Picturing Probability: the poverty of Venn diagrams, the richness of Eikosograms

W.H. Cherry and R.W. Oldford
University of Waterloo

September 10, 2003

Abstract

Diagrams convey information, some intended some not. A history of ringed diagrams including their use by Euler and Venn shows that the information content of these diagrams is consistent and inescapable – they describe abstract interrelations between different entities. This historical consistency predates and would have been known to both Euler and Venn. Venn's use was a true innovation over Euler's and, contrary to what some have recently suggested, Venn's name deserves to be attached to these diagrams.

Venn diagrams visually ground symbolic logic and abstract set operations. They do not ground probability. Their common overuse in introducing probability, especially in teaching, can have undesirable consequences. We define the eikosogram and show it to be semantically coincident with the calculus of probability. The eikosogram visually grounds probability – conditional, marginal and joint – and facilitates its study. To visually ground definition of events, outcome diagrams and outcome trees are recommended.

The eikosogram can be used to develop an axiomatic treatment of probability based on random variables rather than on sets – Kolmogorov's axioms need not be employed but could be derived. Probabilty calulations amount to calculating rectangular areas. Probability statements on random variables can be visually distinguished from those on events as can the separate ideas of disjoint and independent events. Bayes' theorem, the product rule for independence, unconditional and conditional independence all follow simple visual features of the eikosogram.

Eikosograms should be used for the calculus of probability and outcome diagrams and trees to motivate and understand random variables and events. Venn diagrams need no longer be used in teaching probability; if used they should be confined to the consideration of abstract relations between pre-defined events and appear after the introduction of the calculus of probability.

Keywords: Eikosograms, Euler diagrams, Venn diagrams, outcome trees, outcome diagrams, vesica piscis, ideograms, history of probability, logic and probability, understanding conditional probability, probabilistic independence, conditional independence.

1 Introduction

It is now commonplace to use Venn diagrams to explain the rules of probability. Indeed, nearly every introductory treatment has come to rely on them. But this was not always the case. In his book *Symbolic Logic* Venn makes much use of these diagrams, yet in his book on probability, *The Logic of Chance*, they appear nowhere at all![1]

A cursory review of some well known probability texts reveals that the first published use of these diagrams in probability may have occurred as late as 1950 with the publication of Feller's *Theory of Probability* (details are given in the Appendix). Venn diagrams don't seem to have been that much used in probability or, if used, that much appreciated. For example, Gnedenko (1966), a student of Kolmogorov, used Venn diagrams in the third edition of his text *Theory of Probability* but does not refer to them as such until the book's next edition in 1968, and then only as "so-called Venn diagrams". Even by 1969, the published use of Venn diagrams for probability was by no means common.

[1] It is true that Venn's probability book predates his symbolic logic book, however the diagrams only ever appeared in the latter book. This is more interesting given that Venn uses the word 'logic' in the titles of both books and also that Venn's symbolic logic used the numerical values of 1 and 0 to indicate true and false (i.e. certainty and impossibility).

In more recent years, some authors of introductory probability texts have called just about any diagram which marks regions in a plane a ‘Venn diagram’. Others have written that no diagram should be called a ‘Venn diagram’. Dunham (1994), for example, claims that the Venn diagram was produced a century before Venn by Euler and so “If justice is to be served, we should call this an ‘Euler diagram’.” This view is surprisingly commonplace though not everywhere expressed as strongly as Dunham (1994 p. 262) who dismissively writes “Venn’s innovation [over Euler’s diagrams] …might just as well have been discovered by a child with a crayon.” In both cases, the sense of what constitutes a Venn diagram has been lost. In the first case, the Venn diagram is not up to the job and so is stretched beyond its definition, while in the second case it is Euler’s diagram that has been stretched beyond its definition to mistakenly include Venn’s innovative use.

Diagrams convey information and so, like statistical graphics, need to be carefully designed to ensure that the intended information is clearly conveyed, unobscured by unintended interpretation. Diagrams need to be tailored to specific purposes.

Ringed diagrams have a long history in which the information conveyed is consistent and useful in many different contexts. We show their long use in Christian symbolism and hence how the diagrams would have been familiar to both Euler and Venn – attributing originality of ringed diagrams to describe intersection, union or complement to either man would be a mistake. First Euler (in a descriptive post-hoc sense), then Venn (in an active ad hoc sense) would use these and similar diagrams to aid understanding of logic – both uses were important innovations.

The strong coupling of probability and set theory via Kolmogorov’s axioms has led to a tradition where many writers overuse Venn diagrams to introduce probability. Besides showing Venn diagrams to be unnecessary and even inferior to other diagrams for understanding different aspects of probability (esp. the older outcome diagrams and trees and the newer eikosogram) we illustrate some of the negative consequences of their use.

We introduce and develop the diagram called an eikosogram and show how it visually grounds probability and naturally incorporates the rules of probability within its construction. As with the ringed diagrams, the eikosogram is not a new diagram although it has not yet been put to its full use in understanding the probability calculus. It can be used descriptively to identify independence structures at a glance (conditionally and unconditionally) and to distinguish independence statements on events from those on random variables. The distinct notions of disjoint and independent events are easily distinguished visually via eikosograms. Eikosograms can be used actively to derive and to explain probability calculations (e.g. measures of association, determination of marginal and conditional probabilities) and to visually derive theorems (e.g. Bayes theorem).

More formally, the eikosograms provide a visual basis for motivating axioms of probability. As with Kolmogorov these could be based on sets, however we show that eikosograms nicely lend themselves to providing the intuition for axioms based on random variables. As with Whittle’s (1970) axioms for expectation of random variables, set based results could be developed afterwards to yield the Kolmogorov formulation as a special case.

