Fuzzy Knowledge Combination

Bart Kosko
VERAC, Incorporated, 9605 Scranton Road, San Diego, CA 92121

A general answer is given to what one should conclude from disagreeing experts. The answer is generalized further to incorporate the experts' credibility weights. The answer rests on a wide range of intuitively based epistemic axioms, scientific and philosophical conjectures, and formal mathematical relationships. A recurring theme is the making of Bellman–Zadeh fuzzy decisions, wherein a decision is the intersection of fuzzy goal and fuzzy constraint subsets of some space of alternatives. Another result is that measures of central tendency, such as the arithmetic mean, make poor knowledge combination operators. Formally, fuzzy knowledge combination operators are sought. The function space of knowledge combination operators ϕ: $K^n \rightarrow K$ is shrunk by imposing successive axioms. The final shrunken set is said to consist of *admissible* knowledge combination operators. Some of its mathematical properties are explored and a simple admissible operator is finally chosen. Knowledge sources X_i: $S \rightarrow K$ are mappings from epistemic stimuli or questions into a knowledge response set K. The uncertainty of the underlying epistemic situations is captured by the cardinality of K and by the fuzziness of its partial ordering. Admissible knowledge combination operators aggregate knowledge responses in some desirable way. The arithmetic mean is not admissible. Nor in general is a probabilistic framework even definable in the abstract poset setting employed by this theory. The fuzzy knowledge combination theory is extended by associating general credibility weights with the knowledge sources. A new set of weighting axioms is required to satisfy certain intuitions and to satisfy the admissibility axioms. General weighting functions are obtained and thereby weighted admissible operators are obtained. The weighted mean still proves inadmissible. Appendix I contains a technical glossary and summary of the proposed fuzzy knowledge combination theory. Appendix II contains proofs of the probabilistic uncertainty theorems required for the uncertainty testbed used in the theory.

I. THE FUZZY KNOWLEDGE COMBINATION PROBLEM

How can a decision be reached when experts disagree? Take a mean? What if one expert says yes and the other says no? Take a weighted mean? Use thresholds of acceptance? Assume a (certain) structure on the expert uncertainty—assume a probability distribution on the experts' responses?

All knowledge-combination strategies ultimately rest on some epistemological assumptions packed into the assumed knowledge-representation framework or the combination technique. The mathematico–epistemological theory developed in this article rests entirely on uncertainty (fuzzy or random) intuitions. A probabilistic model is explicitly avoided; yet some of the intuitions used are captured in probabilistic relationships and proved to hold almost surely. Some of the intuitions are fuzzy mathematical translations of conservative epistemic beliefs. And some of the intuitions rest on nonparametric statistical biases: Would

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. I, 293-320 (1986)
 CCC 0884-8173/86/040293-28$04.00

you bet more often on the mean or the median? Like all assumptions, these assumptions are ultimately accepted or not on their perceived reasonability—on faith. No attempt is made to persuade the reader of the reasonability of these grounding intuitions other than mathematically elucidating them as clearly as possible and pointing out some of the consequences of denying them.

The fuzzy knowledge-combination problem can be cast as a function-space search. Let S be a set of *query stimuli*. Let K be a partially ordered set of *knowledge responses*. Define *knowledge sources* $X_1, \ldots, X_n$ as functions from query stimuli to knowledge responses, i.e., $X_i: S \rightarrow K$. The fuzziness of the matter is, as always, captured by the common range set K; i.e., the cardinality of K indicates the (worst-case) fuzziness of the knowledge sources: a two-element K indicates certainty and in general an n-element K indicates more fuzziness the larger n is. The partial order on K is assumed to make K suitable for mathematical operations and should not be conceded unquestioningly. More generally, K need only be endowed with a fuzzy partial order; this seems the minimal structural requirement. Pick some query stimulus $s \varepsilon S$. Let $X(s) = (X_1(s), \ldots, X_n(s))$ denote the *knowledge response vector*. Let $k \varepsilon K$ denote the *knowledge* of the *epistemic situation* $(s, X(s))$. Note we assume a point, not interval, estimate of knowledge. The fuzziness of K softens this assumption (which is surely partly why fuzziness culturally evolved). Then the *fuzzy knowledge combination problem* is to find some knowledge-combination function $\emptyset: K^n \rightarrow K$ such that $\emptyset(X(s)) = k$.

II. SHRINKING THE FUNCTION SPACE

Knowledge up, search down goes the AI slogan. We will employ this "intelligent" problem solving strategy to find a fuzzy knowledge combination function $\emptyset$. $\emptyset$ will be trapped between lower and upper knowledge bounds that are intuitively determined. Uncertainty intuitions are invoked to further delimit $\emptyset$'s behavior between the knowledge bounds. This surround-and-march approach amounts to defining a fuzzy decision[1] on the function space of (K-valued) functionals on K^n. The epistemic requirements are goal or constraint sets of functionals. Their successive intersection is what is meant by shrinking the function space.

The combination function ultimately selected in this article is not unique. The epistemic requirements imposed simply do not contain enough structure, a price often paid for generality. A fundamental epistemological question is whether there exists further functional structure compatible with all maps from a cartesian product of a (fuzzy) partial order into itself. Still, the selected knowledge combination function exhibits two properties that surely must be required of general combination operators. First, it is in no way similar to the arithmetic mean of the knowledge responses $X(s)$, which we take as the ever-present benchmark of parametric statistics. In fact the mean is easily barred from candidacy in the class of reasonable knowledge combination functions; indeed it is barred by the very limit theorems that have popularized it. Second, the selected $\emptyset$ turns out to be a generalized confluence operator[1] and hence reduces the Bellman–Zadeh decision theory. The reduction itself is trivial; seeing that such a reduction has occurred is the hard part. This difficulty dissipates, and the reason for requiring such a reduction congeals, when combining knowledge is seen as making an epistemic decision. Knowledge sources, objects that emit knowledge responses when queried, actually are point-to-set mappings. For a given piece of knowledge k depicts some set of actions or some states of affairs, i.e., it forbids some actions or contradicts descriptions about some states of affairs and permits or is consistent with others. The fuzzier the knowledge k, the fuzzier the depicted sets. The act of combining fuzzy knowledge is some act of combining these fuzzy sets, and in the simplest case this means taking their fuzzy intersection, which is precisely the definition of a Bellman–Zadeh fuzzy decision.

Fuzzy decision theory also provides useful geometrical intuitions for combining fuzzy knowledge. To see this, let us begin the function-space search by supposing that a query stimulus s has produced n-many knowledge sets via n-many knowledge sources. What are the natural ways to combine such sets? One way is to take the fuzzy set union. This is a very optimistic act of knowledge combination. An equally intuitive way is pessimistic: take the fuzzy set intersection. Which shall it be? At this point we do not have enough structure assumed to decide, but we do have enough to state the first knowledge combination requirement. That requirement is *boundedness:* $I \subset C \subset U$, i.e., the combined knowledge set C contains the intersection I and is contained in the union U, i.e., the knowledge lies in the middle. The lower bound I is justified simply because it contains exactly the knowledge elements common to all the knowledge sets. The upper bound U is justified because every element of its complement, U^c, is uncommon to all the knowledge sets and because U is minimal with respect to this property, i.e., because U is the intersection of all the sets with this property. These epistemic intuitions remain essentially intact when the knowledge sets involved are fuzzy and the standard fuzzy set operations are used. More precisely, the exactness of the intuitions blurs according to the fuzziness of the knowledge sets.

We return from set to function intuitions, ever aware of their mutual translatability. Let l and m denote the minimum (infimum) and maximum (supremum), respectively, of the knowledge response vector $X(s)$ components:

$$l = \min_i X_i(s),$$

$$m = \max_i X_i(s),$$

where the dependence of the functions $l(s)$ and $m(s)$ on the query stimulus s has been suppressed in the notation for convenience.

The boundedness assumption is that we know *at least* l and *at most* m. We pack this into the first, most plausible, and least binding function space requirement.

$$(A_1) \quad \text{BOUNDEDNESS:} \qquad l \leqslant \phi \leqslant m.$$

The inequalities in A_1 presuppose that the knowledge response set K is not just partially ordered but numerically so. This loss of generality (excluding, for instance, the linguistic quantities of everyday discourse) is only temporary. The ϕ eventually selected extends to most posets, depending on how negation (order reversal) is defined. For the moment we identify K with the unit interval I, i.e., $K = I$. This is also required to define an arithmetic mean A of the knowledge responses:

$$A = \frac{1}{n} \sum_i X_i(s),$$

where again A's dependence on s has been suppressed in the notation.

The space of possible knowledge combination functions delimited by A_1 is obviously too large to determine $\emptyset$. The usefulness of A_1 is to focus attention on l and m, defined on every poset, and to invoke it as a test criterion for candidate combination functions. For instance, in the fashion of independent joint probability density functions, the product

$$\prod_i X_i(s)$$

is immediately denied candidacy in the set of *admissible* knowledge combination functions—as we shall call (and define!) them—except in the degenerate case when all the knowledge responses are equal. The product is too pessimistic; in particular, it ignores the upper bound m in the sense that it is always bounded from above by l (of course m is always a factor in the product) and it tends monotonically toward zero as the number of nonunity knowledge responses increases. Turning from the field operation of multiplication to (normalized) addition of knowledge responses, we see that A_1 is compatible with the mean A since

$$l = \frac{1}{n} \sum_i l \leq A \leq \frac{1}{n} \sum_i m = m.$$

Another candidate combination function compatible with A_1 is the median of the knowledge responses—the benchmark of nonparametric statistics. It turns out, as will be shortly seen, that the median and the mean are judged inadmissable in the same stroke. l and m are also candidates compatible with A_1; in fact any order statistic is. Though l and m seem trivial candidates, this entire theory of fuzzy knowledge combination turns on how the hypothetical question "Should $\emptyset = l$ or $\emptyset = m$?" (where "or" is exclusive) is answered.

