Random Predicate Logic:
A Probabilistic Approach to Vagueness

William A. Dembski

Conceptual Foundations of Science
Baylor University
P. O. Box 7130
Waco, TX 76798

---

1. Old-Fangled Predicates

Predicates are supposed to slice reality neatly in two halves, one for which the predicate holds, the other for which it fails. Yet far from being razors, predicates tend to be dull knives that mangle reality. If reality is a tomato and predicates are knives, then when these knives divide the tomato, plenty of mush remains unaccounted for. Of course some knives are sharper than others, just as some predicates are less vague than others. "*x* is water" is certainly sharper than "*x* is beautiful." But perfect sharpness, perfect boundaries, and perfect separation seem only to obtain in mathematics.

The vagueness inherent in many predicates became particularly evident in the twentieth century. Quantum mechanics, the revival of certain ancient paradoxes, and the philosophy of science all contributed to a growing awareness that vagueness was ineliminable from many predicates. Quantum mechanical superposition seems to allow mutually exclusive simultaneous states. In searching for plausible interpretations of quantum mechanics, some researchers attempted to do away with classical bivalent logic. In its place they substituted multivalent quantum logics. Any logic with more than two values forces its predicates to slice reality into more than two parts.

Ancient paradoxes involving heaps and baldness also pointed up the

vagueness inherent in predicates. These paradoxes cut deeper to the question of vagueness than the quantum logics. The various logics physicists and philosophers invented to interpret quantum mechanics were all embedded in standard predicate logic: correctness of proofs for the new logics had to be verified using classical bivalent logic. Even the predicates of the new logics were sharp in the sense that they divided reality neatly; only now there were more divisions than in the customary binary case. But with these ancient paradoxes boundaries became irremediably fuzzy.

The paradox of the heap illustrates this point. Suppose we are given an unlimited supply of identical pebbles. Set aside a large space and start piling the pebbles on top of each other. Since our supply is unlimited, at some point in the process we shall have a heap. Certain claims are axiomatic:

(1) 0 pebbles cannot constitute a heap.

(2) If n pebbles constitute a heap, then so do n+1.

(3) If n pebbles do not constitute a heap, then neither do n–1.

Claim (3) is the contrapositive of (2). Now there is an obvious, intuitively compelling claim that we are inclined to add:

(4) If n pebbles do not constitute a heap, then neither do n+1.

The rationale here is that if something is not a heap, then adding one tiny pebble isn't suddenly going to transform it into a heap. Together with the principle of mathematical induction, (1) through (4) tell us that there are no heaps—their extension is null.

But this conclusion is clearly unacceptable. With an unlimited supply of pebbles we can fill up the known universe—10^{100} pebbles would certainly do it. Thus for some finite (big) M, M pebbles must constitute a heap. This is the paradox of the heap. If "*x* is a heap" is a binary predicate, we must either retain (4), thereby rendering this predicate vacuous, or we must discard (4), thereby equating the extension of "*x* is a heap" with something like "*x* has at least 1,739,665 pebbles." Neither of these alternatives is appealing. The paradox of baldness is essentially the same—how few hairs must a head possess to count as bald.

Finally, philosophers of science came to appreciate the inherent vagueness of many predicates as they examined the historical development of science. They found predicates evolving over time, getting sharper here,

expanding their extension there, and generally not staying put. Yet as predicates evolved, scientists continued to use the same linguistic forms to designate the predicates. Thus scientists used the predicate "*x* is water" before they knew about H_2O, and they still use "*x* is water" even after learning about atoms and molecules. Moreover, scientists assume some sort of continuity in this predicate: "x is water" has not changed so much that we can't understand ancient manuscripts when they speak of water.

Nor has increased precision and theoretical understanding eliminated vagueness from the predicates of science. Consider the predicate "*x* is H_2O," with its recourse to atomic physics. Does substituting this predicate for "*x* is water" yield a sharp, dichotomous predicate? It seems to. The world of atoms is after all discrete and finite. There are only so many configurations of matter. Either a piece of matter is constituted entirely of H_2O molecules or it isn't. So far so good. But we want a physical tie-in. We want to apply "*x* is H_2O" to objects of experience. Predicates always try to push reality into molds that reality in turn resists. As in ethics so also in science, purity is rare and precious. In material substances purity is statistically unlikely, with absolute purity statistically negligible. It is highly improbable that "*x* is H_2O" applies in a strict sense to any macro-object in the physical world. Any reasonable amount of water will contain impurities that render the predicate, strictly speaking, inapplicable.

To accommodate slight impurities, we might try to install a new predicate to model the concept water. Suppose we decide to go with "x is H_2O within one part per million." Given our technology we can purify water well beyond the tolerance this predicate imposes. But what happens at the limit? What about samples of water that contain more than 10^{20} molecules of H_2O and contain foreign elements so close to one part per million that our technology cannot decide whether the predicate applies? For such borderline cases the predicate will break down.

Still deeper problems remain. Scientists want their predicates well-defined, rationally grounded, readily related to observation statements, metaphysically unobjectionable, simple, natural, beautiful, elegant, etc. But a predicate like "*x* is H_2O within one part per million" is arbitrary, ad hoc, and ugly. What's more, it clearly depends on the more fundamental predicate "*x* is H_2O," which it is supposed to supplant. In the end, neither of these predicates is sharp and dichotomous.

Once philosophers, logicians, and scientists became convinced that the

problem of vagueness would not go away—despite technology's incessant drive toward precision—they turned to alternate logics that might reasonably model vague predicates. The philosophical literature is replete with such logics. Of these, the most popular is the approach to vagueness through fuzzy sets. Fuzzy logic has been a growth industry since the mid 1960s. As an approach to vagueness and uncertainty it has found numerous valuable applications.

I want in this paper to sketch a probabilistic way of handling vague predicates that constitutes not only an extension of fuzzy logic but also a return, or sorts, to traditional bivalent logic. Within fuzzy logic, predicates can assign any value in the unit interval [0,1]. Within the probabilistic logic that I propose, predicates assign only the values 0 and 1, but with varying probabilities. Vague predicates thus become traditional bivalent predicates interpreted against a probabilistic backdrop. As a mathematical formalism this approach subsumes the theory of fuzzy sets (at least first-order fuzzy sets). At the heart of this formalism is a new type of predicate, the *random predicate*. Laying out this formalism is the goal of this paper. Applying and interpreting it will be taken up in a follow-up paper. Let's begin with some probabilistic preliminaries.

