JOHN C. BIGELOW

SEMANTICS OF PROBABILITY

"... push on, *and faith* will follow"
D'ALEMBERT,

A measure of degrees of similarity between possible worlds can be used to generate measures over propositions, or sets of possible worlds. These measures over propositions will count as 'probability measures', at least in the sense that they satisfy the axioms of the probability calculus. In a previous article (Bigelow, 1976), I have outlined one way in which such probability measures can be generated. In the present article I will present a considerably less devious way of generating probability measures. I will draw on two resources. My first resource will be standard techniques of measure theory. I will borrow lavishly from an excellent mathematics textbook by Friedman (1970). My second resource will be provided by concepts originating in modal logic, and also in the analysis of counterfactuals given by David Lewis (1973).

The theory I offer rests on a natural extension of standard techniques used in modal logic for the analysis of concepts of necessity and possibility. The semantics of modal logic rest on a relation, called a (*strict*) *accessibility relation* on possible worlds. Different accessibility relations provide us with analyses of different concepts of necessity and possibility (see Hughes and Cresswell, 1968).

A strict accessibility relation is used to give an analysis of the relationship of *necessitation* which may hold between propositions. We say it is true, in a given possible world, that one proposition necessitates another, when all the accessible worlds in which the first is true are worlds in which the second is true as well.

But there are important relations among propositions which a strict accessibility relation does not illuminate. It may be, in particular, that though one proposition does not strictly *necessitate* another, yet it does nevertheless provide good *inductive support* for it. If the first is true, it may be extremely probable that the second will be true as well. This will be so, I will maintain, when *most* of the worlds in which the first is true are worlds in which the second is true.

Synthese 36 (1977) 459–472. *All Rights Reserved.*

Yet it is not easy to see how we can *measure* the degrees of overlap between propositions, except in the case where the propositions are true in only finitely many possible worlds. The strict accessibility relations of modal logic are not adequate for the task.

I will show how we can measure the degree of overlap between propositions, by appealing to a measure of degrees of similarity between possible worlds. A similarity relation among possible worlds can be seen as a *variable* accessibility relation. The more similar one world is to another, the 'more accessible' it will be. A strict accessibility relation provides us with an analysis of necessitation; a variable accessibility relation provides an analysis of inductive support.

The foundations of probability theory are in an unhappy state. I do not aspire to set them all in order. One of the most vexing problems is an epistemological one. We need to know much more about how to determine the truth values of probability judgements concerning actual states of affairs. The truth values of probability judgements are often unclear, and it is hard to know how we can go about making them clearer. The status of probability judgements is in this respect closely parallel to that of counterfactuals. I offer no help with this epistemological problem.

My aim is not to say how we can *tell* whether a probability judgement is true or false, but rather, to say what it *means* to say it is true or false. I take it as given, that some probability judgements have reasonably determinate truth values, while others do not. A semantic theory for probability judgements should give an account of their meanings, in a way which will enable us, eventually, to explain why those with clear truth values do have those truth values, and why those which are uncertain are so uncertain.

A link between probability theory and possible worlds will shed new light on probability. The entities appealed to in probability theory are already, in fact, possible worlds in all but name. It is a step forward to recognize them for what they are. Note also that world similarity is used for other purposes as well, notably in the analysis of counterfactuals. Thus a world similarity analysis of probability promises to reveal links between probability and counterfactuals.

A link between probability theory and possible worlds is also of interest for another reason. Possible worlds semantics has hitherto seemed isolated from all the natural sciences except linguistics. But probability theory is located well within the mainstream of the natural sciences. By linking possible worlds with probability, we integrate possible worlds semantics with enterprises which have hitherto been independent of it.

It should not be too surprising that probability should be the door through which possible worlds enter into the sciences. In making probability judgements we are, of course, describing features of the actual world. But the way we do so is by envisaging and weighing up various possible alternative states of affairs, or possible worlds. We describe what the world is like, by saying what worlds are like it.

