JOURNAL OF MATHEMATICAL PSYCHOLOGY 37, 63–103 (1993)

Lebesgue Logic for Probabilistic Reasoning and Some Applications to Perception

BRUCE M. BENNETT, DONALD D. HOFFMAN, AND PARESH MURTHY

University of California, Irvine

Reasoning with probabilities is essential to many sciences, such as decision theory, expert systems, neural networks, pattern recognition, and perception in general. In this paper we explore a new logic of probabilities, the Lebesgue logic, in which are defined the logical relations ENTAILS, AND, OR, and NOT on collections of probability measures. In particular, given any two probability measures μ and ν, the Lebesgue logic answers questions such as the following: Does μ entail ν? What is the conjunction of μ and ν, i.e., what is μ AND ν? What is the disjunction of μ and ν, i.e., what is μ OR ν? What is the negation of μ? Several properties of the Lebesgue logic emerge. Among them are (1) the Lebesgue logic is not boolean, in general, but is "locally boolean," (2) the AND of the Lebesgue logic is a generalization of Bayes' rule; (3) one can define probability measures on the Lebesgue logic itself, thereby permitting the representation of probabilistic knowledge without requiring any commitment to a particular probability measure; and (4) many probabilistic inferences can be described as morphisms of the Lebesgue logic, i.e., as maps from one collection of probability measures to another, respecting the Lebesgue logics on both. We close by discussing a concrete problem to which the Lebesgue logic may find application: the problem of sensor fusion in vision and other perceptual modalities.

1. INTRODUCTION

Reasoning with uncertain information is a well known and central feature of many sciences. This has led to an extensive literature on the mathematical foundations of probabilistic inference and related topics (e.g., Adams & Levine, 1975; Beltrametti & Cassinelli, 1981; Dempster, 1968; Gudder, 1988; Nilsson, 1986; Shafer, 1976; Suppes, 1966a; Varadarajan, 1985; Zadeh, 1975). It has also led to the application of this mathematical work in many disciplines, such as artificial intelligence (Duda, Hart, & Nilsson, 1981; Fischler & Firschein, 1987; Pearl, 1988), group decision making (Grofman & Owen, 1986), and quantum physics (Beltrametti & Cassinelli, 1981; Gudder, 1988; Varadarajan, 1985).

Our intent here is not to review this literature, but to investigate a recently discovered logic of probabilities, the Lebesgue logic. We begin with a brief terminological background. The term "logic" is generally used to refer to a set

Correspondence and requests for reprints should be sent to Don Hoffman, Department of Cognitive Science, University of California, Irvine, California 92717.

0022-2496/93 $5.00

480/37/1-5

with a partial order[1]; the elements of the set are the "propositions" and the order relation is "entailment." If a and b are two elements in the set then their least upper bound, denoted $a \vee b$, and their greatest lower bound, denoted $a \wedge b$, correspond respectively to disjunction and conjunction of propositions. A zero element of the logic is an element 0 such that $a \vee 0 = a$ for all a. A unit element is an element 1 such that $a \wedge 1 = a$ for all a. If 1 exists then the "complement" of a, if it exists, is an element a' such that $a \wedge a' = 0$ and $a \vee a' = 1$. The complement of 1 is 0. The basic example is the well-known boolean algebra of subsets of some set. Here the elements of the logic are the subsets, entailment is inclusion, $\vee$ and $\wedge$ correspond, respectively, to union and intersection, and so forth. More generally, there is a class of logics called the "orthocomplemented modular lattices" which are ordinarily considered to be the logics of importance, for example, in quantum mechanics. Briefly, these are logics in which $\wedge$ and $\vee$ exist (for any two elements), 0 and 1 exist, and the complement a' of every element a exists. The boolean algebras are in this class and are essentially characterized therein by having the property that $\wedge$ is distributive over $\vee$. However, the logics in the class are not generally distributive but have the property of "modularity," which generalizes distributivity and which we need not define precisely here (but see Varadarajan, 1985).

When we speak of a "logic of probabilities" we mean a logic in which the elements are probability measures on some fixed space Y. Thus, we view these probability measures themselves as expressing propositions. Intuitively, a probability measure μ expresses a proposition of the form "the probability that the outcome is in event A is $\mu(A)$, the probability that the outcome is in event B is $\mu(B)$, ..." For example, suppose that the points in a probability space correspond to individual receptors, i.e., to rods and cones, on a retina. A probability measure μ on this space might express the probability that at a given subset of receptors and at a given instant of time, a photon is captured by some receptor in the set. Thus for each subset A of receptors, the number $\mu(A)$ represents the probability that some receptor within A captures a photon. In this way μ can be viewed as expressing a proposition of the form "the probability that a photon is captured by some receptor in the collection of receptors A is $\mu(A)$, the probability that a photon is captured by some receptor in the collection of receptors B is $\mu(B)$, ..."

The Lebesgue logic of probabilities follows from a natural partial order on collections of probability measures, the Lebesgue order. This partial order corresponds to *entailment* of the associated propositions, and determines, via the operations of least upper bound and greatest lower bound, the disjunction and conjunction of pairs of probability measures. There are many reasons why the Lebesgue order is a natural order to use in revealing a logical structure on a set of probability measures. These are discussed below. We will see that the structure of this Lebesgue logic differs from that of the ordinary "orthocomplemented modular lattices" in several particulars: Lebesgue logic has a 0, but no 1 and no complements, and $\wedge$

[1] Recall that a partial order on a set is a relation, $\leqslant$, on the set that is reflexive ($\nu \leqslant \nu$), antisymmetric ($\nu \leqslant \mu$ and $\mu \leqslant \nu$ implies $\nu = \mu$), and transitive ($\nu \leqslant \mu$ and $\mu \leqslant \sigma$ implies $\nu \leqslant \sigma$).

and $\vee$ are not always defined. However, the Lebesgue logic has the striking property of "local booleanness," so that 1 and complements and distributivity exist locally.

The Lebesgue logic also gives a natural constraint on probabilistic inferences in the following way. An *inference* can often be conceived of as a function, $f: P \to C$, whose domain P is one set of propositions, called the *premises* of the inference, and whose range C is another set of propositions, called the *conclusions*. By a *probabilistic inference* we mean an inference whose domain P and range C are each collections of probability measures. Suppose now that $f: P \to C$ is a probabilistic inference, and denote the Lebesgue order on P by $\leqslant_P$ and the Lebesgue order on C by $\leqslant_C$. Then, as we discuss later, a natural constraint on f is that f be a *morphism* of the Lebesgue order: this means, in part, that for $p_1, p_2 \in P$, if $p_1 \leqslant_P p_2$ then $f(p_1) \leqslant_C f(p_2)$.

In our discussion of the Lebesgue logic we have decided to freely review elementary but technically significant facts from measure theory when such review is likely to help the reader. We have also decided to present a concrete application of the Lebesgue logic, our intent being to aid the reader in understanding the logic and to make explicit the possible applications of the logic to concrete problems. The concrete problem we have chosen is the integration of probabilistic information, from vision and other perceptual modalities, into a coherent perception of the external world, a problem referred to in the machine perception literature as the problem of "sensor fusion" and in the perceptual psychology literature as "cue integration." Perhaps it goes without saying that we neither intend to nor succeed in resolving this problem. It nevertheless provides an engaging context in which to consider the properties of the Lebesgue logic.

We try to give a complete presentation of the elementary properties of the Lebesgue logic. To this end we give proofs of lemmas and propositions in full detail. We give definitions of many mathematical terms in footnotes. Our goal is to give the reader—even one not expert in measure theory—the tools needed for a complete understanding of the elements of the Lebesgue logic.

2. The Lebesgue Order

Probability measures serve, we just said, as the premises and as the conclusions for probabilistic inferences. To understand the logic of probability measures we begin, in this section, by investigating the logic of *all* probability measures on an arbitrary measurable space. Once this logic is clarified we can then, by restriction, induce logics on various subcollections of probability measures.

Here, in summary, is what we find. There is a natural definition of entailment among probability measures, which can be modeled formally by a partial order—the Lebesgue order—on probability measures. This definition of entailment by itself completely determines a logic on probability measures—the Lebesgue logic. The Lebesgue logic is not boolean, in general, but is "locally boolean." It describes, among other things, (1) when one probability measure entails another, (2) when one can take the AND of two probability measures, (3) what that AND is when it exists, (4) when one can take the OR, and (5) what it is. For many pairs of probability measures the AND exists and is nontrivial in the Lebesgue logic; such pairs are said to be *simultaneously verifiable.* The information from such pairs of measures can be combined into a single probability measure. However, in the Lebesgue logic certain pairs of probability measures do not have a nontrivial AND; such pairs are not simultaneously verifiable and attempts to integrate them must fail. When this happens in human perception, as we shall see, the information from the two measures is not integrated into one unified percept, but instead can give rise to two or more distinct percepts which are perceived one at a time—so-called *multistable percepts.*

We begin our study of the Lebesgue order by introducing some notation.

Notation 1. Let Y be a set. A collection $\mathscr{Y}$ of subsets of Y is called a *σ-algebra* if it contains Y itself and is closed under countable union and complement. Then a *measurable space* is a pair $(Y, \mathscr{Y})$ where $\mathscr{Y}$ is a σ-algebra of subsets of Y. The subsets in $\mathscr{Y}$ are called *events.* A *measure* is a function $\mu: \mathscr{Y} \to [0, \infty]$ such that (1) $\mu(\varnothing) = 0$ and (2) μ is σ-additive, i.e., for any countable collection A_i of pairwise disjoint events, $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$. A *probability measure* is a measure that assigns the value 1 to Y. Then $\mathscr{M}(Y)$, or just $\mathscr{M}$, denotes the collection of positive finite measures, together with the null measure 0, on $(Y, \mathscr{Y})$. $\bar{\mathscr{M}}(Y)$, or just $\bar{\mathscr{M}}$, denotes the collection of probability measures, together with the null measure, on $(Y, \mathscr{Y})$. For $v \in \mathscr{M}$, we define $\|v\| = v(Y) = \int_Y v(dy)$ and, for $v \neq 0$, $\bar{v} = v/\|v\|$.

We will also need the notion of a measure algebra. We recall the following definition.

DEFINITION 2. Let $(Y, \mathscr{Y})$ be a measurable space and μ a measure on Y. We define an equivalence relation $\equiv$ on $\mathscr{Y}$ as follows: For $A, B \in \mathscr{X}$, $A \equiv B$ iff $\mu(A \bigtriangleup B) = 0$ (where $\bigtriangleup$ is symmetric difference, i.e., $A \bigtriangleup B = (A - B) \cup (B - A)$). Let $\mathscr{Y}_\mu$ denote the set of equivalence classes. It is called the *measure algebra modulo μ.*

One easily checks that if $A \equiv A'$ and $B \equiv B'$ then $A \cup B \equiv A' \cup B'$ and $(A')^c \equiv A^c$. It follows that the notions of $\bigcup$ and $\bigcap$ and complement are well defined for equivalence classes of $\equiv$, so that $\mathscr{Y}_\mu$ inherits the structure of a σ-algebra. This justifies the terminology "measure algebra." Let $[A]$ denote the class of A. We will write "$[A] \leqslant [B]$" iff $\mu(A - B) = 0$, i.e., iff $A \subset B$ up to a μ-negligible set. It is clear that this definition is independent of the choice of representative of the equivalence class. Moreover, one checks that the operations of $\bigcap$ and $\bigcup$ on the equivalence classes are, respectively, the greatest lower bound and least upper bound for this relation $\leqslant$.

We want a logic on $\bar{\mathscr{M}}$. A principled way to construct such a logic is to define a relation of entailment among probability measures. Once we have done this the entire logic is fixed; it is a matter of proof, not of further definition, to spell out the notions of conjunction, disjunction, and negation that all follow from the one definition of entailment. The key step to the whole enterprise, then, is to define an appropriate relation of entailment among probability measures. To do this we introduce the *Lebesgue order* on measures, and then stipulate that measure μ entails measure ν if μ is less than ν in the Lebesgue order.

Before introducing the Lebesgue order it will be instructive to consider briefly two other examples of orders on measures. The first is as follows: for $\nu, \mu \in \bar{\mathcal{M}}(Y)$, define that ν is less than μ iff ν is absolutely continuous with respect to μ.[2] The problematic feature of this definition for our purposes is that it defines a partial order on measure classes;[3] for this reason the order is called the *class order*. This means that it ignores a lot of information contained in probability measures. Probability measures assign numerical values to *all* events in the measurable space, but the class order cares *only* about which events are assigned measure zero. Nevertheless these properties of the class order might make it useful as a first coarse approximation to a logic of probabilities.

A second order, unsuitable to our purposes, is the following: ν is less than μ iff for each event $A \in \mathscr{Y}$ it happens that the real number $\nu(A)$ is less than the real number $\mu(A)$. The reason this is unsuitable is that if ν and μ are probability measures the order becomes trivial: one probability measure is less than another only if the two are identical. This order is, of course, nontrivial on the collection of all *finite* measures, but for our purposes, for a logic of probabilities, we need a notion of order that is nontrivial when restricted to the probability measures.

Later we discuss other orders on probability measures (see Remarks 78–81). The conclusion will be that among naturally defined orders on probability measures the Lebesgue order seems to be uniquely suited as a logic for perception.

We now consider the Lebesgue order. To state it we must first recall the Lebesgue decomposition theorem (see, e.g., Royden, 1988) and then introduce some notation.

LEBESGUE DECOMPOSITION THEOREM. *Given any two measures $\nu, \mu \in \mathcal{M}$, the measure μ can be written* uniquely *as the sum of two measures, one measure absolutely continuous with respect to ν and one singular with respect to ν.*

The Lebesgue decomposition is conveniently written in the following notation.

Notation 3. For $\nu, \mu \in \mathcal{M}$, we denote the Lebesgue decomposition of μ with respect to ν by $\mu = \mu_\nu + \mu^\nu$, where μ_ν is absolutely continuous with respect to ν and where μ^ν is singular with respect to ν.

[2] Recall that ν is *absolutely continuous* with respect to μ, written $\nu \ll \mu$, iff for every $A \in \mathscr{Y}$ for which $\mu(A) = 0$ it happens that $\nu(A) = 0$ as well. ν and μ are *mutually singular*, written $\nu \perp \mu$, if there exist events A, $B \in \mathscr{Y}$ such that $Y = A \cup B$ and $\nu(A) = \mu(B) = 0$.

[3] Two measures are in the same measure class if they are mutually absolutely continuous. Measure classes form a partition of the collection of finite measures on a measurable space. The measure class associated to a particular measure ν is denoted $[\nu]$.

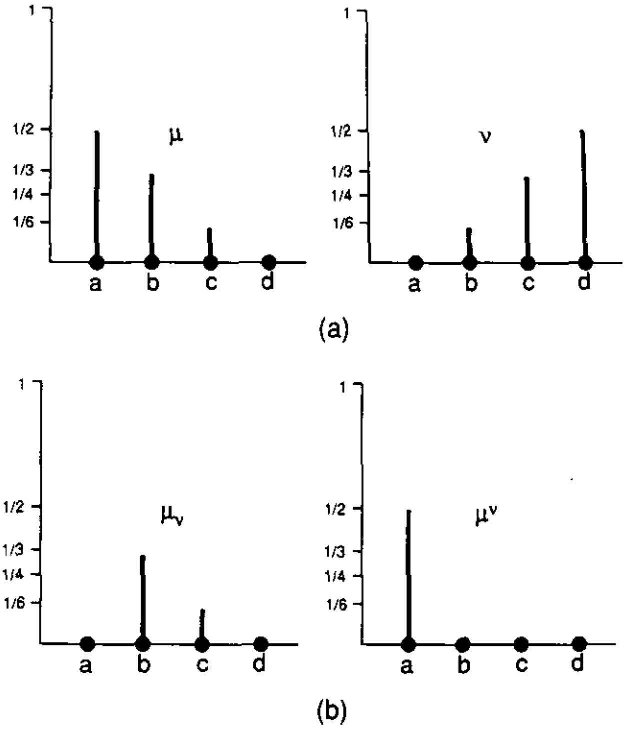

FIG. 1. The Lebesgue decomposition of μ with respect to ν. Part (a) shows the measures μ and ν. Part (b) shows μ_ν, viz., the part of μ that is absolutely continuous with respect to (roughly, overlaps) ν. Part (b) also shows μ^ν, viz., the part of μ that is singular with respect to (roughly, does not overlap) ν.

The Lebesgue decomposition is illustrated in Fig. 1. One reason for Notation 3 is given by the following lemma.