The second function space requirement is that $\emptyset$ must be invariant under permutations of its knowledge response arguments. Note that this is not a veiled assumption of equal likelihood of knowledge response occurrence. It is more an assumption of simultaneity of knowledge source behavior. It simply says that it should not matter how the knowledge sources are labeled. Let $P(n)$ be the symmetric group on n—the set of all permutations of 1, 2, . . . , n—and let $p = (p_1, \ldots, p_n) \in P(n)$ be a permutation. Let $X_p(s) = (X_{p_1}(s), \ldots, X_{p_n}(s))$ be a shuffle of the knowledge response vector components, i.e., the result of exchanging $X_i(s)$ with $X_{p_i}(s)$. Then A_2 can be stated.

$$(A_2) \quad \text{SYMMETRY:} \qquad \emptyset(X(s)) = \emptyset(X_p(s)) \quad \text{for all } p \in P(n).$$

A_2 is compatible with all $\emptyset$ candidates mentioned so far. It plays a little role in the mathematico–epistemological theory developed in this article, namely separating mathematical notation from philosophical disposition. Yet A_2 does rule out continuum-many ϕ candidates, e.g., weighted subtractions of $X(s)$ components.

The next epistemic requirement concerns the *knowledge gap* $m - 1$. This function is a raw measure of the uncertainty in the epistemic situation $(s, X(s))$. Eventually it will be taken as the only measure of uncertainty. As the knowledge gap varies from 0 to 1, the uncertainty varies from minimal to maximal; e.g., when $0 = l \leqslant \phi \leqslant m = 1$ occurs, we know nothing. Hence $1 - (m - l) = 1 - m + l$ is a rough measure of what we know. This simple observation proves to be fundamental when ultimately selecting ϕ. One final note, condition A_1 and the identification of $m - l$ as the knowledge gap is sure to strike some readers as similar to the popular Dempster–Shafer theory of combining evidence. In particular, A_1 is akin to Dempster's[2,3] original notion of multivalued-map-induced probability measures and l and m are interpreted, respectively, as the Support and Plausibility of proposition Q by Shafer.[4] A deeper connection is found in the functional similarity of the relationships

$$\text{Plausibility}(Q) = 1 - \text{Support}(\text{not-}Q)$$

and

$$m(s) = 1 - \min_i(1 - X_i(s)).$$

There seems to be no further conspicuous connections between the present probability-free theory and the quasi-Bayes Dempster–Shafer theory.

The key epistemic question of the present theory is *how should ϕ vary as the knowledge gap $m - l$ varies*? The forthcoming answer is motivated by answering an easier question. If $\phi = l$ and $\phi = m$, pure epistemic pessimism and optimism, are the only choices for all possible stimulus environments, then which should be chosen? A wide range of intuitions, suspicions, and evolved behaviors dictate the conservative answer $\phi = l$, and this is the answer assumed by this theory. Unfortunately, arguments for epistemic conservatism are scarce. For instance, it is standard in statistical decision theory to simply state the maximin risk criterion, the archetype of epistemic conservatism, as one criterion among many. A notable exception, and source of unending argumentation, is the debate in social choice theory over philosopher John Rawls's proposal[5] for a maximin policy of social justice, i.e., for designing social institutions and redistribution schemes so as to maximally improve the lot of those worse off socio–economically. Rawls claims that if you voted on social policy while sitting behind a perfect "veil of ignorance," where you knew neither your future place in society nor your capabilities (i.e., $m - l = 1$ in some sense in all stimulus environments), then in fact you would vote for maximin social policies. The thrust of Rawls's argument, and most of the many counterarguments, has to do with correctly or incorrectly placing odds on perfectly uncertain environments, namely that maximin is the correct placement. This amounts to arguing over which probability distribution ought be assumed on nature, an old subject, and therefore is ignored (unintuited) by this nonprobabilistic epistemic theory.

Epistemic conservatism is central to this theory of fuzzy knowledge combination. Therefore some argument for the choice $\phi = l$ is in order, even if speculative. The present argument appeals to evolutionary biology, to genes, and provides some fuzzy justification for maximin sentiments generally. We cite what zoologist E. O. Wilson[6] calls the *principle of stringency* in population biology. This principle is the ubiquitous induction that the time-energy budgets of organisms evolve only in light of worst-case, or most stringent, environmental parameters, i.e., that expected survival drives adaptation rather than, say, expected reproduction. The principle of stringency explains why predators often ignore killable prey, why foragers often ignore food, and why in general wildly gluttonous, wildly reproductive behavior does not prevail. As Wilson says, "animals and societies do not always live in the midst of plenty" (Ref. 6, p. 143). The claim is that organisms acting to the contrary do not, all else constant, house genes that increase in frequency in the population—extinction too easily occurs. This empirical claim, perhaps the only genetic explanation for laziness, is a conservative minimax claim in the sense that enduring genes abound in organisms who minimize the chance of experiencing maximum loss (death before reproduction). The breadth of the principle of stringency becomes clearer when it is realized that the principle presumably operates on *all* organisms, plant or animal *(et al.)*. For present purposes, we venture that the principal of stringency underlies the observed epistemic conservatism of human individual and group judgment: for instance, the deeply entrenched operative principle that *truth* = *concurrence*, i.e., that the working combined knowledge set is the intersection of informed knowledge source sets i.e., that people are fuzzy decision makers in the Bellman–Zadeh sense. Hence $\phi = l$.

Again let the knowledge gap $m - l$ measure the uncertainty in the epistemic situation $(s, X(s))$. The previous choice of $\phi = l$ over $\phi = m$ can be restated as follows. As the query stimulus variable s ranges over S, all possible stimulus environments, the maximum uncertainty case $m - l = 1$ can be expected to occur, and when it does we know nothing $(0 \leq \phi \leq 1)$, and thus $k = \phi(X(s)) = 0$ in this case. But when $m - l = 1$, $l = 0$, and thus $\phi(X(s)) = l$. Further, in the least uncertain case, when $m - l = 0$ or $m = l$, we know it all $(\phi = m)$, and again $\phi(X(s)) = l$. Let us make explicit that we are focusing on an abstract *sequence* of epistemic situations $(s, X(s))$ indexed by query stimuli $s \in S$. Intuitively it may help to think of s as a continuous parameter. Without loss of generality, let us further suppose that S is ordered so that as s ranges through it the knowledge gap ranges from minimal to maximal, i.e., from $m - l = 0$ to $m - l = 1$. Then, finally, where should ϕ tend, if anywhere, when $m - l$ increases from 0 to 1? *How does ϕ respond to uncertainty*?

Of course ϕ should tend to l as $m - l$ increases; that is the import of choosing $\phi = l$ in the most uncertain case and $\phi = m$ $(= l)$ in the least uncertain case. The continuous intuition is that ϕ monotonically decreases to l as the knowledge gap increases, i.e., uncertainty promotes scepticism. Put another way, the more uncertain the environment, the more we tend to play minimax strategies—the more we tend to play hardball. We now cast this uncertainty intuition as the third function space constraint.

$$(A_3) \qquad \text{CONSERVATISM:} \qquad \phi \downarrow l \quad \text{as} \quad m - l \uparrow 1.$$

Note that in A_3, as in A_1 and A_2, the relevant constituents only depend on the partial order on K.

A_3 implies that any admissible knowledge combination operator is a generalized confluence operator and hence reduces the Bellman–Zadeh fuzzy decision theory in the limiting uncertain case. We recall that a fuzzy decision is confluence of goals and constraints, where the goals and constraints depict fuzzy subsets of some space of alternatives and confluence is interpreted as fuzzy set intersection. Hence $D = G_1 \cap \ldots \cap G_n \cap C_1 \cap \ldots \cap C_m$ with the obvious notation. Therefore the fuzzy set membership function of the fuzzy decision set D is

$$m_D = \min_{i,j} \{m_{G_i}, m_{C_j}\} = l,$$

whence knowledge source response sets are equated with the (epistemic) goal and constraint sets. [More formally, $m_{X_i}(s) = X_i(s)$, and where we recall that $X_i(s)$ picks out some abstract knowledge set, consistent world state descriptions for instance, which can be taken as some fuzzy subset of K.] Hence $\phi \downarrow m_D$, as claimed.

A_3 also implies that m and the median of the knowledge responses are inadmissible. For obviously the choice $\phi = m$ violates A_3. The median is disbarred because in general it is independent of the knowledge gap dispersion. The median need tend nowhere as $m - l$ increases. A similar fate befalls all other order statistics except l, which is obviously admissible. Note that we are saying that a candidate knowledge combination operator is *admissible* only if it satisfies $A_1 - A_3$.

Requirement A_3 admits a random interpretation. As detailed in Appendix II, suppose the knowledge sources $X_1, X_2, \ldots$ are unit-interval valued random variables (measurable with respect to, say, 2^S) on the stimulus space S. We use this random framework to model worst-case behavior of knowledge combination operators, and we use randomness rather than fuzziness here because the available probabilistic theory is so much better explored.

Intuitively, perfectly random knowledge responses represent maximal epistemic entropy. It corresponds to all the queried experts ignoring and shouting at each other. This situation can be largely captured with a random sampling framework. Thus let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.). As demonstrated in Appendix II, it turns out we must also assume that the knowledge-source random variables have density functions and that these density functions are nondegenerate (positive on sets of nonzero Lebesgue measure, a type of compact-support property), a painless assumption to rule out pathologies and one that is satisfied by most popular densities, including the normal density. Judicious application of the Borel–Cantelli Lemma then takes us where we expect to go, and in fact even further. In particular, in the random case we expect to know nothing: $0 \leqslant \phi \leqslant 1$ and, further, $\phi = 0$. As the number n of knowledge sources increases, l should monotonically tend to 0 and m should monotonically tend to 1. Let us call this random version of A_3 requirement $A_3{}^*$:

$$(A_3{}^*) \qquad \phi \downarrow l \quad \text{as} \quad i \to \infty \quad \text{in the random (i.i.d.) case.}$$

Theorem 1 in Appendix II shows that l and m tend to the endpoints of the unit interval with probability one. Indeed Theorem 1 shows much more. It not only shows with probability one that at least one knowledge source will respond with a one and at least one will respond with a zero, but that the knowledge response one will be emitted infinitely often and the knowledge response zero will not be value of l at most finitely often. Since Theorem 1 applies to any nondegenerate subset of the unit interval, it in essence says that in perfect random sampling all the values in the unit interval will be emitted as knowledge responses, infinitely often, and with probability one. For fuzzy theorists, Theorem 1 (and Theorems 2 and 3) amounts to an asymptotic property of the fundamental fuzzy operators supremum and infimum.

Now let us examine the behavior of the arithmetic mean A with respect to A_3 and $A_3{}^*$. Does A tend to l as the sample size grows? Does A tend anywhere? The answer is immediate, unambiguous, and robust. No tampering with the random sampling framework will change it. The answer is no. By the Kolmogorov Strong Law of Large Numbers, with probability one A *converges to the distribution mean* of the knowledge-source random variables. And, in general, the distribution mean is independent of l and m; moreover, in general it is not zero ($\frac{1}{2}$ seems the more intuitive guess), and hence A_3 and $A_3{}^*$ are violated.