Degrees of similarity. Let W be the set of all possible worlds. Any subset of W is called a *proposition*. I introduce a function, d, which will be what is known as an *extended real-valued* function over all pairs of possible worlds. That is, d is a function which maps each pair of possible worlds, i and j, onto an entity $d(i, j)$ which is either a real number, or else an entity called ∞. The entity ∞ is defined to be such that for any real number $x, x < \infty$. The number $d(i, j)$ is to be thought of as representing the degree of similarity between i and j. When $d(i, j) = \infty$, we may think of this as asserting that i and j are so different that we can no longer discriminate degrees of similarity for them.

I will assume, largely for the sake of convenience and brevity of exposition, that d is what is known as a *metric*. That is, I assume that d satisfies the conditions:

D1. $d(i, j) \geqslant 0$
D2. $d(i, j) = 0$ if and only if $i = j$
D3. symmetry: $d(i, j) = d(j, i)$
D4. triangle inequality: $d(i, j) \leqslant d(i, k) + d(k, j)$.

I am inclined to believe that, for my purposes, all I need assume are D1 and D2. But by assuming d to be a metric, I ensure that (W, d) is what is known as a metric space. These have been intensively studied in mathematics. Hence I will have a large body of results which I can draw on. It may be worth enquiring, later, whether or not the results I need essentially depend on the metric properties D3 and D4. I conjecture that they do not.

Using the metric d, we can define a number of useful topological properties of propositions.

Diameter. The diameter of a subset A of W is defined to be

$$d(A) = \sup \{d(i,j): i, j \in A\}$$

That is, $d(A)$ is the least number which is greater than every distance between worlds in A. For the empty set φ, we set $d(\varphi) = 0$.

Distance between sets. We define the distance between two propositions A and B to be

$$d(A, B) = \inf \{d(i,j): i \in A, j \in B\}$$

That is, $d(A, B)$ is the greatest number which is less than every distance from a world in A to a world in B. We set $d(\varphi, B) = d(A, \varphi) = 0$. When $A = \{i\}$ for some $i \in W$, we write $d(i, B)$ for $d(A, B)$.

Open spheres. For any world $i \in W$ and any number $r > 0$, the set

$$S_{(i,r)} = \{j \in W: d(i,j) < r\}$$

is called the open sphere (or open ball) with centre i and radius r.

Accessible worlds. For each possible world i and each number δ (either a real number or ∞), we define the set of δ-accessible worlds from i to be

$$W_{(\delta, i)} = \{j \in W: d(i,j) \leqslant \delta\}.$$

My aim is to calculate the probabilities which various propositions have in any given world i. The propositions I will be interested in will all be subsets of $W_{(\delta, i)}$.

Accessible spheres. Let $\mathscr{K}$ be the class of all open spheres which are subsets of $W_{(\delta, i)}$.

Let $\mathscr{K}_n$ be the class of open spheres in $\mathscr{K}$ with diameter no greater than $1/n$:

$$\mathscr{K}_n = \left\{S \in \mathscr{K}: d(S) \leqslant \frac{1}{n}\right\}.$$

Note that $\mathscr{K}_{n+1} \subseteq \mathscr{K}_n$.

Sequential coverings. A sequence of propositions $\{E_\nu\}$, $\nu = 1, 2, \ldots$ is called a sequential covering of a set A, if

$$A \subseteq \bigcup_{\nu=1}^{\infty} E_\nu.$$

Sequential covering class. A class $\mathscr{K}$ of propositions is a sequential covering class for a set A if, firstly, $\varphi \in \mathscr{K}$, and secondly, for any subset B of A there is a sequential covering of B contained in $\mathscr{K}$.

Weight of spheres. We next require some function λ, with domain $\mathscr{K}$, which is such that $\lambda(\varphi) = 0$, and for each non-empty sphere $S \in \mathscr{K}$, $\lambda(S)$ is either some non-negative real number or else ∞.

The purpose of λ is to assign weights to spheres, so that we can use these spheres to assign weights (and eventually measures) to a wider class of propositions. We cover any given proposition by spheres of a given kind, and then add the weights of these spheres to obtain an approximate measure of the weight of that proposition.

The function λ may be defined using the metric d. I suggest that the weight assigned to a sphere in $W_{(\delta,i)}$ should be proportional to its diameter, and inversely proportional to its distance from world i. The precise definition of λ is not important here, however. I will mention it again briefly later.

Assumption 1. To obtain a probability measure over the propositions in $W_{(\delta,i)}$, we must assume that every class of spheres $\mathscr{K}_n$ is a sequential covering class for $W_{(\delta,i)}$.