LEMMA 4. *If λ, ν, $\mu \in \mathscr{M}$ then one can write, without ambiguity,*

$$\mu^\nu_\lambda, \qquad \mu^{\nu\lambda} = \mu^{\lambda\nu}, \qquad \mu_{\nu\lambda} = \mu_{\lambda\nu}. \tag{5}$$

Proof. Taking the Lebesgue decomposition of μ with respect to λ we obtain

$$\mu = \mu_\lambda + \mu^\lambda.$$

Taking the decomposition of this with respect to ν gives

$$\mu = (\mu_\lambda)^\nu + (\mu_\lambda)_\nu + (\mu^\lambda)^\nu + (\mu^\lambda)_\nu. \tag{6}$$

If instead we decompose μ first with respect to ν and then with respect to λ we obtain

$$\mu = (\mu_\nu)^\lambda + (\mu_\nu)_\lambda + (\mu^\nu)^\lambda + (\mu^\nu)_\lambda. \tag{7}$$

Now, by the uniqueness of the Lebesgue decomposition, the terms in (6) and (7) that are singular with respect to ν must be equal. Thus

$$(\mu_\lambda)^\nu + (\mu^\lambda)^\nu = (\mu^\nu)^\lambda + (\mu^\nu)_\lambda. \tag{8}$$

Again, by uniqueness of the Lebesgue decomposition, the terms in (8) that are singular with respect to λ must be equal, as must those that are absolutely continuous with respect to λ. Thus

$$(\mu_\lambda)^\nu = (\mu^\nu)_\lambda \qquad \text{and} \qquad (\mu^\lambda)^\nu = (\mu^\nu)^\lambda. \tag{9}$$

If we now equate the terms in (6) and (7) that are absolutely continuous with respect to ν we find

$$(\mu_\lambda)_\nu + (\mu^\lambda)_\nu = (\mu_\nu)^\lambda + (\mu_\nu)_\lambda. \tag{10}$$

But by the uniqueness of the Lebesgue decomposition, the terms in (10) that are singular with respect to λ must be equal, as must those that are absolutely continuous with respect to λ. Therefore

$$(\mu_\lambda)_\nu = (\mu_\nu)_\lambda \qquad \text{and} \qquad (\mu^\lambda)_\nu = (\mu_\nu)^\lambda. \tag{11}$$

From (9) and (11) we see that we can drop parentheses, and the result follows. ∎

This lemma indicates that it doesn't matter, in many cases, in which order the Lebesgue decompositions are taken. One useful property of the Lebesgue decomposition that we use in the sequel is the following.

PROPOSITION 12. *Let $\mu, \nu \in L(Y)$. Then μ_ν and ν_μ are in the same measure class.*

Proof. The Lebesgue decomposition of μ_ν with respect to ν_μ is

$$\mu_\nu = (\mu_\nu)_{(\nu_\mu)} + (\mu_\nu)^{(\nu_\mu)}.$$

But $(\mu_\nu)^{(\nu_\mu)} = 0$ since $\mu_\nu^\nu = 0$. Thus $\mu_\nu \ll \nu_\mu$. A similar argument shows that $\nu_\mu \ll \mu_\nu$. ∎

It will also be useful in the sequel to use a refinement of the Lebesgue decomposition due to Hahn, which we call the Hahn presentation.

HAHN PRESENTATION THEOREM. *If μ and ν are positive finite measures on $(X, \mathscr{X})$ then there is a partition of X into three sets—S_μ, $A = A_{\mu,\nu}$, and S_ν—such that $[\mu|_A] = [\nu|_A]$ and $\nu(S_\mu) = 0 = \mu(S_\nu)$.*

Proof. Let A and $B = A^c$ be complementary support sets, respectively, for μ_ν and μ^ν. Let C and $D = C^c$ be complementary support sets, respectively, for ν_μ and ν^μ. By definition of D, $\nu_\mu(D) = 0$ and hence $\mu_\nu(D) = 0$ (by Proposition 12). Therefore we can replace A by $A - D$ and B by $B \cup D$, i.e., we can assume that $\nu^\mu(A) = 0$, and thus that B is a support set for ν^μ and μ^ν. But $\nu^\mu \perp \mu^\nu$ so that B itself can be decomposed into complementary support sets S_μ and S_ν, respectively, for μ^ν and ν^μ. Thus we can express X as a disjoint union of sets A, S_μ, and S_ν, where A is a support set for the mutually absolutely continuous measures μ_ν and ν_μ, S_ν is a support set for ν^μ, and S_μ is a support set for μ^ν. ∎

The Lebesgue decomposition and Hahn presentation are related as follows: $\mu^\nu = \mu|_{S_\mu}$, $\mu_\nu = \mu|_A$, and $\nu^\mu = \nu|_{S_\nu}$.

Using this notation we now introduce, on spaces of positive finite measures, the Lebesgue order (so called because its definition depends essentially upon the Lebesgue decomposition theorem). This is the central definition of the paper.

DEFINITION 13. *Lebesgue Order.* For $\nu, \mu \in \mathcal{M}$,

$$\nu \leqslant \mu \quad \text{iff} \quad \mu_\nu = \alpha\nu, \quad \alpha > 0. \tag{14}$$

In this case we say that ν ENTAILS μ.

If instead of considering all positive finite measures one restricts attention to the probability measures, then this definition can be expressed in terms of normalized restrictions,[4] as indicated in the following proposition.

PROPOSITION 15. *Suppose $\nu, \mu \in \bar{\mathcal{M}}(Y)$. Then $\nu \leqslant \mu$ iff there is an $A \in \mathcal{Y}$ so that $\nu(\cdot) = \mu(\cdot \mid A)$. A is uniquely determined up to a μ-negligible set.*

Proof. Suppose $\nu \leqslant \mu$. Then $\mu = \alpha\nu + \mu^\nu$ with $\mu^\nu \perp \nu$. Since the measures μ^ν and ν are singular, we can choose a set A so that $\mu^\nu(A) = \nu(A^c) = 0$. A is unique up to a μ^ν- and ν-negligible set, i.e., up to a μ-negligible set. Moreover, $\mu|_A = (\alpha\nu + \mu^\nu)|_A = \alpha\nu|_A = \alpha\nu$, so that $\nu = \overline{\mu|_A}$ $(= \mu(\cdot \mid A))$. Conversely, suppose $\nu = \overline{\mu|_A}$, i.e., $\mu|_A = \alpha\nu$ for some $\alpha > 0$. Now $\mu = \mu|_A + \mu|_{A^c}$, i.e., $\mu = \alpha\nu + \mu|_{A^c}$ and it is clear that $\mu|_{A^c} \perp \nu$ $(= \mu|_A)$. By the uniqueness of the Lebesgue decomposition we conclude that $\nu \leqslant \mu$. ∎

$\bar{\mathcal{M}}(Y)$ with the Lebesgue order will henceforth be denoted $L(Y)$, or sometimes just L. If instead of considering all probability measures one restricts attention to any subcollection of probability measures of the form $L_\mu(Y) = \{\nu \in L(Y) \mid \nu \leqslant \mu\}$, then the Lebesgue order has yet a further characterization, as indicated in the following proposition.

PROPOSITION 16. *If $\mathcal{Y}_\mu$ is the measure algebra of $\mathcal{Y}$ modulo μ then the map $f\colon \mathcal{Y}_\mu \to L_\mu$ given by $[A] \mapsto \mu(\cdot \mid A)$ is an order isomorphism.*

[4] Let $\mu \in \mathcal{M}(Y)$ and suppose that A is any event in $\mathcal{Y}$ for which $\mu(A) > 0$ (possibly the event Y itself). The *restriction* of μ to the set A, written $\mu|_A$, is defined by $\mu|_A(\cdot) = \mu(\cdot \cap A)$, where the dot indicates an arbitrary event in $\mathcal{Y}$. The *normalized* restriction of μ to the set A, written $\mu(\cdot \mid A)$, is defined by $\mu(\cdot \mid A) = \mu|_A(\cdot)/\mu(A)$.

Proof. First it is clear that if $A \subset B$ then $\mu(\cdot|A)$ is the normalized restriction to A of $\mu(\cdot|B)$ and therefore, by Proposition 15, $\mu(\cdot|A) \leqslant \mu(\cdot|B)$; thus $[A] \leqslant [B] \Rightarrow \mu(\cdot|A) \leqslant \mu(\cdot|B)$ in the Lebesgue order so that f is an order morphism. Now suppose that $\nu \in L_\mu$, i.e., $\nu \leqslant \mu$. Then $\nu = \mu(\cdot|A)$ for some A (by Proposition [15]. A is well defined up to μ-negligible sets, i.e., there exists a well-defined class $[A]$ associated to ν. In this way we get a map $g: L_\mu \to \mathscr{Y}_\mu$. It is easy to check that $f \circ g$ and $g \circ f$ are the respective identity functions on L_μ and $\mathscr{Y}_\mu$, which concludes the proof. ▮

Note that if ν ENTAILS μ then, intuitively, the support of ν is contained in the support of μ. This comports well with the view that measures describe states of uncertainty—e.g., the uncertainty of photon captures at the retina (mentioned in the introduction). According to this view, the measure with the smaller support embodies, in a qualitative sense, less uncertainty and hence corresponds to a stronger proposition: it represents information which *entails* that of the measure with the larger support. These intuitions about support are made more precise in the following lemma.

LEMMA 17. *Let $\nu, \mu \in L(Y)$. If $\nu \leqslant \mu$ in the Lebesgue order then $\nu \ll \mu$. The converse need not be true.*

Proof. If $\nu \leqslant \mu$ then $\nu = \mu(\cdot|A)$ for some $A \in \mathscr{Y}$ (by Proposition 15) so that $\nu \ll \mu$. To show that the converse need not be true let $\nu = (\rho + \sigma)/2$ and $\mu = \rho/3 + 2\sigma/3$, where ρ and σ are distinct probability measures in $L(Y)$. Then $\nu \ll \mu$ (in fact, $[\nu] = [\mu]$) but by construction $\mu_\nu \neq \alpha\nu$ so that $\nu \nleqslant \mu$. ▮

Two intuitions about the Lebesgue order might be of help. First, if $\nu \leqslant \mu$ in the Lebesgue order then μ can be obtained by taking ν, rescaling it, and then adding a piece "outside of" (i.e., singular to) ν. Thus, as mentioned above, μ has a broader support than ν. Second, recall that a measure ν induces an order on events in its measurable space: event A is less than event B if $\nu(A) \leqslant \nu(B)$. If $\nu \leqslant \mu$ in the Lebesgue order then μ_ν induces the same order on events as does ν.

With these intuitions, we now restrict our attention to probability measures and study properties of the Lebesgue order. Although we henceforth consider the Lebesgue order primarily in the context of probability measures, we should emphasize that the Lebesgue order is based on the linear space structure of the space of all measures, not on a structure that is intrinsic to the convex space of probability measures. However, as the next proposition indicates, if we restrict the Lebesgue order to probability measures it becomes a partial order.

PROPOSITION 18. *The Lebesgue order is a partial order on $\bar{\mathscr{M}}$.*

Proof. For all $\nu \in \bar{\mathscr{M}}$, $\nu_\nu = \nu$, so $\nu \leqslant \nu$, and $\leqslant$ is therefore reflexive. To show transitivity, suppose λ, ν, $\mu \in \bar{\mathscr{M}}$ and $\nu \leqslant \lambda \leqslant \mu$. Then, by definition of $\leqslant$, μ and λ have the Lebesgue decompositions

$$\mu = \alpha\lambda + \mu^\lambda, \tag{19}$$

$$\lambda = \beta\nu + \lambda^\nu. \tag{20}$$

Substituting (20) into (19) gives

$$\mu = \alpha\beta\nu + \alpha\lambda^{\nu} + \mu^{\lambda}. \tag{21}$$

This can be interpreted as the Lebesgue decomposition of μ with respect to ν, where $\mu_\nu = \alpha\beta\nu$ and $\mu^\nu = \alpha\lambda^\nu + \mu^\lambda$. ($\mu^\lambda$ is a term in μ^ν because $\nu \ll \lambda$ and $\mu^\lambda \perp \lambda$ imply $\mu^\lambda \perp \nu$.) Now $\mu_\nu = \alpha\beta\nu$ implies $\nu \leqslant \mu$. We conclude that $\nu \leqslant \lambda \leqslant \mu$ implies $\nu \leqslant \mu$, and $\leqslant$ is therefore transitive. Finally, to show antisymmetry, suppose that $\nu \leqslant \mu$ and $\mu \leqslant \nu$. Since $\mu \leqslant \nu$ we have

$$\nu = \alpha\mu + \nu^{\mu}. \tag{22}$$

But since $\nu \leqslant \mu$ we have $\nu \ll \mu$, so $\nu^\mu = 0$. Hence (22) becomes simply $\nu = \alpha\mu$, whence $\alpha = 1$, since both μ and ν are probability measures. Thus $\mu = \nu$. ∎

Note that it is only in the proof of antisymmetry that we use the hypothesis that ν and μ are *probability* measures; the relation $\leqslant$ is reflexive and transitive on all of $\mathscr{M}$. Note also that if we had, in the definition of the Lebesgue order, restricted the value of α always to lie in the interval $(0, 1]$ then the resulting order would be antisymmetric on all the finite measures, not just on the probability measures.

We close this section with a description of a natural geometric realization of the Lebesgue order on a finite measurable space. Consider a discrete measurable space X with n points $x_1, ..., x_n$. Then we can represent $L(X)$ (with the 0 measure deleted) as the $n-1$ simplex Σ^{n-1} in $\mathbf{R}^n$ consisting of all points $(a_1, ..., a_n)$, $a_i \geqslant 0$, such that $\sum a_i = 1$; such a point corresponds to the probability measure $\sum_i a_i \varepsilon_{x_i}$, where ε_{x_i} is the Dirac measure at x_i. For $m \leqslant n$, let Σ^{m-1} be a subsimplex of Σ^{n-1} determined by a choice of m of the n coordinates. Let F^{m-1} be the interior of Σ^{m-1}. Then for each "edge simplex" Σ^{m-2} of Σ^{m-1} there is an "edge projection" mapping $f\colon F^{m-1} \to \Sigma^{m-2}$ such that if Σ^{m-2} consists of those points in Σ^{m-1} where, say, $a_k = 0$, then f replaces with 0 the kth coordinate of any point in F^{m-1} and then normalizes the result so that the sum of the coordinates is one. Note that for the given F^{m-1} there are m edge simplices and hence m of these maps f. We can now define an order on the points of Σ^{n-1} as follows: it is the smallest transitive relation, $\leqslant$, which satisfies $p \leqslant q$ whenever $p = f(q)$ for some edge projection mapping f as above. This order then corresponds to the Lebesgue order on $L(X)$ (via the correspondence between the probability measures on X and Σ^{n-1}).

3. The Lebesgue Logic

The Lebesgue order induces a logic on probability measures—the Lebesgue logic. We now study the notions of AND and OR in the Lebesgue logic.[5] (Later we apply

[5] Recall that for any partial order the AND of two elements is their greatest lower bound (when it exists); the OR is their least upper bound (when it exists). The greatest lower bound of two probability measures ν and λ is a probability measure σ such that (1) $\sigma \leqslant \nu, \lambda$ and (2) if $\rho \leqslant \nu, \lambda$ is any other probability measure then $\rho \leqslant \sigma$. The greatest lower bound of ν and λ is denoted by $\nu \wedge \lambda$. The least upper bound of ν and λ is a probability measure σ such that (1) $\nu, \lambda \leqslant \sigma$ and (2) if $\nu, \lambda \leqslant \rho$ for ρ any other probability measure, then $\sigma \leqslant \rho$. The least upper bound of ν and λ is denoted by $\nu \vee \lambda$.

these to the problem of sensor fusion in perception.) Two main points emerge. First, for some pairs of probability measures AND is not defined. Those pairs for which it is defined and is not 0 are *simultaneously verifiable*. Second, for some pairs of probability measures OR is not defined—even, sometimes, if the AND is defined. Those pairs for which OR is defined are *compatible*. Compatibility implies simultaneous verifiability, but not vice versa. We now consider this in more detail, beginning first with OR.

PROPOSITION 23. *For the Lebesgue order, the* OR *of two probability measures* μ *and* ν *is defined iff* $\overline{\mu_\nu} = \overline{\nu_\mu} \neq 0$. *When defined, it is given by*

$$\mu \vee \nu = \overline{\frac{\mu}{\|\mu_\nu\|} + \frac{\nu^\mu}{\|\nu_\mu\|}} = \overline{\frac{\nu}{\|\nu_\mu\|} + \frac{\mu^\nu}{\|\mu_\nu\|}}. \tag{24}$$

Proof. One checks immediately that the last equality on the right in (24) holds when $\overline{\mu_\nu} = \overline{\nu_\mu}$. The OR of two probability measures in the Lebesgue logic is, by definition, their least upper bound in the Lebesgue order. We first assume that the least upper bound $\mu \vee \nu$ exists and show that $\overline{\mu_\nu} = \overline{\nu_\mu}$. This follows from the following lemma.