Together, eikosograms, outcome trees and outcome diagrams provide the visual means to ground probability and should form the basis for any introductory treatment. The role of Venn disagrams in probability, if it exists at all, is for the abstract treatment of pre-defined events after the probability calculus is well understood via eikosograms.

The remainder of the paper is organized as follows. Section 2 explores the meaning of diagrams in general and in particular the meaning and use of ring diagrams, historically leading up to their distinctive uses in logic by each of Euler and Venn. The weaknesses of Venn diagrams for teaching probability are discussed in Section 3. In Section 4, we introduce the eikosogram and explore its use for developing the calculus of probability. Section 5 shows how the eikosogram complements other diagrams, notably outcome trees and outcome diagrams, to present a coordinated development of probability. Section 6 wraps up with some concluding remarks.

2 On Diagrams and the Meaning of Venn Diagrams

Good diagrams clarify. Very good diagrams force the ideas upon the viewer. The best diagrams compellingly embody the ideas themselves.

For example, the mathematical philosopher Ludwig Wittgenstein would have that the meaning of the symbolic expression 3×4 is had only by the “ostensive definition” shown by the diagram of Figure 1. ‘What is 3×4?’ can exist as a question only because the diagram provides a schema for determining that $3 \times 4 = 12$. The proof of $3 \times 4 = 12$ is embodied within the definition of multiplication itself and that definition is established diagrammatically by a “perspicuous representation” (e.g. see Wittgenstein (1964) p 66 #27, p. 139, #117 or Glock, 1996, pp. 226 ff., 274 ff., 278 ff.).

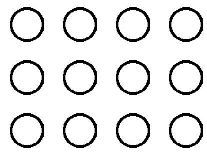


Figure 1: Defining multiplication: This figure *is* the meaning of 3×4.

Diagrams which provide ostensive definitions of fundamental mathematical concepts have a long history. In the Meno dialogue, Plato has Socrates engage in conversation with an uneducated slave boy, asking him questions about squares and triangles ultimately to arrive at the diagram in Figure 2. Although ignorant at the beginning of the dialogue,

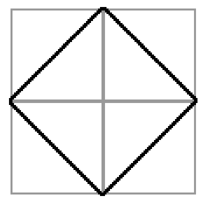


Figure 2: Each small square has area 1. The inscribed square has area of 2 and hence sides of length $\sqrt{2}$.

the slave boy comes to realize that he does indeed know how to construct a square of area 2 (the dialogue actually constructed a square of area 8, or one having sides of length $2\sqrt{2}$). Not having realized this before, nor having been told by anyone, Socrates concludes that the boy's soul must have known this from before the boy was born. With some work, the boy was able to recall this information through a series of questions. From this Socrates concludes that the soul exists and is immortal.

The simpler explanation however is that Socrates led the boy to a diagram (familiar to Socrates) which clearly shows a square of area 2. By showing the existence of the length $\sqrt{2}$, Figure 2 actually gives meaning to the concept of $\sqrt{2}$.

Together, Figures 1 and 2 allow us to pose the question as to whether $\sqrt{2}$ is a rational number. If $\sqrt{2}$ were rational, then it would be possible to draw the square of Figure 2 as a square of circles as in Figure 1, each side having number of circles equal to the numerator of the proposed rational number. That $\sqrt{2}$ is not rational is essentially the same as saying that this cannot be done. Dewdney (1999, pp. 28-29) gives a proof such as the ancient Greeks might have constructed along these lines.

Diagrams can give concrete meaning to concepts which might otherwise remain abstract. Although not always immediately intuitive, like Socrates' guiding of the slave boy, they can be reasoned about until their meaning becomes strikingly clear. Two examples of more interactive diagrams of this nature which one of us has produced are 1. an animation which shows the Theorem of Pythagoras and implicitly its proof (Oldford, 2001a) and 2. a three-dimensional physical construction which gives meaning to the statistical concepts of confounding and the role of randomization in establishing causation (Oldford, 1995). In both cases, the visual representation secures the understanding of otherwise abstract concepts.

Venn-like diagrams have a varied history which long predates Venn's use of them (Venn, 1880, 1881). The diagrams have often been given some mystical or religious significance, yet even then the content is conveyed via the same essential features of the diagrams. The overwhelming features of these diagrams are the union and intersection of individual regions.

2.1 The two-ring diagram

Consider diagram (a) of Figure 3. The simple interlocking rings have been used symbolically to represent the intimate union of two as in the marriage of two individuals, or the union of heaven and earth, or of any two worlds (e.g. see

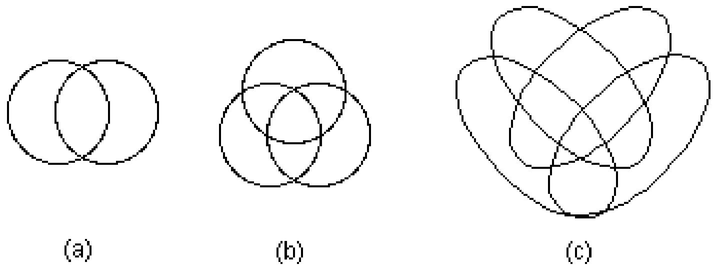


Figure 3: Venn's diagrams.

Liungman, 1991, Mann 1993). The intersection symbolizes where the two become one. This symbolism is of ancient, possibly prehistoric, origin.

