sets of either kind are definable as just those which intersect all sets of the other. The minimal sets of each type (i.e., the sets of that type which have no proper subsets of that type) will prove especially important. The minimal sufficient premise sets for I (m.s.s. for I) will be supposed to be the sets $S_1, \ldots, S_r$ (on occasion r may be 0), and the minimal essential premise sets (m.e.s. for I) will be supposed to be $E_1, \ldots, E_s$ (s may be 0). The size of the smallest m.e. to which a given premise belongs will be significant in that it gives a rough index of the 'weight' of that premise so far as its uncertainty contributes to the maximum uncertainty of the conclusion. *Totally inessential*, of *irrelevant* premises are ones which belong to no m.e.s., and their uncertainties make no contribution to the conclusion's uncertainty.

Continuing with the foregoing inference, $(\phi_1, \ldots, \phi_n)\ I(\psi)$, its associated minimal sufficient and minimal essential *forms* are the sentences $\mathrm{ms}(I)$ and $\mathrm{me}(I)$ defined as follows. Letting $\bigwedge S_j$ be the conjunction of the premises in S_j, or a tautology T if S_j is empty, $\mathrm{ms}(I)$ is the disjunction $\bigwedge S_1 \vee \cdots \vee \bigwedge S_r$ or else is an arbitrary contradiction F if there are no m.s.s. for the inference. Letting $\bigvee E_j$ be the disjunction of the premises of E_j, or F if E_j is empty, $\mathrm{me}(I)$ is the conjunction $\bigvee E_1\ \&\ \cdots\ \&\ \bigvee E_s$, or is T if there are no m.e.s. The two sentences $\mathrm{ms}(I)$ and $\mathrm{me}(I)$ are easily seen to be logically equivalent, but more importantly for our purposes the original conclusion, ψ, can be replaced by either $\mathrm{ms}(I)$ or $\mathrm{me}(I)$, and the resulting inference will be equivalent to the original with respect to uncertainty maximization. Note that we have now reduced the uncertainty maximization problem for an inference with an arbitrary conclusion ψ to that of maximizing the uncertainty of another conclusion $\mathrm{ms}(I)$ or $\mathrm{me}(I)$ which will logically imply ψ but will not in general be logically implied by ψ, and which is of the form of a disjunction of conjunctions of premises or of a conjunction of disjunctions of premises. The reduction can be carried one step farther. If the premises are consistent then each premise can be replaced by a distinct atomic letter with the same replacements being made in $\mathrm{ms}(I)$ or $\mathrm{me}(I)$, and the new inference will be equivalent to the original with respect to uncertainty maximization. The same reduction can also be carried out when the premises are inconsistent, except that in this case it is necessary to add non-logical axioms to the language which specify in effect that sets of atomic formulas which correspond to inconsistent premise sets are inconsistent. It is significant that when this reduction is carried out negation diappears entirely, since premises are atomic formulas and conclusions are conjunctions of disjunctions or disjunctions of conjunctions of premises.

Some special cases should be noted. If ψ is a logical consequence at most of inconsistent premise sets then $\mathrm{ms}(I)$ and $\mathrm{me}(I)$ are both equivalent to a contradiction, F, and in this case any premise probabilities are compatible with ψ's having probability 0. If the conclusion is a logical truth (it is independent of the premises), there are no m.e.s., the only m.s. is the empty set, and both $\mathrm{ms}(I)$ and $\mathrm{me}(I)$ are equivalent to a logical truth. In the case in which the total premise set is sufficient but no proper subset of it is, every individual premise is essential (i.e., the singleton set containing it is essential), and $\mathrm{ms}(I)$ and $\mathrm{me}(I)$ are both equivalent to the conjunction of all of the premises. This is the sound, consistent, *irredundant* inference. It follows easily from the foregoing that any two sound, consistent, irredundant inferences with the same number of premises are equivalent with the respect to uncertainty maximization, no matter how 'strong' or 'weak' their conclusions are. The foregoing generalizes a bit: if all of the premises of an inference are either essential or irrelevant then the conclusion can be replaced by the conjunction of the essential premises and the resulting inference will be equivalent to the original with respect to uncertainty maximization.

Now suppose the total premise set P of an inference I is inconsistent. A subset $P' \subseteq P$ will be called *negatively sufficient for I* if $P \sim P'$ is consistent and sufficient for I. The *minimal* negatively sufficient premises sets for I (m.n.s.s. for I) will be written $NS_1, \ldots, NS_t$. In the special case in which the conclusion of I is entailed at most by inconsistent premise sets there are no m.n.s.s. for I ($t=0$), and in the case in which the total premise set is sufficient for I and consistent, the only m.n.s. for I is the empty premise set.

The m.e.s. and m.n.s.s. for an inference $(\phi_1, \ldots, \phi_n)\, I(\psi)$ are closely related to the minimal subsets of the refutation set $R = \{\phi_1, \ldots, \phi_n, -\psi\}$ which can be falsified in any state of affairs. These minimal falsifiable subsets of R are the complements with respects to R of the maximal consistent subsets of R. If R' is such a subset, it will be an m.e. for I if it does not contain $-\psi$, and if R' does contain $-\psi$ then $R' \sim \{-\psi\}$ is an m.n.s. for I. In any case the m.e.s. and m.n.s.s. for I represent 'minimal falsification states' relative to R, and it proves useful to represent these states in a *minimal falsification matrix* as follows. The rows in the matrix correspond to the m.e.s. and the m.n.s.s. (with the m.e.s. coming first), and the columns correspond to the premises of I, with a final column for the conclusion. '1's and '0's are now written into the cells of the matrix according to the rules: (1) where the cell is the intersection of a premise set row and premise column, a '1' is written in if the premise belongs to the set, and a '0' is written in otherwise, (2) '1's are written into the conclusion column in the m.e. rows, (3) '0's are written into the conclusion column in m.n.s. rows. A simple illustration is the inference of the conclusion '$A \leftrightarrow -B$' from the three premises 'A', 'B', and '$-(A \& B)$' (the premises are both redundant and inconsistent). In this example there are two minimal sufficient premise sets, $S_1 = \{A, -(A \& B)\}$ and $S_2 = \{B, -(A \& B)\}$, two minimal essential premise sets $E_1 = \{A, B\}$ and $E_2 = \{-(A \& B)\}$, and two negatively sufficient premise sets, $NS_1 = \{A\}$, and $NS_2 = \{B\}$. Therefore the minimal falsification matrix is:

	premises			conclusion
	A	B	$-(A \& B)$	$(A \leftrightarrow -B)$
$E_1 = \{A, B\}$	1	1	0	1
$E_2 = \{-(A \& B)\}$	0	0	1	1
$NS_1 = \{A\}$	1	0	0	0
$NS_2 = \{B\}$	0	1	0	0

It is also convenient to regard the entries in the rows of the minimal falsification matrix as the values of *minimal falsification functions* corresponding to these rows, the function corresponding to the row of an m.e., E_j, being symbolized e_j, and the function corresponding to the row of the m.n.s. NS_j being written ns_j. These functions, mapping the premises of the inference plus its conclusion into $\{0,1\}$, are important because they are in fact restrictions of 'extreme' uncertainty functions just to this set of sentences. Uncertainty functions are defined such that $u(\eta) = 1 - p(\eta)$ for some probability function p and for all sentences η of the language. It is trivial that the only uncertainty functions which we need to consider for our sentential languages are generated as convex combinations of *falsity functions* for the language, which are functions f such that for some model, $f(\eta) = 1$ for all sentences η which are false in the model, and $f(\eta) = 0$ for all sentences which are true in the model. In maximizing conclusion uncertainties it is sufficient to consider convex combinations of functions which falsify minimum subsets of the refutation set for the inference, the restrictions of which to the premises and conclusion of the inference are just the minimal falsification functions $e_1 \ldots, e_s$ and $ns_1, \ldots, ns_t$. Accordingly we can restrict attention to uncertainty functions which can be expressed in the form:

$$(1) \qquad u(\eta) = \sum_{j=1}^{s} p_j e_j(\eta) + \sum_{j=1}^{t} q_j ns_j(\eta),$$

where $p_1, \ldots, p_s$ and $q_1, \ldots, q_t$ are non-negative reals summing to 1, and η is either a premise or the conclusion of the inference involved.