LEMMA 25. *If there exists* $\lambda \in L(X)$ *such that* $\nu, \mu \leqslant \lambda$ *then* $\overline{\mu_\nu} = \overline{\nu_\mu}$.

Proof. Since $\nu, \mu \leqslant \lambda$ we can write $\lambda_\nu = \alpha\nu$ and $\lambda_\mu = \beta\mu$ where $\alpha, \beta \in (0, 1]$. Taking the Lebesgue decomposition of λ_ν with respect to μ, and of λ_μ with respect to ν, we obtain

$$\lambda_\nu = \alpha\nu_\mu + \alpha\nu^\mu, \qquad \lambda_\mu = \beta\mu_\nu + \beta\mu^\nu. \tag{26}$$

But we know that these Lebesgue decompositions can also be written

$$\lambda_\nu = \lambda_{\nu\mu} + \lambda_\nu^\mu, \qquad \lambda_\mu = \lambda_{\mu\nu} + \lambda_\mu^\nu. \tag{27}$$

Then by uniqueness of the Lebesgue decomposition we find that $\lambda_{\nu\mu} = \alpha\nu_\mu$ and $\lambda_{\mu\nu} = \beta\mu_\nu$. But by Lemma 4, $\lambda_{\nu\mu} = \lambda_{\mu\nu}$, and the result follows. ∎

Continuing with the proof of Proposition 23, we now show that $\overline{\mu_\nu} = \overline{\nu_\mu}$ implies $\mu \vee \nu$ exists. Let

$$\sigma = \overline{\frac{\mu}{\|\mu_\nu\|} + \frac{\nu^\mu}{\|\nu_\mu\|}} = \overline{\frac{\nu}{\|\nu_\mu\|} + \frac{\mu^\nu}{\|\mu_\nu\|}}. \tag{28}$$

We must show that (i) μ, $\nu \leqslant \sigma$ and (ii) if μ, $\nu \leqslant \rho$ for ρ any other probability measure, then $\sigma \leqslant \rho$. Observe that it is clear from the first equality in (28) that $\sigma_\mu = \alpha\mu$ for some $\alpha > 0$, and from the second equality that $\sigma_\nu = \beta\nu$ for some

$\beta \in (0, 1]$. This proves part (i). Now to show (ii), suppose that $\mu, \nu \leqslant \rho$ for ρ any probability measure. Then we can write

$$\rho = \alpha\nu + \rho^{\nu}, \qquad \alpha \in (0, 1] \tag{29}$$

$$= \beta\mu + \rho^{\mu}, \qquad \beta \in (0, 1]. \tag{30}$$

Substituting (30) into (29) we obtain

$$\begin{aligned} \rho &= \alpha\nu + (\beta\mu + \rho^{\mu})^{\nu} \\ &= \alpha\nu + \beta\mu^{\nu} + \rho^{\mu\nu} \qquad (31) \\ &= \alpha\nu_{\mu} + \alpha\nu^{\mu} + \beta\mu^{\nu} + \rho^{\mu\nu}. \qquad (32) \end{aligned}$$

Similarly, substituting (29) into (30) we obtain

$$\begin{aligned} \rho &= \beta\mu + (\alpha\nu + \rho^{\nu})^{\mu} \\ &= \beta\mu_{\nu} + \beta\mu^{\nu} + \alpha\nu^{\mu} + \rho^{\nu\mu}. \qquad (33) \end{aligned}$$

Equating (32) and (33), and recalling from Lemma 4 that $\rho^{\mu\nu} = \rho^{\nu\mu}$, we find that

$$\alpha\nu_{\mu} = \beta\mu_{\nu}$$

so that

$$\frac{\alpha}{\beta} = \frac{\|\mu_{\nu}\|}{\|\nu_{\mu}\|}. \tag{34}$$

Substituting (34) into (31) we obtain

$$\begin{aligned} \rho &= \beta \frac{\|\mu_{\nu}\|}{\|\nu_{\mu}\|} \nu + \beta\mu^{\nu} + \rho^{\mu\nu} \\ &= \beta \|\mu_{\nu}\| \left(\frac{\nu}{\|\nu_{\mu}\|} + \frac{\mu^{\nu}}{\|\mu_{\nu}\|} + \rho^{\mu\nu} \right). \qquad (35) \end{aligned}$$

Now if $\rho^{\mu\nu} = 0$ then from (35) and the fact that ρ and σ are both probability measures it follows that $\rho = \sigma$. If $\rho^{\mu\nu} \neq 0$ then since $\rho^{\mu\nu} \perp \sigma$ it follows that $\sigma \leqslant \rho$. ∎

Another way to think of the OR is as follows. Let μ and ν be probability measures on $(X, \mathscr{X})$, and let $\{S_{\mu}, A, S_{\nu}\}$ be a Hahn presentation (see section two). If μ and ν admit a common upper bound, and if $A \neq \varnothing$, then $\mu(\cdot \mid A) = \nu(\cdot \mid A)$. In this case μ and ν have a least upper bound, namely,

$$\alpha\mu(\cdot \mid S_{\mu}) + \beta\mu(\cdot \mid A) + \gamma\nu(\cdot \mid S_{\nu}),$$

where α, β, γ are uniquely determined by the conditions

$$\frac{\alpha}{\mu(S_{\mu})} = \frac{\beta}{\mu(A)}, \qquad \frac{\gamma}{\nu(S_{\nu})} = \frac{\beta}{\nu(A)}, \qquad \alpha + \beta + \gamma = 1.$$

If μ and ν admit a common upper bound and $A = \varnothing$ then there is a one-parameter family of minimal upper bounds of the form $a\mu + (1-a)\nu$, $0 < a < 1$ (see Proposition 56).

If $\{\mu_1, ..., \mu_n, ...\}$ is a collection of mutually singular measures, then every strictly positive convex combination $\sum_{i=1}^{\infty} \alpha_i \mu_i$ is a minimal upper bound for the collection.

As we have seen, in the Lebesgue logic certain pairs of probability measures are logically incompatible: the relation OR is not defined for them. If the OR is defined for a pair of probability measures then *all* the logical relations are defined and the two probability measures are in fact part of a boolean sublogic of the Lebesgue logic. Accordingly, we are led to the following definition.

DEFINITION 36. Two probability measures μ and ν are *compatible* if $\overline{\mu_\nu} = \overline{\nu_\mu} \neq 0$. In this case we write $\mu \overset{\vee}{\longleftrightarrow} \nu$.

Using this definition we can say that μ OR ν exists iff $s \overset{\vee}{\longleftrightarrow} \nu$. We can also state this condition using Renyi's equivalence of measures (Renyi, 1955, 1956, 1970; Krauss, 1968). Recall that μ and ν are Renyi equivalent, written $\mu \sim \nu$, iff $\mu = c\nu$ with $0 < c < \infty$. Thus we can say that if μ and ν are not singular then $\mu \overset{\vee}{\longleftrightarrow} \nu$ iff $\mu_\nu \sim \nu_\mu$.

We can state the condition for compatibility of measures in yet another way, using the Hahn presentation. Recall that the Hahn presentation for μ and ν on $(X, \mathscr{X})$ is a partition of X into three sets—S_μ, A, and S_ν—such that $[\mu|_A] = [\nu|_A]$ and $\nu(S_\mu) = 0 = \mu(S_\nu)$. Using this decomposition, we can state that if μ and ν are not mutually singular then $\mu \overset{\vee}{\longleftrightarrow} \nu$ iff $\mu(\cdot|A) = \nu(\cdot|A)$.

Note that the relation of compatibility is not transitive. Furthermore, a collection $\{\mu_\alpha\}$ of pairwise mutually compatible measures need not have a common upper bound except in the case where there is an event B with $\mu_\alpha(B) > 0$ for all α. In this case, if the family is infinite, the upper bound may be forced to be infinite.

The situation for AND is slightly more complicated than that for OR. If μ and ν are compatible then their AND is defined. But their AND can be defined even if they are incompatible.

PROPOSITION 37. *The* AND *of two probability measures μ and ν exists iff there exists at most one real number α with the following property: there exists some event A of ν positive measure such that $\nu|_A = \alpha\mu|_A$. If there is no such α, then μ* AND *ν is 0 (this includes the case where $\mu \perp \nu$). If there is exactly one such α, then μ* AND *ν is $\nu(\cdot|A)$ or, equivalently, $\mu(\cdot|A)$, where*

$$A = \left(\frac{d\nu_\mu}{d\mu}\right)^{-1}(\alpha).^6$$

[6] The symbol $d\nu_\mu/d\mu$ denotes the Radon–Nikodym derivative of ν_μ with respect to μ (see, e.g., Royden, 1988). According to the Radon–Nikodym theorem, given any two σ-finite measures ρ and σ with $\rho \ll \sigma$ there exists a real-valued measurable function ϕ, unique up to sets of σ measure zero, such that for all events A, $\rho(A) = \int_A \phi(x)\,\sigma(dx)$. The function ϕ is called the Radon–Nikodym derivative and is often denoted by $d\rho/d\sigma$.

Proof. Suppose there are numbers $\alpha \neq \beta$ and sets A and B of positive ν measure, such that $\nu|_A = \alpha\mu|_A$ and $\nu|_B = \beta\mu|_B$. We have then that $\nu|_A \leqslant \nu$, $\nu|_A \leqslant \mu$, and also $\nu|_B \leqslant \nu$, $\nu|_B \leqslant \mu$. Let us assume that ν AND μ exists; denote it by λ. In view of the above, $\nu|_A \leqslant \lambda$ and $\nu|_B \leqslant \lambda$, so $\lambda|_{A\cup B} \neq 0$. This fact, together with the fact that $\lambda \leqslant \nu$ implies that there exists $c \neq 0$ such that $\lambda|_{A\cup B} = c\nu|_{A\cup B}$. Similarly, there exists $e \neq 0$ such that $\lambda|_{A\cup B} = e\mu|_{A\cup B}$. In particular $c\nu|_A = e\mu|_A$ and $c\nu|_B = e\mu|_B$. But these imply, respectively, that $\alpha = e/c$ and $\beta = e/c$, so that $\alpha = \beta$. This contradiction shows that λ does not exist.

Now suppose that there is exactly one α such that $\nu|_A = \alpha\mu|_A$ for a set $A = (d\nu_\mu/d\mu)^{-1}(\alpha)$ of positive ν measure. Then $\nu(\cdot|A) \leqslant \nu$ and $\nu(\cdot|A) \leqslant \mu$. We will show $\nu(\cdot|A) = \nu$ AND μ. For this, suppose $\rho \leqslant \nu$ and $\rho \leqslant \mu$ for some $\rho \neq 0$ in $L(X)$. Then there is a set S with $\nu(S) > 0$ and $\mu(S) > 0$ such that $\rho = \nu(\cdot|S)$ and also $\rho = \mu(\cdot|S)$, i.e., $\rho = c\nu|_S$ and $\rho = e\mu|_S$ for some c, $e \neq 0$. But then $\nu|_S = (e/c)\,\mu|_S$ so that $S \subset A$ (and $e/c = \alpha$). Thus $\rho \leqslant \nu|_A$, which gives the result.

Finally, suppose there is no real number α such that for some set A of positive ν measure $\nu|_A = \alpha\mu|_A$. We will show that μ AND ν is the 0 measure. Since the 0 measure is always less than or equal to both ν and μ, it suffices to show that if $\rho \leqslant \nu$ and $\rho \leqslant \mu$, then $\rho = 0$. Suppose, then, that $\rho \leqslant \nu$ and $\rho \leqslant \mu$. If $\rho \neq 0$ there is a set S with $\nu(S)$, $\mu(S) > 0$, such that $\rho = c\nu|_S$ and $\rho = e\mu|_S$ with c, $e \neq 0$. Then, letting $\alpha = e/c$, we get $\nu|_S = \alpha\mu|_S$, a contradiction. Hence, $\rho = 0$. ∎

Another way to think of the AND is as follows. Two measures μ and ν have a common lower bound other than 0 iff there is an event A, with $\mu(A) \neq 0$ and $\nu(A) \neq 0$, such that $\mu(\cdot|A) = \nu(\cdot|A)$. The measures have a greatest lower bound (i.e., an AND) iff such events A are closed upon forming finite unions. In this case there is a largest such A, modulo sets both μ-negligible and ν-negligible. In this case $\mu \wedge \nu = \mu(\cdot|A)$. In general $\mu \wedge \nu = 0$ iff $\mu \perp \nu$ or $\mu(\{f = t\}) = 0$ for all $t \neq 0$ when $\nu_\mu = f_\mu$.

Now one can discuss the joint probability of two propositions, i.e., the probability of their "simultaneous occurrence," if one can form their AND. As we have just seen, it's not always possible to do so in the Lebesgue logic. Even when it is possible, the AND can be 0, indicating that the probability of their simultaneous occurrence is zero. Thus we are led to the following definition.

DEFINITION 38. Two measures μ and ν are *simultaneously verifiable* iff their AND exists and is not 0, i.e., iff there exists *precisely* one real number α such that for some event A of ν positive measure, $\nu|_A = \alpha\mu|_A$. In this case we write $\mu \overset{\wedge}{\leftrightarrow} \nu$.

Note that compatibility implies simultaneous verifiability but not vice versa, and that both compatibility and simultaneous verifiability are intransitive relationships. However, for discrete measurable spaces compatibility and simultaneous verifiability coincide (i.e., if $\nu \wedge \mu$ exists and is not 0 then $\nu \vee \mu$ exists). As we shall discuss later, the existence of probability measures that are not simultaneously verifiable corresponds in perception to the existence of multistable percepts.

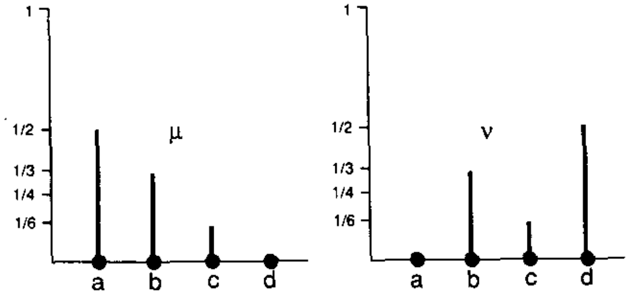

FIG. 2. Two probability measures, μ and ν, which are compatible in the Lebesgue logic.

To develop more intuitions about the Lebesgue logic we now consider two concrete examples in which we compute the AND and OR of probability measures.

EXAMPLE 39. For our examples, the probability space X consists of four points, say (a, b, c, d), and the algebra of events is the full power set. Consider the probability measures $\mu = (\frac{1}{2}, \frac{1}{3}, \frac{1}{6}, 0)$ and $\nu = (0, \frac{1}{3}, \frac{1}{6}, \frac{1}{2})$, as shown in Fig. 2. The first step is to consider the Lebesgue decompositions of each measure with respect to the other. For μ we find that the absolutely continuous part is $\mu_\nu = (0, \frac{1}{3}, \frac{1}{6}, 0)$ and the singular part is $\mu^\nu = (\frac{1}{2}, 0, 0, 0)$. This is illustrated in Fig. 3. For ν we find that $\nu_\mu = (0, \frac{1}{3}, \frac{1}{6}, 0)$ and $\nu^\mu = (0, 0, 0, \frac{1}{2})$. This is illustrated in Fig. 4. Next we compare the absolutely continuous parts, since these determine whether or not the measures are compatible or simultaneously verifiable. We find that $\overline{\mu_\nu} = (0, \frac{2}{3}, \frac{1}{3}, 0) = \overline{\nu_\mu}$, so that μ and ν are compatible, and therefore also simultaneously verifiable. Their AND, then, is just $\overline{\mu_\nu}$, as shown in Fig. 5. (By the way, this method of taking the AND does not assume that μ and ν are independent. Consider, for example, that $\mu(b) = \frac{1}{3}$, $\nu(b) = \frac{1}{3}$, but that $(\mu \text{ AND } \nu)(b) = \frac{2}{3} \neq \mu(b)\,\nu(b)$. This is a clear difference between the Lebesgue AND and other combination rules, such as the Dempster rule of Dempster–Shafer theory (Shafer, 1976; Walley, 1987), which assume independence of the information sources to be combined.) The OR of μ and ν, from Proposition 23, is $(\frac{1}{3}, \frac{2}{9}, \frac{1}{9}, \frac{1}{3})$, as is shown in Fig. 5.

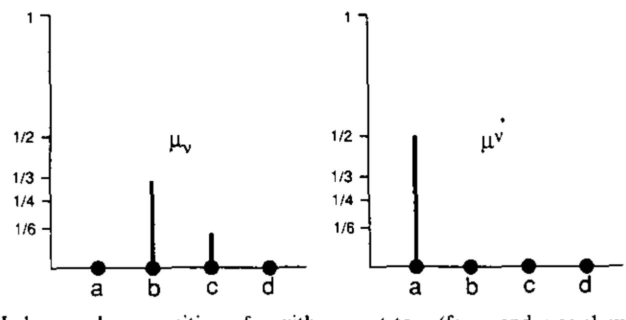

FIG. 3. The Lebesgue decomposition of μ with respect to ν (for μ and ν as shown in Figure 2).