The intersection set, or *vesica piscis* (i.e. fish-shaped container) of Figure 4, has been used by many cultures (the

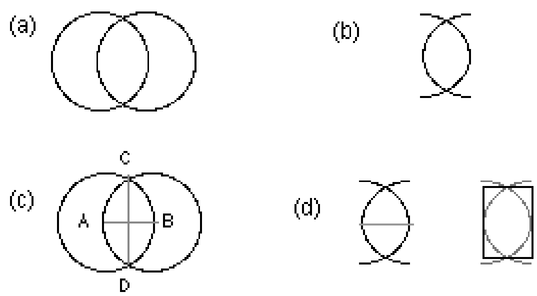


Figure 4: Vesica Piscis.

term vesica piscis is also sometimes used for the whole diagram as in Figure 4 (a)). For example, the cover of the famous chalice well at Glastonbury in Somerset England, whose spring waters have been thought of as sacred since earliest times, is decorated with the vesica piscis as in Figure 4 (a). The figure is formed by two circles of equal radius, each having its centre located on the perimeter of the other.

The mystical interpretation might have been amplified by the practical use of the vesica piscis in determining the location and orientation of sacred structures. According to William Stukely's geometric analysis of Stonehenge in 1726, the stones in the inner horseshoe rings seem to be aligned along the curves formed by vesica pisces as in Figure 4(b) (see Mann, 1993, p. 44). Whether Stonehenge's designers had this in mind or not, that Stukely would consider this possibility indicates at least the mystical import accorded the vesica piscis in 1726.

Orientation according to the cardinal axes of the compass were determined via the vesica piscis as follows. The path of the shadow cast by the tip of an upright post or pillar from morning to night determines a west to east line from A to B of Figure 4 (c). The perpendicular line CD is determined by drawing two circles of radius AB, one centred at A, the other at B - a vesica piscis. A rectangular structure with this orientation (or any other significant orientation, e.g. along a sunrise line) and these proportions is easily formed as in Figure 4 (d). Should a square structure be desired (e.g. Hindu temples for the god Purusha, Mann 1993, p. 72) a second vesica piscis can be formed perpendicular to the first (after first drawing a circle of diameter AB centred at the intersection of the lines AB and CD so as to determine a vertical line of length AB to fix the location of the second vesica piscis – the square is then inscribed by the intersection points of the two vesica pisces).

According to Burkhardt (1967, pp. 23-24) (see also Mann, 1993, pp. 71-75) this means of orientation was universal, used in ancient China and Japan and by the ancient Romans to determine the cardinal axes of their cities. The Lady Chapel of Glastonbury Abbey (1184 C.E.) has both its exterior and interior proportions described exactly by rectangles containing a vesica piscis as in Figure 4 (d) (see Mann. 1993, p. 152) and many of the great cathedrals of Europe were oriented using much the same process.

The mathematical structure of the vesica piscis would have been well known and might itself have contributed something to its mystery. The very first geometrical figure appearing in Euclid's Elements is that of Figure 5. Proposition 1 of the first book asserts that an equilateral triangle ABC can be constructed from the line AB, essentially

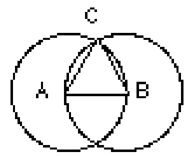


Figure 5: First Figure of Euclid's *Elements*.

by constructing the vesica piscis (see Heath 1908, p. 241).

Interestingly, the equilateral triangle itself has long had a mystical interpretation. According to Liungman (1991), the equilateral triangle is "first and foremost associated with the *holy, divine number of 3*. It is through the tension of opposites that the new is created, the third" (his italics). Xenocrates, a student of Plato, regarded the triangle as a symbol for God. Three appears again in the form of the irrational number $\sqrt{3}$ as the ratio of the length of CD to that of AB in Figure 4 (b). Whether this fact in any way enhanced the mystical significance of the vesica piscis is unknown, although it does seem a plausible speculation – especially for Christian thinkers.

The vesica piscis was adopted as an important symbol in Christianity and appears frequently in Christian art and architecture. Besides the obvious connection with the fish symbol of Figure 4(b) used by early Christians, it came to represent the purity of Christ (possibly through allusion to a stylized womb and so to the virgin birth of Christian scripture). Often the vesica piscis has appeared with a figure of Christ or the Virgin Mary within it (e.g. see Mann, 1993, pp. 24 and 52 for examples from the middle ages). The strength of this symbolism in the Christian faith no doubt significantly contributed to the adoption of the pointed arch (see Figure 6) as a dominant feature in Gothic architecture (e.g. notably in windows and vaults). The vesica piscis continues to be a popular symbol in Christian publications,

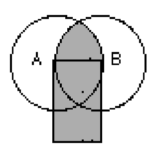


Figure 6: The Gothic arch.

art, and architecture to the present day.

2.2 The three-ring diagram

The three intersecting circles of Venn's diagram in Figure 3(b) is itself an ancient diagram representing a "high spiritual dignity" (Liungman, 1991). As mentioned earlier, the number 3 has long been considered divine. Xenocrates, for example, held the view that human beings had a threefold existence: mind, body, and soul. One can see how, as in the case for two intersecting rings, the union of three different but equal entities each having some attributes in common with another and possibly with all others simultaneously could have a deep mystical or religious appeal.

Certainly, once the holy trinity of the "Father, Son, and Holy Spirit" became established as a fundamental tenet of the Christian faith, the symbols were adopted with the obvious interpretation. The three intersecting rings have long appeared in Christian art and architecture and continue to do so to the present day. Figure 7 shows some variations on the three intersecting rings used in Christian symbolism to represent the holy trinity. The last one, interestingly,

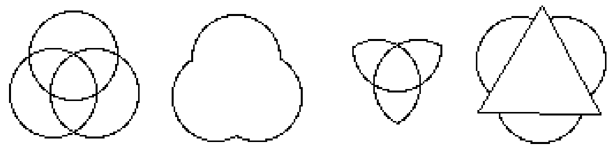


Figure 7: Symbols of the Christian Trinity.

superimposes the equilateral triangle over the three circles thus making use of two ancient spiritual symbols. This symbol is still commonplace on Christian vestments and altar decorations.