2. How to Randomize

The word "random" acts as a universal prefix in mathematics much like the phrase "philosophy of" acts as a universal prefix in philosophy. Just about any discipline can have the phrase "philosophy of" tacked in front, thereby changing the discipline into a meta-discipline. Thus we have philosophy of science, philosophy of history, philosophy of psychology, philosophy of mathematics, etc. The self-referential philosophy of philosophy has yet to catch on, nor has the philosophy of science become sufficiently well-established to merit its own meta-discipline, the philosophy of the philosophy of science.

In mathematics, the adjective "random" is likewise a blanket prefix. To just about any mathematical object one can add the prefix "random" and thereby obtain a new, probabilistically endowed mathematical object. Thus variables become random variables, functions become random functions, operators become random operators, and probability measures—the very objects we require to make all these objects random—become random probability measures. Thus while "philosophy of" is still waiting for its

self-referential completion (i.e., "the philosophy of philosophy"), "random" has long since attained this distinction (see Kallenberg, 1986).

To turn a run-of-the-mill mathematical object into a revved-up random object is straightforward. In mathematics, almost everything is a set or function. Those objects that are not are usually so big and general (cf. categories and classes) that they first have to be converted into sets and functions to do the average mathematician much good. Working within one's favorite brand of set theory (ZFC has been at the top of the charts for some time), the mathematician finds that even the distinction between sets and functions is ultimately illusory (functions are a special type of set, so that functions collapse to sets). Now, traditional predicates are defined on sets and divide a set into those elements for which the predicate obtains and those for which it doesn't. Our task, then, in defining random predicates is to turn ordinary sets into random sets.

There is a perfectly standard way to do this: take a set S that needs to be randomized, take a probability space (Ω,F,**P**), and consider a family of functions from Ω into S, which we denote by $\rho(\Omega,S)$. $\rho(\Omega,S)$ is a randomization of S by the probability space (Ω,F,**P**). Note that $\rho(\Omega,S)$ is any subset of the space of all functions from Ω into S, which is commonly denoted by S^{Ω}. While this approach is perfectly standard, it is usually only implicit in the work of probabilists. Nevertheless, this is what probabilists do all the time, and it is worth considering some examples to illustrate how general this approach really is.

Random variables come to mind first. A real random variable is intuitively some random quantity that ranges over the real numbers **R**. **R** is therefore the commonplace object we need to randomize. Next our recipe calls for a probability space, say (Ω,F,**P**). Finally, we need to define a collection of functions from Ω to **R**, call it $\rho(\Omega,\mathbf{R})$. For convenience, let's restrict $\rho(\Omega,\mathbf{R})$ to the measurable functions from Ω to **R**. Here, then, is the collection of random variables.

Consider next random or stochastic processes. When students of probability are first introduced to stochastic processes, they are typically given the following definition: a stochastic process is an indexed collection of random variables. Thus in the continuous case, a stochastic process is a collection of random variables $\{X_t | t \in [0,\infty)\}$, where t runs over the nonnegative reals and represents time. From this picture it is difficult to see what nonrandom mathematical object was randomized to yield the stochastic process. But as probabilists have long recognized, the

interesting behavior of a random process is not how the random variable $X_t(\omega)$ varies at a fixed time slice t as ω varies in Ω, but rather how $X_t(\omega)$ varies at a fixed random point ω as time t is allowed to run. In effect $X_t(\omega)$ should be conceived as a function of t for fixed ω and not vice versa, as students are initially given to believe.

As a function of t, $X_t(\omega)$ becomes a random path in **R**. This suggests that the objects which our general approach must randomize is a collection of paths in **R**. Thus, our nonrandom object in need of randomization is some subset of $\mathbf{R}^{[0,\infty)}$ (the functions from the nonnegative reals to the reals). For convenience, let us restrict ourselves to continuous paths. Then our nonrandom object is $C([0,\infty),\mathbf{R})$, the continuous functions from $[0,\infty)$ to **R**. To randomize this object we take some probability space $(\Omega,F,\mathbf{P})$ and consider a family of functions from Ω to $C([0,\infty),\mathbf{R})$, call it $\rho(\Omega,C([0,\infty),\mathbf{R}))$. $C([0,\infty),\mathbf{R})$ comes endowed with certain nice topologies which yield nice Borel sets. Thus $\rho(\Omega,C([0,\infty),\mathbf{R}))$ is often taken to be measurable functions X from Ω to $C([0,\infty),\mathbf{R})$. When this is all unwrapped we see that for ω in Ω we get a function $X(\omega)$ in $C([0,\infty),\mathbf{R})$ which in turn for t gives us a real number $X_t(\omega)$. By, on the other hand, first fixing t, we get the original picture of a stochastic process as an indexed family of random variables.

A technical point is worth mentioning here. When a set we want to randomize is in fact a collection of functions, randomization has a convenient alternative formulation. This formulation is entirely equivalent to the one we just described. Thus when the collection S of objects we want to randomize is a collection of functions from U to V (i.e., S is a subset of V^U), then the randomization $\rho(\Omega,S)$ can be reconfigured as $\rho(\Omega xU,V)$. Here $\rho(\Omega xU,V)$ becomes a collection of functions f from ΩxU to V which for fixed $\omega \in \Omega$ is a function in S that sends $u \in U$ to $f(\omega,u) \in V$. In the preceding example $\rho(\Omega,C([0,\infty),\mathbf{R}))$ can be reconfigured as $\rho(\Omega x[0,\infty),\mathbf{R})$, which can be defined as a collection of measurable functions from $\Omega x[0,\infty)$ (assuming an appropriate σ-algebra on the product space) to **R** having continuous paths in the time variable $t \in [0,\infty)$.