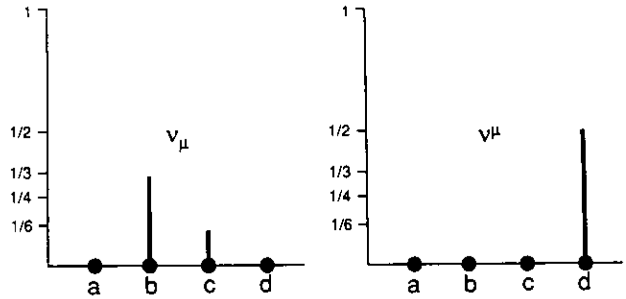

FIG. 4. The Lebesgue decomposition of ν with respect to μ (for μ and ν as shown in Fig. 2).

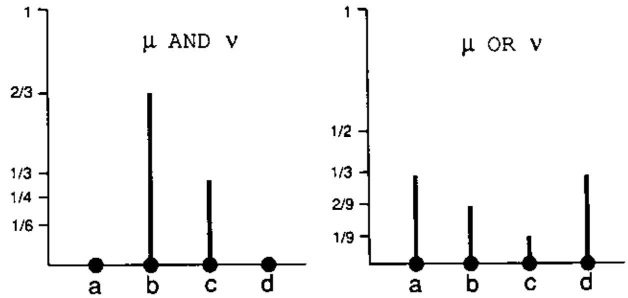

FIG. 5. The AND and OR of μ and ν (for μ and ν as shown in Fig. 2).

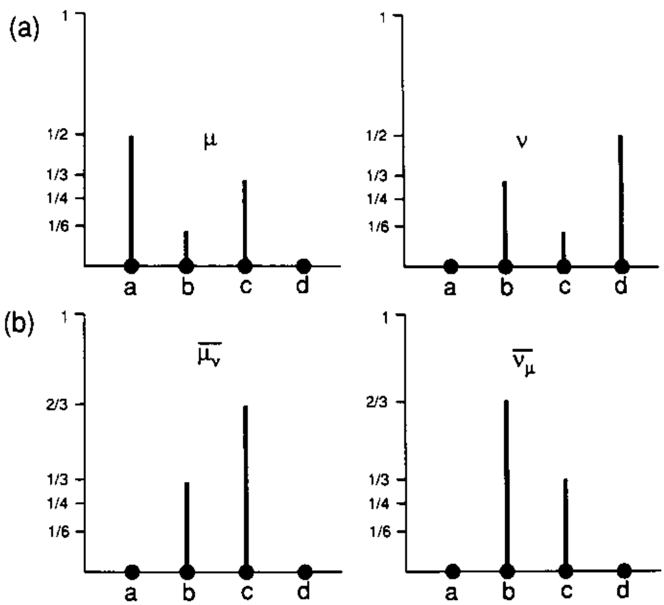

FIG. 6. An example of measures incompatible in the Lebesgue logic. Part (a) shows the probability measures μ and ν. Part (b) shows $\overline{\mu_\nu}$ and $\overline{\nu_\mu}$. Since the two measures in (b) are not equal, μ and ν are incompatible.

EXAMPLE 40. As a second example, let $\mu=(\frac{1}{2}, \frac{1}{6}, \frac{1}{3}, 0)$ and $\nu=(0, \frac{1}{3}, \frac{1}{6}, \frac{1}{2})$, as shown in Fig. 6. Then $\overline{\mu_\nu}=(0, \frac{1}{3}, \frac{2}{3}, 0)$ and $\overline{\nu_\mu}=(0, \frac{2}{3}, \frac{1}{3}, 0)$, as is also shown in Fig. 6. Since $\overline{\mu_\nu}\neq\overline{\nu_\mu}$ the two measures are incompatible and therefore have no OR. To see if the AND exists we must check to see that if there is more than one event A of ν_μ positive measure such that $\nu_\mu(\cdot\,|\,A)=\mu_\nu(\cdot\,|\,A)$, then the Radon–Nikodym derivative $d\mu|_A/d\nu|_A$ does not differ on these events. It happens that there are two disjoint events A for which $\nu_\mu(\cdot\,|\,A)=\mu_\nu(\cdot\,|\,A)$, viz., the events b and c. On b we find $d\mu/d\nu=0.5$ and on c we find $d\mu/d\nu=2$. Since these derivatives are not equal the AND is not defined; the two measures are not simultaneously verifiable.

The AND of the Lebesgue logic generalizes the bayesian updating of probabilities. According to Bayes' rule, if an event H has probability $P(H)$, and subsequently an event A occurs, then the probability of H becomes $P(H\,|\,A)=P(A\,|\,H)\,P(H)/P(A)$. To see the relation between this and the Lebesgue AND, suppose one is given information in the form of a probability measure μ, and that later one is given further information in the form of a probability measure ν. To combine these two sources of information into a single probabilistic statement using the Lebesgue logic one might, depending on the nature of the information sources, decide to take their Lebesgue AND, $\mu\wedge\nu$. Now by definition $\mu\wedge\nu\leqslant\mu$ so that, according to Proposition 15, there is a set A such that $\mu\wedge\nu(\cdot)=\mu(\cdot\,|\,A)$. Thus taking the AND of μ with ν is equivalent to conditionalizing μ on the event A, just as in Bayes' rule. The Lebesgue logic provides a coherent framework within which to understand the role of the bayesian updating of probabilities: it plays the role of the AND of the Lebesgue logic.

Having discussed AND and OR, we consider NOT. The Lebesgue logic has no globally defined NOT because it has no global supremum, no "unit." (It does have a global infimum, the null measure 0.) It would be natural to let the NOT of a probability measure ν be the "biggest" probability measure ρ such that $\nu\perp\rho$. But the lack of a unit makes the notion of biggest undefined. However, one can pick a probability measure λ such that $\nu\leqslant\lambda$ and let it be the "unit." With respect to this unit one can then define the NOT of ν—indeed of any μ such that $\mu\leqslant\lambda$. Therefore we recall the following notation introduced in Section 2:

Notation 41. For any $\lambda\in L(X)$, we let $L_\lambda(X)=\{\mu\in L(X)\,|\,\mu\leqslant\lambda\}$. We sometimes write just L_λ.

PROPOSITION 42. *L_λ forms a boolean sublogic of the Lebesgue logic.*

Proof. This follows immediately from Proposition 16 since $\mathscr{X}_\lambda$ is boolean. ∎

The logical operations of AND and OR, when restricted to such a boolean sublogic, can be given particularly simple characterizations, as indicated in the following proposition.

PROPOSITION 43. *If* $\nu, \mu \in L_\lambda$, *so that* $\lambda = \alpha\nu + \lambda^\nu = B\mu + \lambda^\mu$, *then*

$$\nu \text{ OR } \mu = \overline{\alpha\nu + \lambda^\nu_\mu} = \overline{\beta\mu + \lambda^\mu_\nu} = \overline{\lambda_{\nu+\mu}}, \tag{44}$$

$$\nu \text{ AND } \mu = \overline{\nu_\mu} = \overline{\mu_\nu} = \overline{\lambda_{\nu\mu}}, \tag{45}$$

$$\text{NOT } \nu = \overline{\lambda^\nu}. \tag{46}$$

Proof. Via Proposition 15 we can identify each measure ν in L_λ with a support set S_ν (modulo λ) in such a way that $\nu = \lambda(\cdot \mid S_\nu)$. The formulas above follow directly from this correspondence. ∎

The boolean sublogics L_λ have a geometric interpretation, as the following example shows.

EXAMPLE 47. Let X be the measurable space of four points. Then $L(X)$ is a tetrahedron plus one point for the 0 measure. The center of mass of the tetrahedron corresponds to the uniform measure, call it λ, on the four points. The measure λ is greater, in the Lebesgue order, than the centers of mass of each of the four faces of the tetrahedron. The center of mass of a face, in turn, is greater in the Lebesgue order than the centers of mass of the three edges that border it. And the center of mass of an edge is greater, in the Lebesgue order, than the two vertices at its ends. Thus L_λ consists of 16 measures: one for the center of mass of the tetrahedron (λ itself), four for the centers of mass of the faces, six for the centers of mass of the edges, four for the vertices, and the 0 measure.

The boolean sublogics allow us to define a notion of complement. We can also define relative complements. Suppose that μ and ν have an upper bound λ. Then the relative complement $\mu \neg \nu$ is defined in the boolean algebra L_λ by $\mu \neg \nu = \mu \wedge \nu'$ where ν' is the complement of ν in L_λ, i.e., $\nu' = \overline{\lambda^\nu}$ (by Proposition 43). In fact it is not hard to see that this definition of relative complement is actually independent of the choice of λ (provided that $\mu, \nu \leqslant \lambda$). Thus the notion of the relative complement $\mu \neg \nu$ is invariantly defined provided that λ exists.

The sublogics L_λ are boolean and, therefore, satisfy the distributivity laws. For measures that are not all in a single sublogic, distributivity can fail to hold. We consider four cases and give examples. (1) Consider the distributive equation

$$\lambda \wedge (\mu \vee \nu) = (\lambda \wedge \mu) \vee (\lambda \wedge \nu).$$

In the Lebesgue logic, the existence of $\lambda \wedge (\mu \vee \nu)$ does not imply the existence of $(\lambda \wedge \mu) \vee (\lambda \wedge \nu)$. For example, suppose we have a discrete measurable space Y consisting of three points, and consider the measures $\mu = (\frac{2}{3}, \frac{1}{3}, 0)$, $\nu = (0, \frac{1}{3}, \frac{2}{3})$, and $\lambda = (\frac{1}{2}, 0, \frac{1}{2})$. Then $\overline{\mu_\nu} = \overline{\nu_\mu} = (0, 1, 0)$, so that $\mu \vee \nu$ exists, and is $(\frac{2}{5}, \frac{1}{5}, \frac{2}{5})$. We then compute that $\lambda \wedge (\mu \vee \nu) = \lambda$. Now $\lambda \wedge \mu = (1, 0, 0)$ and $\lambda \wedge \nu = (0, 0, 1)$, so that $\lambda \wedge \mu$ is singular to $\lambda \wedge \nu$. Therefore, according to Proposition 23, $(\lambda \wedge \mu) \vee (\lambda \wedge \nu)$ does not exist (instead there is a one-parameter family of minimal upper bounds, as will be shown in Proposition 56). (2) The existence of $(\lambda \wedge \mu) \vee (\lambda \wedge \nu)$ does not imply the existence of $\lambda \wedge (\mu \vee \nu)$. For example, let Y consist of four points, and consider the measures $\mu = (\frac{1}{8}, \frac{1}{3}, \frac{1}{6}, \frac{3}{8})$, $\nu = (\frac{3}{8}, \frac{1}{3}, \frac{1}{6}, \frac{1}{8})$, and $\lambda = (0, \frac{2}{3}, \frac{1}{3}, 0)$. (3) Now consider the distributive equation

$$\lambda \vee (\mu \wedge \nu) = (\lambda \vee \mu) \wedge (\lambda \vee \nu).$$

In the Lebesgue logic the existence of $\lambda \vee (\mu \wedge \nu)$ does not imply the existence of $(\lambda \vee \mu) \wedge (\lambda \vee \nu)$. For example, let Y consist of three points, and consider the measures $\mu = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6})$, $\nu = (0, 0, 1)$, and $\lambda = (\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$. (4) The existence of $(\lambda \vee \mu) \wedge (\lambda \vee \nu)$ does not imply the existence of $\lambda \vee (\mu \wedge \nu)$. For example, let Y consist of three points and let $\mu = (\frac{2}{3}, \frac{1}{3}, 0)$, $\nu = (0, \frac{1}{3}, \frac{2}{3})$, and $\lambda = (\frac{1}{2}, 0, \frac{1}{2})$.

We now consider the compatibility relation in more detail. Any pair of measures in a boolean sublogic L_λ are compatible (and *a fortiori* simultaneously verifiable). Indeed, compatibility is intimately linked with the boolean sublogics, as indicated by the following proposition.

PROPOSITION 48. *Two probability measures are in a boolean sublogic of the Lebesgue logic iff they are compatible or singular.*

Proof. The case of singular measures is straightforward and left to the reader. We first show that compatibility of two measures implies that the two are in a boolean sublogic. Let $\nu, \mu \in L(X)$, and suppose that $\overline{\nu_\mu} = \overline{\mu_\nu} \neq 0$. Then, by Proposition 23, $\nu \vee \mu$ exists and therefore $\nu, \mu \in L_{\nu \vee \mu}$. Now we show the converse. Assume that $\nu, \mu \in L_\rho$, for some $\rho \in L(X)$. Then the Lebesgue decomposition of ρ with respect to ν and then μ is

$$\rho = \alpha\nu + \rho^\nu \tag{49}$$

$$= \alpha\nu_\mu + \alpha\nu^\mu + \rho^\nu_\mu + \rho^{\nu\mu} \tag{50}$$

and the Lebesgue decomposition of ρ with respect to μ and then ν is

$$\rho = \beta\mu + \rho^\mu \tag{51}$$

$$= \beta\mu_\nu + \beta\mu^\nu + \rho^\mu_\nu + \rho^{\mu\nu}. \tag{52}$$

By uniqueness of the Lebesgue decomposition the parts of (50) and (52) that are absolutely continuous with respect to ν must be equal. Thus

$$\alpha\nu_\mu + \alpha\nu^\mu = \beta\mu_\nu + \rho^\mu_\nu. \tag{53}$$

However, from (49) we have that $\rho_\nu = \alpha\nu$ so that $\rho^\mu_\nu = \alpha\nu^\mu$. Thus (53) becomes

$$\alpha\nu_\mu + \alpha\nu^\mu = \beta\mu_\nu + \alpha\nu^\mu. \tag{54}$$

Therefore $\alpha\nu_\mu = \beta\mu_\nu$. ∎

So we can think of the Lebesgue logic as being comprised, in general, of an infinite collection of "local" boolean logics, namely the L_μ's, which are pieced together so that some are mutually compatible and others are not. We have seen in Proposition 16 that L_μ is order isomorphic to the measure algebra $\mathscr{X}_\mu$. We now further clarify the structure of the L_μ's by showing that they are also isomorphic to $C_\mu(X)$, viz., the lattice of measure classes consisting of measures absolutely continuous with respect to μ. For this purpose it suffices to prove

PROPOSITION 55. *$C_\mu(X)$ is canonically order isomorphic to $\mathscr{X}_\mu$.*

Proof. We let g be the map $g: \mathscr{X}_\mu \to C_\mu$ that sends $[A]$ to $[\mu|_A]$ (where brackets indicate equivalence classes). The map g is well defined since if A and C differ by a μ-negligible set we have $\mu|_A = \mu|_C$. If $[A] \leqslant [B]$ in $\mathscr{X}_\mu$, i.e., if A is contained in B up to a μ-negligible set, then clearly $\mu|_A \ll \mu|_B$ so that g is order preserving. To show that g is surjective, suppose $\nu \ll \mu$. We then have $\nu = \nu_\mu$ so that $[\nu] = [\nu_\mu] = [\mu_\nu]$ by Proposition 12. But $\mu_\nu = \mu|_A$ for a suitable set A. Hence $[\nu] = [\mu|_A]$, i.e., $[\nu] = g[A]$. To show injectivity we simply note that if $\mu|_A = \mu|_B$ then A and B differ by a μ-negligible set. ∎

We summarize the results of Propositions 16 and 55 on the representations of the local boolean sublogics L_μ. For $\mu \in L(X)$ we have order isomorphisms as shown:

$$\begin{tikzcd} L_\mu(X) \arrow[r, phantom, "="] \arrow[d, "f"'] & L_\mu(X) \arrow[d, "h = g \circ f"] \\ \mathscr{X}_\mu \arrow[r, "g"] & C_\mu(X) \end{tikzcd}$$

where: For $\nu \in L_\mu(X)$, let $A \in \mathscr{X}$ be a set such that $\nu = \mu(\cdot \mid A)$ (by Proposition 15). Then $f(\nu) = [A]$. For $[A] \in \mathscr{X}_\mu$, $g([A]) = [\mu(\cdot \mid A)]$. For $\nu \in L(X)$, $h(\nu) = [\nu]$. Note that in contrast to h, both f and g are not defined globally (i.e., they are not restrictions to $L_\mu(X)$ and $\mathscr{X}_\mu$, respectively, of maps defined on $L(X)$ and $\mathscr{X}$).