Mathematically, if the circles are drawn (as with the vesica piscis) so that their centres are at the three corners of the intersection set, then the intersection set shares a curious geometric property with a circle – the figure, called a Reuleaux triangle (e.g. see Santalo, 1976, p 8 ff), has constant width through its centre. That is, parallel tangent lines have the same distance between them, wherever they are positioned on the boundary.

2.3 The logic diagrams of Euler

Over the course of one year from 1760 to 1761, the natural scientist and mathematician Leonhard Euler wrote a series of letters to a German princess in which he presented his thoughts on a variety of scientific and philosophical topics with such clarity and generality that the letters were to sweep Europe as "a treasury of science" (Condorcet, p. 12, 1823 preface to Euler) accessible to the reader without much previous knowledge of the subjects addressed.

In the 1823 preface to the third English edition, Euler is regarded as "a philosopher who devote[d] himself to the task of perspicuous illustration." When Euler comes to explain Aristotelian logic to the princess, he makes use of a series of diagrams, diagrams which were to become known in logic as "Eulerian diagrams".

Euler was educated in mathematics as a child by his father, himself a Protestant minister educated in theology and a friend of the great mathematician Johann Bernoulli (e.g. see O'Connor and Robertson, 2001). The plan had been for the younger Euler to study theology at university and this he did, until Bernoulli convinced the father of the young man's formidable mathematical talents. A devout Christian all his life and one-time student of theology, it is hard to imagine that Euler would not have been well aware of the pervasive Christian symbols.

Whatever the source, the diagrams he presented the princess to better explicate Aristotelian logic would be familiar to someone both trained in mathematics and aware of Christian symbolism. The four basic propositions of Aristotle as shown by Euler appear in Figure 8. The diagrams make the points by the intersection (or not) of the circular areas, by

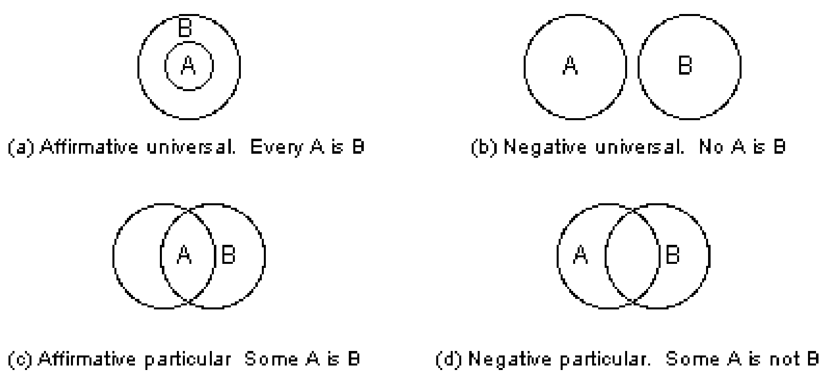


Figure 8: Basic Euler diagrams for the four Aristotelian propositions.

containment (or not) of circular areas, and by containment of the letters A and B – the letter placement allowed Euler to indicate the two "particular" propositions of Figure 8 (c) and (d).

Euler went on to show how all of the Aristotelian syllogisms might be demonstrated in the same way. For example, Figure 9 shows how these diagrams illustrate a relatively simple syllogism.

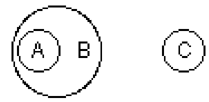


Figure 9: Euler diagram for the syllogism: No B is C; All A is B; ∴ no A is C.

Some syllogisms might need more than one diagram. Figure 10 shows all possible configurations for one such syllogism. Each diagram is itself consistent with the whole of the information contained in the propositions and

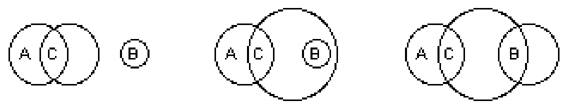


Figure 10: Euler diagrams which are each consistent with the syllogism: No A is B; Some C is A; ∴ some C is not B.

hence in the conclusion of the syllogism. While any one would explain the syllogism, it might be misleading in other respects. Consequently, Euler would completely enumerate the different cases which generate a given syllogism and present them all – nowhere in his letters to the German princess does Euler make use of the three ring diagram of Figure 3(b).

Unfortunately, not all syllogisms can be represented this way. As Venn (1881, pp. 523-4) pointed out even a fairly straightforward proposition such as "All A is either B or C only (i.e. not both)" cannot be expressed with the circles of an Euler diagram. One might attempt to do so via a collection of diagrams as we have done in Figure 11, but individually these do not contain the complete information available in the syllogism and seemingly contradict one another as to what that information might be.

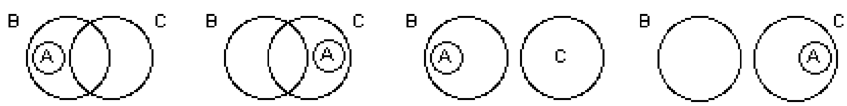


Figure 11: Euler diagrams which collectively express the single proposition: A is either B or C only.

2.4 The logic diagrams of Venn

John Venn graduated from Cambridge University in 1857, was ordained as a Christian priest two years later, and returned to Cambridge in 1862 as a lecturer in "Moral Science" where he studied and taught logic and probability (O'Connor and Robertson, 2001).

Venn was keenly interested in developing a symbolic logic and wanted a diagrammatic representation to go with it. Euler's diagrams were well known and had widespread appeal by the time of his writing in 1881:

> "Until I came to look somewhat closely into the matter I had not realized how prevalent such an appeal as this had become. Thus of the first sixty logical treatises, published in the last century or so, which were consulted for this purpose:- somewhat at random, as they happened to be most accessible:- it appeared that thirty-four appealed to the aid of diagrams, nearly all making use of the Eulerian Scheme."
> John Venn, *Symbolic Logic*, 1881 (page 110 of the 2nd Edition, 1894).