An immediate hint that A is not a general knowledge combination operator is that it requires real numbers for its definition. It is not even a closed operator on the set of integers. And it certainly does not generalize to arbitrary partially ordered sets. Much of the problem lies in the *normalization* factor $1/n$. It requires a division operation and hence divisible quantities, and it amounts to implicitly giving each quantity equal rank. The latter property is largely responsible for the Strong Law of Large Numbers, at least intuitively, in that as n increases each quantity eventually becomes smaller relative to n until each quantity becomes virtually negligible, and all the quantities have left are their shared distribution properties.

Lest one think A_3 artificially rules out the mean by choosing $\emptyset = l$ over $\emptyset = m$, suppose $\emptyset = m$ were chosen instead. Then the corresponding A_3-like condition would still rule out the mean (and median), and for the same reason A_3 does, namely because the Strong Law sends A somewhere dependent only on the underlying probability distribution and in general different from m. A more robust claim can be made. If an admissible knowledge combination operator is at all a function of the epistemic uncertainty of the epistemic situation $(s, X(s))$, i.e., if it depends at all on the knowledge gap $m - l$, the arithmetic mean and like measures of central tendency are inadmissible. The point is dispersion prevails because it measures uncertainty and central tendency does not. And uncertainty measures must surely be incorporated in admissible combination operators.

The next function space constraint summarizes the disposition that *all* we ever really know about the epistemic situation $(s, X(s))$ is the knowledge gap $m - l$. We feel good if it is small, bad if it is large. More generally, given any poset K, the only variables of interest—at any rate the only variables we can trust—are the lower and upper bounds l and m. This strong nonparametric (distribution-free) assumption is akin to the special case of picking (betting on) the median over the mean for arbitrary stimulus environments if those are the only two choices. Intuitions accustomed to mean-squared-error benchmarks, for example, will find it too stringent. On the other hand, nonparametric intuitions accustomed to order statistics frequently proving to be sufficient statistics should find it quite natural. In any event, we codify it as function space assumption A_4.

$$(A_4) \qquad \text{NONPARAMETRICISM:} \qquad \varnothing(X(s)) = \varnothing(l, m).$$

Of course A_4 rules out the mean A (and the median) immediately. For that reason it is listed after A_3, which in principle could follow A_4 in decreasing plausibility and increasing tractability. For it is important to see that limit properties such as the Strong Law of Large Numbers, not epistemic gerrymandering, preclude A's admissibility.

But upon reflection it appears a new mean may be admissible, namely the operator $l + m / 2$. More generally, consider the family of convex combinations $\lambda l + (1 - \lambda)m$ where $0 < \lambda < 1$; we denote this family of functions, which has the cardinality of the continuum, by C. Clearly every element of C satisfies A_1 and A_2. But how about A_3? Again we invoke Theorem 1 from the Appendix. Since, in the random setting, l tends to 0 and m tends to 1, it follows that any convex combination tends to $1 - \lambda$, and hence is inadmissible according to A_3^*/A_3, even though A_4 is clearly satisfied.

There is final property of knowledge combination operators that needs to be mentioned but will not be assumed as a function space constraint. (Hence this mathematico–epistemological theory defines a knowledge combination operator as *admissible* if and only if it satisfies $A_1 - A_4$.) It can be called the *preferred endpoint* property. This property holds that one endpoint of the knowledge response set K is preferred to the other, and in fact m is preferred to l. The raw intuition is that we care about nonfalse propositions and nonempty sets more than their opposites, even though logic does not care (about anything for that matter)—i.e., if the unit interval, say, is taken as the set of truth values or the set of degrees of fuzzy set membership, then more attention is paid to quantities near 1 than 0, and in general it should cost functional procedures more to yield results near 1 than 0. For current purposes this means that how the knowledge gap $m - l$ affects $\varnothing$ should depend on how close m and l are to sup K, e.g., on how small $1 - m + l$ is. The closer m is to 1, the more $\varnothing$ should tend toward l. Likewise, the closer m is to 0, the more $\varnothing$ should tend toward m and thus be more lenient, since then there is in some sense less to lose with optimism. We call this preferred endpoint property *leniency* in the present context.

$$\text{LENIENCY:} \qquad \varnothing(l, m) \uparrow m \quad \text{as} \quad 1 - m \uparrow 1.$$

A less general, more intuitive version of the leniency condition is the following rejoinder to A_3:

$$\varnothing \uparrow m \quad \text{as} \quad m - l \downarrow 0.$$

Note that the above condition does not imply A_3, or vice versa. Note also the stringency of the leniency condition: it rules out all knowledge combination candidates so far discussed, except possibly m, which is ruled out by A_3.

III. SELECTING A KNOWLEDGE COMBINATION OPERATOR

The function space Φ of admissible knowledge combination operators—i.e., those maps $\emptyset$: $K^n \rightarrow K$ that satisfy conditions $A_1 - A_4$ of Section II—is still too large to easily and uniquely select a $\emptyset$. So we must guess a $\emptyset \in \Phi$ and see how it behaves. But then the problem arises over which guess is best.

We assume the simplest guess is best justified. And the simplest guess is just $\emptyset = l$. Indeed $\emptyset = l$ enjoys a distinguished status in Φ because of its equivalence with the Bellman–Zadeh fuzzy decision operator and because l is the upper bound on two function spaces of triangular or t-norms, to be defined shortly, namely those n-placed t-norms defined on K^n and those two-placed t-norms defined on all (l, m) pairs of K^2. But $\emptyset = l$ is to simple. It ignores the uncertainty measured by the knowledge gap $m - l$.

So let us guess as follows. Since $m - l$ is a rough measure of what we do not know, let us take its negation $1 - m + l$ as a rough measure of what we do know, i.e., guess $\emptyset = 1 - m + l$. This operator satisfies A_2 since it satisfies A_4. More importantly, this operator satisfies A_3, and satisfies it linearly, since it is inversely related to the knowledge gap $m - l$. A_3* is likewise satisfied, as application of Theorem 1 of Appendix II immediately shows. A problem occurs, though, with the boundedness condition A_1. For if $(l, m) = (0.2, 0.4)$, say, then $1 - 0.4 + 0.2 = 0.8 > 0.4$, and this cannot be.

For $1 - m + l \leq m$, $\emptyset = 1 - m + l$ behaves as desired. The simplest way to deal with the offending cases is to put $\emptyset = m$ in those cases. Hence we are led to select the following knowledge combination operator:

$$\emptyset(l, m) = \min(m, 1 - m + l).$$

Does this $\emptyset$ satisfy the rest of the boundedness condition? Does $\emptyset \geq l$ always hold? Suppose not: $\emptyset < l$. Then $1 - m + l < l$. Subtracting l from both sides then gives $m > 1$, a contradiction. Hence $\emptyset \in \Phi$.

$\emptyset$'s behavior on the unit square is displayed in Table I. $\emptyset$ is lenient because it penalizes more for the same absolute difference in the knowledge gap $m - l$ occurring nearer 1 than nearer 0: e.g., both $(0.1, 0.4)$ and $(0.6, 0.9)$ have a 0.3 gap, but $\emptyset(0.1, 0.4) = 0.4$ while $\emptyset(0.6, 0.9) = 0.7$. Underlying this observed leniency is a fundamental decomposition of $\emptyset$. Recall from elementary real analysis the identity $\min(x, y) = \frac{1}{2}(x + y - |x - y|)$ for real numbers x and y. Hence $\emptyset(l, m) = \frac{1}{2}(1 + l - |2m - (1 + l)|)$. Hence if $m \geq \frac{1}{2}(1 + l)$, then $|2m - (1 + l)| = 2m - 1 - l$; otherwise, $|2m - (1 + 1)| = 1 + l - 2m$. Hence

$$\emptyset(l, m) = \begin{cases} 1 - m + l & \text{if } m \geq \frac{1}{2}(1 + l) \\ m & \text{if } m \leq \frac{1}{2}(1 + l). \end{cases}$$

The leniency of $\emptyset$ is due to putting $\emptyset = m$ when $m \leq \frac{1}{2}(1 + l)$. More generally we may wish to exploit the knowledge gap behavior when $m \leq 1 - m$ and thus we may select $\emptyset^*$:

$$\emptyset^*(l, m) = \begin{cases} \min(m, 1 - m + l) & \text{if } m \geq 1 - m \\ \min(l, m - l) & \text{if } m \leq 1 - m. \end{cases}$$

Table I. Behavior of ø$(1, m) = \min(m, 1 - m + l)$ on the unit square.

							m					
	ø(l, m)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	0	0	0.1	0.2	0.3	0.4	0.5	0.4	0.3	0.2	0.1	0
	0.1	*	0.1	0.2	0.3	0.4	0.5	0.5	0.4	0.3	0.2	0.1
	0.2	*	*	0.2	0.3	0.4	0.5	0.6	0.5	0.4	0.3	0.2
	0.3	*	*	*	0.3	0.4	0.5	0.6	0.6	0.5	0.4	0.3
	0.4	*	*	*	*	0.4	0.5	0.6	0.7	0.6	0.5	0.4
l	0.5	*	*	*	*	*	0.5	0.6	0.7	0.7	0.6	0.5
	0.6	*	*	*	*	*	*	0.6	0.7	0.8	0.7	0.6
	0.7	*	*	*	*	*	*	*	0.7	0.8	0.8	0.7
	0.8	*	*	*	*	*	*	*	*	0.8	0.9	0.8
	0.9	*	*	*	*	*	*	*	*	*	0.9	0.9
	1	*	*	*	*	*	*	*	*	*	*	1

*Indicates that ø is undefined on (l, m) pairs where $l > m$.

Note that ø $= 1 - m + l$ if $m \geq \frac{1}{2}(1 + l)$, and ø $= m$ if not. Note further that ø $= m$ when $m \leq 1 - m$, and hence when $m \leq \frac{1}{2}$. Finally, the piecewise continuity of ø implies that ø values between consecutive row or column entries are between those entries (except on rows where ø changes direction at the midpoint between two consecutive and equal entries).

The dual-like portion $\max(l, m - 1)$ allows the knowledge gap to be taken as the knowledge k of the epistemic situation $(s, X(s))$ except when it falls below the lower bound l. Note that ø* satisfies A_1 since if $\max(l, m - l) > m$, then $l \leq m$ implies that $m - l > m$, which implies that $l < 0$, impossible. Note further that ø* ≤ ø. Does this property generalize? Is ø the upper bound on Φ?