That is, we must assume that any proposition in $W_{(\delta,i)}$ can be covered by a countable sequence of spheres of less than any given diameter.

(This property of $W_{(\delta,i)}$ is something like the topological concept of compactness. A space is compact if every cover of it contains a finite cover.)

Note that I do *not* assume that for *every* space $(W_{(\delta,i)}, d)$, $\mathscr{K}_n$ is a sequential covering class. My claim is that it is *only* when this condition is satisfied, that probability measures are definable over $(W_{(\delta,i)}, d)$, using the present construction. I will return to this point later.

Adding weights. When Assumption 1 is satisfied for a space $(W_{(\delta,i)}, d)$, we define a function μ_n^* corresponding to each $\mathscr{K}_n$, as follows. We require that $\mu_n^*(\varphi) = 0$, and for any non-empty subset A of $W_{(\delta,i)}$, we have

$$\mu_n^*(A) = \inf\left\{ \sum_{\nu=1}^{\infty} \lambda(S_\nu) : \{S_\nu\},\ \nu = 1, 2, \ldots, \right.$$

$$\text{is a subset of } \mathscr{K}_n, \text{ and}$$

$$\left. A \subseteq \bigcup_{\nu=1}^{\infty} S_\nu \right\}.$$

That is, μ_n^* measures the 'weight' of a proposition A by adding the 'weights' of the spheres used in its most economical covering chosen from $\mathscr{K}_n$.

Outer measures. It can be proved (Friedman, Theorem 1.4.1) that each function μ_n^*, defined as above, is what is called an *outer measure* over $W_{(\delta,i)}$. That is, each μ_n^* is such that

1. The domain of μ_n^* is the class of all subsets of $W_{(\delta,i)}$
2. μ_n^* is non-negative and $\mu_n^*(\varphi) = 0$.
3. μ_n^* is *monotonic*: that is, if A and B are in the domain of μ_n^* and $A \subset B$, then

   $$\mu_n^*(A) \leqslant \mu_n^*(B).$$

4. μ_n^* is *countably subadditive:* that is, whenever $\{A_\nu\}, \nu = 1, 2, \ldots,$ is a sequence of sets in the domain of μ_n^*, we have

   $$\mu_n^*\left(\bigcup_{\nu=1}^{\infty} A_\nu \right) \leqslant \sum_{\nu=1}^{\infty} \mu_n^*(A_\nu)$$

Taking limits. Since $\mathscr{K}_{n+1} \subseteq \mathscr{K}_n$, we have

$$\mu_n^*(A) \leqslant \mu_{n+1}^*(A).$$

Hence the sequence $\{\mu_n^*(A)\}$, $n = 1, 2, \ldots,$ is monotone increasing and bounded above (at least by ∞). It follows that this sequence tends to a limit.

DEFINITION Let μ^* be that function such that for any subset A of $W_{(\delta,i)}$,

$$\mu^*(A) = \lim_{n \to \infty} \mu_n^*(A).$$

THEOREM. It can be proved (Friedman, Theorem 1.9.1) that μ^* is what is known as a *metric outer measure*. A metric outer measure is an outer measure which meets the further requirement that whenever

$$d(A, B) > 0$$

we have

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

Measurable sets. Given any outer measure μ^*, we say that a proposition A is μ^*-measurable if, for any other proposition B in the domain of μ^*,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B - A).$$

Pictorially, the idea behind μ^*-measurability seems to be this. A μ^*-measurable proposition A is one which has a sharp and regular boundary in logical space. Any such proposition A will, as it were, make a clean cut across any other proposition B which it intersects. It will divide B into two portions, $(B \cap A)$ and $(B - A)$, with a sharp boundary between them. Because these two portions have a sharp boundary, the outer measure on the whole of B will be the sum of the outer measures on each of the portions, $(B \cap A)$ and $(B - A)$, taken separately. As the size of spheres used in coverings decreases, the best coverings of the two portions $(B \cap A)$ and $(B - A)$ separately will come closer and closer to forming the best cover of B as a whole.

A special case of a μ^*-measurable set will be the case of a set A which is such that

$$d(A, -A) > 0,$$

where $-A$ is the complement of A in $W_{(\delta, i)}$. Call any such set *isolated*.