Still more general examples exist. Thus mathematicians run random paths within curved geometric spaces, e.g., Brownian motions on Riemannian manifolds. They randomize linear operators between Banach spaces. And yes, even the probabilities that make objects random can be randomized. To summarize, if S is an arbitrary set and $(\Omega,F,\mathbf{P})$ an arbitrary probability space, then by definition any collection of functions $\rho(\Omega,S)$

from Ω to S is a *randomization of* S. If S is a collection of functions from a domain U to a target V, then this randomization can be taken as a corresponding collection of functions $\rho(\Omega \times U, V)$ from $\Omega \times U$ to V, each of which for fixed $\omega \in \Omega$ becomes a function in S.

I want to make one last preliminary remark, which concerns the probability space $(\Omega, F, \mathbf{P})$. The formalism I've just sketched for randomizing mathematical objects incorporates an arbitrary probability space. An ordered triple like $(\Omega, F, \mathbf{P})$ is the basic object of study in probability theory. The notations Ω and $\mathbf{P}$ are virtually standard. Ω is any old set, F is a σ-algebra on Ω, and $\mathbf{P}$ is a probability measure on F. Nevertheless, mathematicians conceive of $(\Omega, F, \mathbf{P})$ as representing all the randomness and information about the world that might interest us in a given application. As more information becomes available or as circumstances change, $(\Omega, F, \mathbf{P})$ does not have to be altered or augmented. If we forgot to include anything in $(\Omega, F, \mathbf{P})$, say some class of events (Λ, L), we can form the product space $(\Omega \otimes \Lambda, F \otimes L, \mathbf{P}')$, where $\mathbf{P}'$ is a suitably augmented probability on the product whose marginal distribution on Ω is just $\mathbf{P}$ (here $\Omega \otimes \Lambda$ is the Cartesian product of Ω and Λ, and $F \otimes L$ is the σ-algebra induced on this Cartesian product by F and L). But then we pretend that $(\Omega \otimes \Lambda, F \otimes L, \mathbf{P}')$ was the object we were studying all along, which we accomplish by renaming $(\Omega \otimes \Lambda, F \otimes L, \mathbf{P}')$ as the original $(\Omega, F, \mathbf{P})$. In this way $(\Omega, F, \mathbf{P})$ captures all the randomness in the world relevant to any given inquiry, and in particular to the random predicates we shall be defining.

3. Random Predicates

A collection of random predicates is a randomized collection of old-fashioned predicates. For a universe of objects U, an old-fashioned predicate *P* is an indicator function on U, i.e., a function from U to {0,1} whose counterimage at 1 is the extension of the predicate. From this definition it follows that relations can be subsumed under predicates. An n-place relation can then be defined as any indicator function *R* on the n-fold Cartesian product of U, i.e., a function from U^n to {0,1}. It follows that an n-place relation on U is a predicate on U^n. Given a collection of predicates on U^n, call it C, we can randomize C as follows: take the probability space $(\Omega, F, \mathbf{P})$ and consider a collection of functions from Ω to C, call it $\rho(\Omega, C)$. For Φ in $\rho(\Omega, C)$ and for ω in Ω, $\Phi(\omega)$ is therefore a

predicate (in this case, an n-place relation). With this requirement $\rho(\Omega,C)$ becomes a collection of random predicates, and each such Φ is a random predicate.

For theoretical purposes this formulation is fine. But for practical purposes we want a slightly different formulation of random predicates, one that exploits the fact that predicates are functions (in this case, indicator functions). As we saw in the last section, when randomizing functions from U to V by means of the probability space $(\Omega,F,\mathbf{P})$, randomization means taking a Cartesian product of the probability space with the domain U and considering functions from that product space into V. For any natural number n we therefore define an *n-place random predicate* on U as a function P from $\Omega x U^n$ to $\{0,1\}$. Thus an n-place random predicate is such that for fixed $\omega \in \Omega$, $P(\omega,x_1,...,x_n)$ is a 0-1-valued function of $x_1,...,x_n$ (each of the x_is belonging to U).

It will help to follow certain notational conventions to streamline our discussion of random predicates. Henceforth italic capital Roman letters *P, Q, R,* etc. will denote random predicates with respect to the probability space $(\Omega,F,\mathbf{P})$ and the universe U. ω will denote a generic element of Ω. Italic small Roman letters *a, b, c, d,* and *e* will denote constants in U. Italic small Roman letters *u, v, w, x, y,* and *z* will denote variables ranging over U. We can think of a 1-place random predicate P in three ways:

(1) As a joint 0-1-valued function of ω in Ω and u in U, P sends (ω,u) to $P(\omega,u)$.

(2) For a fixed a in U $P(\omega,a)$ is a 0-1-valued random variable in ω (ignore measurability problems). In this case we write P_a for the random variable in ω.

(3) For a fixed ω in Ω $P(\omega,u)$ is a 0-1-valued function in u. In this case we write P_ω for the old-fashioned predicate on U induced by ω.

(3) provides perhaps the clearest representation of P as a random predicate. In standard probabilistic fashion we suppress the ω in P_ω and write P for P_ω whenever ω is unspecified. If in fact P_ω does not vary in ω (i.e., is independent of ω), then P is just an old-fashioned predicate. In this way random predicate logic subsumes ordinary predicate logic.

Similar observations hold for higher-place predicates. Thus, if Q is a 2-place random predicate, then for fixed ω, Q_ω is a genuine 2-place predicate; for fixed a and b, $Q_{a,b}$ is a 0-1-valued random variable on Ω; and for fixed a,

Q(ω,a,u) is a 1-place random predicate in ω and *u*. We refer to ω as the *chance variable* and to u as the *predicate variable*. When the chance variable is fixed, the random predicate is just an ordinary predicate. When the predicate variable is fixed, the random predicate is just a 0-1-valued random variable. Thus a 0-place random predicate is just a 0-1-valued random variable.