Let us replace the minimum with product in ø to get ø**$(l,m) = m - m^2 + ml$. Then ø** ≤ ø trivally. Admissibility follows since, first, $m - m^2 + ml > m$ implies $l > m$ and, second, $m - m^2 + ml < l$ implies $m(1 - m) < l(1 - m)$ and thus, for positive m, implies $m < l$, absurd. (A_3 follows since in the uncertain limit ø** $= 1 - 1 + 0 = 0$ by Theorem 1.) This result is further evidence for the conjecture that ø is the largest of all admissible knowledge combination operators.

But the conjecture is false. A less general version is true, though. For let $\Theta(l, m) = \min(m, \sqrt{1 - m + l})$. Then Θ is admissible since if $\Theta < l$, then $1 - m + l < l^2$, and then $1 - m < l(l - 1) < 0$, which implies $m > 1$, impossible. But note that ø$(0.04, 1) = 0.04 < 0.2 = \theta(0.04, 1)$.

The difference between ø** and Θ is that ø** modified the operation ø used on its arguments while Θ modified its arguments. The distinction is critical. We will show variations such as ø** that are reasonable in the sense of being t-norms always lie between our simpliest guesses that ø $= l$ and ø $= \min(m, 1 - m + l)$. This theorem will complete the first part of this theory of knowledge combination; the second part will be completed when this theorem is extended to account for weighted knowledge sources. The first task is to give a working definition of t-norms and t-conorms.

A *triangular norm* or *t-norm* T is an n-to-1 function T: $K^n \to K$ that generalizes minimum (infimum) on K. For present purposes we will restrict ourselves to the two-dimensional case (the n-dimensional case is an immediate extension of notation) and to the poset I, the unit interval, the case most extensively studied in the t-norm literature. Then a t-norm T: $I^2 \to I$ generalizes min in that it satisfies $T_1 - T_4$:

$$(T_1) \quad \text{BOUNDARY:} \quad T(x, 1) = x,$$

$$(T_2) \quad \text{SYMMETRY:} \quad T(x, y) = T(y, x),$$

$$(T_3) \quad \text{MONOTONICITY:} \quad T(x, z) \leq T(y, z) \quad \text{if } x \leq y,$$

$$(T_4) \quad \text{ASSOCIATIVITY:} \quad T(x, T(y, z)) = T(T(x, y), z)$$

for all $x, y, z \in I$. A *t-conorm* S: $I^2 \to I$ generalizes max by satisfying $S_1 - S_4$, where S_1 is $S(x, 0) = x$ and $S_2 - S_4$ are identical with $T_2 - T_4$ when T is replaced by S in the definitions.

All t-norms and t-conorms satisfy two general properties. First, they are, respectively, bounded above and below by min and max:

$$T(x, y) \leq \min(x, y) \leq \max(x, y) \leq S(x, y)$$

for all T and S. A simple proof of this fact is given in Yager.[7] Recall that, most generally considered, negation is an *order-reversing* operation on a poset. (In Klement,[8] several unit-interval characterizations of negation are surveyed.) On I the most popular negation operator N: $I \to I$ is simply $N(x) = 1 - x$, where it happens that N^2 is the identity operator. The second property is that, for every negation operator, every t-norm T and t-conorm S has associated with it a *De Morgan dual* t-conorm S' and t-norm T', respectively, given by

$$S'(x, y) = N(T(N(x), N(y)))$$

and

$$T'(x, y) = N(S(N(x), N(y))).$$

The proof is trivial and essentially amounts, in the case of S', to using the order-reversing property of N on T_3 to give $S'(x, 0) = x$ instead of T_1, and similarly in the case of T'. (Note here, though, that $N^2(x) = x$ must hold.) This property is simply a generalization of the relationship $\min(x, y) = 1 - \max(1 - x, 1 - y)$ and similarly for max. The important consequence is that a De Morgan-dual relationship in general does not hold between arbitrary t-norms and t-conorms.

Before exploring the function space of t-norms let us observe two properties of $\phi = l$. First, consider the space of all (l, m) pairs in I^2, i.e., the upper triangular region of the unit square. Then there is exactly one t-norm on this space that satisfies A_1, namely $\phi = l$, since $T(x, y) \leq \min(x, y) = l$ for all t-norms. Next, let us return to the knowledge response vector $X(s)$ and its associated product space K^n. Then by a direct extension to n-place t-norms of the above two-place argument, $T(X_1(s), \ldots, X_n(s)) = l$ is again the only admissible t-norm. This motivates the conjecture that $\phi = l$ enjoys a special status in Φ.

What is the smallest t-norm on I^2? The smallest t-norm T_s is, as expected, always zero except when one of the t-norm axioms $T_1 - T_4$ must be satisfied, and the only axiom that is not satisfied in the zero-degenerate case is the boundary condition T_1. Hence

$$T_s(x, y) = \begin{cases} \min(x, y) & \text{if } \max(x, y) = 1 \\ 0 & \text{if } \max(x, y) < 1. \end{cases}$$

Then clearly for every t-norm T, $T_s \leqslant T \leqslant \min$ holds. But this function space of t-norms is too large for our purposes. For it includes discontinuous t-norms like T_s, i.e., when ϕ is I-valued, it must be a continuous function of its arguments. More generally, whenever our poset K is equipped with a metric, small changes in the knowledge response vector $X(s)$ should only produce small changes in $\phi(X(s))$, i.e., $c(X, X')$ small should imply $d(\phi(X), \phi(X'))$ small where d is the poset metric and c is a compound metric defined in terms of d.

So what is the smallest *continuous* t-norm T_c? As discussed in Goodman and Nguyen,[9] T_c is given by

$$T_c(x, y) = \max(0, x + y - 1).$$

T_c is quite ubiquitous. It is not only the lower bound on the continuous t-norms, but is also the lower bound on the two-place *copulas*, which are essentially bivariate distribution functions with uniform marginal distributions (again see Ref. 9). Copulas are essentially continuous t-norms that are not associative but instead satisfy a positive increasing monotone property; they, like continuous t-norms, are natural phenomena on probabilistic metric spaces, where prefuzzy theory mathematicians first studied both (see, for instance, the pioneering works of Menger,[10] Schweizer and Sklar,[11,12] and Ling[13]). Then there are the Frank t-norms. Noting that $x + y = \min(x, y) + \max(x, y)$ holds for all real x and y, Frank showed,[14] among other things, that the class of t-norms that satisfy $x + y = T(x, y) + S'(x, y)$, where S' is the De Morgan dual of T, are bounded according to $T_c \leqslant T \leqslant \min$. As Prade points out,[15] Frank t-norms are required for probability-functional operators of probabilistic conjunctions, i.e.,

$$\text{Prob}(A \text{ and } B) = T(\text{Prob}(A), \text{Prob}(B)).$$

Then there are the Yager t-norms T_p defined by

$$T_p(x, y) = 1 - \min(1, [(1 - x)^p + (1 - y)^p]^{1/p})$$

and parametrized by $p \geqslant 0$. Note that $T_1 = T_c$ by De Morgan's Law. Moreover, as discussed in Yager,[16] the T_p are increasing with p; in fact, $T_c = T_1 \leqslant T_p \leqslant T_\infty = \min$ (indeed T_p decreases to T_s as p decreases to 0), again revealing T_c's esteemed status in the space of t-norms.

Decompose the function space of t-norms Γ on I^2 into the following two components:

$$\Gamma_1 = \{T\colon T(x, y) \geqslant \max(0, x + y - 1)\}$$

and

$$\Gamma_2 = \{T\colon T(x, y) < \max(0, x + y - 1)\}.$$

In what follows we are only concerned with t-norms taking m and $1 - m + l$ as their arguments. The t-norms need not be continuous; the point of the previous paragraph was to focus attention on T_c as a pivotal t-norm. Further, the following theorem, which says that Γ_1 is admissible and Γ_2 is inadmissible, is understood as counting copula operators as admissible.

THEOREM.

IF the t-norms in Γ take only m and $1 - m + l$ as arguments,
THEN $\Gamma_1 \subset \Phi$ and $\Gamma_2 \subset \Phi^c$.

PROOF. For all $T \varepsilon \Gamma$, A_4 is satisfied by assumption. Hence A_2 is satisfied.

Next, note that the monotonicity condition T_3 implies that as the knowledge gap increases, $T \downarrow T_c$ for $T \varepsilon \Gamma_1$ and $T \downarrow T_s$ for $T \varepsilon \Gamma_2$. The thrust of the proof is the following equality:

$$T_c = \max(0, m + 1 - m + l - 1) = \max(0, l) = l.$$

Thus in one stroke all Γ_1 t-norms satisfy A_1 and A_3 and no Γ_2 t-norms do, and the result follows. Q.E.D.

Hence all "interesting" knowledge combination t-norms T are such that $l \leq T \leq \min(m, 1 - m + l)$. The robustness of this claim depends on the extent to which the difference operator "$-$" generalizes to different K posets.

Finally, a fuzzy-set interpretation of the selected knowledge combination operator $\emptyset(l, m) = \min(m, 1 - m + l)$ is instructive. Let L and M represent the intersection and union of knowledge sets as defined by membership functions l and m. Then

$$\begin{aligned}\emptyset(L, M) &= M \cap (M - L)^c \\ &= M \cap (M \cap L^c)^c \\ &= M \cap (M^c \cup L) \\ &= (M \cap M^c) \cup L \quad \text{since } M \cap L = L.\end{aligned}$$

Hence $\emptyset \downarrow L$ as $M \uparrow X$, where X is the entire space (e.g., $X = S$), again exhibiting subsumption of the Bellman–Zadeh fuzzy decision theory. This result is only interesting for fuzzy sets since only then can $M \cap M^c \neq \emptyset$ occur. The smaller M is relative to M^c, the greater contribution M makes to $\emptyset$. Again the generality of this argument depends on the generality of the fuzzy set difference operator.