The 'weights' $p_1, \ldots, p_s$ and $q_1, \ldots, q_t$ in Equation (1) are those attaching to the corresponding minimal falsification functions $e_1, \ldots, e_s$ and $ns_1, \ldots, ns_t$, and indirectly to the sets $E_1, \ldots, E_s$ and $NS_1, \ldots, \mathrm{NS}_t$. Thus, the weights $p_1, \ldots, p_s$ will be called the *m.e. weights* which together with the *m.n.s. weights* $q_1, \ldots, q_t$, *generate* the function u defined by Equation (1). Note that the uncertainty of the conclusion ψ given by the uncertainty function defined in this way is just the sum of the m.e. weights, $p_1, \ldots, p_s$, and the uncertainty of any premise is equal to the sum of the weights of the m.e.s. and m.n.s.s. to which it belongs.

Bounds on the uncertainties of the premises in the inference $I = \langle\langle \phi_1, \ldots, \phi_n \rangle, \psi\rangle$ will be represented by non-negative real vectors $\boldsymbol{\varepsilon} = \langle \varepsilon_1, \ldots, \varepsilon_n \rangle$, which will be called *premise uncertainty bound vectors for* I (p.u.b.v.s. for I) in the *space* of all such vectors (the p.u.b.v. space of I). An uncertainty function, u, for the language is *consistent* with $\boldsymbol{\varepsilon}$ if $u(\phi_i) \leqslant \varepsilon_i$ for $i = 1, \ldots, n$, and $\boldsymbol{\varepsilon}$ is said to be *consistent for* I if there exists an uncertainty function which is consistent with it, and otherwise *inconsistent* for I. If the total premise set is consistent then all p.u.b.v.s. are consistent for I, but if the total premise set is inconsistent then 'sufficiently small' p.u.b.v.s. will be inconsistent. *Uniform* p.u.b.v.s. are those whose components are all equal.

If $\boldsymbol{\varepsilon} = \langle \varepsilon_1, \ldots, \varepsilon_n \rangle$ is a p.u.b.v. which is consistent for the inference $(\phi_1, \ldots, \phi_n)\, I(\psi)$, then there is obviously a maximum value of $u(\psi)$ for uncertainty functions u which are consistent with $\boldsymbol{\varepsilon}$, and accordingly we define

$$\mu(I; \boldsymbol{\varepsilon}) = \mu(I; \varepsilon_1, \ldots, \varepsilon_n)$$

to be equal to that maximum (and we define $\mu(I; \varepsilon)$ to be the same as $\mu(I: \boldsymbol{\varepsilon})$ for p.u.b.v.s. with uniform components ε). For fixed I, $\mu(I; \boldsymbol{\varepsilon})$ and $\mu(I; \varepsilon)$ are clearly real valued functions, which will be called the *conclusion uncertainty maximization function* (c.u.m.f.) and the *uniform conclusion uncertainty maximization function* (u.c.u.m.f.), respectively, for I, which are defined for all consistent p.u.b.v.s. These are the functions whose properties will concern us in what follows. Obviously both functions are monotone increasing (though not strictly monotone increasing) in all of their arguments over their domains of definition, and in the case in which the premises are consistent and entail a non-logically true conclusion, $\mu(I; 0)=0$ and $\mu(I; 1)=1$ (logically implied consequences of perfectly certain premises are perfectly certain and consequences *depending* on 'perfectly uncertain' premises are perfectly uncertain). To investigate these functions in detail, it proves helpful to utilize results from the theory of Linear Programming (see especially Goldman and Tucker, 1956).

Restricting ourselves to uncertainty functions of the form (1), our problem becomes that of maximizing the conclusion uncertainty

$$(2)\qquad u(\psi) = p_1 + \cdots + p_s$$

(if $s=0$, $u(\psi)$ must equal 0) subject to the 'primary constraints':

$$(3)\qquad u(\phi_i) = \sum_{j=1}^{s} p_j e_j(\phi_i) + \sum_{j=1}^{t} q_j ns_j(\phi_i) \leqslant \varepsilon_i,$$

for $i=1,\ldots,n$, where m.e. weights $p_1,\ldots,p_s$ and m.n.s. weights $q_1,\ldots,q_t$ are non-negative and sum to 1. The dual *minimization* problem is that of minimizing the linear form

$$(4)\qquad \boldsymbol{w}\cdot\boldsymbol{\varepsilon} + v = w_1\varepsilon_1 + \cdots + w_n\varepsilon_n + v$$

subjects to the 'dual constraints'

$$(5.1)\qquad \sum_{i=1}^{n} w_i e_j(\phi_i) + v \geqslant 1 \quad \text{for} \quad j = 1,\ldots,s,$$

and

$$(5.2)\qquad \sum_{j=1}^{n} w_i ns_j(\phi_i) + v \geqslant 0 \quad \text{for} \quad j = 1,\ldots,t,$$

where $w_1,\ldots,w_n$ are non-negative reals, and v is an arbitrary real.

Observe that the 'weights', $w_1, \ldots, w_n$ in the *dual weighting system* $\langle w, v\rangle = \langle\langle w_1, \ldots, w_n\rangle, v\rangle$ entering into (4), (5.1), and (5.2) above correspond to the individual premises $\phi_1, \ldots, \phi_n$, and for this reason we will call each w_i the ith *premise weight* of the dual weighting system. For reasons which will become apparent later, it is appropriate to call the parameter v the *consistency index* of the dual weighting system. If the system $\langle \boldsymbol{w}, v\rangle$ satisfies the dual constraints (5.1) and (5.2) we will say that it is *consistent* with them. These inequalities can be restated in words as follows: the necessary and sufficient condition for $\langle \boldsymbol{w}, v\rangle$ to be consistent with the dual constraints is that the *sum of the weights of the premises in each minimal essential set plus the consistency index must be at least* 1, *and the sum of the weights of the premises in each minimal negatively sufficient set plus the consistency index must be at least* 0. Whether or not the dual weighting system is consistent with the dual constraints, it generates a linear functional $\boldsymbol{w}\cdot\boldsymbol{\varepsilon}+v$ according to (4) over the space of p.u.b.v.s. $\boldsymbol{\varepsilon}$, which will be called the *conclusion uncertainty bound function* (c.u.b.f.) *generated by* $\langle \boldsymbol{w}, v\rangle$.

The essential facts about the maximization and dual minimization problems and their interconnections are as follows. The dual system of constraints (5) is always consistent, since the dual weighting system $\langle\langle 0, \ldots, 0\rangle, 1\rangle$ is always consistent with it, but for a fixed p.u.b.v. $\boldsymbol{\varepsilon}$ there will be a *minimum* of form (4) if and only if the primary constraints (3) are consistent. If the uncertainty function u is consistent with the primary constraints, and the dual weighting system $\langle \boldsymbol{w}, v\rangle$ is consistent with the dual constraints, then

$$(6) \qquad u(\psi) \leqslant \mu(I; \boldsymbol{\varepsilon}) \leqslant \boldsymbol{w}\cdot\boldsymbol{\varepsilon} + v\,.$$

Thus, the conclusion uncertainty bound functions $\boldsymbol{w}\cdot\boldsymbol{\varepsilon}+v$ do in fact give upper bounds to conclusion uncertainties. If the primary constraints are consistent, then there exist non-negative $p_1, \ldots, p_s$ and $q_1, \ldots, q_t$ summing to 1 consistent with these constraints and generating an uncertainty function u, and there exists a dual weighting system $\langle \boldsymbol{w}, v\rangle$ consistent with the dual constraints such that $u(\phi) = \boldsymbol{w}\cdot\boldsymbol{\varepsilon}+v$, and therefore according to (6), $\mu(I; \boldsymbol{\varepsilon})$ is equal to this value. Thus, our attention turns to minimum c.u.b.f.s., since their values also give conclusion uncertainty maxima, and more generally to c.u.b.f.s. which are 'informative' about uncertainty maxima.