For simplicity, let us make the following general rule concerning measurability: *When in doubt, assume measurability*. Thus, for the random predicate *P*, P_a is a random variable for all *a* in U. If U itself has a measure-theoretic structure, we can assume that *P(ω,u)* is simultaneously measurable in ω and *u*. Dependencies among the random variables P_a are of especial interest. Suppose, for instance, that *P(x)* is the 1-place random predicate "*x* is a heap." Suppose U contains all tightly packed collections of pebbles with all pebbles the same size. Let *a* denote a collection of a thousand pebbles and *b* denote a collection of a million pebbles. Clearly, whenever P_ω makes *a* out to be a heap, P_ω must also make *b* out to be a heap. It follows that the random variable P_a is always less than the random variable P_b. Thus $\mathbf{P}[P_a = 1] \leq \mathbf{P}[P_b = 1]$ and $\mathbf{P}[P_a \leq P_b] = 1$. Thus to understand the vague predicate "*x* is a heap" when interpreted as a random predicate *P*, it is essential to consider dependencies among random variables like P_a and P_b.

Since for fixed ω, random predicates are just old-fashioned predicates, old-fashioned logic continues to apply. Moreover, since predicates are indicator functions, logical connectives and quantifiers take on a particularly simple form. We may summarize the logical operations of random predicates as follows (suppressing chance and predicate variables wherever possible):

(1) $\neg P := 1 - P$

(2) $P \wedge Q := \min(P,Q)$

(3) $P \vee Q := \max(P,Q)$

(4) $P \rightarrow Q := (\neg P) \vee Q$

(5) $\forall x P(x) := \inf_{x \in \mathrm{U}} P(x)$

(6) $\exists x P(x) := \sup_{x \in U} P(x)$.

To these we add the following derived operations:

(7) $P \leftrightarrow Q := (P \rightarrow Q) \wedge (Q \rightarrow P)$

(8) $P \oplus Q := (\neg P \wedge Q) \vee (P \wedge \neg Q) = P + Q \mod 2$

(9) $\exists! x P(x) := \exists x [P(x) \wedge \forall y (x \neq y \rightarrow \neg P(y))]$.

(7) is just the biconditional, (8) is the symmetric difference, which can be defined via modular addition, and (9) is unique existence. Observe that in (9) the predicate $x \neq y$ is conceived as a random two-place predicate: define $E(\omega,x,y) = 1$ if x = y and 0 otherwise. Then (9) can be rewritten as,

(9′) $\exists! x P(x) := \exists x [P(x) \wedge \forall y (\neg E(x,y) \rightarrow \neg P(y))]$,

or alternatively,

(9″) $\exists! x P(x) := \exists x [P(x) \wedge \forall y (P(y) \rightarrow E(x,y))]$.

With random predicates in hand it is natural to ask whether some predicates are more random than others. Since our primary task is to model vagueness, this is a question of ranking predicates along a vagueness-sharpness dimension. Some predicates will be sharper than others. Since old-fashioned predicates are just random predicates for which the chance variable is irrelevant, these nonrandom predicates constitute a natural prototype for sharp predicates. Thus, for any partial ordering $\leq$ of random predicates where $S \leq T$ means T is sharper than S, the maximal elements under $\leq$ should comprise the standard nonrandom predicates (i.e., those random predicates P for which $P(\omega,u)$ is constant in ω).

The minimal elements of such a partial ordering, however, are not as straightforward. There is in fact a single minimal element which distinguishes itself as the vaguest predicate. But to understand how this vaguest of random predicates achieves this distinction, it is necessary to understand the special role that the probability 1/2 plays in the theory of probability. Consider, for instance, a random walk on the integers which

has probability 1/2 of moving either up or down a single integer. Such a random walk will run through all the integers (this happens with probability 1). But as soon as the probability p of moving up is different from 1/2, the random walk drifts, either continually up if p is greater than 1/2, or continually down if p is less than 1/2 (this happens with probability 1). In neither case does the full range of integers get exhausted as when p equals 1/2.

Similarly, with a random predicate P, if for a given a in U the random variable P_a is such that $\mathbf{P}[P_a = 0] = \mathbf{P}[P_a = 1] = 1/2$, then it is impossible by repeated independent applications of the random variable P_a to determine whether the predicate tends to affirm or deny a. Half the time P_a will say yes to a (i.e., equal 1), the other half it will say no (i.e., equal 0). If the probability differed from 1/2—if say $\mathbf{P}[P_a = 1] > 1/2$—then independent, identically distributed samples of P_a will indicate a tendency of P to say yes to a. This is just a consequence of the strong law of large numbers, which says that sample averages approximate population averages, with the approximation becoming better and better as the sample size increases.

For a predicate P, dependencies between objects a and b in U can complicate matters. Suppose $\mathbf{P}[P_a = 1] = \mathbf{P}[P_b = 1] = 1/2$, but $\mathbf{P}[P_b = 1|P_a = 1] > \mathbf{P}[P_b = 1]$. Then knowing P holds for a makes it more likely that P holds for b. Such dependencies can be exploited. Hence, even though for b taken by itself we cannot by repeated sampling decide whether P holds (on average) for b, in conjunction with a a decision is possible. Thus for P to be vague we shall want not only $\mathbf{P}[P_x = 1] = 1/2$ for all x in U, but also $\{P_x : x \in \mathrm{U}\}$ to be a collection of independent random variables. There is only one random predicate satisfying these constraints. We shall denote it by H. H is the vaguest random predicate.[1]

This analysis marks the extremes of the vagueness-sharpness dimension. For any partial ordering $\leq$ that ranks random predicates according to sharpness, H is not simply a minimal element: every other random predicate is strictly sharper than H. As for the nonrandom, old-

---

[1] H has the following mathematical construction: Take $(\Omega,\mathrm{F},\mathbf{P})$ so that Ω is the collection of all functions from U into {0,1}—this is just the (possibly infinite) Cartesian product of {0,1} with itself "U times." Let F be the Borel sets induced by the product topology and let $\mathbf{P}$ be the infinite product measure on Ω of the measure on {0,1} which assigns equal mass (i.e., 1/2) to {0} and {1} (see Bauer 1981). Then ω in Ω is just a function from U to {0,1} (i.e., an old-fashioned predicate), and H is just the evaluation map: $H(\omega,x) := \omega(x)$.

fashioned predicates embedded among the genuinely random predicates, these are maximal elements of $\leq$. To examine what happens between these extremes, I want first to take an excursion into the theory of fuzzy sets. The probabilistic intuitions which are present in random predicates, but absent from fuzzy sets, make it possible distinguish random predicates according to vagueness and sharpness in a way not possible for fuzzy sets. As we shall see, random predicates entail a more powerful mathematical theory than fuzzy sets. In fact, random predicates readily incorporate fuzzy sets.