IV. WEIGHTING THE KNOWLEDGE SOURCES

Some knowledge sources are more credible than others. Every knowledge combination theory must deal with this. On the present theory, all knowledge sources $X_1, \ldots, X_n$ have been assumed equally (and maximally) credible. And this heroic assumption presents immediate problems. For instance, if there are 50 knowledge sources and 49 respond with a 0.9 but one responds 0.1, then $\emptyset(0.1, 0.9) = 0.2$—a sort of rotten apple property, or so it appears. In practice one expects such outliers to have very high or, usually, very low credibility. We require of course that low-credibility knowledge responses make proportionately less contribution to $\emptyset$ than high-credibility knowledge sources. But, in passing, let it be clear that the popular heuristic of *equally weighting knowledge sources in uncertain environments* is a justifiably expensive heuristic. For this amounts to making experts of everyone. Increasing numbers of knowledge source responses can be expected to produce increasing epistemic entropy, and from expert entropy nothing follows (what if one doctor says operate and a second says otherwise?), i.e., $\emptyset \downarrow 0$ just as surely as if coins were flipped.

Let $w_i \varepsilon K$ denote the *weight* of X_i and let $w = (w_1, \ldots, w_n)$ denote the *knowledge source weight vector*. Then how should X and w be combined by $\emptyset$?—i.e., what is $\emptyset^w$?

The present theory assumes $\emptyset^w(l, m) = \emptyset(l^w, m^w)$. Hence the task is to find weighting functions for l and m. This has the consequence of leaving the Theorem of the previous section intact. We further assume the following problem formulation. We seek *weighting functions* f_i^l and f_i^m such that

$$l^w = \min_i f_i^l(w_i, X_i),$$
$$m^w = \max_i f_i^m(w_i, X_i),$$

where again dependence on $s \varepsilon S$ has been suppressed in the notation for convenience. These assumptions have the virtues of simplicity and (apparent!) tractability; they also permit immediate generalization of the previous knowledge combination results.

Once again we engage in a function space search. And once again we appeal to smoothly varying intuitions. In particular, where should f_i^l and f_i^m tend as w_i traverses the partial order, especially as w_i moves across I? Surely at maximal credibility, when $w_i = 1$, we require $f_i^l = f_i^m = X_i$, and this relationship should be approached as w_i approaches 1. Call this requirement W_1.

$$(W_1) \quad f_i^l(w_i, X_i) \to X_i \text{ and } f_i^m(w_i, X_i) \to X_i \text{ as } w_i \uparrow 1.$$

The situation is more delicate as w_i tends to its lower bound. Consider f_i^l. When $w_i = 0$, $f_i^l(w_i, X_i)$ should make no contribution to l^w (unless $w_i = 0$ for all i) since knowledge source X_i is literally incredible. But since l^w is a minimum operator, we must have $f_i^l(w_i, X_i) = 1$ to annihilate its effect on l^w. More generally, as the credibility w_i of X_i decreases, we require that the weighted response $f_i^l(w_i, X_i)$ increases so that it makes progressively less contribution to l^w. Hence

$$(W_2) \quad f_i^l(w_i, X_i) \uparrow 1 \text{ as } w_i \downarrow 0.$$

For sake of generality, and for simplicity, let us restrict the constituent terms of $f_i^l(w_i, X_i)$ to w_i, X_i min, max, and negation. More specific terms can be substituted—e.g., t-norms for min and $\sqrt{w_i}$ for w_i—without in general violating W_2. Then a simple examination of cases leads to the following specification for f_i^l:

$$f_i^l(w_i, X_i) = \max(1 - w_i, X_i),$$

where we take $1 - w_i$ for not-w_i for ease of examples. Clearly this f_i^l satisfies W_1 and W_2. Suppose, however, $\min(1 - w_i, X_i)$ were chosen instead. Then neither W_1 nor W_2 is satisfied. Suppose, as many are apt to guess in general for a weighting function, the choice is $\min(w_i, X_i)$. Then f_i^l tends to X_i as w_i tends to 1 in accord with W_1, but f_i^l follows w_i to 0 in violation of W_2. The latter consequence is a genuine bad apple property: it implies that if any knowledge source is *in*credible, they all are in the sense that then $l^w = 0$. Suppose next that the choice is $\max(w_i, X_i)$. Then neither W_1 nor W_2 is satisfied.

The choice $f_i^l(w_i, X_i) = \max(1 - w_i, X_i)$ has been arrived at in a different way in a related context by Yager.[17] Yager is concerned with how to condition $C(O)$, the degree to which decision object O satisfies constraint or criterion C, on $w(C)$, the weight of the constraint. Yager simply interprets this conditioning process as asserting the conditional IF $w(C)$, THEN $C(O)$ in a fuzzy logic. A classical material-implication interpretation of the conditional leads to not-$w(C)$ OR $C(O)$, which in turn leads to the Kleene[18] implication operator max(not-$w(C)$, $C(O)$), or $\max(1 - w(C), C(O))$ on I, which is just f_i^l. In Ref. 17 Yager points out several interesting properties of this weighting function; in particular, it behaves as a thresholding function in the fashion of classroom grading schemes—no discrimination of grades occurs below the D-minus grade, i.e., F grades. Yager then forms the fuzzy weighted decision function $m_D^w(O) = \min_i \max(1 - w(C_i), C_i(O))$. Hence $m_D^w = l^w$. Further, according to the immediate weighted extension of A_3, call it $A_3{}^w$, $\emptyset^w \downarrow m_D^w$, subsuming Yager's generalized Bellman–Zadeh fuzzy decision theory.

The next task is to select f_i^m. Again let w_i tend to its lower bound. Symmetric with f_i^l, when $w_i = 0$, $f_i^m(w_i, X_i)$ should make no contribution to m^w. Since m^w is a maximum operator, this means $f_i^m(w_i, X_i) = 0$ must hold since then and only then the weighted response is ignored by m^w. More generally, as w_i decreases, f_i^m should decrease as well and thus make progressively less contribution to m^w. Hence

$$(W_3) \quad f_i^m(w_i, X_i) \downarrow 0 \text{ as } w_i \downarrow 0.$$

Examination of cases quickly leads to the selection

$$f_i^m(w_i, X_i) = \min(w_i, X_i).$$

Suppose we chose f_i^m as we chose f_i^l: $f_i^m(w_i, X_i) = \max(1 - w_i, X_i)$. Then W_1 is satisfied but W_3 is not. The latter violation is a symmetric bad apple property: a single $w_i = 0$ forces $m^w = 1$, and the lower any knowledge source's credibility, the more m^w tends to 1. The remaining general choices for f_i^m are ruled out because, similar to the alternative choices for f_i^l, they violate W_1 or W_3. Finally, note that now m^w generalizes the common procedure of weighted combination, namely multiplying knowledge responses by their credibility weights as in the weighted mean

$$A^w = \frac{1}{n} \sum_i^n w_i X_i.$$

Observe that the unweighted knowledge combination theory is subsumed as the special case when $w = (1, 1, \ldots, 1)$, or more generally when w is the vector of all maximal values in the poset K. This follows of course since f_i^l and f_i^w satisfy

W_1. Let us explore an example. Suppose we are given the unweighted (i.e., maximally weighted) knowledge response vector $X = (0.2, 0.6, 0.7, 0.6, 0.8)$ in response to some query. Then the knowledge gap is 0.6 and $\emptyset(0.2, 0.8) = 0.4$. Here the sample mean A is 0.58 with a (biased) sample deviation of 0.20396 or roughly 0.2. Knowledge source X_1 is the outlier here. Suppose we find out X_1 has little credibility, is particular, suppose we are now given $w = (0.3, 0.8, 1, 0.7, 0.9)$. Then the weighted knowledge gap is 0.2, significantly less than the unweighted case, since $l^w = \min(0.7, 0.6, 0.7, 0.6, 0.8) = 0.6$ and $m^w = \max(0.2, 0.6, 0.7, 0.6, 0.8)$. Hence $\emptyset^w = 0.8$. But $A^w = 0.476$, strictly less than A, with an associated sample deviation of roughly 0.24, which is strictly more than in the unweighted case and which further suggests that the variance, defined in terms of A, is an inaccurate measure of epistemic uncertainty. These peculiarities arise from the mean's delicate definition in terms of the special field operations of addition and multiplication (division).

Problems arise as the weight vector w approaches the zero vector. For suppose $w = (0, 0, \ldots, 0)$. Then $l^w > m^w$! The problem can already be seen, as can its solution, in the previous example where X_1 is assigned the relatively low weight of 0.3 and consequently $f_1^l(w_1, X_1) > f_1^m(w_1, X_1)$. Yet in the example $l^w < m^w$. The explanation of course is that *the weighting scheme of $\emptyset^w$ depends on sample size*. This sample size dependency is packed into assumptions $W_1 - W_3$. The smaller the number n of knowledge sources in the epistemic situation, the less chance the min l and max m have to discount low-credibility responses. Now it will always be the case that $\emptyset^w \leq m^w$ holds, but the underlying knowledge-gap arguments still require that $A_1{}^w$ hold: $l^w \leq \emptyset^w \leq m^w$. We need a limit theorem.

Theorem 2 in Appendix II guarantees that $A_1{}^w$ holds for large numbers of knowledge sources when $\emptyset^w = \min(m^w, 1 - m^w + l^w)$. For it says that in the random sampling (i.i.d.) case, l^w tends to 0 and m^w tends to 1 with probability one, and hence in the limit the weighted knowledge gap is always positive. This requires some interpretation. In Theorem 2 the Borel–Cantelli Lemma is used to show that, with probability one, the event $\min((w_k, X_k) = 1$ occurs infinitely often and the event $\max(1 - w_j, X_j) = 0$ fails to occur at most finitely often. These are extremely strong probabilistic statements and it takes a result as solid and fundamental as the Borel–Cantelli Lemma to prove them. Theorem 3 generalizes this to weighting schemes of the form $\min(w_k^1, w_k^2, \ldots, w_k^m; X_k)$ and $\max(w_j^1, w_j^2, \ldots, w_j^m; X_j)$, where any of the weights w_j^i or w_k^i can be of the form w_i or $1 - w_i$ for any finite m. Theorem 3 of course subsumes Theorems 1 and 2 when all $w_j^i = w_k^i = 1$ $(Y_i = 1)$. These are new random sampling theorems involving infinite matrices of i.i.d. (and nondegenerate density) random variables. The thrust of the theorems is that the intuitive endpoint-degeneracy property per row of Theorem 1 seeps over to any new sequence of random variables formed by taking pairwise min or max combinations of finitely many rows in the m-by-∞ matrix of i.i.d. random variables, and this occurs infinitely often with probability one. Put another way, the biggest of the pairwise smallest weighted random responses will be one and the smallest of the pairwise biggest will be zero. As discussed in Appendix II, these theorems can be generalized by dropping the needlessly strong condition of identical distribution from the i.i.d. framework provided the sequence of integrated density functions diverges; independence, however, must be maintained to invoke the Borel–Cantelli Lemma or even, perhaps, the Kolmogorov Zero-One Law (which, without further hypotheses, only says that the functions lim inf X_i and lim sup X_i will be constants in the tail sigma-algebra).