Venn's logic, like Boole's, was mathematical in nature. For example, $xy\bar{z} = 0$ indicates that the simultaneous condition x and y and not z cannot occur. The mathematics allowed propositions such as this to accumulate and inferences to be drawn as the information became available. Venn's diagrams had to serve in the same way. In his words:

> "Of course we must positively insist that our diagrammatic scheme and our purely symbolic scheme shall be in complete correspondence and harmony with each other. *The main objection of the common or Eulerian diagrams is that such correspondence is not secured. ...* But symbolic and diagrammatic systems are to some extent artificial, and *they ought therefore to be so constructed as to work in perfect harmony together*."
> John Venn, *Symbolic Logic*, 1881 (page 139 of the 2nd Edition, 1894).
> Italic emphasis is added.

Besides the failings alluded to in the previous section, Euler's diagrams required considerable thought in the construction – all possibilities needed to be followed as the diagrams were constructed. If you know the answer, as is the case for simple syllogisms, the diagrams are easy to construct; if you don't they can be considerable work.

Euler diagrams were designed to *demonstrate* the known content of a syllogism; Venn's diagrams were designed to *derive* the content. Remarkably, this profound distinction between the two diagrams can be missed by some mathematical popularizers, notably Dunham (1994 p. 262) who imagines Venn's innovation being discovered by any "child with a crayon".

Given his religious training, it would be surprising if Venn were unaware of the Christian symbolism of at least the three ring diagram he was to introduce to the study of logic. This three-ring diagram was to be employed to record the logical content of each proposition *as it became available*.

Figure 12 illustrates this use for a simple syllogism – one shades out the regions which correspond to impossible conditions as they become known. In this way, information accumulates by being added to the diagram as it becomes

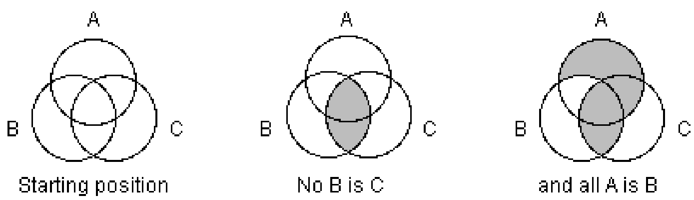


Figure 12: No B is C; All A is B; therefore no A is C.

available. At any point one can see the consequences of the information to date – only the unshaded regions (including the region outside all three circles: not A not B not C) are possible.

Figure 13 illustrates a more complicated syllogism which requires Venn's diagram of Figure 3(c) (which seems to be original to Venn) in order to render the logic diagrammatically. Left to right the diagrams show the effect of adding each new piece of information to what is known. Carrying out the logic via Euler diagrams would be considerably more difficult.

Besides their *active* use in the analysis of logical structure, Venn's diagrams differ from Euler's in another important respect. Each region represents a class; unshaded it remains possible, shaded it becomes impossible. There is no provision for indicating the particular "Some A is B" – it remains indistinguishable from "A and B has not been ruled out". Venn sees no need to explicitly distinguish these possibilities; they remain only because of the historical dominance of Aristotelian logic.

2.5 The essence of Venn diagrams

Throughout their long history, Venn-like diagrams seem to be put to similar use, albeit in different contexts. The diagrams compel one to think in terms of identifying different entities, what they have in common, and how they differ from one another and possibly from everything else. As formal set theory developed, the same figures were used to

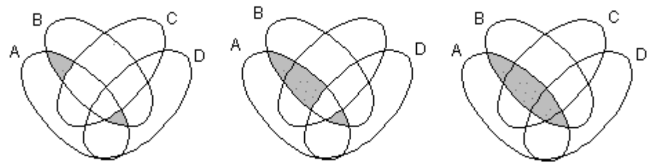


Figure 13: A complex syllogism – the information of each statement is added to the diagram by progressively shading those regions which the statement excludes. From left to right the cumulative effect of the following statements can be read from the diagrams: i. All A is either B and C, or not B; ii. If any A and B is C, then it is D; and iii. No A and D is B and C. From the last figure we see that together these statements imply that no A is B.

naturally embody the properties of sets – intersection, union, complement. However, just as some ideas can be given meaning only by a diagram, a diagram can be incapable of easily producing anything but these ideas.

3 Criticism of Introducing Probability via Venn Diagrams

Venn diagrams (and Euler diagrams) are a useful tool in logic where conditions are either possible or impossible, can occur together or cannot; consequently they are also useful in set theory to illustrate relations between sets. Since Kolmogorov's axioms are now routinely used and are based on events as sets, Venn diagrams would seem to be well suited to use in developing and illustrating probability concepts. Indeed, as indicated in the Introduction and in the Appendix, their use has now come to dominate introductory treatments of probability over the past two decades or so. While successfully grounding the nature of interrelationships between *predetermined* events, Venn diagrams fare less well on grounding the probability calculus itself. The principal shortcoming of their uncritical use lies in this distinction – their semantic coincidence with set relations will always weight the balance of discussion of probability toward events, especially as sets, and away from the probability values themselves. This can have subtle and undesirable side effects.

3.1 Inability to quantify probabilities

Venn's symbolic logic uses the numerical values one and zero to represent possibility and impossibility, respectively; multiplication of the values of different states correctly determine the possibility or not of their simultaneous occurence. While the extension to the certain *probabilities* of one and zero seems short, it is in fact substantive.