We indicated in Proposition 23 that the OR of two probability measures μ and ν exists iff $\overline{\mu_\nu} = \overline{\nu_\mu} \neq 0$. If $\mu \perp \nu$ then $\mu_\nu = \nu_\mu = 0$ so that μ and ν have no OR, i.e., no least upper bound. However, in this case they do have a one-parameter family of minimal upper bounds, as the next proposition indicates.

PROPOSITION 56. *If μ, $\nu \in L$ and $\mu \perp \nu$ then there is a one-parameter family u_t of minimal, mutually incomparable, upper bounds given by $u_t = t\mu + (1-t)\,\nu$, where $0 < t < 1$. The u_t belong to the same measure class. If μ, $\nu \in L_\lambda$ then there is a unique t' such that $u_{t'} \leqslant \lambda$.*

Proof. We first show that each u_t is a minimal upper bound of the measures ν, μ. Recall that this means that if $\rho \in L(X)$ satisfies ν, $\mu \leqslant \rho \leqslant u_t$ then $u_t = \rho$. Suppose then that there is a $\rho \in L(X)$ such that $\nu, \mu \leqslant \rho \leqslant u_t$. Then we can write the Lebesgue decompositions

$$\rho = a\nu + \rho^\nu, \qquad a \in (0, 1) \tag{57}$$

$$= b\mu + \rho^\mu, \qquad b \in (0, 1) \tag{58}$$

$$u_t = c\rho + u_t^\rho, \qquad c \in (0, 1]. \tag{59}$$

Substituting (58) into (57), and recalling that $\mu^{\nu} = \mu$ (since $\mu \perp \nu$) we obtain

$$\rho = a\nu + b\mu + \rho^{\mu\nu}. \tag{60}$$

But since $u_t^{\mu\nu} = 0$ and $\rho \leqslant u_t$, it follows that $\rho^{\mu\nu} = 0$. Thus

$$\rho = a\nu + b\mu. \tag{61}$$

Since $a, b > 0$, it follows that ρ is in the same measure class as u_t and that therefore $u_t^{\rho} = 0$. So from (59) and the fact that u_t and ρ are probability measures it follows that $u_t = \rho$.

That the u_t are all in the same measure class and are mutually incomparable is clear. We proceed to show that if $\mu, \nu \in L_{\lambda}$ then there is a unique t' such that $u_{t'} \leqslant \lambda$. Suppose that $\mu, \nu \in L_{\lambda}$ so that

$$\lambda = a\nu + \lambda^{\nu}, \qquad a \in (0, 1) \tag{62}$$

$$= b\mu + \lambda^{\mu}, \qquad b \in (0, 1). \tag{63}$$

Substituting (63) into (62) and recalling that $\mu^{\nu} = \mu$ we obtain

$$\lambda = a\nu + b\mu + \lambda^{\mu\nu}.$$

Thus

$$\lambda_{u_t} = a\nu + b\mu. \tag{64}$$

Now $u_t \leqslant \lambda$ iff $\lambda_{u_t} = cu_t$ for some $c \in (0, 1]$, i.e., using (64), iff $a\nu + b\mu = c(t\nu + (1-t)\mu)$. This implies that $t/(1-t) = a/b$, i.e., that $t = a/(a+b)$. ▌

Having discussed the AND and OR of the Lebesgue logic in some detail, we now consider the "atoms" of the Lebesgue logic. An atom of the Lebesgue logic is a probability measure that is greater than (i.e., entailed by) no other probability measure in the logic, except for the 0 measure (which is less than all measures). Understanding the atoms of a logic is central to understanding the structure of the logic itself. We begin with a definition.

DEFINITION 65. Let $(X, \mathscr{X})$ be a measurable space. A *Lebesgue indecomposable* measure on X is a measure $\mu \in \mathscr{M}(X)$ such that for all $T \in \mathscr{X}$ either $\mu(T) = \mu(X)$ or $\mu(T) = 0$.

PROPOSITION 66. *A measure μ is Lebesgue indecomposable iff it has no nontrivial Lebesgue decompositions, i.e., iff there does not exist a measure ρ such that $\mu \neq \mu_{\rho} \neq \mu^{\rho} \neq 0$.*

Proof. We first show that if μ is Lebesgue indecomposable then μ has no nontrivial Lebesgue decompositions. Suppose that $\mu = \mu_{\rho} + \mu^{\rho}$ is a nontrivial Lebesgue decomposition of μ. Since $\mu_\rho \perp \mu^\rho$ there exists a measurable set A such that $\mu_\rho(A) = \mu^\rho(A^c) = 0$. Therefore

$$\begin{aligned}\mu(X) &= \mu(A) + \mu(A^c)\\ &= \mu_\rho(A) + \mu^\rho(A) + \mu_\rho(A^c) + \mu^\rho(A^c)\\ &= \mu^\rho(A) + \mu_\rho(A^c).\end{aligned}$$

We find then that

$$\mu(A) = \mu^\rho(A) = \mu(X) - \mu_\rho(A^c) < \mu(X)$$

since $\mu_\rho(A^c) \neq 0$. (If $\mu_\rho(A^c)$ were equal to 0, then since also $\mu_\rho(A) = 0$, it would follow that $\mu_\rho(A \cup A^c) = \mu_\rho(X) = 0$. But this contradicts our assumption that the Lebesgue decomposition of μ with respect to ρ is nontrivial.) Thus μ is not Lebesgue indecomposable. Now we show that if μ has no nontrivial Lebesgue decompositions then μ is Lebesgue indecomposable. Suppose μ is not Lebesgue indecomposable, i.e., suppose that there is a set $T \in \mathscr{X}$ such that $\mu(T) < \mu(X)$. Let $\rho = \mu|_T$. Then $\mu = \mu_\rho + \mu^\rho$ is a nontrivial Lebesgue decomposition of μ. ∎

These Lebesgue indecomposable measures are the atoms of the Lebesgue logic. Before showing this we first give a precise definition of atom.

DEFINITION 67. An element μ of a partially ordered set is an atom iff $0 \leqslant \nu \leqslant \mu$ implies either $\nu = 0$ or $\nu = \mu$.

Remark 68. For a given measurable space $(X, \mathscr{X})$, all and only Lebesgue indecomposable measures are atoms in the Lebesgue logic $L(X)$.

Proof. We first show that all Lebesgue indecomposable measures are atoms. Suppose $\mu \in L(X)$, $\mu \neq 0$, is a Lebesgue indecomposable measure and that $0 \leqslant \nu \leqslant \mu$ for some $\nu \in L(X)$. Then, by definition of Lebesgue order, $\mu = \alpha\nu + \mu^\nu$ is the Lebesgue decomposition of μ with respect to ν. ν might, of course, be 0. If $\nu \neq 0$ then since μ is Lebesgue indecomposable, $\mu^\nu = 0$, so that $\mu = \alpha\nu$. But μ and ν are in $L(X)$, so therefore $\alpha = 1$ and $\mu = \nu$. Thus μ is an atom. We now show that all atoms are Lebesgue indecomposable measures. Suppose μ is not Lebesgue indecomposable. Then there is a measure ρ such that $\mu \neq \mu_\rho \neq \mu^\rho \neq 0$. Since $0 \leqslant \mu_\rho \leqslant \mu$ but $\mu_\rho \neq 0$ and $\mu_\rho \neq \mu$ we conclude that μ is not an atom of $L(X)$. ∎

We close this section with a few technical remarks, mostly aimed at suggesting some further directions for investigation of the Lebesgue logic.

Remark 69. If $(X, \mathscr{X})$ is a metric space with σ-algebra generated by the metric topology then the only Lebesgue indecomposable measures, and therefore the only atoms of $L(X)$, are Dirac measures. (A Dirac measure gives a weight of one to a single point and a weight of zero to all measurable sets not containing that point.) As an example of a Lebesgue indecomposable probability measure that is not a

Dirac measure, consider an uncountable space X with σ-algebra generated by (1) all countable subsets of X and (2) by two uncountable sets A and A^c. Then a measure μ, that assigns the number 1 to the set A and the number 0 to A^c and to all countable measurable sets, is Lebesgue indecomposable but not Dirac.

Remark 70. The maps $\mu \mapsto \mu_\nu$ and $\mu \mapsto \mu^\nu$ are linear operations on $\mathscr{M}$.

Remark 71. If μ, λ, $\nu \in \mathscr{M}$ then $\mu_{\lambda_\nu} = \mu_{(\lambda_\nu)}$ and $\mu_{(\lambda^\nu)} = \mu_\lambda^\nu$.

Remark 72. Suppose $\nu \neq \lambda$. Then $\nu \perp \lambda$ iff $\nu, \lambda \leqslant \overline{\nu + \lambda}$.

Remark 73. Suppose $\mu_1, \mu_2 \in L(X)$ and that $\mu_1 = \nu_1 + \lambda_1$ and $\mu_2 = \nu_2 + \lambda_2$ are decompositions of μ_1 and μ_2 such that $\nu_i \perp \lambda_i$. If $\mu_1 \wedge \mu_2$ exists then $\nu_1 \wedge \nu_2$, $\nu_1 \wedge \lambda_2$, $\lambda_1 \wedge \nu_2$, and $\lambda_1 \wedge \lambda_2$ each exist. (They might be 0.)

Remark 74. Let $\mu_1 \times \mu_2$ denote the product measure (i.e., $\mu_1 \times \mu_2(A \times B) = \mu_1(A)\,\mu_2(B)$). Then $\mu_1 \times \mu_2 \leqslant \nu_1 \times \nu_2$ iff $\mu_i \leqslant \nu_i$.

Remark 75. Suppose $\nu, \lambda, \mu \in L(X)$ and $\nu \perp \lambda$. If $\mu \wedge \nu$ and $\mu \wedge \lambda$ each exist then there exist c_1, $c_2 \in \mathbf{R}_0$ (positive reals) such that for all $t \in [0, 1]$ we have $(t\nu + (1-t)\,\lambda) \wedge \mu = c_1 \nu \wedge \mu + c_2 \lambda \wedge \mu$.

Remark 76. Let Y_1, Y_2, and X be measurable spaces with Lebesgue logics, respectively, $L(Y_1)$, $L(Y_2)$, and $L(X)$. Let $\phi_1 : L(Y_1) \to L(X)$ and $\phi_2 : L(Y_2) \to L(X)$ be logic morphisms. Denote the Lebesgue logic of product measures on $Y_1 \times Y_2$ by $L(Y_1) \otimes L(Y_2)$. Then the map $\phi_1 \wedge \phi_2 : L(Y_1) \otimes L(Y_2) \to L(X)$ given, where it is defined, by $(\phi_1 \wedge \phi_2)(\mu \times \nu) = \phi_1(\mu) \wedge \phi_2(\nu)$ is a logic morphism. (Such a construction might be of interest for the integration of "observers," discussed later.)

Remark 77. If $(X, \mathscr{X})$ is a topological space, then the Lebesgue order can be *completed.* In the completed Lebesgue order we define $\nu \leqslant \mu$ iff (1) $\mu_\nu = \alpha\nu$, for $\alpha > 0$, or (2) ν is supported[7] on a μ measure zero *subset* of the support of μ. The completed Lebesgue order is a completion of the Lebesgue order in the following sense: If ν is supported on a μ measure zero subset of the support of μ then, under suitable conditions, ν is expressible as a limit of sequences of normalized restrictions of μ over a family of sets which contract down to the support of ν. In the logic associated to the completed Lebesgue order μ AND ν exists iff (1) $\mu \perp \nu$, in which case it is 0, or (2) $\overline{\mu_\nu} = \overline{\nu_\mu}$, in which case it is $\overline{\mu_\nu}$.

Remark 78. In the view of Marr (1982) and of many other researchers, perception is best understood as *information processing.* From this viewpoint, a natural condition for probability measures to be relevant to the study of perception, is that the logic be related nontrivially to the information content of the measures. Specifi-

[7] In general *the* support of a measure is not well defined. However, for any probability measure ν on a topological space the support is the intersection of all closed sets A for which $\nu(A) = 1$. See, e.g., Royden (1988).

cally, this viewpoint suggests that the entailment relation between measures should respect the information content, or, equivalently, the entropy, of the measures. Recall that the entropy of a measure ν on a finite measurable space $(X, 2^X)$ is

$$\mathrm{Ent}(\nu) = -\sum_{x \in X} \log[\nu(x)]\, \nu(x).$$

The Lebesgue order respects entropy: if $\lambda \leqslant \nu$ in the Lebesgue order then $\mathrm{Ent}(\lambda) \leqslant \mathrm{Ent}(\nu)$. In addition to respecting the information content of measures in this quantitative fashion, the Lebesgue order also respects their information content in a more qualitative fashion, i.e., by respecting the absolute continuity of measures: if $\lambda \leqslant \nu$ in the Lebesgue order then $\lambda \ll \nu$ (but the converse is not true—see Lemma 17). Intuitively, this means that if λ entails ν in the Lebesgue logic then the support of λ is properly contained in the support of ν, and this is one way in which λ can be more informative, in a qualitative sense, than ν. There are several interesting orders on measures that do not respect either the qualitative or quantitative information content of probability measures and which are therefore of less relevance to perception. The next three remarks mention three such orders.

Remark 79. The Riesz order on finite measures can be specified by first defining $[\mu \vee \nu](A) = \mathrm{Sup}\{\mu(B) + \nu(C) : B \cup C = A \text{ and } B \cap C = \varnothing\}$, and then stipulating that $\mu \leqslant \nu$ iff $\mu \vee \nu = \nu$. The Riesz order, restricted to probability measures, becomes $\mu \leqslant \nu$ iff $\mu = \nu$, and therefore generates no useful logic on probabilities.

Remark 80. The Gleason parts order on probability measures can be specified by first defining $\mu \leqslant_a \nu$ iff $\nu = a\mu + (1-a)\lambda$ for some probability measure λ, then letting $s = \mathrm{Sup}\{a : \mu \leqslant_a \nu\}$, and finally stipulating that $\mu \leqslant \nu$ iff $\mu \leqslant_s \nu$. The Gleason parts order does not respect entropy.

Remark 81. The collection of finite measures on a compact convex metric space can be given the Choquet order. In this order, $\mu \leqslant \nu$ iff $\mu(f) \leqslant \nu(f)$ for all continuous convex functions f on X. The Choquet order is not well defined on discrete measurable spaces. Moreover, the Choquet order does not respect absolute continuity.

Remark 82. Let $P_\sigma : L(X) \to L(X)$ be the projection operator $\mu \mapsto \sigma \wedge \mu$. Define the *domain* of P_σ to be $D(P_\sigma) = \{\mu \in L(X) \mid \mu \overset{\triangle}{\leftrightarrow} \sigma\}$. Given this notation, then (1) If, for all $\mu \in D(P_\rho) \cap D(P_\sigma)$, it is true that $P_\sigma P_\rho \mu = P_\rho P_\sigma \mu \neq 0$, then $\sigma \overset{\triangle}{\leftrightarrow} \rho$. Moreover, (2) if $\sigma \overset{\triangle}{\leftrightarrow} \rho$ then for all $\mu \in D(P_\rho) \cap D(P_\sigma)$ it is true that $P_\sigma P_\rho \mu = P_\rho P_\sigma \mu$. Thus in the Lebesgue logic, as in quantum logic, compatible propositions are intimately linked with commuting operators.

Remark 83. Of importance to the structure theory of quantum logics are the *blocks*, i.e., maximal boolean sublogics. In the Lebesgue logic, no boolean sublogic $L_\mu(X)$ is a block if $\mathscr{X}$ is infinite, unless $\mathscr{X}$ is isomorphic to 2^N with N countable. In this case we may assume $X = N$, and we find that for any probability measure μ strictly positive on 2^N, the boolean sublogic $L_\mu(X)$ is a block. If $\mathscr{X}$ is not isomorphic to 2^N and μ is an *infinite* countably additive measure then the Renyi full conditional probability (Renyi, 1970; see also Fine, 1973) $\{\mu(\cdot|A): 0<\mu(A)<\infty\}$ is a family of measures mutually compatible in the Lebesgue logic.

Remark 84. Sometimes probabilistic systems can exhibit periodic behavior; *quasi-compact markov chains are one example* (Revuz, 1984). This *periodic* probabilistic behavior can be conveniently expressed using complex measures (Revuz, 1984). The Lebesgue decomposition can easily be extended to such measures: if μ and ν are complex measures, then $\mu=\mu_\nu+\mu^\nu$, where μ_ν is absolutely continuous with respect to the total variation of ν, and μ^ν is singular to the total variation of ν (see, e.g., Royden, 1988). The Lebesgue order is then simply $\nu\leqslant\mu$ iff $\mu_\nu=\alpha\nu$, $\alpha>0$. (α is real, not complex, otherwise the resulting order would not be a partial order.) The character of the logic that results, and its application to periodic probabilistic systems, has yet to be investigated systematically.

Remark 85. If $[\mu_1]=[\mu_2]$ but $\mu_1\neq\mu_2$ then μ_1 and μ_2 are not comparable in the Lebesgue order. However, in general $L_{\mu_1}\cap L_{\mu_2}\neq\emptyset$.