Some remarks are also in order about the weighted mean A^w. First, from the previous example we already see that A^w violates $A_1{}^w$ since there $A^w = 0.476 < 0.6 = l^w$. $A_4{}^w$, $\emptyset^w(X(s)) = \emptyset(l^w, m^w)$, is also immediately violated by definition of A^w. Likewise, $A_2{}^w$ is immediately satisfied by the commutativity of multiplication and addition. $A_3{}^w$ is violated because ultimately A^w behaves independently of the weighted knowledge gap $m^w - l^w$. The Kolmogorov Strong Law of Large Numbers, and several variations of it, can be applied both to the sequence of weighted i.i.d. (finite-variance) random variables $w_j X_1$, $w_2 X_2$, . . . or to the sequence of product i.i.d. random variables $W_1 X_1$, $W_2 X_2$, . . . to show probability-one convergence to a weight-adjusted mean or to a joint-distribution mean, etc. Yet another mean-related convergence can be observed from the cross-correlation ergodicity of the stochastic processes W_i and X_i if they are jointly stationary processes. But weighted central tendency still ignores weighted dispersion.

Finally, the t-norm Theorem of the previous section is unaffected by weighting the knowledge sources according to f_i^l and f_i^m. For instance, we still have $\max(0, x + y - 1) = l^w$ when m^w and $1 - m^w + l^w$ are substituted for x and y. The reason is simply that the t-norms (and copulas) in question modify how $\emptyset^w$ combines its weighted arguments, not how the arguments are weighted. Positive weighted knowledge gap limit behavior is assumed.

V. CONCLUSIONS

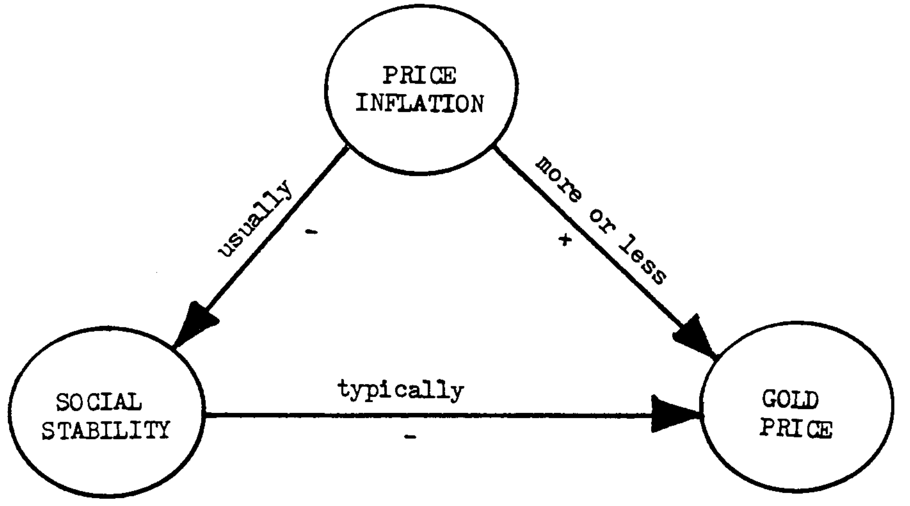
The epistemic uncertainty of an epistemic situation is captured by the cardinality of the knowledge response set K that the knowledge sources map into when queried and by the fuzziness of the partial order on K. The larger K's cardinality or the less precise its partial order, the more uncertain the sequence of epistemic situations $\{(s, X(s))\}_{s \in S}$. Throughout this article we have assumed that K's partial order is crisp (nonfuzzy). This is seldom the case in daily discourse; humans are habitually inconsistent and intransitive in their utterances and somewhat less so in their dispositions. The theory of fuzzy n-place relations (see Ref. 19) is the only mathematical theory currently available for directly representing vague posets; in particular, it naturally permits degrees of reflexivity, anti-symmetry, and transitivity in the fuzzy binary relation case. Most of the present theory remains intact in this general setting since it is built upon the poset structure of K. If a difference operator can be defined, then fuzzy knowledge gaps can be defined and axioms A_3 and $A_3{}^*$, and their weighted extensions, will be in force. Otherwise, $\emptyset = l$, or $\emptyset^w = l^w$, must do, i.e., even here, where knowledge responses such as "I tend to somewhat recommend it" abound, the Bellman–Zadeh fuzzy decision theory applies.

Pushing the level of abstraction further, into the domain of analytic philosophy, how does knowledge relate to truth? The dominant view in modern philosophy, indeed one of its few (near) invariants, is the logical empiricist view that *knowledge is justified true belief.*[20-24] Debates center around justification, truth, and belief—and primarily just these three notions—in philosophical epistemology. Justification is seen as scientific practice, truth as a property of statements or utterances, and beliefs as biological dispositions,[25,26] or as Russell put it,[27] as suspended reactions. The present theory does not commit to this view. It simply, and abstractly, associates a knowledge quantity k with every epistemic situation $(s, X(s))$. There are two ties with truth, though. First, the primacy of the Bellman–Zadeh fuzzy decision theory can be viewed as encapsulating the operative doctrine that *degree of truth* corresponds to *degree of concurrence* among informed knowledge sources in a particular problem domain. This conventionalism is exalted into a lone secular commandment by the influential philosophical logician Willard Van Orman Quine in his "naturalized epistemology,"[28] and generally abounds in the epistemology and philosophy of science literature. The second tie with truth is that $\phi(l, m) = \min(m, 1 - m + l)$ can be viewed as a generalized Łukasiewicz implication (continuous truth) operator.[29] In the multivalued truth framework where $t(S) \in I$ is the amount of truth of statement S and A and B are statements, the truth of the conditional statement C, "IF A, THEN B," can be given by $t(C) = \min(1, 1 - t(A) + t(B))$. The recurring requirement of any fuzzy truth operator is that the truth of the antecedent be bounded by the truth of the consequent, i.e., $t(A) \leqslant t(B)$, i.e., truth cannot imply falsehood, in the old tongue. If m and l are interpreted as statements, then we only have complete truth in the certain case where $m = l$ since otherwise $m > l$. Finally, the philosophical setting provides an opportunity to examine the ontological consequences of a statistical interpretation of knowledge. Here the point estimation framework[30] views the knowledge responses as values taken on by random variables (something we have already done to construct an uncertainty testbed). The random variables have associated distribution functions or probability measures. The family of joint probability (finite product) measures is then assumed indexed by some knowledge parameter k. The point-estimation task is to estimate the parameter k. This is related to truth directly in terms of how accurately k is estimated, but k itself is not truth. k is, as Quine[31] might call it, some absolute fact of the epistemic matter, some privileged translation manual for going from question–answer pairs (epistemic situations) to a single answer in any environment, i.e., knowledge k exists, hanging in the air like a Platonic solid.

Now let us turn to the practical aspects of this knowledge combination theory. The immediate utility of this approach is that it permits *multiple domain experts* to contribute to the knowledge engineering tasks of knowledge acquisition, representation, and, of course, combination in the expert system paradigm. It equally facilitates multiperson multiobjective decision making, but here we limit the discussion to knowledge base building. This utility stems from the poset generality and computational simplicity of this approach. It allows arbitrary many knowledge sources with arbitrary levels of credibility to participate in the knowledge base building process. (If, in a sequence of epistemic situations, some knowledge source X_i does not participate, then $f_i^l = 1$ and $f_i^m = 0$ in l^w and m^w.) The only operations that need be used are min, max, and negation, which are computationally simpler than multiplication, addition, and difference. This poset generality allows more knowledge sources to be queried and queried in a metalanguage closer to the fuzzy object language of the knowledge sources. This in turn allows knowledge responses to be more easily acquired and, further, greater knowledge source concurrence can be expected (since, for instance, in terms of quantifiers the nonfuzzy quantifier *all* implies the fuzzy quantifer *most* but not conversely, see Zadeh[32] for further examples). The standard interrogation of the single domain expert by the knowledge engineer can be substantially generalized. Mass-mailing of questionnaires or expert document transcribing, for instance, can replace one-on-one verbal interrogation, or at least supplement it.

Standard expert systems consist of a collection of condition-action rules or implications with associated uncertainty factors (and other peripheral units!). A powerful application of the present fuzzy knowledge combination theory is to allow several weighted experts, not just one, to determine the uncertainty weights. Different subsets of experts with different credibility weights can be used for different rules. Again the generality obtained from a poset structure tends to increase expert concurrence. And multiple weighted experts concurring on an if-then relationship offers a partial answer, perhaps the best available, to the recurrent expert-system question "Where do the uncertainty numbers come from?" a question that drives many AI researchers to despair.[33,34] Some of the advantages of acquiring knowledge from multiple experts have recently been reported by Dym and Mittal.[35]

This fuzzy knowledge combination theory is especially useful for combining arbitrary fuzzy cognitive maps.[36] In fact, this theory was initiated by just this application. The idea is to let knowledge sources draw and amend causal pictures, and have them drawn and amended for them, to form huge connected knowledge bases in arbitrary problem domains. A fragment of such a fuzzy causal picture from the soft knowledge domain of international sociology might be:



Here the nodes represent variables causal concept nodes, or simply fuzzy sets. A variable quantity like SOCIAL STABILITY can be a node but SOCIETY cannot be (this can be relaxed). Arbitrary data, like newspaper reports, activate all the nodes to different degrees. The directed fuzzy edges represent causality. Fuzzy poset values or weights like *usually* indicate degree of causality. Plus (+) indicates causal increase and minus (−) indicates causal decrease (e.g., if price inflation increases, then social stability usually decreases; if price inflation decreases, then social stability usually increases). On a numeric restriction on the fuzzy causal weights, dynamic (adaptive feedback) fuzzy cognitive maps obey a non-Hebbian learning law and further subsume standard rule-set inference engine and neural net models.[37,38] Dynamic fuzzy cognitive maps also exhibit many associative memory properties. Some of these applications, including a general application of the present fuzzy knowledge combination theory, often require that negative causality be transformed into positive causality by introducing *dis-concepts*[36] (e.g., replace "Price inflation usually decreases social stability" with "Price inflation usually increases social *in*stability"). If arbitrary many knowledge sources, suitably weighted, are queried about the strengths of arbitrary causal connections, then the present fuzzy knowledge combination theory can be applied to quickly produce a single fuzzy cognitive map. Such maps can easily possess millions, even billions, of causal concept nodes and orders of magnitude more causal connections. They can further be augmented and modified as more knowledge sources respond, as they change their responses in light of evidence or other responses, or as they are dynamically reweighted. At this point, fuzzy poset versions of many *artificial neural system*[39–41] properties—adaptive resonance, avalanche activation, rapid spatiotemporal pattern classification—can occur, often without any one mind (or computer) perceiving their occurrence. In principle the knowledge of the ages could be stored in such huge fuzzy cognitive maps. Indeed such knowledge would only initiate, not culminate, the dynamic map learning process.