Venn criticised De Morgan, LaPlace and others for regarding probability as an extension of formal logic (Venn 1888, VI, Sect. 4 ff). One of the innovations of Venn's logic diagrams over those of Euler was the disposal of the Aristotelian *particular* propositions of Figure 8(c) and (d) which Venn viewed to be the concern of the 'science of Probability' and not of formal logic (Venn 1888, I, Sect. 2 ff) – that Venn did not then use his diagrams for explicating probability is not surprising. Venn diagrams are not intended to quantify uncertainty between zero and one.

Yet it is clearly desirable to have a diagram which does – if one looks in some texts (or worse, searches on the internet) Venn diagrams can be seen to appear with counts or probability values attached to the different regions. Unfortunately, because the semantic content of these diagrams is not up to the task of showing probabilities, the diagram becomes no more than a visual key with which to record and to extract these numbers. Few probabilistic insights are thus had visually, no more than would be available from Venn diagrams unadorned. Instead, further insight (e.g. probabilistic independence) must depend on calculations using the numerically labelled Venn diagrams as a convenient lookup table for probabilities of various events and their intersections.

We seem to be missing a diagram designed for the calculus of probability. This is the purpose of eikosograms – developed and explored in Section 4.

3.2 Disjoint versus independent events

Intuition for the development of probability via Kolmogorov's axioms is based entirely on how well one understands sets (and, added for mathematical convenience, such abstractions as countable infinity and limits). Disjoint events are critically important in this development and are well served by the Venn diagram of Figures 14(a). The primary

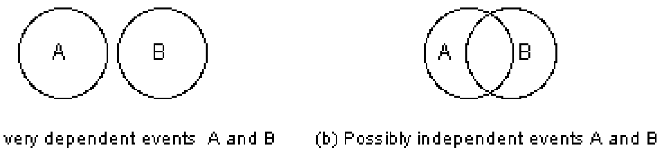


Figure 14: A mixed visual message: (a) represents disjoint but dependent events, (b) overlapping but possibly independent events

purpose of disjoint sets is to set up the additivity axiom which every measure must satisfy. Aside from being a finite and particularly normed measure, probability is distinguished from other measures by capturing such concepts as probabilistic independence and conditional probability. Not being part of Kolmogorov's axioms these concepts require separate and subsequent definition.

When independence is introduced to students, the Venn diagrams of Figures 14 are typically revisited to carefully explain the difference between disjoint and independent events. Even so, some students continue to occasionally confuse the two concepts. This, and the fact that the revisit and careful explication are necessary at all, seems indicative of a problem.

The difficulty is that the semantic content of the diagram of Figure 14(a) is one of *separation* which could easily be associated with either '*disjoint*' or '*independent*'. Tradition, via sets, has the separation associated with disjointness and denied to independence; tradition offers no similar picture for independence. Revisiting Figure 14(a) after independence is introduced seems necessary because despite the immediate visual message to the contrary the events of Figure 14(a) are highly dependent – if one occurs the other cannot. Similarly, the diagram of Figure 14(b) visually shows *connectedness*, matching perfectly the sense of *overlapping* events but consequently mismatching a potential visual association with the *dependence* of events – the events of Figure 14(b) could very well be independent.

Having traditionally associated Figures 14(a) and (b) with disjoint and intersecting events (concepts related to sets, regardless of their measure), we have no diagram for the independence and dependence of events (concepts related to probability). We seem to be in the curious position of providing no visual aid for the concepts most peculiar to probability. The dilemma is resolved via eikosograms which dramatically show independence and visually distinguishes disjoint from independent events.

3.3 Independence again: events and random variables

Probability developed from Venn diagrams is essentially confined to events and so most introductory treatments now introduce probabilistic independence by way of only two events, A and B which are said to be independent when $Pr(AB) = Pr(A)Pr(B)$. This is not how Kolmogorov chose to define the independence of events, with good reason.

Instead, Kolmogorov (1933, pp. 8-12) first defined independence for random variables[2] having finitely many distinct values and only afterwards defined independence of events in terms of this definition. The random variables X and Y are defined to be independent if $Pr(X = x_i, Y = y_j) = Pr(X = x_i)Pr(Y = y_j)$ for all i and j which index the possible values of X and Y. Events A and B are said to independent if two binary random variables, each corresponding to that event's occurrence or not, are detemined to be independent.

For two events A and B to be independent, only one of

$$\begin{aligned} Pr(AB) &= Pr(A)Pr(B)\,, \quad Pr(AB^c) = Pr(A)Pr(B^c) \\ Pr(A^cB) &= Pr(A^c)Pr(B)\,, \quad Pr(A^cB^c) = Pr(A^c)Pr(B^c) \end{aligned}$$

[2]Kolmogorov actually used the word experiment whose outcomes constitute a finite partition of the basic set, or sample space. This is equivalent to a finitely many valued random variable for which all values are possible and at least one must occur. Kolmogorov's definition is also for the *mutual independence* of n variables rather than just two.

need be demonstrated as the others will follow mathematically. So the first could be, and often has been, taken as the definition of independence of events.

This Venn diagram based approach reverses Kolmogorov's order, beginning with the independence of events as defined by the first equation above and then later introducing independence of random variables. The transfer of this idea from events on a Venn diagram to random variables requires some care. It is not unheard of for a student to have found $Pr(X = x_i, Y = y_j) = Pr(X = x_i)Pr(Y = y_j)$ hold for one i, j pair and then to mistakenly declare the independence of the random variables X and Y (particularly if, as here, the equality has been expressed entirely symbolically). The problem becomes more pronounced when conditional independence is considered and the condition as well as the random quantities can be an event or a variable.