4. Fuzzy Sets

We have been considering random predicates on the universe U. U also supports fuzzy sets. A fuzzy set on U is just a function $\mu: U \rightarrow [0,1]$.[2] A crisp set is a fuzzy set whose only values are 0 and 1, i.e., an old-fashioned set, or equivalently, an indicator function. $\mu(a)$ denotes the degree or grade of membership in the fuzzy set μ. Thus $\mu(a) = 0$ signifies total exclusion from μ, whereas $\mu(a) = 1$ signifies total inclusion in μ. Interpreting the intermediate values of μ, however, is less perspicuous. Certainly $\mu(a) < \mu(b)$ means b has a higher degree of membership in μ than a. Thus μ gives us reliable ordinal information. Still, we expect more than just ordinal properties from a function taking values in the unit interval.[3]

If ordinal considerations were the sole concern in distinguishing degree of membership, then for any strictly increasing function f on the unit interval ($f : [0,1] \rightarrow [0,1]$ where $f(s) < f(t)$ for $s < t$), $f \circ \mu$ does as much effective work as the fuzzy set μ. Clearly we want to exploit more structure than just the ordinal properties of [0,1]. For simplicity let us take U to be the unit interval [0,1] as well. Consider the crisp set $\mu = 1_{[1/2,1]}$, the indicator function on the set [1/2,1]. If any ordinal transformation of μ is permissible, then for a very small positive ε, $\varepsilon\mu = \varepsilon 1_{[1/2,1]}$ is equivalent to μ ($\varepsilon\mu$ is the product of ε and μ). This, however, is counterintuitive. $\varepsilon\mu$ is no longer crisp and assigns zero or

---

[2] Actually, μ is a membership function. If $\mu: U \rightarrow M$, i.e., μ is a function from U to some membership set M, then μ induces the fuzzy set $\{(x,\mu(x)) : x \in U\}$. Thus the fuzzy set is the graph of μ. Since μ incorporates all mathematical information of interest, I won't distinguish between membership functions and fuzzy sets. See Kaufmann (1975, 6) for more details.

[3] For the difference between ordinal and interval scales, see Torgerson (1958).

negligible value to all elements of U = [0,1]. Thus $\varepsilon\mu$ is virtually without members whereas μ has half the space U for its members. Even worse is a strictly increasing transformation like $\varepsilon\mu + 1/2$ for very small positive ε. Now every member of U is assigned a value arbitrarily close to 1/2. Thus any crisp set can be taken to a fuzzy set for which every element has only a middling degree of membership.

I am belaboring this point because I want to stress the need for fixed references. $\mu(a) = 0$ means a is definitely not a member of μ; $\mu(b) = 1$ means b is definitely a member of μ. Thus any strictly increasing function f on the unit interval must have the additional property of sending 0 to 0 and 1 to 1 ($f(0) = 0$ and $f(1) = 1$). Even this requirement, however, is inadequate for characterizing fuzzy sets. With U still equal to [0,1], if we take the fuzzy set ν equal to $\varepsilon 1_{[0,1/2)} + (1-\varepsilon)1_{[1/2,1]}$ for an arbitrarily small positive ε, then intuition suggests that ν and $\mu = 1_{[1/2,1]}$ are virtually identical. Yet if we consider the function $f(x) = x^n$ for sufficiently large positive n, $f \circ \mu = \mu$ whereas $f \circ \nu$ is virtually zero—the intuition for $f \circ \mu$ and $f \circ \nu$ are completely different.

If an ordinal structure that preserves the extremes 0 and 1 is the sole consideration in scaling fuzzy sets, our intuitions regarding fuzziness are not only unsatisfied but also unsatisfiable. In saying that they are unsatisfiable, I mean that a coherent notion of sharpness for fuzzy sets is not possible. To the sharpness of fuzzy sets corresponds a straightforward intuition. For fuzzy sets σ and τ, τ is sharper than σ if any element of U with a high degree of membership in σ has an even higher degree of membership in τ, whereas any element of U with a low degree of membership in σ has an even lower degree of membership in τ.

Implicit here is the question of a third reference point. 0 and 1 have already been fixed. But what is the cutoff above which degree of membership is high but below which degree of membership is low? Probabilities offer a natural cutoff—1/2. Games where the probability of winning is 1/2 (resp. less than 1/2 and greater than 1/2) are fair games (resp. losing and winning games). This phenomenon comes up in random walks, with Brownian motion, and thus on Wall Street and in Las Vegas. Probabilities different from 1/2 can be exploited for profit. Probabilities equal to 1/2 cannot.[4] Now fuzzy sets offer no such natural cutoff. As Klir and Folger (1988, 11) observe in their text on fuzzy sets, "it is important

[4]There are some deep results from the theory of martingales that confirm this claim.

to realize that membership grades are *not* probabilities." Thus fuzzy sets cannot look to probability for such a cutoff. Yet without a cutoff no coherent notion of sharpness for fuzzy sets is possible.

To see this, consider the crisp set $\mu = 1_{[1/2,1]}$ and the fuzzy set $\pi_{\varepsilon\delta} = \varepsilon 1_{[1,1/2)} + (1-\delta)1_{[1/2,1]}$ on U = [0,1]. ε and δ are both positive real numbers strictly between 0 and 1. Now, if ε and δ are both very small, then our intuition is that on [0,1/2) $\pi_{\varepsilon\delta}$ has a low degree of membership while on [1/2,1] $\pi_{\varepsilon\delta}$ has a high degree of membership. Moreover, since the crisp set μ has still a higher degree (resp. lower degree) of membership where $\pi_{\varepsilon\delta}$ has a high degree (resp. low degree) of membership, μ is sharper than $\pi_{\varepsilon\delta}$. This is as it should be. But if now we start fiddling with ε and δ, we find that sharpness must remain a murky intuition that cannot be formalized. Any formalization of sharpness must certainly involve a partial ordering: if τ is sharper than σ and σ is sharper than ρ, then τ must be sharper than ρ. This follows from the sharper fuzzy set indicating a higher (resp. lower) degree of membership where the duller fuzzy set indicates a high (resp. low) degree of membership.