To conclude our summary of the relevant linear programming results, it is to be noted that the class of all dual weighting systems $\langle \boldsymbol{w}, v\rangle$ consistent with the dual constraints (5) forms a convex set in $n+1$ dimensional vector space, and this convex set has only a finite number of *extreme points* (consistent dual weighting systems which are not convex combinations of other consistent dual weighting systems), which we may designate $\langle \boldsymbol{w}^1, v^1\rangle, \ldots, \langle \boldsymbol{w}^m, v^m\rangle$. These will be called the *characteristic* dual weighting systems for the inference I, since if there exists a minimum value of $\boldsymbol{w}\cdot\boldsymbol{\varepsilon}+v$ for arbitrary $\langle \boldsymbol{w}, v\rangle$ consistent with the dual constraints, this value must be assumed at one of the extreme points, and therefore we can write:

$$(7)\qquad \mu(I; \boldsymbol{\varepsilon}) = \min(\boldsymbol{w}^1\cdot\boldsymbol{\varepsilon} + v^1, \ldots, \boldsymbol{w}^m\cdot\boldsymbol{\varepsilon} + v^m).$$

The values $\boldsymbol{w}^k\cdot\boldsymbol{\varepsilon}+v^k$ for $k=1,\ldots,m$, therefore determine $\mu(I; \boldsymbol{\varepsilon})$ in the range of its definition. In certain cases these values also give information about this range, since the fact that $\boldsymbol{w}^k\cdot\boldsymbol{\varepsilon}+v^k$ is actually *negative* means that $\boldsymbol{\varepsilon}$ is not a consistent p.u.b.v. for I. However, the fact that all of the values $\boldsymbol{w}^k\cdot\boldsymbol{\varepsilon}+v^k$, $k=1,\ldots,m$, are non-negative is not always a sufficient condition for $\boldsymbol{\varepsilon}$ to be consistent for I.

More terminology. If $\langle \boldsymbol{w}^k, v^k\rangle$, $k=1,\ldots,m$, are the characteristic dual weighting systems for an inference I, then the functions $\boldsymbol{w}^k\cdot\boldsymbol{\varepsilon}+v^k$ which are defined over the p.u.b.v. space for the inference will be called the *characteristic functions* for the inference. For a particular $\boldsymbol{\varepsilon}$, the characteristic function or functions which minimize $\boldsymbol{w}^k\cdot\boldsymbol{\varepsilon}+v^k$ will be called the *applicable* functions for $\boldsymbol{\varepsilon}$, and the associated weighting system or systems $\langle \boldsymbol{w}^k, v^k\rangle$ will be the applicable weighting system of systems for $\boldsymbol{\varepsilon}$. The class of all $\boldsymbol{\varepsilon}$ for which a particular characteristic function is applicable is always a convex (possibly empty) set of p.u.b.v.s. which will be called the *applicable domain* of the function (and of its associated weighting system), and the restriction of the function to its applicable domain will be called the applicable *part* of the function.

Some general properties of $\mu(I; \boldsymbol{\varepsilon})$ follow immediately. This function must be continuous and piece-wise linear over its domain of definition. Piece-wise linearity follows from the fact that $\mu(I; \boldsymbol{\varepsilon})$ is the finite union of the applicable parts of the characteristic functions for I, restricted to the set of consistent p.u.b.v.s., where each characteristic function is itself linear. Now consider a fixed p.u.b.v., $\boldsymbol{\varepsilon}=\langle\varepsilon_1,\ldots,\varepsilon_n\rangle$ and a particular premise ϕ_i. The *applicable* weight or weights of ϕ_i for $\boldsymbol{\varepsilon}$ will be just the ith components w^k of the applicable dual weighting system or systems for $\boldsymbol{\varepsilon}$. These weights have the following significance. If the uncertainty bound ε_i is *increased* by a small amount δ, all other uncertainty bounds remaining the same, then the conclusion uncertainty maximum $\mu(I;\varepsilon)$ will increase by the amount $w_i^k\delta$, where w_i^k is the *smallest* of the applicable ith premise weights for $\boldsymbol{\varepsilon}$. If the uncertainty bound ε_i is *decreased* by a small amount δ then, provided the resulting uncertainty bounds are consistent, the maximum conclusion uncertainty will decrease by the amount $w_i^k\delta$, where now w_i^k is the *largest* of the applicable ith premise weights. Thus, the applicable ith premise weights afford a 'locally applicable index' of the importance of ϕ_i so far as its uncertainty influences the maximum uncertainty of the conclusion. That $\mu(I;\boldsymbol{\varepsilon})$ increases in proportion to smallest applicable premise weights but decreases in proportion to largest applicable premise weights means, roughly, that the more uncertain a premise is, the less will changes in its uncertainty affect the maximum uncertainty of the conclusion. More roughly still, the more probable premises will be the ones whose uncertainties most importantly affect the maximum conclusion uncertainty.

All of the foregoing applies *mutatis mutandis* to uniform p.u.b.v.s. and to the values of the uniform c.u.m.f. $\mu(I;\varepsilon)$. Here all that matters are *sums* of premise weights and consistency indices in characteristic dual weighting systems, since it follows trivially from (7) that in the uniform bound case,

$$(8)\qquad \mu(I;\varepsilon)=\min(s(w^1)\,\varepsilon+v^1,\ldots,s(w^m)\,\varepsilon+v^m)$$

where $s(w^k)$ is the sum of the premise weights in the characteristic system $\langle w^k, v^k\rangle$. Those systems $\langle w^k, v^k\rangle$ which minimize (8) for particular ε may be termed *applicable* for ε, and so on. Trivially, $\mu(I;\varepsilon)$ will be continuous, piece-wise linear, and increase with smallest applicable premise-weight sums and decrease with largest applicable premise weight sums. In the case in which the premises are consistent and entail the conclusion it will be seen that the consistency index is either 1 or 0, and

$$(9)\qquad \mu(I;\varepsilon)=\min(s(w^k)\,\varepsilon, 1)$$

where $\langle w^k, v^k\rangle$ is the characteristic system with consistency index 0 and smallest premise weight sum. Note also that except in the case in which the conclusion is a logical truth, all premise weight sums must be greater than or equal to 1, and therefore $\mu(I;\varepsilon)$ cannot be less than ε unless the conclusion is a logical truth.

Most of the results to follow concern the consistent and sufficient premise case, and these are easy generalizations from two simple examples. The first is the inference with premises 'A', 'B', and 'C', and conclusion '$A \,\&\, (B \vee C)$'. This conclusion is already in minimal essential form, the two minimal essential premise sets being $E_1 = \{A\}$, and $E_2 = \{B, C\}$. As in all consistent sufficient premise cases, the only minimal negatively sufficient premise set is the empty set: i.e., $NS_1 = \wedge$. The minimal falsification matrix therefore is:

	premises			conclusion
	A	B	C	$A \& (B \vee C)$
$E_1 = \{A\}$	1	0	0	1
$E_2 = \{B, C\}$	0	1	1	1
$NS_1 = \wedge$	0	0	0	0

The primary uncertainty maximization problem is that of maximizing the uncertainty $u(A \,\&\, (B \vee C)) = p_1 + p_2$ subject to the primary constraints $u(A) = p_1 \leqslant \varepsilon_1$, $u(B) = p_2 \leqslant \varepsilon_2$, and $u(C) = p_2 \leqslant \varepsilon_3$. This problem is trivially solvable without going over to the dual system, but consideration of the dual problem is still illuminating.

The dual problem in the example is that of minimizing the linear form

$$w \cdot \boldsymbol{\varepsilon} + v = w_1\varepsilon_1 + w_2\varepsilon_2 + w_3\varepsilon_3 + v$$

subject to the dual constraints

$$w_1 \cdot 1 + w_2 \cdot 0 + w_3 \cdot 0 + v \geqslant 1$$
$$w_1 \cdot 0 + w_2 \cdot 1 + w_3 \cdot 1 + v \geqslant 1$$
$$w_1 \cdot 0 + w_2 \cdot 0 + w_3 \cdot 0 + v \geqslant 0 .$$

The correspondence between the above inequalities and the rows in the minimal falsification matrix is obvious. The extreme weighting systems consistent with the dual constraints, which are obtained by 'maximizing equated constraints' (turning as many inequalities as possible into equalities, where the *a priori* inequalities $w_1 \geqslant 0$, $w_2 \geqslant 0$ and $w_3 \geqslant 0$ are included), are:

$$\langle w^1, v^1\rangle = \langle\langle 0,0,0\rangle, 1\rangle$$
$$\langle w^2, v^2\rangle = \langle\langle 1,1,0\rangle, 0\rangle$$
$$\langle w^3, v^3\rangle = \langle\langle 1,0,1\rangle, 0\rangle .$$

In this case, therefore, Equation (7) reduces to:

$$\mu(I; \varepsilon_1, \varepsilon_2, \varepsilon_3) = \min(1, \varepsilon_1 + \varepsilon_2, \varepsilon_1 + \varepsilon_3) .$$