We note also that we find independence to be more naturally understood and so better defined in terms of conditional probability rather than as above. So too, perhaps, did Kolmogorov (1933, p. 11) who wrote:

> "In introducing the concept of independence, no use was made of conditional probability. Our aim was to explain as clearly as possible, in a purely mathematical manner, the meaning of this concept. Its applications, however, generally depend upon the properties of certain conditional probabilities."

Even though Kolmogorov had already defined conditional probability (p. 6) prior to independence (p. 9), a set-theoretic (or Venn diagram) starting point seems to have encouraged the less natural definition.

3.4 Almost mute on the nature of relationships

One of the most important uses of probability is to describe the relationships between different events or random variables. The conditional probability of one event or random variable given another summarizes that relationship in a way which matches intuition and experience. Although helpful for explaining conditioning of events, the direction and strength of the conditional probability relation is beyond helpful illustration via Venn diagrams.

3.5 Effect on axiomatic treatment and interpretation of events and probability

There are other axioms for probability than Kolmogorov's; there are other applications of the probability calculus than those of Venn's experiential relative frequency. Venn diagrams seem to carry baggage that is seen by some writers to be inappropriate for either the axioms or the interpretation of probability.

For example, Shafer (2001) criticises the timelessness associated with events in the usual Kolmogorov set-up. In Shafer (1996) a theory of probability and causality is built which depends critically on the temporal ordering of situations. These are presented in trees whose nodes are situations arrived at over time and where branching indicates different possibilities which follow. Events must be situated in time. To quote Shafer (2001, p. 11): "An event in this new dynamic theory is a partial slice across the tree, formed by one or more nodes". In this theory, an event is not naturally taken to be a subset of a sample space and Venn diagrams have little to offer.

Contrary to Venn, Jaynes (1996) develops probability as a normative extension of formal logic. Also contrary to Venn, Jaynes (1996, pp. 221-222) declares that Venn's diagrams actively mislead when applied to propositions. Because regions of intersection suggest subregions and because probability is additive on disjoint regions, Venn diagrams lead to a viewpoint that logical propositions can always be refined to a "disjunction of mutually exclusive sub-propositions." The points in the diagram "must represent some ultimate elementary propositions ω_i into which A can be resolved." So tied to sets are the diagrams that they quickly lead us

> "…to the conclusion that the propositions to which we assign probabilities correspond to sets of points in some space, that the logical disjunction $A + B$ stands for the union of sets, the conjunction AB for their intersection, that the probabilities are an additive measure over those sets. But the general theory we are developing has no such structure; all these things are properties only of the Venn diagram."

Jaynes' point here is that looking at Venn diagrams in relation to a normative theory of logic can easily lead to absurdity, particularly in the notion that there will exist some elementary sub-propositions into which the propositions of interest can be meaningfully resolved.[3] This criticism is more properly directed at using sets to provide the intuitive basis for probability; the Venn diagram unavoidably suggests relations between abstract sets and so becomes the target.

[3] The problem of refinement and resolution is one that has concerned many writers (e.g. Savage (1954, pp. 82-91), Fine (1973, pp. 60-1)) and does not go away by developing events by way of trees (e.g. Shafer (1996, pp. 275-297)).

Fine (1973, 61-4) criticises the Venn diagram as a basis for probability (actually, the σ-field on which Kolmogorov's axioms are based) because it suggests that if events A and B are of interest, then so too must be their individual complements, their union and their intersection (all closure properties of σ-fields). For example, when A is a logical proposition then its complement is of interest because it represents *not* A; however, when A is the event that an experiment (or random variable) has a particular outcome (or value) then each alternative individual outcome is of interest – rarely would the complement be of any intrinsic interest. Similarly, it can happen that events A (e.g. grainy photograph) and B (e.g. dark photograph) might be of interest and that probabilities of these events could be reasonably determined, yet that neither their intersection nor their union be of any intrinsic interest and that these probabilities could not be reasonably determined. Fine (1973) takes these cases to show that the Venn diagram (or σ-field) imposes more structure on probability than is needed or is reasonable for some applications.

A number of authors have developed axioms for entirely different concepts, based on the relevant intuition, which are then shown to produce a quantitative concept which obeys Kolmogorov's rules of probability (when applied to sets). None of these axiomatic approaches were based on the set theoretic intuition contained in a Venn diagram, yet all ultimately produced a quantitative concept which obeyed the rules, or calculus, of probability.

For example, Whittle (1970) develops axioms for the expectation of a random variable based on intution about averages; in a reversal of Kolmogorov, probability is then defined as expectation applied to indicator variables. Somewhat farther removed is the approach of Savage (1954) where reasonable postulates for personal preference are developed to produce what might be described as a theory for rational personal judgement; as the rules developed for preference ultimately coincide with those of probability, Savage (1954) calls this preference "personal" probability. De Finetti (1937, 1974) asserts that probability does not exist and instead develops postulates for prices based on rational bets; again a quantity which follows Kolmogorov's rules results, although this too is interpreted as a personal probability (alternatively, de Finetti's "prevision" mathematically , if not conceptually, matches Whittle's "expectation"). Jaynes (1996, A-5) eschewed this personalistic approach as "belonging to the field of psychology rather than probability theory" and instead developed a normative extension of formal logic on "desiderata of rationality and consistency" for the notion of the plausibility of logical propositions; again the quantitative plausibility so produced obeys Kolmogorov's axioms when sets are used in place of propositions.

Because a quantitative concept resulted which followed the rules of probability, most of these authors took their postulates (for different concepts) to give meaning to probability – some suggesting theirs is the only meaning; others that theirs is one of many. This produced much fodder for polemical argument.