If we now reconsider μ and $\pi_{\varepsilon\delta}$, we find that for small ε and δ, μ must—if our intuition means anything—be sharper than $\pi_{\varepsilon\delta}$. Moreover, for very small ε and δ, if we augment these slightly, say to ε' and δ', then $\pi_{\varepsilon\delta}$ must be sharper than $\pi_{\varepsilon'\delta'}$—again, if our intuition means anything. By transitivity μ will be sharper than $\pi_{\varepsilon'\delta'}$. But how much can we vary ε and δ before running into trouble? Eventually we shall run into trouble, for if we choose ε and δ close to 1 then $\pi_{\varepsilon\delta}$ will have much more in common with $1_{[0,1/2)}$, the complement of μ, than with μ itself. For ε and δ close to zero, augmenting these slightly to ε' and δ', we find $\pi_{\varepsilon\delta}$ sharper than $\pi_{\varepsilon'\delta'}$. But for ε and δ close to one, augmenting these to ε' and δ' so that they become still closer to one, we find $\pi_{\varepsilon\delta}$ less sharp than $\pi_{\varepsilon'\delta'}$ ($\pi_{\varepsilon'\delta'}$ is now approaching the complement of μ, i.e., $1_{[0,1/2)}$). Something goes wrong as we keep increasing ε and δ. The problem lies with the absence of a cutoff. Unless there is a cutoff which tells us when an element of U has a high (resp. low) degree of membership which can be made higher (resp. lower) by taking the membership function closer to 1 (resp. 0), sharpness cannot be a transitive relation on fuzzy sets.

We have two reference points for fuzzy sets, 0 and 1. If we should add a third reference point α (where $0 < \alpha < 1$), then it is possible to define a coherent sharpness relation for the fuzzy sets on U. Thus for fuzzy sets σ and τ, τ is sharper than σ if for all $u \in$ U

(1) $\tau(u) \geq \sigma(u)$ whenever $\sigma(u) > \alpha$ and

(2) $\tau(u) \leq \sigma(u)$ whenever $\sigma(u) < \alpha$.

Since grade of fuzzy-set membership is not reducible to probability or any other more primitive notion, α is arbitrary. Moreover, points $u \in U$ for which $\sigma(u) = \alpha$ become indifference points, and irrelevant to the sharpness relation. Still, this definition provides a coherent notion of randomness for fuzzy sets, and introduces an idea that will help us to model sharpness for random predicates in the next section. Because probabilities undergird random predicates, sharpness relations on random predicates will have a natural cutoff, namely, $\alpha = 1/2$.

In closing this discussion of fuzzy sets, I want to demonstrate that the theory of random predicates subsumes the theory of fuzzy sets. The result is stated for families of 1-place predicates, but can be extended to families of finite-place predicates more generally.

Proposition 4.1. Let U be an arbitrary nonempty set. Suppose $\{\mu_\alpha: \alpha \in J\}$ is an arbitrary collection of fuzzy sets on U, i.e., $\mu_\alpha: U \rightarrow [0,1]$ for each α in the indexing set J. Then there is a probability space $(\Omega,F,\mathbf{P})$ and a family of random predicates $\{Q_\alpha: \alpha \in J\}$ on U such that $\mu_\alpha(u) = \mathbf{P}\{\omega \in \Omega: Q_\alpha(\omega,u) = 1\}$ for all $u \in U$ and $\alpha \in J$.

Proof. Let Ω be the product space $\{0,1\}^{JxU}$ and let F be the Borel sets induced by the product topology. Let $p_{\alpha u}$ be the (unique) probability on $\{0,1\}$ satisfying $p_{\alpha u}(\{1\}) = \mu_\alpha(u)$. Define $\mathbf{P}$ as the product measure on $\Omega = \{0,1\}^{JxU}$ of all the $p_{\alpha u}$s for $\alpha \in J$ and $u \in U$. $\omega \in \Omega$ is a function from JxU into $\{0,1\}$. Hence we can define $Q_\alpha(\omega,u) = \omega(\alpha,u)$. It is immediate that $\mu_\alpha(u) = \mathbf{P}\{\omega \in \Omega: Q_\alpha(\omega,u) = 1\}$. «

The Q_αs that model the μ_αs are all stochastically independent. Thus the entire collection of random variables $Q_\alpha(\omega,u)$ (thinking of ω as variable, and α and u as indexing parameters) are independent. This is a direct consequence of $\mathbf{P}$ being a product measure. Hence for any collection of fuzzy sets $\{\mu_\alpha: \alpha \in J\}$, the degree of membership in μ_α of some $u \in U$ has no influence, causally or otherwise, on the degree of membership in μ_β of some $v \in U$ (α and β distinct). Because dependencies, conditioning, and causal influence is the stuff of probability theory, we can rightly expect a richer theory of vagueness from random predicates than from fuzzy sets.

5. Sharpness

The random predicate T is sharper than the random predicate S if whenever S has a high (resp. low) probability of obtaining for $u \in$ U, then T has an even higher (resp. lower) probability of obtaining for u. Since we are dealing with probabilities, we use 1/2 as a cutoff. Thus for S to have a high (resp. low) probability of obtaining for u means $\mathbf{P}[S_u = 1] > 1/2$ (resp. $< 1/2$). Those u for which $\mathbf{P}[S_u = 1] = 1/2$ will constitute indifference points: a sharper predicate T can either favor u in its extension (i.e., $\mathbf{P}[T_u = 1] > 1/2$), or tend to place it in the extension of its complement (i.e., $\mathbf{P}[T_u = 0] = \mathbf{P}[\neg T_u = 1] > 1/2$), or continue to be indifferent about u (i.e., $\mathbf{P}[T_u = 1] = 1/2$). Note that we take the relations "being sharper than" and "having higher probability than" to permit equality (i.e., weakly, not strictly). Thus technically speaking, S will be sharper than itself, and T will have a higher probability of obtaining for u than S even if $\mathbf{P}[T_u = 1] = \mathbf{P}[S_u = 1]$. This convention avoids certain inconveniences that arise with strict inequality.