Among the things which generalize easily from the example, the following may be noted. The *total uncertainty weighting system*, $\langle\langle 0,\ldots,0\rangle, 1\rangle$, with all premise weights 0 and consistency index 1, is always among the characteristic weighting systems, and the associated characteristic function has the constant value 1. For each minimal sufficient premise set, S_j, there is a corresponding *unitary* characteristic system $\langle w^k, 0\rangle$, where the components w_i^k are 0's or 1's and are equal to 1 for those i such that ϕ_i belongs to S_j. In the example there were two minimal sufficient premise sets, $S_1 = \{A, B\}$ and $S_2 = \{A, C\}$, and the corresponding unitary characteristic systems were $\langle w^2, v^2\rangle = \langle\langle 1,1,0\rangle, 0\rangle$ and $\langle w^3, v^3\rangle = \langle\langle 1,0,1\rangle, 0\rangle$. The characteristic functions which correspond to these minimal sufficient premise sets are simply the sums of the uncertainty bounds of the premises in the sets to which they correspond. In the present case the only characteristic functions were the total uncertainty function and the unitary functions corresponding to minimal essential premise sets. In cases of this kind the maximum conclusion uncertainty either equals 1 (total uncertainty) or else equals the sum of the premise uncertainty bounds in the 'least uncertain' (least premise uncertainty bound sum) of its sufficient premise sets. Such inferences act as though their conclusions depend solely on their least uncertain sufficient premise sets.

In the consistent, sufficient, irredundant case, the *only* sufficient premise set is the total premise set, and in this case it is evident that the maximum conclusion uncertainty is either 1 or else equals the sum of all of the premise uncertainty bounds, whichever is least. This combines the two 'theorems' of elementary probability stated in the introduction.

The only minimal negatively sufficient set in the consistent and sufficient premise case being the empty premise set, it follows trivially that the consistency index must always be non-negative, and in fact will equal 0 in all but the total uncertainty weighting system. In effect, then, we can eliminate the consistency index from our considerations so long as the premises are consistent, and concentrate on the 'reduced' dual constraints which result from the original constraints when v is set equal to 0. The extreme solutions to the reduced constraints then give the premise weights in all characteristic systems but the total uncertainty system, and we get the complete set of characteristic weighting systems by simply adding the total uncertainty system to the extreme solutions to the reduced constraints. When we come to inconsistent premise sets it will be seen that consistency indices play a more important role.

The final generalization illustrated in our first example has to do with the fact that the conclusion of the inference was a conjunction, each of whose conjuncts could be looked on as the conclusion of a 'sub-inference'. That is, our original inference had the form $(A, B, C)\ I(A\ \&\ (B \vee C))$, and the two sub-inferences can be represented as $(A, B, C)\ I(A)$ and $(A, B, C)\ I(B \vee C)$. In this particular case, the subinferences are *separate* in the following sense: each premise is relevant to at most one of the subinferences, and is irrelevant (totally inessential) to the other. Where an inference with consistent premises and a conjunctive conclusion can be separated in this way, the non-total uncertainty characteristic functions of the compound inference will always be *sums* of the non-total uncertainty characteristic functions of the sub-inferences. It is trivial that the only non-total uncertainty characteristic function for $I(A)$ is ε_1 (the maximum uncertainty of its conclusion, 'A', is equal to the that of its first premise), and the only non-total uncertainty characteristic functions for $I(B \vee C)$ are ε_2 and ε_3. In virtue of the fact that the sub-inferences are separable, it follows that the nontotal uncertainty characteristic functions of $I(A\ \&\ (B \vee C))$ must be $\varepsilon_1 + \varepsilon_2$ and $\varepsilon_1 + \varepsilon_3$.[4]

So far we have only encountered premise weights of 0 and 1, and this is essential because we have arrived at characteristic functions which are constructible from 'unitary' functions by just two operations: minimization, which corresponds roughly to disjunction, and addition, which corresponds roughly to conjunction. We will now see that there is another possibility: redundant but not irrelevant premises can lend 'statistical support' in limiting the maximum uncertainty of conclusions. As our example we will take the inference with premises 'A', 'B', and 'C', as before, but with the conclusion ψ='$(A \& B) \vee (A \& C) \vee (B \& C)$'. The conclusion expresses the fact that at least two of the premises are true, and there is obvious redundancy among the premises – in fact no single premise is essential. Here there are three minimal essential premise sets, $E_1 = \{A, B\}$, $E_2 = \{A, C\}$, and $E_3 = \{B, C)$. The minimal falsification matrix is:

	premises			conclusion
	A	B	C	ψ
$E_1 = \{A, B\}$	1	1	0	1
$E_2 = \{A, C\}$	1	0	1	1
$E_3 = \{B, C\}$	0	1	1	1
$NS_1 = \Lambda$	0	0	0	0.

The 'reduced' system of dual constraints (arrived at by setting $v=0$) is:

$$w_1 \cdot + w_2 \cdot 1 + w_3 \cdot 0 \geqslant 1$$
$$w_1 \cdot 1 + w_2 \cdot 1 + w_3 \cdot 0 \geqslant 1$$
$$w_1 \cdot 0 + w_2 \cdot 1 + w_3 \cdot 1 \geqslant 1 .$$

The extreme solutions to these inequalities are:

$$w^1 = \langle 1, 1, 0 \rangle$$
$$w^2 = \langle 1, 0, 1 \rangle$$
$$w^3 = \langle 0, 1, 1 \rangle$$
$$w^4 = \langle \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2} \rangle .$$

It follows that

$$\mu(I; \varepsilon_1, \varepsilon_2, \varepsilon_3) = \min [1, \varepsilon_1 + \varepsilon_2, \varepsilon_1 + \varepsilon_3, \varepsilon_2 + \varepsilon_3, \tfrac{1}{2}(\varepsilon_1 + \varepsilon_2 + \varepsilon_3)] .$$

The first four weighting systems and charasteristic functions above are of kinds already encountered, the second, third, and fourth being the functions corresponding to the three minimal sufficient premise sets for the inference. The last characteristic function, $\frac{1}{2}\varepsilon_1 + \frac{1}{2}\varepsilon_2 + \frac{1}{2}\varepsilon_3$, and associated premise weighting system, $\langle \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \rangle$, are of a new type which we wish to consider in detail. Observe that the domain of application of this function is the set of p.u.b.v.s. $\langle \varepsilon_1, \varepsilon_2, \varepsilon_3 \rangle$ such that $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 \leqslant 2$, and such that no single weight exceeds the sum of the other two. This includes the uniform bound ε as a special case, and in this case the new characteristic function gives a maximum uncertainty bound of $1.5\,\varepsilon$, which is smaller than the uncertainty bounds of 2ε which are given by the unitary characteristic functions. The fact that the maximum conclusion uncertainty here does not depend on the premise uncertainties in any unique minimal sufficient set, but rather is contributed to by all, suggests the appropriateness of calling these characteristic functions and weighting systems *statistical.*

The statistical premise weights of $\frac{1}{2}$ in the example are significant in that they are reciprocals of the sizes of the m.e.s. in the example. This generalizes as follows. Let c_i be the size of the smallest m.e. to which a given premise ϕ_i belongs, if it belongs to any, and let $w_i = 1/c_i$ if c_i is defined, and otherwise let $w_i = 0$. Then the weights $\langle w_1, \ldots, \varepsilon_n \rangle$ are easily seen to satisfy the reduced dual constraints for the inference in question, hence by (7).

$$(10) \qquad \mu(I; \varepsilon_1, \ldots, \varepsilon_n) \leqslant \frac{\varepsilon_1}{c_1} + \cdots + \frac{\varepsilon_n}{c_n},$$

where the sum on the right is taken over those ε_i/c_i for which ϕ_i is relevant hence c_i is defined. There is also a partial converse of (10) which applies in the uniform bound case. Let $D_1, D_2, \ldots$ be arbitrary essential (not necessarily minimal essential) premise sets having the property that every premise belongs to the same number of these essential sets as every other. Letting c be the size of the largest of these sets, it is not hard to show by considering the primary constraints that:

$$(11) \qquad \frac{n\varepsilon}{c} \leqslant \mu(I; \varepsilon).$$

Both (10) and (11) therefore relate maximum conclusion uncertainties to reciprocals of sizes of essential premise sets.