When A is isolated, then for any set B we have

$$d(B \cap A, B - A) > 0.$$

Hence, since μ^* is a *metric* outer measure, we have

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B - A),$$

and so A is measurable.

I think it is reasonably plausible to suppose that some intuitively natural propositions are isolated. Later, I will mention some other relatively 'natural' propositions which are measurable.

Measure space. Let $\mathscr{A}_{(\delta, i)}$ be the class of all μ^*-measurable sets in $W_{(\delta, i)}$; and let μ be the restriction of μ^* to $\mathscr{A}$.

Then it can be proved (Friedman, Theorem 1.3.1) that $\mathscr{A}_{(\delta, i)}$ is what is known as a *σ-algebra* over $W_{(\delta, i)}$. That is, $\mathscr{A}_{(\delta, i)}$ meets the following conditions:

1. $W_{(\delta, i)} \in \mathscr{A}_{(\delta, i)}$
2. If $A, B \in \mathscr{A}_{(\delta, i)}$, then $A - B \in \mathscr{A}_{(\delta, i)}$
3. $\mathscr{A}_{(\delta, i)}$ is *σ-additive*: if $\{A_\nu\}$, $\nu = 1, 2, \ldots$, is a subset of $\mathscr{A}_{(\delta, i)}$, then

$$\bigcup_{\nu=1}^{\infty} A_\nu \in \mathscr{A}_{(\delta, i)}.$$

And it can further be shown that μ, the restriction of μ^* to $\mathscr{A}_{(\delta, i)}$, is what is known as a *measure* over $\mathscr{A}_{(\delta, i)}$. That is, μ is such that

1. The domain of μ is a σ-algebra
2. μ is non-negative on $\mathscr{A}_{(\delta, i)}$
3. $\mu(\varphi) = 0$
4. μ is *σ-additive* (or *completely additive*) on $\mathscr{A}_{(\delta, i)}$: that is, if $\{A_\nu\}$, $\nu = 1, 2, \ldots$, is a sequence of *disjoint* sets in $\mathscr{A}_{(\delta, i)}$, then

$$\mu\left(\bigcup_{\nu=1}^{\infty} A_\nu\right) = \sum_{\nu=1}^{\infty} \mu(A_\nu).$$

A measure space is a triple $(W, \mathscr{A}, \mu)$ such that W is a set, $\mathscr{A}$ a σ-algebra of subsets of W, and μ a measure over $\mathscr{A}$. The space we have constructed, $(W_{(\delta, i)}, \mathscr{A}_{(\delta, i)}, \mu)$ is a measure space.

Assumption 2. To obtain a probability space we must assume that in the measure space $(W_{(\delta, i)}\ \mathscr{A}_{(\delta, i)}, \mu)$ we have

$$0 < \mu(W_{(\delta, i)}) < \infty.$$

I will briefly discuss this assumption later. I do not assume the assumption to be true for all worlds i; I assert only that the present construction yields a probability measure over $W_{(\delta,i)}$ only in the case of worlds i for which the assumption is satisfied.

Probability spaces. If $0 < \mu(W_{(\delta,i)}) < \infty$, we define a function $P_{(\delta,i)}$ as follows: $P_{(\delta,i)}(\varphi) = 0$ and for any non-empty set $A \in \mathscr{A}_{(\delta,i)}$, we have

$$P_{(\delta,i)}(A) = \frac{\mu(A)}{\mu(W_{(\delta,i)})}.$$

It can be proved that $P_{(\delta,i)}$ is what is known as a ***probability measure***: that is, a measure whose range lies between 0 and 1. Thus

$$(W_{(\delta,i)}, \mathscr{A}_{(\delta,i)}, P_{(\delta,i)})$$

is a probability space: that is, a measure space whose measure is a probability measure.

Semantics. The probability spaces generated in the above manner can be used to provide an interpretation of sentences in a language suitable for talking about probability.

For simplicity, let our language L be the propositional calculus with one embellishment. We obtain modal logic from the propositional calculus by adding one sentential operator, $\Box$, which makes a sentence, $\Box\,\alpha$, out of any sentence α. I will add more than a single sentential operator to the propositional calculus. For each real number x between 0 and 1, there will be a sentential operator P_x. Thus for any sentence α,

$$P_x(\alpha)$$

will be a sentence, to be interpreted as asserting that the probability of α is x.