4. Probabilistic Inferences: Morphisms of Lebesgue Logics

A probabilistic inference can often be conceived of as a map from a space of premises, with its logic, to a space of conclusions, with its logic. This map should not be arbitrary. Instead it should map premises to conclusions in a manner that respects the logics of both: If premise ν entails premise μ then the conclusion associated to ν should entail the conclusion associated to μ. Such a map is called a *morphism*. Intuitively, if an inference is modeled by a morphism, then as one increases the resolution of the premises one correspondingly increases the resolution of the conclusions. These ideas underlie the following definition.

Definition 86. Let $L(Y)$ and $L(X)$ be the Lebesgue logics of probability measures on measurable spaces $(Y,\mathscr{Y})$ and $(X,\mathscr{X})$ respectively. Then a map $\phi: L(Y)\to L(X)$ is a *Lebesgue morphism* iff for $\nu,\mu\in L(Y)$, (1) $\nu\leqslant\mu$ implies $\phi(\nu)\leqslant\phi(\mu)$ and (2) $\nu\perp\mu$ implies $\phi(\nu)\perp\phi(\mu)$.

An example of a Lebesgue morphism, and one which shows that such morphisms need not be linear, is the following. Let $Y=\{a,b\}$ and $X=\{c,d\}$ be measurable spaces each having just two points. The probability measures on both spaces are just the Dirac measures and their convex combinations. Consider the map $\phi: L(Y)\to L(X)$ given by $\alpha\varepsilon_a+(1-\alpha)\,\varepsilon_b\mapsto\alpha^2\varepsilon_c+(1-\alpha^2)\,\varepsilon_d$, where $0\leqslant\alpha\leqslant 1$ and ε_x denotes Dirac measure at the point x. The map ϕ is a Lebesgue isomorphism.

We now consider a few properties of Lebesgue morphisms.

Proposition 87. *If $\phi: \mathscr{M}(\mathscr{Y})\to\mathscr{M}(\mathscr{X})$ is a linear map that preserves singularity of measures then ϕ is a Lebesgue morphism.*

Proof. If ν, $\mu \in \mathcal{M}(Y)$ satisfy $\nu \leqslant \mu$ then we can write $\mu = \alpha\nu + \mu^\nu$ where $\nu \ll \mu_\nu$ and $\mu^\nu \perp \nu$. Then $\phi(\mu) = \alpha\phi(\nu) + \phi(\mu^\nu)$. It follows from this equation that $\phi(\nu) \ll \phi(\mu)$. We also know that $\phi(\mu^\nu) \perp \phi(\nu)$. So by the uniqueness of the Lebesgue decomposition $\phi(\nu) \leqslant \phi(\mu)$. ∎

PROPOSITION 88. *Let $\eta: Y \times \mathscr{X} \to [0, 1]$ be a markovian kernel*[8] *that preserves singularity of probability measures. Then η is an injective Lebesgue morphism.*

Proof. By Proposition 87, η is a Lebesgue morphism. Assume that η is not injective. Then there exist measures μ, $\nu \in L(Y)$ with $\mu \neq \nu$ but such that $\mu\eta = \nu\eta$, i.e., such that $(\mu - \nu)\eta = 0$. Let $p - q$ be the Jordan decomposition of $\mu - \nu$, i.e., p, $q \in \mathcal{M}(Y)$ and $p \perp q$. Then $(p - q)\eta = (\mu - \nu)\eta = 0$, which implies that

$$\frac{p}{\|p\|}\eta = \frac{q}{\|q\|}\eta.$$

But $p \perp q$, so η does not preserve singularity. ∎

We return to the subject of Lebesgue morphisms later in our discussions of observer theory.

5. STATES ON LEBESGUE LOGICS

It might seem to some that the language of probability measures is sufficiently cautious for dealing with uncertainty. After all, it allows us to avoid bald assertions to the effect that such and such an event certainly obtained or certainly did not obtain, and encourages instead more circumspect assertions regarding the probabilities of events. But, one could argue, even this might be overly brash: How can one know the precise probabilities of the various events? Perhaps committing to a particular probability measure is itself insufficiently cautious.

It turns out that there is a useful notion, already well developed for the case of ordinary logics, which bears on this issue: the notion of *state*. Here we use the term "ordinary logic" informally to mean one in which $\vee$ and $\wedge$ exist for all pairs of elements, in which there is a unit element, and in which, as a consequence, there is a global notion of complement. For such a logic $\mathscr{L}$, a state is defined to be a *measure* on $\mathscr{L}$, namely, a function $p: a \to p(a)$ for $a \in \mathscr{L}$ such that (i) p is real valued and $0 \leqslant p(a) \leqslant 1$ for all $a \in \mathscr{L}$, (ii) if a countable collection of elements $a_i \in \mathscr{L}$ satisfy $a_i \wedge a_j = 0$ for $i \neq j$ then $p(\bigvee_i a_i) = \sum_i p(a_i)$. As we have seen there is no globally defined unit in the Lebesgue logic, and $\vee$ and $\wedge$ are not defined for all

[8] Recall that if $(X, \mathscr{X})$ and $(Y, \mathscr{Y})$ are measurable spaces, a *kernel* on $Y \times \mathscr{X}$ is a map $\eta: Y \times \mathscr{X} \to \mathbf{R} \cup \{\infty\}$ such that (1) for every $y \in Y$, the mapping $A \mapsto \eta(y, A)$ is a measure on X, denoted $\eta(y, \cdot)$, and (2) for every $A \in \mathscr{X}$, the mapping $y \mapsto \eta(y, A)$ is a measurable function on Y, denoted $\eta(\cdot, A)$. Given a measure $\nu \in L(Y)$, the measure $\nu\eta$ given by $\nu\eta(A) = \int_Y \nu(dy)\,\eta(y, A)$, $A \in \mathscr{X}$, is a measure in $L(X)$. Thus η can be viewed as a map from $L(Y)$ to $L(X)$ given by $\nu \mapsto \nu\eta$.

pairs of measures. Still this definition of state literally makes sense for the Lebesgue logic, and we formalize it in the following definition.[9]

DEFINITION 89. A *state* on a Lebesgue logic $L(X)$ is a map $S: L(X) \to [0, 1]$ such that

(i) $S(0)=0$; $1 \in \text{Range}(S)$;

(ii) If a countable collection of measures $v_i \in L(X)$ satisfy $v_i \wedge v_j = 0$, for $i \neq j$, then $S(\sum_i \alpha_i v_i) = \sum_i S(v_i)$, where $\alpha_i \in (0, 1]$ and $\sum_i \alpha_i = 1$.

The striking consequence of this definition is that the resulting states are essentially discrete, as indicated in the following proposition (for which we are indebted to an anonymous reviewer).

PROPOSITION 90. *Let S be a state on $L(X)$ in the sense of Definition 89. Then there is an atomic measure ρ whose set of atoms is a collection of sets $\{A_i\}$ with $\rho(A_i) = \rho_i$, satisfying the following property: For any $\lambda \in L(X)$, $S(\lambda) = \sum_{j \in I_\lambda} \rho_j$, where $I_\lambda = \{j: A_j$ is an atom of $\lambda\}$.*

Proof. If S is a state then according to Definition 89 there is a probability measure $\mu \in L(X)$ with $S(\mu)=1$. If $v \wedge \mu = 0$ then $S(\mu + v) = S(\mu) + S(v) = 1$ so $S(v)=0$. If $v \perp \mu$ then $S(v)=0$. As a result, for any $v \in L(X)$, $S(v) = S(v_\mu)$. Write $v_\mu = f_1 \mu + f_2 \mu$ where $\mu(\{f_2 = t\}) = 0$ for all $t > 0$ and f_1 is a step function $\sum_{i=1}^\infty \alpha_i I_{A_i}$. (Here I_{A_i} denotes the indicator function of the set A_i.) Since $(f_2 \mu) \wedge \mu = 0$ we have $S(v_\mu) = S(f_1 \mu)$. That is, S is essentially determined by its values on μ-step measures $\sum_{i=1}^\infty \alpha_i \mu(\cdot \mid A_i)$. If μ is not atomic there is an A_0 with $\mu(A_0) > 0$ such that $\mu(\cdot \mid A_0)$ is non-atomic. On A_0 there is a function $0 < f < 1$ with $\mu(\{f = t\}) = 0$ if $0 < t < 1$. Write $\mu(\cdot \mid A_0) = f\mu(\cdot \mid A_0) + (1-f)\, \mu(\cdot \mid A_0)$. Since $f\mu(\cdot \mid A_0) \wedge \mu = 0 = (1-f)\, \mu(\cdot \mid A_0) \wedge \mu$ one has $S(f\mu(\cdot \mid A_0)) = S((1-f)\, \mu(\cdot \mid A_0) = 0$. This, together with the hypothesis on f implies that $f\mu(\cdot \mid A_0) \wedge (1-f)\, \mu(\cdot \mid A_0) = 0$ so $S(\mu(\cdot \mid A_0)) = 0$. A_0 may be chosen so that $\mu(\cdot \mid A_0^c)$ is atomic. We have $S(\mu(\cdot \mid A_0^c)) = 1$. Replace μ by $\mu(\cdot \mid A_0^c)$. Let $\{A_i : i \in \mathbf{N}\}$ be μ-atoms with $\sum_{i=1}^\infty \mu(A_i) = 1$. Let $\rho_i = S(\mu(\cdot \mid A_i))$. The proposition now follows directly. ∎

Definition 89 yields states which assign nonzero values to atomic measures only. However, we can weaken part (ii) of the definition to allow "generalized states." The natural way to do this is to require only that states be additive over convex combinations of probability measures which are *singular*, as opposed to requiring that additivity hold whenever the greatest lower bound is 0. (Recall that in the Lebesgue logic if $v_i \perp v_j$ then $v_i \wedge v_j = 0$ but not conversely.)

[9] The notion of a state on the Lebesgue logic is quite similar in spirit to the notion of probability measure on a σ-additive class discussed by Gudder (1988). Gudder, following Suppes (1966b), argues that σ-algebras are unnecessarily restrictive structures for the domain of definition of probability measures. He argues instead for a more general class of structures in which the meet and join of two arbitrarily chosen members need not exist.

DEFINITION 91. A *generalized state* on a Lebesgue logic $L(X)$ is a map $S: L(X) \to [0, 1]$ such that

(i) $S(0)=0$; $1 \in \text{Range}(S)$;

(ii) If a countable collection of measures $\nu_i \in L(X)$ satisfy $\nu_i \perp \nu_j$, for $i \neq j$, then $S(\sum_i \alpha_i \nu_i) = \sum_i S(\nu_i)$, where $\alpha_i \in (0, 1]$ and $\sum_i \alpha_i = 1$.

Here is an important intuition: If we restrict our attention to the local sublogics $L_\mu(X)$ then the condition that $\nu_i \wedge \nu_j = 0$ and the condition that $\nu_i \perp \nu_j$ are identical. Therefore, we may think of a generalized state as a global object which "glues together" genuine states of the local boolean logics L_μ, much as the Lebesgue logic itself is a generalized boolean logic in that it "glues together" the boolean logics L_μ.

We now give several examples of ways to generate generalized states; they correspond to ways of strengthening condition (ii) in Definition 91 by constraining $S(\nu)$, for $\nu \ll \mu$, in terms of some property of the relationship between μ and ν.

EXAMPLE 92. Choose μ and choose $\rho \ll \mu$. We then define S by $S(\nu) = \rho(A_{\nu,\mu})$ (where, as usual $A_{\nu,\mu}$ is the domain of mutual absolute continuity in the Hahn presentation—see Section 2). Note that for any $A \in \mathscr{X}$ we have $S(\mu(\cdot \mid A)) = \rho(A)$. Since $\rho \ll \mu$ we can view ρ as a measure on $\mathscr{X}_\mu$, the boolean measure algebra modulo μ (as discussed in section two, $\mathscr{X}_\mu$ is canonically isomorphic to C_μ, the lattice of measure classes absolutely continuous with respect to μ, and is also canonically isomorphic to L_μ, the sublattice of the Lebesgue logic consisting of measures less than μ.) Thus we may view this state S as the composition

$$L(X) \xrightarrow{\phi_\mu} \mathscr{X}_\mu \xrightarrow{\rho} [0, 1],$$

where ϕ_μ is the map $\phi_\mu(\nu) = A_{\mu,\nu}$.

EXAMPLE 93. Choose $\mu \in L(X)$. Then for any $\nu \in L(X)$ consider $\nu(\cdot \mid A_{\nu,\mu})$. Write $\nu(\cdot \mid A_{\nu,\mu}) = \nu_1 + \nu_2$ where $\nu_1 = \sum_i \alpha_i \mu(\cdot \mid A_i)$ and $\nu_2 \wedge \mu = 0$. Then let $S(\nu) = \sum_i S(\mu(\cdot \mid A_i))$. These generalized states are in one-to-one correspondence with pairs (μ, ρ), where $\mu \in L(X)$ and ρ is a probability measure on $\mathscr{X}_\mu$ (i.e., $\rho \ll \mu$), via the composition

$$L(X) \xrightarrow{\psi_\mu} \mathscr{X}_\mu \xrightarrow{\rho} [0, 1]. \tag{94}$$

Here ψ_μ is the map $\psi_\mu(\nu) = B_{\nu,\mu}$ where $B_{\nu,\mu} \subset A_{\nu,\mu}$ is the set on which ν is a "μ-step measure," i.e., on which $d\nu_\mu/d\mu$ is locally constant a.e. μ. We have then simply that $S(\nu) = \rho(B_{\nu,\mu})$.

We remark that *globally* the generalized states (Definition 91), as well as the Lebesgue states themselves (Definition 89), fail to have certain attractive properties (such as regularity and existence of supports—see, e.g., Beltrametti & Cassinelli, 1981), but that *locally*, on the boolean sublogics L_μ, these properties obtain. Finally, we remark that the notion of a state, or generalized state, on the Lebesgue logic corresponds intuitively to a notion of a measure on probability measures.

6. A Concrete Application: Perception

Although much remains to be done in developing the formalism and mathematical implications of the Lebesgue logic, we now turn to consider a concrete application. One would guess that there are many applications since there are many disciplines that require probabilistic reasoning. But we have chosen to consider an application in the scientific discipline of perception, largely because it was examination of observer theory, a formal theory of perception, that led to the development of the Lebesgue logic. We sketch the relationship of the Lebesgue logic to observer theory, in particular to a formal structure called an "observer," and then suggest how the Lebesgue logic might apply to the perceptual problem of sensor fusion. But first it is appropriate to indicate, at least briefly, why probabilities and probabilistic reasoning are central to perception. For this we refer to Fig. 7 and the following principles.

Principle 1. *No Perception without Representation.* Why do you perceive a cube when you look at Fig. 7? Quite simply, most perceptual theorists now agree, because your visual processing of the figure eventuates, at some point, in a representation of a cube. Such visual processing might also eventuate in other, unperceived, representations. Thus, although there is no perception without representation, there might be representation without (conscious) perception.

Principle 2: *No Representation without Representations.* When you see the cube you also see less than the cube: you see a flat line drawing. And when you see the line drawing you also see less than the line drawing: you see an array of light and dark. (Similarly, *mutatis mutandis*, for the staircase and other drawings.) But by the first principle, what you see is what you represent. So you represent the line drawing and the array of light and dark. You can not, in general, have one representation without entertaining others as well. Perceptual processes generate many distinct representations.

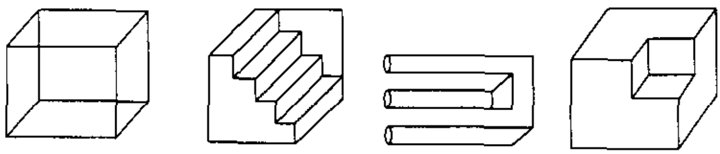

FIG. 7. Some multistable figures (from left to right): the Necker cube, Schröder's staircase, the devil's tuning fork, and a notched cube.

Principle 3: *No Representations without Inferences.* The figure is flat, but cubes are not. So how do you see a cube when you look at the figure? You infer it. Your representation of the flat line drawing (discussed in Principle 2) is a premise from which you infer the representation of the three-dimensional cube as a conclusion. Special properties of the drawing that are unlikely by accident, such as the precise coincidence of the tips of different line segments, provide human vision the license it needs to make the inference. Disrupt these properties, disturb the coincidences, and the cube disappears. The inference is no longer licensed and the conclusion no longer drawn. Of course in most cases you are not aware of the inferential process, only of (some of) its conclusions. Perception, as Helmholtz put it, involves *unconscious* inferences.