As a first application of (10) and (11) consider an inference $(\phi_1, \ldots, \phi_n)\, I(\psi)$ in which ψ is *not* entailed by any premise set with a or fewer members, and *is* entailed by every premise set with more than b members. Then no essential premise set can have less than $n-b$ members, and it follows from (10) that

$$(12) \qquad \mu(I; \varepsilon_1, \ldots, \varepsilon_n) \leqslant \frac{\varepsilon_1 + \cdots + \varepsilon_n}{n - b}.$$

Every $n-a$ member premise set is essential, and since each premise belongs to the same number of $n-a$ member premise sets, (11) applies and we get:

$$(13) \qquad \frac{n\varepsilon}{n-a} \leqslant \mu(I;\varepsilon).$$

The special statistical case is that in which the conclusion is not entailed by any premise set with a members, and is entailed by any premise set with more than a members. In this case the conclusion is equivalent to the proposition 'more than a premises are true', and (12) and (13) combine to imply:

$$(14) \qquad \mu(I;\varepsilon) = \frac{\varepsilon}{1-\frac{a}{n}}.$$

The conclusion uncertainty maximum will never be less than the uniform premise bound, but that it may still be 'acceptably small' if ε is small and $1-a/n$ is not too close to 0. For example, if the premises are 'survey data' items 'person 1 is a Democrat', 'person 2 is a Democrat' and so on up to 1,000, each item of which has an *a priori* uncertainty of 0.02, and the conclusion is that more than 900 of the persons surveyed are Democrats, then the maximum conclusion uncertainty will be $0.02/(1-0.9)=0.2$. This 'smallish' bound is of course much higher than would be expected intuitively, and much higher than the bound which follows if errors are assumed independent. Nonetheless it shows that not all of the confidence reposed in such statistical conclusions depends on tacit inductive assumptions.

A slight generalization of the simple statistical case is that in which the premises $\phi_1,\ldots,\phi_n$ can be put into the form $\phi_i = A_i$ for $i=1,\ldots,k$, and $\phi_i = -A_i$ for $i=k+1,\ldots,n$, and the conclusion is 'the number of A_i's which are true lies between k_1 and k_2 (exclusive)'. This is a case in which the conclusion can be expressed as a conjunction, 'more than k_1 A_i's are true' *and* 'less than k_2 A_i's are true' where this conjunction *separates* the premises so that the positive premises $A_1,\ldots,A_k$ are relevant only to the first conjunct while the negative premises $-A_{k+1},\ldots,-A_n$ are relevant only to the second. Letting I_1 and I_2 be the sub-inferences whose conclusions are the first and second conjuncts, respectively, I_1 and I_2 are separately of the 'simple' statistical form previously discussed (after their irrelevant premises have been deleted), and Equation (14) implies:

$$\mu(I_1; \varepsilon) = \frac{\varepsilon}{1 - \frac{k_1}{k}}$$

and

$$\mu(I_2; \varepsilon) = \frac{\varepsilon}{1 - \frac{n - k_2}{n - k}}.$$

In virtue of the separation of the sub-inferences, the maximum uncertainty of the compound inference, $\mu(I; \varepsilon)$, must equal the sum of the two component conclusion uncertainty maxima above.

A final statistical example leads to a possibly surprising result. In this case the premises of the inference are identities of the form $a_i = a_j$ for all $1 \leqslant i < j \leqslant n$ (there are $\frac{1}{2}n(n-1)$ such premises) and the conclusion is $a_1 = a_2 = \cdots = a_n$ ($a_1, \ldots, a_n$ are all equal). The inference is symmetric in the premises, and it is sufficient to consider the essential sets to which the premise $a_1 = a_2$ belongs. One such is the $n-1$ member set $D_1 = \{a_1 = a_2, a_1 = a_3, \ldots, a_1 = a_n\}$, which is easily seen to be not only essential, but to be the smallest essential set to which $a_1 = a_2$ belongs. Setting each $c_i = n-1$ and each $\varepsilon_k = \varepsilon$ in (10), and recalling there are $\frac{1}{2}n(n-1)$ premises in all, we then get:

$$\mu(I; \varepsilon) \leqslant \tfrac{1}{2}n\varepsilon.$$

On the other hand, permuting a_1 and any other a_i in the premise set D_1 (and reversing the order of the constants in the equations where necessary) yields a new m.e., D_i, and the set of $D_1, \ldots, D_n$ is such that each $a_i = a_j$ belongs to the same number of these essential sets. Hence (11) applies, and gives $\frac{1}{2}n\varepsilon \leqslant \mu(I; \varepsilon)$. Thus, the inequality runs in both directions, so $\mu(I; \varepsilon) = \frac{1}{2}n\varepsilon$. This result is counterintuitive in particular cases: for instance, where n 'individuals' are each independently 'pairwise identified'. This counterintuitiveness is closely related to the fact that the maximum conclusion uncertainty of the inference with the same premises and conclusion $a_1 = a_2$ is ε for uniform premise uncertainty bound ε; in this case the additional identity data lend no reinforcement in reducing the maximum uncertainty of the conclusion.

A variant on the above example leads to similar somewhat surprising results. Here we may take our $\frac{1}{2}n(n-1)$ items of data to be *inequalities* of the form $a_i<a_j$ for $1\leqslant i<j\leqslant n$, where '$<$' is a strict ordering relation (in a language with an axiom to this effect), and the conclusion is $a_1<a_2<<\cdots<a_n$. This case is one in which each premise is either essential or irrelevant, the premises of form $a_i<a_{i+1}$ being essential and the rest irrelevant. The maximum uncertainty of the conclusion is therefore equal to the sum of the uncertainty bounds on its relevant premises, hence in the uniform uncertainty bound case the conclusion uncertainty maximum is $(n-1)\,\varepsilon$. Once again the maximum conclusion uncertainty is intuitively too high, and this is connected with the fact that none of the premises $a_i<a_j$ for $j-1>1$ contributes to reducing the maximum conclusion uncertainty. We will comment briefly on these discrepancies between our formal results and intuitive expectations in Section 3.

We conclude this section by noting some conclusion uncertainty phenomena in two inferences with inconsistent premise sets. The first is one already cited: to infer '$A\leftrightarrow -B$' from the three premises 'A', 'B', and '$-(A\,\&B)$'. The two m.e.s. were $\{A, B\}$ and $\{-(A\,\&B)\}$, and the two minimal negatively sufficient premise sets were the singleton sets $\{A\}$ and $\{B\}$. The dual constraints, which come from the minimal falsification matrix already given, are:

$$w_1\cdot 1 + w_2\cdot 1 + w_3\cdot 0 + v \geqslant 1$$
$$w_1\cdot 0 + w_2\cdot 0 + w_3\cdot 1 + v \geqslant 1$$
$$w_1\cdot 1 + w_2\cdot 0 + w_3\cdot 0 + v \geqslant 0$$
$$w_1\cdot 0 + w_2\cdot 1 + w_3\cdot 0 + v \geqslant 0\,.$$

The four extreme solutions are:

$$\langle w^1, v^1\rangle = \langle\langle 0, 0, 0\rangle, 1\rangle$$
$$\langle w^2, v^2\rangle = \langle\langle 1, 0, 1\rangle, 0\rangle$$
$$\langle w^3, v^3\rangle = \langle\langle 0, 1, 1\rangle, 0\rangle$$
$$\langle w^4, v^4\rangle = \langle\langle 1, 1, 2\rangle, -1\rangle\,.$$

It follows that the c.u.m.f., which is defined for all $\varepsilon_1+\varepsilon_2+\varepsilon_3\geqslant 1$, is:

$$\mu(I;\,\varepsilon_1, \varepsilon_2, \varepsilon_3) = \min(1, \varepsilon_1 + \varepsilon_3, \varepsilon_2 + \varepsilon_3, \varepsilon_1 + \varepsilon_2 + 2\varepsilon_3 - 1)\,.$$

Among the things to note in the example are the following. The fourth characteristic function and weighting system have negative consistency index $v^4 = -1$. This is a system in which all premises are essential (i.e., all have positive weight), but the totality are inconsistent. The domain of applicability of this system is the set of premise bounds such that both $\varepsilon_1 + \varepsilon_3$ and $\varepsilon_2 + \varepsilon_3$ do not exceed 1. This includes the uniform bound case in which ε is between $\frac{1}{4}$ and $\frac{1}{2}$. Note that the conclusion uncertainty bound in this case is $4\varepsilon - 1$, and that this may be significantly less than that given by the two 'consistent' characteristic functions, both of whose values are 2ε. The third premise '$-(A \& B)$' has applicable weight 2 in the inconsistent characteristic function, and this means that where that function applies, the maximum conclusion uncertainty is 'doubly dependent' on the uncertainty of the final premise. In one extreme case, that in which this premise is *certain*, it is possible for the two remaining premises to be highly uncertain while the conclusion is certain: if $\varepsilon_1 = \varepsilon_2 = \frac{1}{2}$ while $\varepsilon_3 = 0$, then $\mu(I) = 0$. This again is strange, to say the least, and we shall comment on its significance in the following section.