And this brings us finally to the popular AI topic of conditioning hypotheses on evidence. For if the knowledge of the ages can be dynamically stored and dynamically modified in fuzzy cognitive map structures, then surely conditioning somewhere occurs. In fact it occurs in several places but we restrict attention to the following. When it is realized that conditioning is nothing more than making one thing a function of another, then the act of combining weighted knowledge responses is a generic act of conditioning. For if the epistemic situation $(s, X(s))$ occurs, then we get $k = \emptyset(X(s))$. More generally, if $(s, X(s))$ occurs and we are given weight vector w, then we get $k = \emptyset^{w}(X(s))$. Standard conditioning then results if the antecedent of the latter conditional is interpreted as a description of evidence and the consequent is interpreted as the going hypothesis.

This article is the result of a multiknowledge-source dialectical process. The author is especially grateful for the responses from Ian Abramson, David Brown, Gaston Jennet, Allan McKenzie, and Rod Taber.

APPENDIX I: TECHNICAL GLOSSARY AND SUMMARY

Glossary

S = set of query stimuli, questions
s = an element of S, a question
K = partially ordered set (poset) of knowledge responses, of answers
X_i = ith knowledge source, X_i: $S \rightarrow K$
$X(s) = (X_1(s), \ldots, X_n(s))$, the knowledge response vector
$(s, X(s))$ = an epistemic situation (e.g., question–answer pair)
$k \varepsilon K$ = the knowledge of the epistemic situation $(s, X(s))$
ϕ: $K^n \rightarrow K$ = a knowledge combination function such that $\phi(X(s)) = k$
I = intersection of given knowledge sets
C = combination of given knowledge sets
U = union of given knowledge sets
$l = \min_i X_i(s)$, the least operator
$m = \max_i X_i(s)$, the most operator
$A = \frac{1}{n} \sum_i X_i(s)$ the arithmetic mean
$\prod_i X_i(s)$ = product of the knowledge responses
$P(n)$ = the symmetric group on n, all permutations of $\{1, 2, \ldots, n\}$
$p = (p_1, \ldots, p_n) \varepsilon P(n)$, a permutation
$X_p(s) = (X_{p_1}(s), \ldots, X_{p_n}(s))$, a shuffle of $X(s)$
$m - l$ = the knowledge gap
$D = G_1 \cap \ldots \cap G_n \cap C_1 \cap \ldots \cap C_m$ a Bellman–Zadeh fuzzy decision
C_i = ith constraint set, subset of some space of alternatives
G_j = jth goal set, subset of some space of alternatives
$m_D = \min_{i,j} \{m_{G_i}, m_{C_j}\}$, fuzzy decision membership function
i.i.d. = independent and identically distributed (random variables)
Φ = set of admissible knowledge combination operators
$I = [0, 1]$
T: $I^2 \rightarrow I$ = a triangular norm or t-norm if $T_1 - T_4$ hold:
(T_1) BOUNDARY: $T(x, 1) = x$,
(T_2) SYMMETRY: $T(x, y) = T(y, x)$,
(T_3) MONOTONICITY: $T(x, z) \leq T(y, z)$ if $x \leq y$,
(T_4) ASSOCIATIVITY: $T(x, T(y, z)) = T(T(x, y), z)$
S: $I^2 \rightarrow I$ = a t-conorm
N: $I \rightarrow I$ = a negation operator (e.g., $N(x) = 1 - x$)
$\Gamma_1 = \{T\text{: } T(x, y) \geq \max(0, x + y - 1)\}$
$\Gamma_2 = \{T\text{: } T(x, y) < \max(0, x + y - 1)]$
$w_i \varepsilon K$ = credibility weight of knowledge source X_i
$w = (w_1, \ldots, w_n)$
ϕ^w = weighted knowledge combination operator
l^w = weighted least operator = $\min_i f_i^l(w_i, X_i)$

m^w = weighted most operator = $\max_i f_i^m(w_i, X_i)$
f_i^l = weighted knowledge response in l^w
f_i^m = weighted knowledge response in m^w
$A^w = \frac{1}{n} \sum_i^n w_i X_i$ the weighted mean
$m^w - l^w$ = the weighted knowledge gap

Summary of the Fuzzy Knowledge Combination Theory

(1) Fuzzy Knowledge Combination Problem: Find $\varnothing$ (that is admissible)
(2) $\varnothing \in \Phi$, i.e., $\varnothing$ is *admissible* if and only if $A_1 - A_4$ hold:

(A_1) BOUNDEDNESS: $l \leq \varnothing \leq m$,

(A_2) SYMMETRY: $\varnothing(X(s)) = \varnothing(X_p(s))$ for all $p \in P(n)$,

(A_3) CONSERVATISM: $\varnothing \downarrow l$ as $m - l \uparrow 1$,

$(A_3{}^*)$ $\varnothing \downarrow l$ as $i \rightarrow \infty$ in the random (i.i.d.) case,

(A_4) NONPARAMETRICISM: $\varnothing(X(s)) = \varnothing(l,m)$.

(3) It is desirable but not required that any admissible $\varnothing$ be lenient in one of the following two ways:

LENIENCY: $\varnothing(l,m) \uparrow m$ as $1 - m \uparrow 1$,
$\varnothing \uparrow m$ as $m - l \downarrow 0$.

(4) We choose $\varnothing$ as $\varnothing(l,m) = \min(m, 1 - m + 1)$.
(5) *THEOREM.*

IF the t-norms in Γ take only m and $1 - m + l$ as arguments,
THEN $\Gamma_1 \subset \Phi$ and $\Gamma_2 \subset \Phi^c$.

(6) Weight $\varnothing$ according to $\varnothing^w(l, m) = \varnothing(l^w, m^w)$.
(7) Weight l and m separately:

$$l^w = \min_i \ f_i^l(w_i, X_i),$$

$$m^w = \max_i \ f_i^m(w_i, X_i).$$

(8) Choose f_i^l and f_i^m so that $W_1 - W_3$ hold:

(W_1) $f_i^l(w_i, X_i) \rightarrow X_i$ and $f_i^m(w_i, X_i) \rightarrow X_i$ as $w_i \uparrow 1$,

(W_2) $f_i^l(w_i, X_i) \uparrow 1$ as $w_i \downarrow 0$,

(W_3) $f_i^m(w_i, X_i) \downarrow 0$ as $w_i \downarrow 0$.

(9) Select $f_i^l(w_i, X_i) = \max(1 - w_i, X_i)$ and $f_i^m(w_i, X_i) = \min(w_i, X_i)$.
(10) Hence $l^w = \min_i \max(1 - w_i, X_i)$,

$m^w = \max_i \min(w_i, X_i)$,

$\varnothing^w(l, m) = \min(m^w, 1 - m^w + l^w)$.

APPENDIX II: PROOFS OF PROBABILISTIC UNCERTAINTY THEOREMS

If $X_i(\omega)$ is the response of the ith knowledge source to evidence or query ω, then worst-case properties of knowledge combination schemes used to build knowledge bases can be modeled by viewing $X_1, X_2, \ldots$ as independent identically distributed (i.i.d.) random variables. Independence reflects knowledge source individuality. Identical distribution reflects problem domain focus. Randomness reflects professed knowledge.

More generally, associated with knowledge source X_i is some credibility measure Y_i. The sequence $Y_1, Y_2, \ldots$ can also be viewed, for worst-case analysis, as i.i.d. random variables with similar interpretations of independence, identical distribution, and randomness. The problem, which is both a stochastic and artificial intelligence problem, is to determine the worst-case behavior of the acquired knowledge when the (X_i, Y_i) pairs are combined.

Knowledge responses can be combined in many ways. Whatever the combination technique, the combined responses will have some sort of lower and upper bound. The present view is that these bounds are the limits inferior and superior of the response sequence. We know at least the lim inf, at most the lim sup. So if the X_i take their values in [0, 1], then surely $\liminf_{i\to\infty} X_i = 0$ and $\limsup_{i\to\infty} X_i = 1$, i.e., in the random case we know nothing. When will this be true? More generally, if, in the fashion of fuzzy set theory and logic knowledge source X_i's credibility Y_i is incorporated into X_i's knowledge response in min/max manner, as $X_i \wedge Y_i$ or as $X_i \vee Y_i$, then when will $\liminf_{i\to\infty} X_i \vee Y_i = 0$ and $\limsup_{i\to\infty} X_i \wedge Y_i = 1$ almost surely?

The conditions and proofs needed to answer the questions are surprisingly nontrivial. i.i.d. is not sufficient for almost-sure convergence. Some type of positivity is needed on X_i's and Y_i's distributions. The Borel–Cantelli Lemma then leads to the results. The first question is answered in Theorem 1, the second in Theorem 2, and a generalized question in Theorem 3. Although the answer to the third question answers the second, and the second answers the first, it is easier to prove the theorems in the other direction.

THEOREM 1. IF $(X_i)_{i=1}^{\infty}$ is a sequence of random variables on some probability space (Ω, σ, P) such that for all i:

(1) $X_i\colon \Omega \to [0, 1]$,

(2) the $(X_i)_{i=1}^{\infty}$ are independent and identically distributed,

(3) the corresponding sequence of probability density functions

$$(f_i)_{i=1}^{\infty} \text{ obeys } \int_I f_i(x)dx > 0,$$

where integration is with respect to Lebesque measure, for every interval I of the form $[k, 1]$ or $[0, 1-k]$ for every $k \in (0, 1)$,

THEN (A) $\limsup_{i\to\infty} X_i = 1$ almost surely,

(B) $\liminf_{i\to\infty} X_i = 0$ almost surely.