Let a *frame* be a triple, (W, d, δ), where W is a set, d a metric on W, and δ some real number (used to give a standard of accessibility).

An *interpretation* of our language L on a frame will be a function V which assigns a subset of W to each sentence of the language.

A *model* for the language L is a quadruple (W, d, δ, V), where (W, d, δ) is a frame, and V an interpretation.

In a model (W, d, δ, V), the value of a sentence

$$P_x(\alpha)$$

will be a subset of W,

$$V(P_x(\alpha)) \subset W.$$

The value of this sentence is to be the set of worlds $i \in W$ such that the induced probability space

$$(W_{(\delta, i)}, \mathscr{A}_{(\delta, i)}, P_{(\delta, i)})$$

is *defined*, and

$$P_{(\delta, i)}(V(\alpha)) = x.$$

DISCUSSION

The interpretation I have suggested for probability statements has some purely formal interest. It welds together modal logic and probability theory in a way which produces minimal distortions in both.

Another point worth noting, is that in the present framework the main points of emphasis are different from the more usual ones. In particular, I have emphasized the fact that a statement about probabilities will have different truth values in different possible worlds. Its truth value in a possible world will depend on facts about 'what that world is like', on what alternative states of affairs there are for that world, and the 'ease' with which they could have been brought about in that world.

These features of the theory make it attractive enough to warrant elaboration of its consequences, and comparison with its rivals (frequency, propensity, and subjectivist theories for instance). But I will restrict myself here to a brief discussion of a number of crucial steps in my construction of a probability measure from the world-similarity metric.

The fundamental question is this. The above construction gives us probability measures: but only if d is a suitable sort of metric over W. Even when d does generate probability measures, it does so only for some possible worlds i, and only over some of the propositions accessible from i – the 'measurable' ones.

But I want my construction to associate a probability measure with a world if and only if there are propositions which have determinate probabilities in that world. Furthermore, I want the 'measurable' propositions to be the ones with determinate probabilities. And I want the probabilities which the construction assigns to them to be the same as the probabilities which they really have in that world.

The metric d has to ensure all of this; and on top of it all, d must be intuitively plausible as a measure of degrees of world similarity. That is a tall order.

One way of softening our problems is by beginning with an attempt to find *any* d which is formally adequate to give the right results. We begin by searching for a d which generates the right truth-conditions for probability statements. We may then argue that because it gives the right results, this must be the function d which best represents degrees of world-similarity.

Too much use of this evasive manoeuver robs the theory of informativeness. But there is nothing wrong with letting at least some considerations about desirable consequences filter back into our account of the world-similarity measure. I seriously maintain that the very existence of probability statements with relatively clear truth-conditions provides strong evidence that the world-similarity measure does have all the properties I need.

Measurability One property I need d to have is the following. I need d to be such that the induced class of μ^*-measurable sets for a world i contains all and only the propositions which have probabilities in world i.

I defend the claim that d meets this condition, by illustrating the wide range of sets which will be μ^*-measurable. I have already mentioned isolated sets; but there are more important measurable sets than this.

A set A is called an *open set* if for any $i \in A$ there is some $r > 0$ such that

$$S_{(i,\, r)} \subseteq A.$$

A set is *closed* if its complement is open.

The μ^* corresponding to a world i is a metric outer measure over $W_{(\delta,\, i)}$. It can thus be proved (Friedman, Theorem 1.8.2) that every *closed* subset of $W_{(\delta,\, i)}$ is μ^*-measurable.

The class of *Borel sets* of $W_{(\delta,\, i)}$ is defined to be the smallest σ-algebra of subsets of $W_{(\delta,\, i,)}$ which contains all open sets in $W_{(\delta,\, i)}$. This means that the class of Borel sets of $W_{(\delta, i)}$ is the class of all countable unions of open sets in $W_{(\delta,\, i)}$.

It can be proved (Friedman, Corollary 1.8.3) that every Borel set in $W_{(\delta, i)}$ is μ^*-measurable.

The unmeasurable sets will be very strange indeed. I do not mind if we are unable to assign them probabilities.