As a last example, consider the inference whose premises are n atomic formulas $A_1, \ldots, A_n$ and whose conclusion is 'more than a of the premises are true', where our language will now be assumed to contain a non-logical axiom equivalent to 'not more than b of the premises are true' for some $b > a$. If b is less than n the total premise set is inconsistent. This inference is easily analyzed along the lines already indicated, and we will only state the results concerning the uniform uncertainty bound case. Here the domain of definition of $\mu(I; \varepsilon)$ is the class of $\varepsilon \geqslant 1 - b/n$ (if $b = n$ all bounds are consistent), the premise weights are all equal to $1/(b-a)$, the consistency index is $-(n-b)/b-a)$, and for all consistent ε,

$$\mu(I; \varepsilon) = \min\left(1, \frac{b - n(1 - \varepsilon)}{b - a}\right).$$

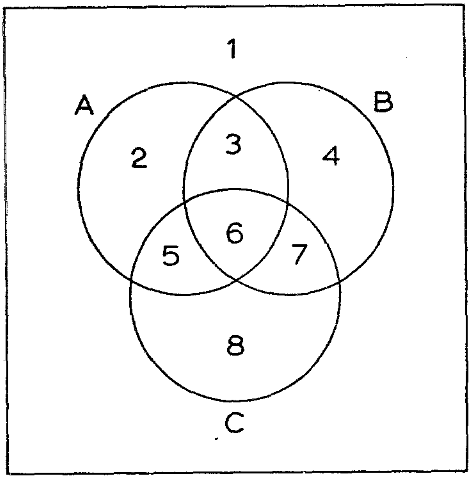
Observe that the consistency index, which depends on a, b, and n, can have fractional and arbitrarily large negative values. Finally, note a somewhat paradoxical result analogous to one encountered with the first inference from inconsistent premises. This is that as the 'amount of inconsistency' increases from 'no inconsistency' ($b = n$) to 'maximal inconsistency' ($b = n(1 - \varepsilon)$), the uncertainty bound $\mu(I; \varepsilon)$ actually decreases from

$n\varepsilon/(n-a)$ to 0. Once again we have mathematical results in search of interpretation, and it is time to turn to questions of significance and other issues of methodology.

3. Logical significance of uncertainty bounds

In this and the succeeding two sections we make some methodological observations more with the aim of explaining the significance of our formal results than to resolve any of the difficult issues involved. First we want to ask of what import it is to the logician to know that the maximum uncertainty which the conclusion of the inference scheme $(\phi_1, \ldots, \phi_n)\, I(\psi)$ can have compatible with some given premise uncertainty bounds is equal to the value $\mu(I; \varepsilon_1, \ldots, \varepsilon_n)$. Let us take the following as plausible: the foregoing means that there are actual propositions, $p_1, \ldots, p_n$ and q of the forms of $\phi_1, \ldots, \phi_n$ and ψ, respectively, (i.e., $p_1, \ldots, p_n$ and q would be properly symbolized as $\phi_1, \ldots, \phi_n$ and ψ, repsectively), and some occasion on which it would be rational to estimate each p_i as having uncertainty no greater than ε_i for $i=1, \ldots, n$, while the uncertainty of q would be rationally estimated as equal to $\mu(I; \varepsilon_1, \ldots, \varepsilon_n)$. Now consider a logician sitting in expert judgment when someone asks him concerning propositions $p'_1, \ldots, p'_n$ and q' which are also of the forms of $\phi_1, \ldots, \phi_n$ and ψ, respectively: is it rational for me to conclude q' on the basis of $p'_1, \ldots, p'_n$? The logician can answer that without further information he cannot tell how certain his interlocutor is of his premises, but given bounds $\varepsilon_1, \ldots, \varepsilon_n$ which are plausible for premises of the sorts involved, the conclusion's uncertainty cannot exceed $\mu(I; \varepsilon_1, \ldots, \varepsilon_n)$, and furthermore there are premises and conclusions of the same form in which the conclusion's uncertainty would actually equal the value $\mu(I; \varepsilon_1, \ldots, \varepsilon_n)$. Thus, granted only the plausible assumption that the premise uncertainties do not exceed $\varepsilon_1, \ldots, \varepsilon_n$, all that can be assured concerning the conclusion's uncertainty *in virtue of the inference's being of the from* of $(\phi_1, \ldots, \phi t)\, I(\psi)$ is that it cannot exceed $\mu(I; \varepsilon_1, \ldots, \varepsilon_n)$. Of course a deductive logician querried aboud the rationality of inferring q' from $p'_1, \ldots, p'_n$ may reply that deductive logic is concerned only with possible truth values and that questions about degrees of certainty of conclusions are properly addressed to inductive logicians. Whether in fact we have entered the domain of inductive logic, or have at any rate blurred the distinction between deductive and inductive logic in introducing probabilities and uncertainties, is a matter which will be returned to in Section 5.

Now consider more closely the nature of the 'extreme' uncertainty functions, u, which maximize the conclusion uncertainty $u(\psi)$ subject to premise uncertainty bounds, $u(\phi_i) \leqslant \varepsilon_i$ for $i=1, \ldots, n$, and what the significance is of the facts that $u(\psi)$ is always at least ε in the uniform premise uncertainty bound case, with consistent premises, while $u(\psi)$ can be smaller than ε when the premises are inconsistent. Let us reconsider the inference $(A, B, C)\, I(A \,\&\, B \vee A \,\&\, C \vee B \,\&\, C)$ already discussed in Section 2, but where we now allow the possibility of adding a non-logical axiom entailing the inconsistency of the premises. Possible truth and probability functions relevant to this inference are perspicuously represented in the following Venn diagram, where 'A', 'B', and 'C' are represented as circles (and their truth-conditional combinations are represented in obvious ways), and where probability functions can be represented as distributions of non-negative probabilities summing to 1 into the eight minimal subregions marked '1', ..., '8' in the diagram:



Note that the minimal subregions correspond to possible truth-functions, or 'possible truth-conditional states of affairs' while a distribution of probabilities into them corresponds to a 'possible probabilistic state of affairs'. Region 6 is of importance since it is the one in which all premises are true, and if the premises are consistent then any probability from 0 to 1 can be put into this region. Regions 2, 4, and 8 are important because these are the ones in which as many premises as possible are true while the conclusion is false. In a sense, these correspond to the minimal essential falsification functions for our inference.

Now suppose that the premises are consistent and we want each of them to have a probability at least $1-\varepsilon$ (hence an uncertainty no greater than ε) while keeping the conclusion's probability as low as possible. One way of assuring probability at least $1-\varepsilon$ for all premises at once is to put $1-\varepsilon$ probability into region 6, their region of intersection where all of them are true. If the remaining probability ε is now distributed into regions 1, 2, 4, and 8 in which the conclusion is false, then the conclusion will have a probability no greater than $1-\varepsilon$. Hence we know that we can make all premises have probability at least $1-\varepsilon$ while the conclusion's probability is no greater than this value. The conclusion can be made still more improbable while retaining the preassigned premise probability bounds if all probabilities outside of the premise intersection are distributed into regions 2, 4, and 8 where maximal numbers of premises are true while the conclusion is false. This in turn allows taking some probability out of the premise intersection, and in fact it is easily seen that the way to get maximum conclusion uncertainty compatible with premise probabilities at least $1-\varepsilon$ is to put probabilities of $\frac{1}{2}\varepsilon$ into regions 2, 4, and 8 and a probability of $1-\frac{3}{2}\varepsilon$ into region 6.