I want next to introduce four sharpness relations on random predicates: *sharper*, *indistinguishably sharper*, *almost certainly sharper*, and *stochastically sharper*. These relations differ in the way they exploit the underlying probabilistic structure of random predicates. The way I have listed them, they become progressively weaker in the sense of logical implication. Thus "T sharper than S" implies "T indistinguishably sharper than S" implies "T almost certainly sharper than S" implies "T stochastically sharper than S." Throughout our discussion the underlying probability space is $(\Omega,\mathrm{F},\mathbf{P})$ and the universe on which the random predicates are defined is U.

Definition 5.1. T is *sharper* than S if for all $u \in$ U

(1) whenever $\mathbf{P}[S_u = 1] > 1/2$, $T_u \geq S_u$,

(2) whenever $\mathbf{P}[S_u = 0] > 1/2$, $T_u \leq S_u$.

In this case we write $S \angle T$.

Definition 5.2. T is *indistinguishably sharper* than S if there is some set Z of probability zero in Ω (i.e., $\mathbf{P}(\mathrm{Z}) = 0$) such that for all $u \in$ U

(1) whenever $\mathbf{P}[S_u = 1] > 1/2$, $T_u \geq S_u$ on Ω–Z,

(2) whenever $\mathbf{P}[S_u = 0] > 1/2$, $T_u \leq S_u$ on Ω–Z.

In this case we write $S \angle_{\mathrm{i}} T$. Note that the set Z is entirely independent of

any given u. In effect, if we could throw away Z from Ω, definitions 5.1 and 5.2 would coincide.

Definition 5.3. T is almost certainly sharper than S if for all u ∈ U

(1) whenever $\mathbf{P}[S_u = 1] > 1/2$, $\mathbf{P}[T_u \geq S_u] = 1$,

(2) whenever $\mathbf{P}[S_u = 0] > 1/2$, $\mathbf{P}[T_u \leq S_u] = 1$.

In this case we write $S \angle_{\text{ac}} T$. Note that for each u there is a **P**-null set Z_u off of which $T_u \geq S_u$ (resp. $T_u \leq S_u$). If U is uncountable, the union of all these Z_u's can have positive probability.

Definition 5.4. T is stochastically sharper than S if for all u ∈ U

(1) whenever $\mathbf{P}[S_u = 1] > 1/2$, $\mathbf{P}[T_u = 1] \geq \mathbf{P}[S_u = 1]$

(2) whenever $\mathbf{P}[S_u = 0] > 1/2$, $\mathbf{P}[T_u = 0] \geq \mathbf{P}[S_u = 0]$.

In this case we write $S \angle_{\text{s}} T$. Note that with stochastic sharpness T_u and S_u can no longer be compared directly for different ω in Ω. There can be sets A and B of positive **P**-probability such that for ω ∈ A $S_u(\omega) > T_u(\omega)$ and for ω ∈ B $T_u(\omega) < S_u(\omega)$. The one requirement is that when $\mathbf{P}[S_u = 1] > 1/2$ (resp. $\mathbf{P}[S_u = 0] > 1/2$), then T_u is on average bigger (resp. smaller) than S_u.

Proposition 5.5. $\angle \Rightarrow \angle_{\text{i}} \Rightarrow \angle_{\text{ac}} \Rightarrow \angle_{\text{s}}$, where each implication is strict.

Proof. $\angle$ is just $\angle_{\text{i}}$ where the null probability set Z is empty. Thus the first implication holds. For the second implication, if Z is **P**-null, then $\mathbf{P}(\Omega\text{–}Z) = 1$ implying that $T_u \geq S_u$ on Ω–Z (resp. $T_u \leq S_u$ on Ω–Z) entails $\mathbf{P}[T_u \geq S_u] = 1$ (resp. $\mathbf{P}[T_u \leq S_u] = 1$). For the final implication, if for a fixed u, $\mathbf{P}[T_u \geq S_u] = 1$ (resp. $\mathbf{P}[T_u \leq S_u] = 1$), then there is a **P**-null set Z_u (note the dependence of this set on u) off of which $T_u \geq S_u$ (resp. $T_u \leq S_u$), i.e., on Ω–Z_u, $T_u \geq S_u$ (resp. $T_u \leq S_u$). Thus the collection of ωs from Ω–Z_u for which $S_u(\omega) = 1$ is contained in the collection of ωs from Ω–Z_u for which $T_u(\omega) = 1$, i.e.,

$$\{\omega \in \Omega\text{–}Z_u\colon S_u(\omega) = 1\} \subset \{\omega \in \Omega\text{–}Z_u\colon T_u(\omega) = 1\}.$$

Because probabilities are monotone it follows that

$$\mathbf{P}(\{\omega \in \Omega\text{–}Z_u\colon S_u(\omega) = 1\}) \leq \mathbf{P}(\{\omega \in \Omega\text{–}Z_u\colon T_u(\omega) = 1\}).$$

But this just says $\mathbf{P}[T_u = 1] \geq \mathbf{P}[S_u = 1]$. The opposite case of $T_u \leq S_u$ is handled similarly. Thus the third implication holds.

To see that these implications are strict we construct three counter-

examples. In each case the probability space (Ω,F,**P**) is ([0,1],B([0,1]),λ), i.e., Lebesgue measure on the unit interval, and the universe U is [0,1] as well. Our random predicates *S* will be (Borel measurable) indicator functions on the Cartesian product [0,1]x[0,1].

We consider random predicates *P*, *Q*, *R*, *S*, and *T*:

(1) $P(\omega,u) := 0$

(2) $Q(\omega,u) :=$ 1 if ω is rational, 0 otherwise

(3) $R(\omega,u) :=$ 1 if $\omega = u$, 0 otherwise

(4) $S(\omega,u) :=$ 1 if $-1/8 \leq \omega - u \leq 1/8$, 0 otherwise

(5) $T(\omega,u) :=$ 1 if $1/8 \leq \omega - u \leq 1/4$, 0 otherwise.