PROOF. By the Borel–Cantelli Lemma, if $(E_i)_{i=1}^{\infty}$ is a sequence of independents events in σ, then

$$P(\limsup_{i\to\infty} E_i) = P\left(\bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} E_i\right) = 1 \quad \text{if} \sum_{i=1}^{\infty} P(E_i) = \infty.$$

Since, by (2), the X_i are *identically* distributed, it follows from (3) that

$$\sum_{i=1}^{\infty} \int_I f_i(x)dx = \infty.$$

So, for $I = [k, 1]$, (A) follows by putting

$$E_i = \{\omega \ \varepsilon \ \Omega: X_i(\omega) \ \varepsilon \ I\},$$

and thus

$$P(E_i) = \int_I f_i(x)dx.$$

(B) follows from (A) by applying the identity

$$\liminf_{i\to\infty} X_i = -\limsup_{i\to\infty} -X_i$$

and using the reflected sequence $(-X_i)_{i=1}^{\infty}$. Q.E.D.

THEOREM 2. IF the sequences of random variables $(X_i)_{i=1}^{\infty}$ and $(Y_i)_{i=1}^{\infty}$ each satisfy the hypotheses of Theorem 1 and are pairwise independent,*

THEN (A) $\limsup_{i\to\infty} X_i \wedge Y_i = 1$ almost surely,

(B) $\liminf_{i\to\infty} X_i \vee Y_i = 0$ almost surely.

PROOF. Let $Z_i = X_i \wedge Y_i$, $H_i(x) = P(z_i \leq x) = \int_0^x h_i(w)dw$, $F_i(x) = P(X_i \leqq x)$, and $G_i(x) = P(Y_i \leqq x)$. For convenience put $F_i(S) = \int_S dF_i$ and similarly for $G_i(S)$. Then

$$\begin{aligned} H_i(x) &= P(Z_i \leqq x) \\ &= 1 - P(Z_i > x) \\ &= 1 - P(X_i > x \text{ AND } Y_i > x) \\ &= 1 - P(X_i > x)\, P(Y_i > x) \quad \text{by independence,} \\ &= 1 - (1 - F_i(x))\,(1 - G_i(x)). \end{aligned}$$

So

$$\begin{aligned} h_i(x) &= f_i(x)(1 - G_i(x)) + g_i(x)(1 - F_i(x)) \\ &= f_i(x)G_i(x, 1] + g_i(x)F_i(x,1]. \end{aligned}$$

*Else $Y_i = 1 - X_i$ yields a counterexample.

Hence the series

$$\sum_{i=1}^{\infty} \int_I h_i(x)dx = \sum_{i=1}^{\infty} \int_I f_i(x)G_i(x, 1]dx + \sum_{i=1}^{\infty} \int_I g_i(x)F_i(x, 1]dx$$

converges only if

$$\lim_{i \to \infty} \int_I f_i(x)G_i(x, 1]dx = 0,$$

which contradicts the assumptions of identical distribution and

$$\int_I f_i(x)dx > 0.$$

So, as in the proof of Theorem 1, the Borel–Cantelli Lemma implies (A) if

$$E_i = \{\omega \in \Omega: Z_i(\omega) \in I\}$$

and thus

$$P(E_i) = \int_I h_i(x)dx.$$

(B) follows from (A) using the reflected sequences $(-X_i)_{i=1}^{\infty}$ and $(-Y_i)_{i=1}^{\infty}$ and the relationship

$$\liminf_{i \to \infty} X_i \vee Y_i = -\limsup_{i \to \infty} -X_i \wedge -Y_i. \quad \text{Q.E.D.}$$

A direct extension of the proof of Theorem 2 yields the following general result.

THEOREM 3. IF the pairwise independent sequences of random variables

$$(X_i^1)_{i=1}^{\infty}, (X_i^2)_{i=1}^{\infty}, \ldots (X_i^n)_{i=1}^{\infty}$$

each satisfy the hypotheses of Theorem 1 with (1) generalized to $X_i^j: \Omega \to [a, b]$ (and the interval I suitably modified),

THEN (A) $\limsup_{i \to \infty} X_i^1 \wedge X_i^2 \ldots \wedge X_i^n = b$ almost surely.

(B) $\liminf_{i \to \infty} X_i^1 \vee X_i^2 \ldots \vee X_i^n = a$ almost surely.

The infinitely often endpoint degeneracy of Theorem 3 fails for $n = \infty$, since as n gets large the n-factor product in each summed integral in the joint density's decomposition gets arbitrarily small. The hypotheses of Theorem 3 can be weakened by dropping the requirement of identical distribution in (2) (as stated in Theorem 1) and replacing condition (3) with

$$\sum_{i=1}^{\infty} \int_I h_i(x)dx = \infty.$$

To use the Borel–Cantelli Lemma, the latter condition and independence are necessary.

References

1. R. E. Bellman and L.A. Zadeh, "Decision making in a fuzzy environment," *Management Science*, **17**, 141–164 (1970).
2. A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Annals of Mathematical Statistics*, **38**, 325–339 (1967).
3. A. P. Dempster, "A generalization of Bayesian inference," *Journal of the Royal Statistical Society*-B, **30**, 205–247 (1968).
4. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
5. J. Rawls, *A Theory of Justice*, Harvard University Press, Cambridge, MA, 1971.
6. E. O. Wilson, *Sociobiology: The New Synthesis*, Belknap Press of Harvard University Press, Cambridge, MA, 1975.
7. R. R. Yager, "A note on textured sets," *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-11**, 730–731 (1981).
8. E. P. Klement, "Operations on fuzzy sets and fuzzy numbers related to triangular norms," *Proceedings of the 11th International Symposium on Multi-Valued Logic*, 1981, 218–225.
9. I. R. Goodman and H. T. Nguyen, *Uncertainty Models for Knowledge-Based Systems: A Unified Approach to the Measurement of Uncertainty*, North-Holland, New York, 1985.
10. K. Menger, "Statistical metric spaces," *Proceedings of the National Academy of Sciences-USA*, **USA 28**, 1942, 535–537.
11. B. Schweizer and A. Sklar, "Statistical metric spaces," *Pacific Journal of Mathematics*, **10**, 313–334 (1960).
12. B. Schweizer and A. Sklar, *Probabilistic Metric Spaces*, North-Holland, New York, 1983.
13. C. H. Ling, "Representation of associative functions," *Publ. Math. Debrecen*, **12**, 189–212 (1965).
14. M. J. Frank, "On the simultaneous associativity of F(x, y) and x + y − F(x, y)," *Aequat. Math.*, **19**, 194–226 (1979).
15. H. Prade, "A computational approach to approximate and plausible reasoning with applications to expert systems," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **PAMI-7**(3), 260–283 (1985).
16. R. R. Yager, "Some procedures for selecting operators for fuzzy operations," Tech. Rep. RRY-79-05, Iona College, New Rochelle, NY, 1979.
17. R. R. Yager, "A new methodology for ordinal multiobjective decisions based on fuzzy sets," *Decision Sciences*, **12**(4), 589–600 (October 1981).
18. B. R. Gaines, "Foundations of fuzzy reasoning," *International Journal of Man–Machine Studies*, **8**, 623–688 (1976).
19. A. Kaufmann, *Introdution to the Theory of Fuzzy Subsets—Volume I*. Academic, New York, 1975.
20. A. J. Ayer, "Philosophy and knowledge," in *The Problem of Knowledge*, The Chaucer Press, London, England, 1956, Chap. I.
21. E. Gettier, "Is justified true belief knowledge?" *Analysis*, **XXIII.6**(96), 121–123 (June 1963).
22. B. Skyrms, "The explication of 'X knows that p'," *The Journal of Philosophy*, **LXIV**(12), 373–389 (June 1967).
23. R. Chisholm, "Knowledge," in *Theory of Knowledge*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ, 1977, Chap. 6.
24. M. Kaplan, "It's not what you know that counts," *The Journal of Philosophy*, **LXXXII**(7), 350–363 (July 1985).
25. P. M. Churchland, "Eliminative materialism and propositional attitudes," *Journal of Philosophy*, **LXXVIII**(2), 67–90 (February 1981).

26. W. V. O. "States of mind," *The Journal of Philosophy*, **LXXXII**(1), 5–8 (January 1985).
27. B. Russell, *Human Knowledge: Its Scope and Limits*, Simon and Schuster, New York, 1948.
28. W.V.O. Quine, "Epistemology naturalized," in *Ontological Relativity & Other Essays*, Columbia University Press, New York, 1969, Chap. 3.
29. W. Bandler, and L. Kohout, "Fuzzy power sets and fuzzy implication operators," *Fuzzy Sets and Systems*, **4**, 13–30 (1980).
30. E. L. Lehmann, *Theory of Point Estimation*, Wiley, New York, 1983.
31. W. V. O. Quine, "Translation and meaning," in *Word and Object*, M.I.T. Press, Cambridge, MA, 1960, Chap. II.
32. L. A. Zadeh, "A theory of commonsense knowledge," in *Aspects of Vagueness*, Skala, Termini, and Trillas, Eds., Reidel, Dordrecht, The Netherlands, 1984, 257–295.
33. J. Doyle, "Methodological simplicity in expert system construction: the case of judgments and reasoned assumptions," *AI Magazine* **3**(2), 39–43 (1984).
34. P. R. Cohen and M. Sullivan, "An endorsement-based plan recognition program," *Proceedings of the IJCAI-85*, Los Angeles, CA, 1985, 475–479.
35. C. L. Dym and S. Mittal, "Knowledge acquisition from multiple experts," *AI Magazine*, **6**(2) (1985).
36. B. Kosko, "Fuzzy cognitive maps," *International Journal of Man–Machine Studies*, **24**, 65–75 (1986).
37. B. Kosko, "Adaptive inference," in review.
38. B. Kosko and J. S. Limm, "Vision as causal activation and association," *Proceedings of the SPIE (Society for Photo-optical Instrumentation Engineers) Conference on Intelligent Robots and Computer Vision*, September 1985, volume 579, 104–109.
39. S. Grossberg, "Adaptive Resonance in Development, Perception, and Cognition," in *Mathematical Psychology and Psychophysiology*, S. Grossberg, Ed., American Mathematical Society, Providence, RI, 1981.
40. R. Hecht-Nielsen, "Neural analog processing," *Proc. SPIE*, **360**, 180–189 (1983).
41. R. Hecht-Nielsen, "Nearest matched filter classification of spatiotemporal patterns," *Applied Optics* (in press).