Countable covers. In order to obtain a probability measure for a world i, I had to assume (Assumption 1) that every subset of $W_{(\delta, i)}$ can be covered by a countable number of spheres, no matter how small. Furthermore, implicit in Assumption 2 is the requirement that we can always cover a set in such a way that the *sum* of weights of the spheres used is *finite*. Indeed, as we decrease the size of spheres used in the coverings, we need the sum of their weights to tend to some finite limit. What do we need to assume about d in order to ensure these assumptions are met? I will consider the second assumption first.

I suggest that we need to assume very little about d in order to ensure the truth of Assumption 2. It is the function λ which assigns weights to spheres; hence λ is the function which we must ensure to be of a suitable kind.

Roughly, what λ is needed for, is to reflect the *number* of spheres required to cover a proposition. We could make λ a simple function which assigns the number 1 to every non-empty sphere. Then we would calculate the weight of a proposition by simply *counting* the spheres needed to cover it.

But this would not be a suitable definition for λ. As the covering spheres decrease in size, more of them will be required to complete the covering. Hence the sum of their weights will increase as their size decreases. Taking the limit of progressively finer coverings, we will not in general obtain a finite outer measure for the proposition we are considering. Indeed, the outer measure of a proposition will be finite only if the proposition is true in only finitely many possible worlds.

We must therefore choose a definition of λ which ensures that as the spheres used in a covering decrease in size, the weights assigned to them will decrease also. That is, we want to ensure that although finer coverings will require more spheres, these spheres will be of decreasing weight. Then it need not follow that the sum of their weights increases without limit as their size decreases.

I suggest that the weight assigned to a sphere should be proportional to its diameter: the smaller the sphere, the smaller its weight. And furthermore, the weight assigned to a sphere relative to world i should be inversely proportional to its distance from i: the closer to i, the greater its weight.

We might therefore consider defining λ by:

$$\lambda(S) = \frac{d(S)}{d(i, S)}.$$

However this definition is still unsuitable, since as a sphere S gets closer and closer to i, $\lambda(S)$ will tend to infinity, and when $d(i, S) = 0$, for instance when $i \in S$, $\lambda(S)$ will be undefined.

Hence, I suggest we define λ by:

$$\lambda(S) = \frac{d(S)}{1 + d(i, S)}.$$

Then when $i \in S$ we simply have $\lambda(S) = d(S)$. When $d(i, S) = 0$ we have $\lambda(S) = d(S)$. When $d(i, S) > 0$ we have $\lambda(S) < d(S)$.

This definition of λ ensures that as we progress through finer and finer covers of a set, the weights of spheres used in the cover will decrease. In finer coverings,the spheres will have smaller diameters; and in addition, many of them will be more distant from world i. Both these factors will serve to decrease the weights assigned to them by λ.

This definition of λ may still not be adequate. It is not sufficient that the weights of covering spheres should decrease as their number increases. The sum of their weights must tend to a limit as their number increases. What must λ and d be like for this to happen? I do not know.

Another question I cannot answer, is: What has to be the case in order for every subset of $W_{(\delta, i)}$ to be coverable by a countable sequence of spheres, no matter how small? That is, what must be the case for Assumption 1 to be satisfied?

Yet I see no reason for thinking it would take a very strange and implausible metric d to meet Assumption 1. On the contrary, it is more likely to require a very strange d to make Assumption 1 fail – especially for worlds rather like our own. Of course, many possible worlds are very strange; and the similarity relations in their region may be very strange too. But the possible worlds similar to the actual world are not so strange. So the similarity relations of worlds near the actual world will not be as strange as might be feared. And after all, it is primarily for worlds rather like the actual world that we want the probabilities of propositions to be definable.

Victoria University of Wellington,
New Zealand

BIBLIOGRAPHY

Bigelow, J. C., 'Possible Worlds Foundations for Probability', *Journal of Philosophical Logic* 5 (1976), 299–320.

Friedman, A., *Foundations of Modern Analysis*, Holt, Rinehart and Winston, N.Y., 1970.

Hughes, G. E. and Cresswell, M. J., *An Introduction to Modal Logic*, Methuen, London, 1968.

Lewis, D.K., *Counterfactuals*, Blackwell, Oxford, 1973.