Principle 4. *No Inferences without Risks.* When you see the cube you are, in an obvious sense, wrong. For, as we said, drawings are flat but cubes are not. To perceive the cube is therefore to misperceive, to hallucinate a shape with depth when in fact, in the flat figure, there is none. Those special properties of the figure which normally indicate a proper premise for a legitimate inference of three-dimensional shape have, in this case, misled the system. Perception does not traffic in clean deductive inferences, whose only risk is error by inadvertence, but in inferences whose conclusions typically go well beyond the premises, and whose risk, therefore, is intrinsic. Perceptual inferences are, by nature, risky business.

Principle 5: *No Risks without Probabilities.* When you see the cube you see more than a cube, you see two cubes, and you flip from one to the other. Both are compatible with the drawing but, since at most one object can be in one place at one time, at most one cube can be the right interpretation. Therefore accepting either interpretation entails the risk of being wrong. A rational solution, and one apparently adopted by human vision, assesses the risks and assigns probabilities of correctness. For the cube the two interpretations are deemed equally likely and, accordingly, you see each interpretation about half the time. For the staircase the two interpretations are not deemed equally likely and, accordingly, you see each interpretation with differing frequency.[10] In these and many other cases, the conclusion of the inference is not *one* interpretation, but a probability measure over possible interpretations. In those cases in which just one interpretation is selected, the probability measure gives a weight of one to that interpretation. Thus perceptual conclusions are, in all cases, probability measures. These conclusions, in turn, are often premises for further perceptual inferences. For this reason (and because of that ever-present nemesis—noise) perceptual premises are also probability measures. Perceptual inferences, then, take probabilistic premises to probabilistic conclusions.

Motivated in part by these principles, Bennett, Hoffman, &Prakash (1987, 1989, 1991) proposed a formalism to model the probabilistic inferences typical of

[10] The temporal behavior of multistable percepts is studied in Ditzinger and Haken (1989) and Mueller and Blake (1989).

perception. This formalism is called observer theory, and the aspect of this theory of primary interest here is its definition of an observer. We recall this definition (schematized in Fig. 8).

DEFINITION 95. An *observer* is a six-tuple (X, Y, E, S, π, η) where

1. X and Y are measurable spaces. E is an event of X. S is an event of Y. Points of X and Y are measurable.

2. π is a measurable map from X onto Y such that $\pi(E) = S$.

3. $\eta: Y \times \mathscr{X} \to [0, 1]$ is a kernel such that, for $s \in S$, $\eta(s, \pi^{-1}(s) \cap E) = 1$, and for $y \in Y - S$, $\eta(y, X) = 0$.

We illustrate this definition with a concrete example. But first it is appropriate to note how the six components of an observer model a perceptual inference. X is a space of "elementary conclusions." The set of all probability measures on X is the set of all possible conclusions of the observer's inference. E is a subset of X containing the "distinguished" elementary conclusions. It embodies the constraint used by the observer to interpret ambiguous sensory information. Probability measures on E are conclusions in accord with this constraint. Y is the space of "elementary premises." The set of all probability measures on Y is the set of all possible premises of the observer. S is a subset of Y containing the "distinguished"

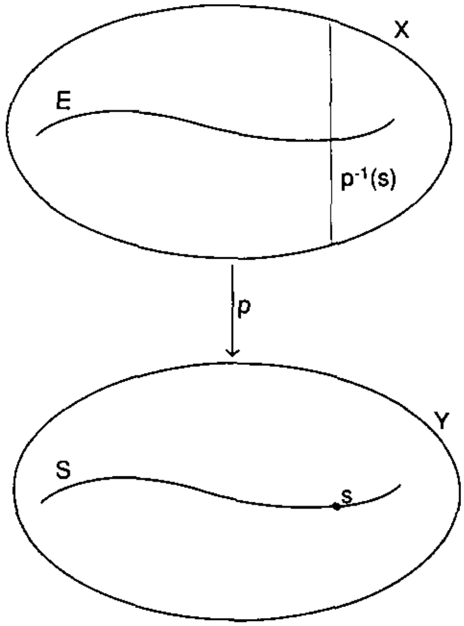

FIG. 8. An illustration of the definition of observer.

elementary premises. Probability measures on S are premises for which the observer can reach a conclusion in accord with its constraint. π is a function from X to Y which specifies, for each *elementary* premise (i.e., each point of Y) the set of compatible *elementary* conclusions (i.e., points of X). η is the inference map. To each probability measure on S (each distinguished premise) η associates a probability measure on E (a distinguished conclusion).

For a concrete example we turn to Ullman's (1979) theory of structure from rigid motion. It is a remarkable capacity of human vision that it often can, by viewing dynamic two-dimensional images, such as those that appear on a television screen, perceive the three-dimensional structures and motions which generated the two-dimensional images. To explain this ability, Ullman proves that a reasonable inference about 3D structure can be made on the basis of three distinct orthographic (parallel projected) views of four noncoplanar points. Under such conditions, Ullman proves, the number of rigid three-dimensional interpretations of the views is generically at most two (the two being mirror reflections of each other about the image plane) and the probability of false targests (roughly, the probability that the views are compatible with a rigid 3D interpretation when in fact such an interpretation is false) is zero. For this visual inference the observer structure is as follows. Y, the elementary premises, is the space of all sets of three views of four points, i.e., $Y = \mathbf{R}^{24}$ (4 points by 3 views by 2 coordinates per point). X, the 3D interpretations, is all triples of four points in $3D$ space, i.e., $X = \mathbf{R}^{36}$ (4 points by 3 instants of time by 3 coordinates per point). $E \subset X$ is the rigid interpretations. π from X to Y is given by orthographic projection. $S \subset Y$ are those three views of four points that are compatible with at least one rigid interpretation. And to each premise (point) of S the kernel η assigns a probability measure supported on the two rigid interpretations that are compatible, according to Ullman's theorem, with that premise.

This example bears on the relationship between observer theory and standard information theory, as follows. Ullman's theory predicts that if one presents a human subject with an appropriate physical stimulus, the subject will see not just one rigid interpretation, but two. And in fact this is what happens; the subject sees first one rigid interpretation, then the other, and then continues to alternate between them. The subject does not know which of the two interpretations is correct. Indeed *both interpretations may be incorrect*, for one can easily show that the mapping from the external world to the retinal images is infinite to one, and that therefore *the set of possible three-dimensional interpretations is infinite*. Thus the uncertainty here is of a fundamentally different kind than that treated by the customary notion of an information channel (as introduced by Shannon, 1948). To put this somewhat differently, no recoding of the physical situation can eliminate the equivocation here. The retinal information is intrinsically insufficient, even in the absence of noise, to allow a nonarbitrary choice from the infinite set of possible three-dimensional interpretations. This uncertainty is fundamental, an uncertainty which no improvement in resolution can remove. A crucial task of the perceptual psychologist, therefore, is to discover what background assumptions a sensory system employs in choosing certain interpretations in the presence of this fundamental uncertainty. It is exactly the structure of these assumptions in the presence of fundamental uncertainty that is not accessible to classical information-theoretic analyses and which is the *raison d'etre* of observer theory. The fibres of the map π, for example, express the fundamental uncertainty, and E, the distinguished interpretations, express the background assumptions used by the observer to overcome this uncertainty. By contrast, the role of information theory here would be to study the manner in which noise "perturbs" π. Strictly speaking, in the observer formalism one does not consider X to model a source, only a class of interpretations. The source may be regarded as separate from both X and Y, and one may imagine that this source and Y are related by a channel if one wishes. One may then use information theory to study stochastic properties of stimulus reception. The interpretive part of the perceptual processing, however, with its fundamental uncertainty and biases, is encoded in π and η, i.e., in the observer structure.

More concrete examples of observers and further details about observer theory are given by Bennett *et al.* (1989, 1991). Our interest here is to describe the relationship between Lebesgue morphisms and observers. For this, note that each observer has a collection of probability measures which serve as its premises, viz., the set $L(Y)$ of all probability measures on Y with the Lebesgue order. Moreover, each observer has a collection of probability measures which serve as its conclusions, viz., the set $L(X)$ of all probability measures on X with the Lebesgue order. The observer's inference is a function, $\bar{\eta}$, from its Lebesgue logic of premises $L(Y)$ to its Lebesgue logic of conclusions $L(X)$. The inference function $\bar{\eta}$ derives from the observer's kernel η as follows. First, recall that any kernel can be viewed as a linear operator on measures, $\nu \mapsto \nu\eta$, defined by

$$\nu\eta(A) = \int_Y \nu(dy)\,\eta(y, A). \tag{96}$$

Then

DEFINITION 97. $\bar{\eta}: L(Y) \to L(X)$ is given by

$$\nu\bar{\eta} = \begin{cases} \overline{(\nu\eta)} & \text{if } \nu\eta \neq 0 \\ 0 & \text{if } \nu\eta = 0. \end{cases}$$

Now for the main point. The key relationship between observers and Lebesgue morphisms is this: the map $\bar{\eta}$, it happens, respects the Lebesgue logics of premises and conclusions; it is a logic morphism. The proof of the following theorem can be found in Bennett, Hoffman, and Murthy (1993).

THEOREM 98. *Let $O = (X, Y, E, S, \pi, \eta)$ be an observer, and let $L(Y)$ and $L(X)$ denote the Lebesgue logics of probability measures on Y and X respectively. Then $\bar{\eta}: L(Y) \to L(X)$ is a Lebesgue morphism.*

This theorem indicates that the inferences of observers are morphisms from Lebesgue logics of premises to Lebesgue logics of conclusions. The question of the converse naturally arises: If an inference is given by a mapping between Lebesgue logics, can it be described by a canonically chosen observer? The answer, in many cases, is yes. We state two theorems which are proved in Bennett, Hoffman, and Murthy (1993).

THEOREM 99. *If* $(Y, 2^Y)$ *is a discrete measurable space,* $(X, \mathscr{X})$ *any measurable space, and* $\phi: L(Y) \to L(X)$ *an injective morphism of Lebesgue logics, then there exists a canonical representation of* ϕ *by an observer.*

THEOREM 100. *Let* $(Y, \mathscr{Y})$ *and* $(Z, \mathscr{Z})$ *be metric spaces with their associated measurable structures. Let* $\mathscr{B}(Z)$ *denote the positive measures* μ *such that* $\mu(Z) \leqslant 1$, *i.e., the "subprobability" measures. Let* $\phi: L(Y) \to \mathscr{B}(Z)$ *be a convex, vaguely continuous map. Then there exists a canonical observer representation of* ϕ.

These theorems point to the centrality of observers for information-processing theories of perception (and for inductive inferences more generally). If one admits, as is now generally accepted, that perceptual information processing is probabilistic, and if one admits, as seems reasonable, that perceptual inferences should respect the Lebesgue logic of probabilities then, according to these theorems, for a large class of perceptual inferences there is a canonically associated observer: the definition of observer provides a normal form of description. (The "observer thesis" then makes the further assertion that *every* perceptual inference can be described by an observer. This is discussed in Bennett *et al.*, 1989.)

Observables

One can use the fact that $\bar{\eta}$ is a Lebesgue morphism to give a principled definition of "observable" in the language of observer theory. This is of interest in its own right, but we devote attention to it now because it is a necessary step toward applying the Lebesgue logic to the problem of sensor fusion. Our discussion of observables is guided in part by the work of Varadarajan (1985).

Recall that to each classical mechanical system $\mathscr{G}$ there is associated a phase space $\mathscr{S}$ whose points are in one-to-one correspondence with the states of $\mathscr{G}$. The dynamics of $\mathscr{G}$ is specified by a Hamiltonian, a smooth function $H: \mathscr{S} \to \mathbf{R}$. The observables of $\mathscr{G}$ are described by real valued functions on $\mathscr{S}$. For instance, if $\mathscr{G}$ is a single particle of mass m which moves along a single line under a potential field V, then $\mathscr{S}$ is $\mathbf{R}^2$ and the Hamiltonian H is given by

$$H(x, p) = \frac{1}{2m} p^2 + V(x), \tag{101}$$

where x is the position along the line and p the momentum. The function $s \mapsto p^2/2m$ is the kinetic energy of the particle and the function $s \mapsto V(x)$ its potential energy.

In general, if $g: \mathscr{S} \to \mathbf{R}$ describes an observable, then $g(s)$ is the value of that observable when $\mathscr{G}$ is in the state s. An observation results in an observation statement: "the value of g is in the measurable set A of $\mathbf{R}$." The set of all such statements (for the observable g) is, of course, in one-to-one correspondence with the measurable sets of $\mathbf{R}$. The inclusion relation among such sets naturally models the relation of entailment among the statements. From this it follows that the logic of the statements is boolean. Now given a statement of the form "the value of g is in the measurable set A of $\mathbf{R}$" one can *infer* the statement: "the state of the system lies in the set $g^{-1}(A)$." These are the most general empirically verifiable statements that can be made about $\mathscr{G}$. Since they are in one-to-one correspondence with certain subsets of $\mathscr{S}$ (viz., subsets of the form $g^{-1}(A)$) their logic is also boolean. If $g: \mathscr{S} \to \mathbf{R}$ is an observable then the map g^{-1} is easily seen to be a morphism from the boolean logic of observation statements (i.e., measurable sets of $\mathbf{R}$) into the boolean logic of empirically verifiable statements about $\mathscr{G}$ (i.e., measurable sets of $\mathscr{S}$). The morphism g^{-1} is informationally equivalent to g and is therefore also called an observable. We conclude that the observables of a classical mechanical system are morphisms from the boolean logic of (measurable sets of) $\mathbf{R}$ into a boolean logic of empirically verifiable statements about the system.

When we consider the observables of a quantum mechanical system the story is similar except in one important respect: the logic of empirically verifiable statements about a quantum system is not boolean. The observables are still morphisms from the boolean logic of $\mathbf{R}$ into a logic of empirically verifiable statements about the system, but this latter logic is no longer assumed to be boolean. It is, in fact, assumed to be a nondistributive lattice, the lattice of subspaces of a Hilbert space. From this assumption about the logic flow the counterintuitive features of quantum theory.

Taking our lead from these examples we can define, in the language of observer theory, an observable for a perceptual inference. Denote the Lebesgue logic of premises for the inference by $L(Y)$ and the Lebesgue logic of conclusions by $L(X)$. Then an observable for the inference is a morphism from a boolean sublogic of $L(Y)$ into the logic $L(X)$. Here we have substituted a boolean sublogic of $L(Y)$ for the boolean logic of measurable sets of $\mathbf{R}$. And we have substituted the Lebesgue logic $L(X)$ for the logic of empirically verifiable statements. We now make this more precise using the language of observer theory.

DEFINITION 102. Let $O = (X, Y, E, S, \pi, \eta)$ be an observer, and let $L(Y)$ and $L(X)$ denote the Lebesgue logics of probability measures on Y and X respectively. For each $\nu \in L(Y)$ denote by $L_\nu(Y)$ the boolean sublogic associated to ν (see Proposition 42). Then an *observable* for O is a logic morphism $\eta_\nu: L_\nu(Y) \to L(X)$ obtained by restricting the domain of $\bar{\eta}$ to L_ν.

Intuitively, an observable is the restriction of a perceptual inference to a "nice" set of possible premises, viz., premises that are all mutually compatible.

There is clearly a one-to-one correspondence between observables for a perceptual inference and elements of $L(Y)$. If we make the natural definition that $\eta_\nu \leqslant \eta_\mu$ iff $\nu \leqslant \mu$ then we see that the logic of observables is isomorphic to $L(Y)$. This means that we can talk of compatible observables and simultaneously verifiable observables, just as we have done for premises and conclusions. It also means that the observables associated to perceptual inferences are much closer in spirit to the observables of quantum theory than to those of classical mechanics: the logic of observables for perceptual inferences is not boolean and admits observables that are not simultaneously verifiable (for a discussion of quantum observables see, e.g., Wheeler & Zurek (1983)). This suggests that nonboolean logics of observation are not restricted to the submicroscopic domain. Even in the macroscopic world of normal perception the logic of observation is not boolean.

Sequential Observables

In the Lebesgue logic not all propositions are simultaneously verifiable. As we have seen earlier, this property of the logic, when combined with observer theory, predicts that some pairs of observables are not simultaneously observable, even in normal (i.e., macroscopic) vision. Pairs that are not simultaneously observable are *sequential observables.* Sequential observables in human vision are not hard to find. One example is binocular rivalry. Normally the images falling at the left and right eyes are quite similar, and human vision uses any small disparities between them to infer a unified perception of depth. But if the two eyes are shown very dissimilar images a unified percept becomes impossible. One perceives the two images in succession; first one for a while, then the other. During a transition one sometimes sees a composite image, a mosaic of patches from both images. An example is shown in Fig. 9.

Another example of sequential observables can be seen in line drawings of "impossible objects" such as the "devil's tuning fork" of Fig. 7. You can either perceive the horizontal lines as occluding contours of cylinders or as edges of polyhedra. You do not perceive both interpretations simultaneously; instead you flip back and forth between them.