Of course the foregoing conclusion uncertainty maximization procedure won't work if the premises are inconsistent, since in that case we can only put 0 probability into their region of intersection. Here instead of putting probability into region 6 we try to put sufficient probability into maximal consistent sets of premises, corresponding to regions 3, 5, and 7 (whose falsity functions are the negative sufficient falsity functions for our inference). If it is possible to get enough probability into 3, 5, and 7 to give each premise a probability of at least $1-\varepsilon$, then the remaining probability is distributed into regions falsifying the conclusion, as before. But evidently it will not always be possible to get enough probability into 3, 5, and 7 to make each premise have probability at least $1-\varepsilon$ and still have enough left over to keep the conclusion's probability below $1-\varepsilon$ (and it may not be possible to get enough into 3, 5, and 7 to make the premises have probability at least $1-\varepsilon$, no matter what the conclusion's probability is), and this will be the situation when $\mu(I;\varepsilon)$ is smaller than ε.

A very important thing to note about the 'extreme' probabilities and uncertainties maximizing $u(\psi)$ is that these are also ones involving extremes of *non-independence*. This is most evident in the consistent premise case in the present example, from the fact that the *conditional* probability of any premise given the *falsity* of any other premise is equal to zero in the extreme probabilistic state of affairs. This is not the state of affairs which we normally envisage when someone asserts that he accepts 'premises' of such and such forms – say of forms $\phi_1, \ldots, \phi_n$ – where we are apt to imagine instead that the premises represent items of independently acquired information and where the falsity of any one item would not necessarily call any other item into question. Perhaps it would even be somewhat misleading for a person to describe the sorts of highly interdependent systems of propositions which we need to consider in arriving at conclusion uncertainty maxima as 'premises'. Be that as it may, the fact that extreme probabilistic states of affairs involve extreme interdependence suggests that we might expect logically possible conclusion uncertainty bounds to differ greatly from the conclusion uncertainties which follow if 'normal' independence assumptions are made. This is not the place to enter in detail into the error probabilities which follow from independence assumptions (this being a standard aspect statistics), but a couple of remarks are in order by way of comparing logically maximum conclusion uncertainties with those following under independence assumptions.

The easiest uncertainty bound comparison can be made where the premises of the inference are n independent propositions symbolizable by atomic formulas 'A_1',…, 'A_n' and the conclusion is equivalent to the assertion that more than a of the propositions are true. Abbreviating the conclusion as '$M(a, n)$' (more than a out of the n premises are true) we have the inference $I(M(a, n)) = (A_1, \ldots, A_n)\, I(M(a, n))$, where we have already seen that the uniform conclusion uncertainty bound function is given by:

$$\mu(I(M(a, n)); \varepsilon) = \frac{\varepsilon}{1 - \frac{a}{n}}.$$

The inference just considered was the special case in which $n=3$ and $a=1$

where we have

$$\mu(I(M(1, 3)); \varepsilon) = \tfrac{3}{2}\varepsilon .$$

If $\varepsilon = 0.1$, for instance, the logically maximum conclusion uncertainty will be 0.15. However, when the premises are assumed to be statistically independent the probability of any subset of them being false is equal to the product of their individual probabilities of falsity, which is here assumed to be ε. Writing the independent uncertainty function which assigns probability ε as the probability of falsity for each individual premise as $\boldsymbol{u}_\varepsilon$, we have

$$\boldsymbol{u}_\varepsilon (M(a, n)) = \sum_{j=0}^{a} \binom{n}{j} (1 - \varepsilon)^j \varepsilon^{n-j},$$

where $\binom{n}{j}$ is the binomial coefficient $n!/j!(n-j)!$ Where $n = 3$ and $a = 1$ we have

$$\boldsymbol{u}_\varepsilon (M(1, 3)) = \varepsilon^3 + 3\varepsilon^2 (1 - \varepsilon) .$$

If $\varepsilon = 0.1$, for instance, $\boldsymbol{u}_\varepsilon (M(1, 3))$ must be 0.028 which is much smaller than the maximum logically possible conclusion uncertainty of 0.15. The difference between two values gives a measure of the degree to which the conclusion depends on unexpressed independence assumptions, which must be regarded as empirical in character since the same kinds of assumptions can be used to justify the obviously inductive inference of the conclusion $M(a, n)$ from the single premise $M(a-1, n)$ when a and n are large enough (this is the inference of 'more than a of the premises are true' from 'at least a of the premises are true').

A more striking result emerges when we reconsider the inference of the conclusion $a_1 = a_2 = \cdots = a_n$ from the $n(n-1)/2$ premises $a_i = a_j$ for $1 \leqslant i < j \leqslant n$, where we have seen that the uniform uncertainty bound is given by $\tfrac{1}{2}n\varepsilon$. For instance, if $n = 10$ and $\varepsilon = 0.1$ then the maximum conclusion uncertainty will be 0.5, which is much larger than intuition would lead us to expect in an inference of this kind. Independence assumptions are tricky to formulate in the present situation because of the manifest logical dependencies among the premises, but a plausible approach is to assume some *a priori* probability distribution over 'possible equational states of affairs' from which posterior probabilities and uncertainties are computed on the basis of somewhat uncertain 'data' which have the form of *reports* that some of the a_i's are equal to others, and where it is assumed that the probabilities of error in the reports are independent. Proceeding in this way, we might assume that all possible equational states of affairs among, say, the ten items $a_1, \ldots, a_{10}$ were equally likely *a priori* (each such equational state of affairs would correspond to a partition of the ten items into mutually exclusive subsets), and that the probability of any individual report's being correct was some value, say π. Under this assumption, even if 0 were only as large as $\frac{2}{3}$, the posterior uncertainty of the conclusion $a_1 = a_2 = \cdots = a_{10}$ from the given 'dubious data' could be no larger than 0.02. This shows that our intuition about the correctness of this conclusion *is* justified if independence assumptions of an essentially empirical character are justified. This is clearly an inference which it is appropriate to call 'deceptively deductive'.

4. The Possible Significance of Probability Change

In our formal analysis of conclusion uncertainty maxima we have treated probabilities and uncertainties as though they were like static truth-values, and have disregarded the fact that, unlike truth-values, they are subject to change with circumstances, and in particular when new information is acquired. On the other hand some of our informal remarks have suggested the relevance of probability change to the analysis of conclusion uncertainties, and in this section we want to discuss briefly some of the ways in which systematic consideration of probability change might be expected to affect our picture of inferential probabilities. Let us here leave entirely aside the effect on probabilities of deductive discoveries, which is a largely unexplored subject (except, see Hintikka (1970)), and concentrate on probability change consequent on empirical discoveries.

The usually assumed law of probability change consequent on the acquisition of a new item ϕ of *a posteriori* certain information is that the new probability of any other factual proposition ψ is equal to what was previously the *conditional probability of* ψ *given* ϕ. Assuming the 'conditionalization' law of probability change, it would follow that systematic analysis of conclusion uncertainty bounds in a 'changing probabilities' framework cannot restrict itself to factual propositions alone, as we have done here, but must bring conditionals explicitly into the picture. In other words, the key to the analysis of factual (non-conditional) inference uncertainties within a changing probability picture is the analysis of conditional inference uncertainties in a static probability framework – which is just what we have earlier noted as an essential limitation on our present investigation.

There are two matters of particular importance upon which probability change considerations can be expected to throw light. One has to do with the *weights* of premises of deductive inferences described in Section 2, where the weight of a premise gives an index of the degree to which the maximum uncertainty of the conclusion of the inference depends on the uncertainty of that premise. Recall that *irrelevant* premises always have weight 0, which means that the conclusion does not depend on those premises' probabilities, *essential* premises have weight 1, which means that these premises' uncertainties always contribute their total values to the maximum conclusion uncertainty (unless this is already equal to 1), while redundant but not irrelevant premises may have intermediate weights. What this suggests is that where reasoners are primarily concerned with ultimate conclusions, they should take greatest pains to assure the certainty of premises with the largest weights. However, 'assuring the certainty' of the premise is most plausible looked upon as an operation which *increases* the probability of that premise, and this is a matter of probability change. In particular, we must worry about the possibility that assuring one premise's probability may have side effects on the probabilities of other premises, so that conclusion uncertainty maxima will not only be directly affected by a change in the first premise's probability, but will be indirectly affected by the other premise probability changes. But the only case in which the first premise's probability change will not affect the probabilities of other premises is that in which the other premises are statistically independent of the first premise, and statistical independence is excluded in extreme probabilistic states of affairs. What this suggests is that while some sort of *weight* is probably a useful concept in terms of which to describe the contribution of a premise's uncertainty to that of the conclusion of an inference, the static weights we have considered in this paper tell only a part of the story, which becomes considerably more complicated when probability changes are taken account of.