It is immediate that $P \angle_i Q$ without $P \angle Q$, $Q \angle_{ac} R$ without $Q \angle_i R$, and $S \angle_s T$ without $S \angle_{ac} T$. «

Our four characterizations of sharpness have analogues in the theory of stochastic processes. There two stochastic processes X_t and Y_t defined on (Ω,F,**P**) are *indistinguishable* if for almost every $\omega \in \Omega$, the entire paths (i.e., functions in t) given by $X_t(\omega)$ and $Y_t(\omega)$ are identical. Alternatively X_t and Y_t are *modifications* of each other if for every fixed t, X_t and Y_t agree off of a set of probability zero. These two notions correspond respectively to $\angle_i$ and $\angle_{ac}$. Moreover, the idea of stochastic convergence motivates $\angle_s$.[5]

Sharpness in its purest sense is given by $\angle$. Thus when we speak of sharpness, we shall mean $\angle$. $\angle_i$ and $\angle_{ac}$ are clearly derivative notions. $\angle_i$ gives us $\angle$ if we just excise a set of probability zero from Ω. Note that if the universe U is countable, then $\angle_{ac}$ and $\angle_i$ coincide: if $S \angle_{ac} T$, then for each $u \in$ U there is a **P**-null set Z_u for which the right dominance relation holds. The grand union of these Z_u's is a set Z which is also **P**-null because it is the countable union of sets of measure zero. This is the Z we need so that $S \angle_i T$. $\angle_s$ was included for completeness.

Let us next prove a monotone convergence result for $\angle$. To state this result we need a definition.

Definition 5.6. The *indifference set* of a random predicate *S* is the set of

---

[5]See Elliott (1982, 13) for the difference between indistinguishable stochastic processes and processes that are modifications of each other. Indistinguishability is a much stronger notion. For stochastic convergence (also known as convergence in probability) see Bauer (1981, 93).

all $u \in$ U such that $\mathbf{P}[S_u = 1] = 1/2$. Denote this set by IN($S$). If the sequence of random predicates $\{S^i\}$ is monotone increasing in $\angle$, i.e., $S^0 \angle S^1 \angle S^2 \angle \ldots$, then $\text{IN}(S^i) \supset \text{IN}(S^{i+1})$ for all i. Thus we can extend the notion of an indifference set to the monotone sequence $\{S^i\}$:

$$\text{IN}(\{S^i\}) := \bigcap_{0 \le i < \infty} \text{IN}(S^i).$$

Proposition 5.7. Let $\{S^i\}$ be a sequence of random predicates on the probability space (Ω,F,**P**) and the universe U. Suppose $\{S^i\}$ is monotone increasing in $\angle$. Then there is a random predicate S which is sharper than all the S^i's, and to which the sequence $\{S^i\}$ converges pointwise off of the indifference set IN($\{S^i\}$), i.e., for $\omega \in \Omega$ and $u \in$ U – IN($\{S^i\}$), $S^i(\omega,u) \rightarrow S(\omega,u)$.

Proof. If $u \in$ U – IN($\{S^i\}$), then there is some m (depending on u) such that $\mathbf{P}[S^m{}_u = 1]$ is different from 1/2. Let us assume that $\mathbf{P}[S^m{}_u = 1] > 1/2$ (the case $\mathbf{P}[S^m{}_u = 1] < 1/2$ is handled similarly). Since the S^is are monotone increasing in $\angle$, it follows that for all $q > p \ge 0$, $\mathbf{P}[S^{m+q}{}_u = 1] \ge \mathbf{P}[S^{m+p}{}_u = 1] > 1/2$, and $S^{m+q}{}_u \ge S^{m+p}{}_u$. Since this last inequality holds for all $\omega \in \Omega$ it follows that $S^i{}_u$ is monotone increasing after a certain point and therefore converges (pointwise). This defines S off of IN($\{S^i\}$). On IN($\{S^i\}$) S can be arbitrarily defined. «

At the end of section 3 we identified the extremes of any sharpness relation on random predicates. The sharpest predicates were the nonrandom (old-fashioned) predicates embedded among the random predicates. These serve as maximal elements. Alternatively, the dullest predicate H was one where $\{H_u : u \in \text{U}\}$ is a stochastically independent family of random variables such that $\mathbf{P}[H_u = 1] = 1/2$ for all u. As we can now see, these intuitions are confirmed in $\angle$. The nonrandom predicates are indeed maximal elements of $\angle$, and for any random predicate S, $H \angle S$. Actually, the assumption that all the H_us be stochastically independent is unnecessary; it is enough that $\mathbf{P}[H_u = 1] = 1/2$ for all u. Note that $\angle$ is strictly speaking a preorder, not a partial order, inasmuch as it is possible that $S \angle T$ and $T \angle S$ without S and T being equal.

References

Bauer, H. 1981. *Probability Theory and Elements of Measure Theory*. London: Academic Press.

Billingsley, P. 1979. *Probability and Measure*. New York: Wiley.

Dembski, W. A. 1990. "Uniform Probability," *Journal of Theoretical Probability*, 3(4): 611–626.

1991. Randomness by design. *Nous* 25(1): 75–106.

Elliott, R. J. 1964. *Stochastic Calculus and Applications*. New York: Springer-Verlag.

Fine, K. 1975. "Vagueness, Truth and Logic," *Synthese* 30: 265–300.

Goguen, J. A. 1968–69. "The Logic of Inexact Concepts," *Synthese* 19: 325-373.

Kallenberg, Olav. 1986. *Random Measures*, 4th ed. London: Academic Press.

Kaufmann, A. 1975. *Introduction to the Theory of Fuzzy Subsets*, Vol. 1. Orlando, Florida: Academic Press.

Klir, G. J. and T. A. Folger. 1988. *Fuzzy Sets, Uncertainty, and Information*. Englewood Cliffs, N.J.: Prentice Hall.

Torgerson, W. S. 1958. *Theory and Methods of Scaling*. New York: Wiley.