Structure from motion provides another example. In many displays of structure from motion, human subjects perceive, sequentially, two different interpretations of the 3D structure and motion. (Each interpretation is a mirror reversal, about the image plane, of the other.) This example differs in an interesting fashion from the examples of binocular rivalry and impossible objects: whereas the latter examples can be attributed to *premises* that are not simultaneously verifiable (incompatible images in the case of rivalry and incompatible cues to depth and form in the case of impossible objects) the example of structure from motion cannot: here multistable percepts arise *even when there is but one premise.* Rather, the multistable

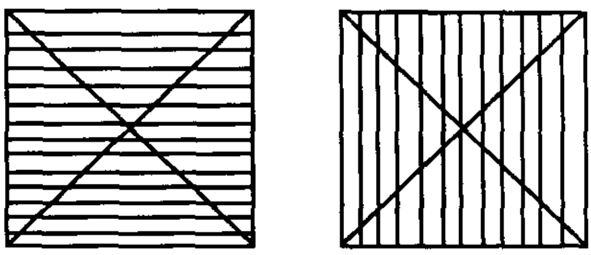

FIG. 9. An illustration of binocular rivalry. If you cross your eyes so that the large X's fuse, you will see the horizontal and vertical lines appear sequentially in time. This is an example of sequential observables in human vision.

percept is due to the probability measure describing the conclusion. This deserves brief discussion.

Let us return once again to Ullman's theory of structure from motion. We said earlier that according to this theory each premise compatible with a rigid interpretation is, in fact, compatible with two (the two being reflections of each other about the image plane). Accordingly, for each such premise s the associated conclusion $\eta(s, \cdot)$ is a probability measure supported on two separate points in X corresponding to the two interpretations. Now $\bar{\eta}$, according to Theorem 98, is a morphism between premises and conclusions: to a premise λ, the morphism $\bar{\eta}$ assigns a probability measure $\lambda\bar{\eta}$ as the conclusion. But $\lambda\bar{\eta}$ in this case is the sum of two subprobability measures, say μ_1 and μ_2, whose supports are *mutually disjoint.* Now not all probability measures can be written as a sum of measures having disjoint supports. A gaussian, for instance, cannot. Cut it however you like into two measures, the supports of those measures, being closed sets, will not be disjoint (they will overlap on their boundaries). So it is notable when a measure can be written as the sum of disjoint parts. We will call such a measure *reducible.*

Any time a measure is reducible, the (normalized) component measures are mutually singular and therefore not simultaneously verifiable in the Lebesgue logic (since the AND of two mutually singular probability measures is 0). In Ullman's case, $\lambda\bar{\eta} = \mu_1 + \mu_2$ where $\overline{\mu_1}$ and $\overline{\mu_2}$ are not simultaneously verifiable. So we do not see the interpretations corresponding to μ_1 and μ_2 simultaneously, but sequentially. The relative frequencies with which the first and second interpretations are seen are given respectively by $\|\mu_1\|$ and $\|\mu_2\|$. This is one role played by the probabilities in $\lambda\bar{\eta}$. But, intriguingly, there is a second role as well. The "variance" of μ_1 can, of course, change without changing $\|\mu_1\|$. And this variance corresponds to the precision or noisiness of the corresponding interpretation; no variance means a noise free interpretation, high variance a noisy interpretation. Similarly for $\|\mu_2\|$.

So any time a probability measure is reducible, the total weight of each part is its relative likelihood vis-à-vis the other parts, and the dispersion of each part is a measure of its noisiness or uncertainty.[11] Because the parts are not simultaneously verifiable in the Lebesgue logic they are seen sequentially.

[11] This perceptual interpretation of probability measures, in terms of reported frequencies of percepts and measured variance in subject responses, provides a way to connect the formalism of the Lebesgue logic with empirical data from psychophysical experiments. Thus the Lebesgue logic has empirical content for this domain. Nonetheless it is natural to question the need to use probabilities at all for the representation and calculus of uncertainties. As we noted before, the notion of state on a Lebesgue logic may allow us to back away from commitment to a particular probability measure. But another direction can be pursued as well: the use of orders rather than measures. Let $(X, \mathscr{X}, \mu)$ be a measure space, and let $\langle\mu\rangle$ denote the order induced on $\mathscr{X}$ by μ. (For $A, B \in \mathscr{X}$, we set $A \leqslant B$ in the order $\langle\mu\rangle$ if $\mu(A) \leqslant \mu(B)$ in the natural order on the real line.) It might be that the representation of uncertainty used, say, in human vision, is not a measure such as μ but an order such as $\langle\mu\rangle$. Premises of visual inferences, as well as conclusions, might be orders, not probability measures. In this case the inferencing mechanism probably cannot be a kernel η for the following reasons. First $\langle\nu\rangle = \langle\alpha\mu\rangle$ does not in general imply that $\langle\nu\eta\rangle = \langle\mu\eta\rangle$. Second, suppose $\eta: Y \times \mathscr{X} \to [0, 1]$ is a kernel, and let $\langle\eta\rangle = \{\langle\eta(y, \cdot)\rangle | y \in Y\}$. Then $\langle\eta_1\rangle = \langle\eta_2\rangle$ does not imply that $\langle\nu\eta_1\rangle = \langle\nu\eta_2\rangle$. Thus a different inferencing mechanism seems to be required, certainly an interesting direction for study.

Cue Integration

The Lebesgue logic, as we have just seen, predicts failures of integration for cues that are not simultaneously verifiable. It also provides an algorithm for integration of cues that are simultaneously verifiable. Formal properties of this algorithm have already been discussed. Now let us turn to apply the Lebesgue logic to a real example of cue integration in vision. Here, since there has been much recent research on the topic, we have many examples from which to choose. There are empirical studies of the visual integration of stereo with motion parallax (Graham & Rogers, 1982; Rogers & Collett, 1989), stereo with monocular perspective and other monocular cues (Holway & Boring, 1941; Stevens & Brookes, 1988; van der Meer, 1979), stereo with specularities (Blake & Bülthoff, 1990), stereo with texture (Buckley, Frisby, & Mayhew, 1988, 1989), vision with speech (Massaro & Cohen, 1983), and various combinations of size, occlusion, motion parallax, and height in the picture plane (Bruno & Cutting, 1988; Ono, Rivest, & Ono, 1986). There are also theoretical studies of cue integration based on multigrid methods (Terzopoulos, 1986), markov random fields (Aloimonos & Shulman, 1989; Poggio, 1985, 1987), algebraic models (Anderson, 1974), and probabilistic first order logic (Nilsson, 1986). We will consider the integration of stereo with structure from motion, as discussed by Richards (1984).

Richards invites us to recall the ambiguities of 3D interpretation that arise from stereo and motion when the two are used separately. From stereo one can infer the 3D structure of an object up to a single scale factor which depends on the fixation distance. Suppose, for instance, that one is looking at four points, $p_1, ..., p_4$, whose depths relative to the eyes are $d_1, ..., d_4$. Using small angle approximations, the horizontal disparity, δ_i, between point p_1 and p_i is given by $\delta_i = (d_1 - d_i)(I/d_1^2)$, where I is the interpupil separation between the eyes (or cameras) and where p_1 is the point being fixated. To use this equation to compute the d_i's from the δ_i's requires knowing the fixation distance d_1. The effect of different choices for d_1 is to scale in depth the 3D structure assigned to the remaining points: the structure stretches or shrinks in depth as one changes the value assumed for the fixation distance. Thus the conclusion of a stereo inference can, in this case, be considered a probability measure, say ν, on the one-parameter family of possible 3D interpretations.

Now recall that the ambiguity of interpretation for structure from motion, as in Ullman's analysis, is of a very different kind. Here the ambiguity is between a 3D interpretation and its mirror reflection. Thus the conclusion of a structure-from-motion inference, as we discussed before, can be considered a reducible measure $\mu = \mu_1 + \mu_2$.

The ambiguities for stereo and structure from motion are different. And here is Richards' key point: Of the two interpretations that arise from a motion analysis, *only one of them is, in general, a member of the one-parameter family of interpretations that arise from stereo.* Translating this into the language of probabilities, this means that ν (from stereo) is singular with respect to one component of μ (from motion), say the component μ_1, but not to the other, viz., μ_2. Consequently, if we try to integrate the cues ν and μ by taking their Lebesgue AND, μ_1 drops out in the process; the mirror ambiguity of structure from motion is eliminated. Put differently, μ AND ν reduces to μ_1 AND ν. (This reduction in ambiguity would also obtain were we to use the AND of the measure class order—discussed in Section 2—rather than the Lebesgue order.) What further reduction in ambiguity might occur depends on the precise forms of the measures μ_1 and ν. If, for example, the fixation distance is as likely to be underestimated as overestimated in the case of stereo, and if the errors in the motion analysis are as likely to stretch 3D structures as to compress them, then μ_1 and ν could have a nontrivial AND, resulting in a further reduction of ambiguity.

The Lebesgue AND can be used in a precisely analogous fashion to explain the integration of shading and texture together with relative disparities of specularity. One can devise displays of surfaces (Blake & Bülthoff, 1990) in which the monocular shading and texture information leads to a multistable percept: the perceived surface can be either convex or concave. When one then adds, in stereo, a specular highlight, subjects choose that interprepation of the surface which is consistent with the specularity. Here, as in the case of stereo and motion, the Lebesgue AND eliminates one component of a reducible probability measure. The two components of the reducible measure, in the present case, represent the multistability (concave versus convex) that arises from monocular texture and shading. The measure arising from stereo specularity is singular to one of the components, so that this component is eliminated in the Lebesgue AND.

7. Concluding Remarks

Probabilistic reasoning is central to many sciences. A probabilistic inference can often be viewed as a map from one collection of probability measures to another collection of probability measures. In this article we have shown that the collection of all probability measures on a given measurable space has a natural order, the Lebesgue order, and a concomitant logic, the Lebesgue logic. We have suggested that a natural condition on a map representing a probabilistic inference is that it respect this order, i.e., that it be a morphism of Lebesgue logics. We have also shown that there is an intimate relationship between observers—a formalization of inductive inference introduced by Bennett *et al.* (1989)—and maps that respect the Lebesgue logic. Finally, we have sketched a possible application of the Lebesgue logic to the problem of sensor fusion in perception.

Acknowledgments

We thank Marc Albert, Joe Arpaia, Jack Beusmans, Mike Braunstein, David Honig, Xiangen Hu, Ram Kakarala, Jin Kim, Alain Lewis, Jeff Liter, Louis Narens, Frank Pollick, Chetan Prakash, Whitman

Richards, Scott Richman, John Sinek, Brian Skyrms, and Peter Woodruff for helpful discussions. We also thank two anonymous reviewers for careful and insightful comments which greatly improved the paper. This work was supported by National Science Foundation Grant IRI-8700924 and by Office of Naval Research Contract N00014-88-K-0354.

REFERENCES

ADAMS, E., & LEVINE, H. (1975). On the uncertainties transmitted from premises to conclusions in deductive inferences. *Synthese*, **30**, 429–460.

ALOIMONOS, J.,& SHULMAN, D. (1989). *Integration of visual modules*. New York: Academic Press.

ANDERSON, N. H. (1974). Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Measurement, psychophysics, and neural information processing* (pp. 236–305). San Francisco: Freeman.

BELTRAMETTI, E., & CASSINELLI, G. (1981). *The logic of quantum mechanics*. Cambridge: Cambridge Univ. Press.

BENNETT, B., HOFFMAN, D., & PRAKASH, C. (1987). Perception and computation. *Proceedings of the IEEE First International Conference on Computer Vision, London* (pp. 356–364). Washington, DC: Computer Society Press.

BENNETT, B., HOFFMAN, D., & PRAKASH, C. (1989). *Observer mechanics: A formal theory of perception*. New York: Academic Press.

BENETT, B., HOFFMAN, D., & MURTHY, P. (1993). Observer representations of probabilistic inferences. In preparation.

BENNETT, B., HOFFMAN, D., & PRAKASH, P. (1991). Unity of perception. *Cognition*, **38**, 295–334.

BLAKE, A., & BÜLTHOFF, H. (1990). Does the brain know the physics of specular reflection? *Nature*, **343**, 165–168.

BRUNO, N., & CUTTING, J. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, **117**, 2, 161–170.

BUCKLEY, D., FRISBY, J., & MAYHEW, J. (1988). Interaction of texture and stereo cues in the perception of surface slant. *Perception*, **17** (Abstract).

BUCKLEY, D., FRISBY, J., & MAYHEW, J. (1989). Integration of stereo and texture cues in the formation of discontinuities during three-dimensional surface interpolation. *Perception*, **18**, 563–588.

DEMPSTER, A. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society, B*, **30**, 205–247.

DITZINGER, T., & HAKEN, H. (1989). Oscillations in the perception of ambiguous patterns. *Biological Cybernetics*, **61**, 279–287.

DUDA, R., HART, P., & Nilsson, N. (1981). Subjective Bayesian methods for rulebase inference systems. In B. Webber and N. Nilsson (Eds.) *Readings in artificial intelligence*. Palo Alto, CA: Tioga.

FINE, T. (1973). *Theories of probability*. New York: Academic Press.

FISCHLER, M., & FIRSCHEIN, O. (1987). *Intelligence: The eye, the brain, and the computer*. Reading, MA: Addison–Wesley.

GRAHAM, M., & ROGERS, B. (1982). Simultaneous and successive contrast effects in the perception of depth from motion and stereoscopic information. *Perception*, **11**, 247–262.

GROFMAN, B., & OWEN, G. (1986). *Information pooling and group decision making*. London: JAI Press.

GUDDER, S. (1988). *Quantum probability*. New York: Academic Press.

HALMOS, P. R. (1974). *Measure theory*. New York: Springer-Verlag.

HOLWAY, A., & BORING E. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology*, **54**, 21–37.

KRAUSS, P. H. (1968). Representation of conditional probability measures on Boolean algebras. *Acta Math. Acad. Sci. Hungaricae*, **19**, 229–241.

Massaro, D., & Cohen, M. (1983). Evaluation and integration of visual and auditory information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, **9**, 753–771.

Mueller, T. J., & Blake, R. (1989). A fresh look at the temporal dynamics of binocular rivalry. *Biological Cybernetics*, **61**, 223–232.

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

Ono, M., Rivest, J., & Ono, H. (1986). Depth perception as a function of motion parallax and absolute distance information. *Journal of Ecperimental Psychology: Human Perception and Performance*, **12**, 331–337.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann.

Poggio, T. (1985). *Integrating vision modules with coupled MRF's.* (Working Paper 285). Cambridge, MA: MIT AI Lab.

Poggio, T. (1987). MIT progress in understanding images. *Proceedings of the Image Understanding Workshop* (pp. 41–54).

Renyi, A. (1955). On a new axiomatic theory of probability. *Acta Math. Acad. Sci. Hungaricae*, **6**, 285–355.

Renyi, A. (1956). On conditional probability spaces, generated by a dimensionally ordered set of measures. *Theory Prob. Appl.* **1**, 61–71.

Renyi, A. (1970). *Foundations of probability.* San Francisco: Holden-Day.

Revuz, D. (1984). *Markov chains.* Amsterdam: North-Holland.

Richards, W. (1984). Structure from stereo and motion. *Journal of the Optical Society of America, A*, **2**, 343–349.

Rogers, B., & Collett, T. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *The Quarterly Journal of Experimental Psychology*, **41A**, 697–717.

Royden, H. (1988). *Real analysis.* New York: Macmillan.

Shafer, G. (1976). A mathematical theory of evidence. Princeton: Princeton Univ. Press.

Shannon, C. E. (1948). The mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423 and 623–656.

Stevens, K., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, **28**, 371–386.

Suppes, P. (1966a). Probabilistic inference and the concept of total evidence. In Hintikka and P. Suppes (Eds.), *Aspects of inductive logic.* Amsterdam: North-Holland.

Suppes, P. (1966b). The probabilistic argument for a nonclassical logic for quantum mechanics. *Philosophy of Science*, **33**, 14–21.

Terzopoulos, D. (1986). Integrating information from multiple sources. In A. P. Pentland (Ed.), *From pixels to predicates* (pp. 111–142). Norwood, NJ: Ablex.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Van der Meer, H. (1979). Interrelation of the effects of binocular disparity and perspective cues on judgments of depth and height. *Perception & Psychophysics*, **26**, 481–488.

Varadarajan, V. S. (1985). *Geometry of quantum mechanics.* New York: Springer-Verlag.

Walley, P. (1987). *Statistical reasoning with imprecise probabilities.* Chapman.

Wheeler, J., & Zurek, W. (1983). *Quantum theory and measurement.* Princeton: Princeton Univ. Press.

Zadeh, L. (1975). Fuzzy logic and approximate reasoning. *Synthese*, **30**, 407–428.

Received: October 2, 1990