Finally, we may expect probability change considerations to be especially significant so far as concerns *conflict* or *inconsistency* phenomena. This is not to say that simultaneous acceptance of inconsistent propositions is not possible or unimportant (recall the 'preface paradox'), but one of the authors has argued elsewhere (Adams (1975)) that the more common type of real life propositional conflict involves one proposition being accepted at one time and then countervailing evidence coming to hand which forces its retraction. Thus, the more common real life situation in which someone is represented as accepting all of $A_1, \ldots, A_n$ *and* $-(A_1 \& \ldots \ldots \& A_n)$ will be one in which certain of these propositions are accepted at one time and others at other times. Again, when we ask what uncertainties are compatible with accepting all of these premises, we will need to take explicit account of the *times* of their acceptance which is something which falls outside of our static analysis.

5. Probability, deduction, and rational inference

We want to do two things in this section: (1) to argue that the sorts of probability-uncertainty considerations we are concerned with here properly belong to the province of deductive logic, and (2) to disclaim the suggestion that the probabilities we have been dealing with give us certain deeper insights into the nature of *rationality* which some have hoped to gain *via* probabilistic analysis. Concerning the first claim, it could suffice to note that all of our probabilistic results have to do with the 'space' of *logically possible probability functions* which can be defined so as to apply to premises and conclusions of inferences, and this space is in turn completely defined by the possible combinations of truth-values which these propositions can have. Possible probabilistic states of affairs are uniquely determined by possible truth-conditional states of affairs, and possible truth-conditional states of affairs are the subject matter of deductive logic. It might be argued that while possible probabilistic states of affairs are indeed deductively determined, we have nevertheless gone beyond the bounds of deductive logic in implicitly suggesting the appropriateness of a non-truth-conditional 'criterion' of rationality for inferences: namely that in some sense the fact that the premises of an inference are all *probable* should guarantee that the conclusion is also *probable*. This implicit criterion is sufficiently vaguely formulated to allow of possibly non-deductive interpretations, but we would suggest that as we have interpreted it here it still falls naturally into the 'deductive' side of the deductive-inductive partition. In particular, the sense we attach to 'guarantee' in the criterion is logical – it should not be logically possible for premises to be probable while conclusions are improbable. Furthermore, when so interpreted (and made suitably precise in other ways), our criterion becomes if anything *stricter* than the standard requirement that it should not be possible for premises to be true while conclusions are false. In short, while we use the probabilisitic concepts which some have come to regard as the hallmark of inductive inference, we do not regard this use as supporting (*or* countering) claims such as that the deduction-induction distinction is untenable.

The foregoing probably makes it clear that *our* use of probabilities does not throw much light on certain deep issues relating to rationality and rational inference which one might hope to deal with in probabilistic terms, but some further remarks on this are in order. One might wish to focus an investigation of rationality in an analysis of the *reason relation*:

$R(S; t; p_1, ..., p_n; q) =$ person S's knowledge or belief in propositions $p_1, ..., p_n$ at time t gives him good reason for believing q.

It should be immediately obvious that our 'criterion' that the high probability of $p_1, ..., p_n$ should guarantee the high probability of q can be neither a necessary nor a sufficient condition for $R(S; t; p_1 ..., p_n; q)$ to hold for arbitrary S and t. Non-necessity follows from the fact that $p_1, ..., p_n$ can only 'probabilistically entail' q according to our criterion if these premises logically imply q, and we should at least intuitively wish $R(S; t; p_1, ..., p_n; q)$ to hold *sometimes* when $p_1, ..., p_n$ only furnish something like good *inductive* grounds for believing q. This is connected with the fact that $R(S; t; p_1, ..., p_n; q)$ should in some way depend on the person S who is making the inference, and the time (occasion) t when he makes it. This dependence is reflected in the fact that whether $p_1, ..., p_n$ furnish S with good reason for believing q on an occasion will depend not just on $p_1, ..., p_n$ and q, but on *what else* S knows at time t. A falling barometer gives *me*, who knows something about the meteorological significance of barometric indications, good reason for thinking a change in the weather is coming, but it would not give a person ignorant of the meteorological facts grounds for such an inference. In restricting ourselves to purely 'logical' truth-conditional *or* probabilistic relations, we simultaneously exclude inductive reasons *and* reasons which depend on the 'reasoner's situation' from consideration (perhaps these come to the same thing).

That the fact that $p_1, \ldots, p_n$ 'probabilistically entail' q is not a sufficient condition for $R(S; t; p_1, \ldots, p_n; q)$ to hold follows from the fact that it is possible for $p_1 \ldots, p_n$ to entail q 'remotely' in such a way that a reasoner accepting $p_1, \ldots, p_n$ would not 'see' that q followed, and in this case we would not want to say that that person was justified in believing q (perhaps he could *have* good reasons but not be justified in believing q because he would not *know* that his reasons were good). Of course this is simply to reiterate the well known observation that logical entailment is not a sufficient condition for rational inference if the reasoner is not able to 'see the connection' between his premises and his conclusion. It is interesting to note in this connection, though, that whereas it is possible to spell out at least in outline what 'seeing the connection' is in the case of truth-conditional soundness – namely finding and being able to reproduce an acceptable *derivation* of the conclusion from the premises – we as yet lack any derivational complement to our probabilistic requirement that the high probability of premises should assure that of the conclusion. What this suggests is the desirability of both descriptive and normative studies of practically applicable procedures which reasoners do and/or can apply in order to assure that they are not led from highly probable premises to 'insufficiently probable' conclusions. This is unfinished logical business, however, and it leads to our final observation on the significance of the positive findings of the previous sections. The contribution of our present studies to ongoing research into inference processes and their rationality is probably this: we now know at least roughly *how much* 'objective uncertainty' is passed on to conclusions because of uncertainties in the premises, and what we need to find out is how persons do and/or can follow reasoning *procedures* which will always assure them 'sufficient certainty' in the conclusions they arrive at.

University of California, Berkeley

NOTES

[1] See Harman (1967) and Kyburg (1967) for discussions of the Lottery Paradox and its implications.

[2] Adams (1965) and Suppes (1966) give proofs of this theorem, which is apparently well known to probability theorists.

[3] Probabilistic aspects of the logic of conditionals are discussed in Adams (1975).
[4] A sufficient condition for it to be the case that all of the characteristic functions of an inference which is already in 'normal form' (premises atomic formulas and conclusions in minimal essential or sufficient form) should be either total uncertainty or unitary functions is that the conclusion be equivalent to a formula built from atomic formulas by just conjunction and disjunction, and where each atomic formula occurs at most once. This condition is not necessary for 'unitary representability', but it is an interesting unsolved problem to give necessary and sufficient conditions for this.

REFERENCES

Adams, E. W., 'On the Logic of Conditionals', *Inquiry* **8** (1965), 166–197.

Adams, E. W., 'Prior Probabilities and Counterfactual Conditionals', to appear in C. W. Harper and W. Hooker (eds.), *Proceedings of an International Conference on the Foundations of Statistics*, D. Reidel Publishing Company, Dordrecht and Boston, 1975.

Goldman, A. J. and Tucker, A. W., 'Theory of Linear Programming', in H. W. Kuhn and A. W. Tucker (eds.), *Linear Inequalities and Related Systems, Annals of Mathematics Study*, Vol. 38, Princeton, N.J., 1956, pp. 53–98.

Harman, G., 'Detachment, Probability, and Maximum Likelihood', *Nos* **1** (1967), 401–411.

Hintikka, J., 'Surface Information and Depth Information', in J. Hintikka and P. Suppes (eds.), *Information and Inference*, D. Reidel Publishing Company, Dordrecht, 1970, pp. 263–297.

Kyburg, H., 'Probability, Rationality, and the Rule of Detachment', in Y. Bar-Hillel (ed.), *Logic, Methodology, and Philosophy of Science*, North-Holland Publishing Company, Amsterdam, 1967, pp. 301–310.

Suppes, P., 'Probabilistic Inference and the Concept of Total Evidence', in J. Hintikka and P. Suppes (eds.), *Aspects of Inductive Logic*, North-Holland Publishing Company, Amsterdam, 1966, pp. 49–65.