For our purposes, it is sufficient to notice two things. First, all agree on the rules of probability if not on the objects to which they apply – there is a common calculus. Second, when first developed Kolmogorov's (1933) set-based axioms (and through them, the Venn diagram) were meant to provide this common basis, but subsequently were found to be unnecessary and/or unappealing with respect to some later developments.

3.6 Potential for meaningless probability examples

The concerns of the previous section can show themselves in the teaching of introductory probability. Venn diagrams can skew the teaching towards what are fundamentally problems in set theory or logic which only involve probability incidentally.

As Fine (1973) pointed out, Venn diagrams suggest that all regions are of interest and consequently that so too must be the associated probabilities. Immediate from a Venn diagram is the set-theoretic (or logical) result $A \cup B = A + B - A \cap B$, now relating events. Immediately following Kolmogorov's axioms, the probabilistic version, $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$, is thus given a prominence in introductory treatments it might not otherwise enjoy – sometimes even elevated to being called a *principle* (the "inclusion-exclusion principle").

Teachers can be hard-pressed to produce meaningful concrete illustrations of this 'principle' of probability and often fall unwaryingly into the trap which Fine (1973) discussed. It is not unusual to see examples and problems where the intersection and union of events are either intrinsically uninteresting or are such that their probabilities could not be reasonably determined in practice. Two typical examples taken from published sources are:

> Illustration 1: *Paul and Sarah both apply for jobs at a local shopping centre; the probability Paul gets a job is 0.4, the probability Sarah gets a job is 0.45 and the probability they both get jobs is 0.1.*
> *What is the probability at least one of them is employed?*

> Illustration 2: *Suppose that 75% of all homeowners fertilize their lawns, 60% apply herbicides and 35% apply insecticides. In addition, suppose that 20% apply none of these, 30% apply all three, 56% apply herbicides and fertilizer, and 33% apply*

insecticides and fertilizer.
What percentage apply (a) herbicides and insecticides; (b) herbicides and insecticides but not fertilizer?

In the first illustration, it is not at all clear what the probabilities mean, if anything. Are they personal probabilities? And if so whose? Are they measures of the plausibility of logical propositions? If so, how were they arrived at? Are they experiential relative frequencies? If so, based on what data? And how are the characteristics of Paul and Sarah related to those of the data so as to permit the different probabilities and their application here? In particular, how would the joint probability of 0.1 (no independence) have been produced?

The second illustration at least has numbers which have self-evident meaning. They are characteristics of the population of homeowners and we can imagine them having been produced by census or survey. As relative frequencies, they have the appearance of probability but need not be interpreted as probability; measures of total acreage subjected to each of the possibilities could have served the same purpose and arguably would have been more interesting. Part (a) of the question is typical of such illustrations. It is constructed by revealing only part of the information which would have been contained in the background data; it is hard to imagine background information which would have captured "herbicides and fertilizer" and "insecticides and fertilizer" without also having captured "herbicides and insecticides". It is only the artificial *selective* revelation of data characteristics that allow the problem to be posed. Part (b) is a logical and calculational exercise, and that for an event of no obvious intrinsic interest.

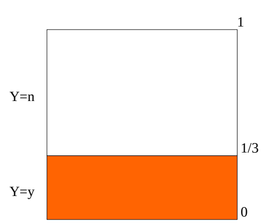
An inherently meaninglessness context obscures the concept of probability and irrelevance trivializes its statistical application. That such illustrations can be easily (and are regularly) produced seems, again, indicative of a problem.

4 Eikosograms

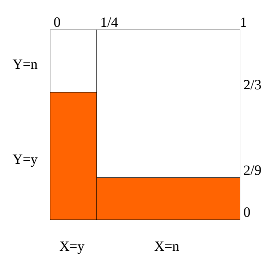
In developing his system for symbolic logic Venn promoted and followed the dictum that "…symbolic and diagrammatic systems …be so constructed as to work in perfect harmony together" (Venn, 1881, p.139). The result was his system of diagrams tightly coupled with a symbolic representation for logic. That these same diagrams then fall short of satisfying this dictum for probability should come as no surprise.

A diagram tailored to probability and one which arguably fulfills Wittgenstein's notion of an "ostensive definition" for probability (especially for conditional probability) is the *eikosogram* – a word[4] constructed to evoke 'probability picture' from classical Greek words for probability (*eikos*) and drawing or writing (*gramma*).

In an eikosogram, rectangular regions match events and areas match probabilities. All eikosgrams are built on a unit square whose unit area represents the probability 1, or certainty. The region is divided horizontally into non-overlapping strips, one for each distinct value that a random variable can take. In the case of a single event which either occurs or does not, i.e. a binary random variable, there will be only two horizontal strips. Figure 15(a) shows



(a) Eikosogram for Y



(b) Eikosogram for Y given X

Figure 15: Eikosograms with binary random variables Y and X each taking two values: y and n for "yes" and "no" and having $Pr(Y = \mathsf{y}) = 1/3$, the area of the shaded regions in either (a) or (b). $Pr(X = \mathsf{y}) = 1/4$, $Pr(Y = \mathsf{y}|X = \mathsf{y}) = 2/3$.

an example (throughout, values of y and n will indicate the binary values "yes" and "no"); shading or colouring helps distinguish variable values. When a conditioning variable (or event) is introduced, the square is first divided into vertical strips of width equal to the probability of each distinct value of the conditioning variable. Then, within

[4]This construction was kindly suggested by our colleague Prof. G.W. Bennett.

each vertical strip, horizontal strips are introduced so as the height of each matches the corresponding conditional probability. All resulting rectangular blocks have areas equal to the probabilities involved.

Just as the ring diagrams were not new to Venn, so too this diagram has seen use before – variants of it have been used to describe observed frequencies for centuries (at least as early as 1693 by Halley; see Friendly 2002 for some history on these variants). Recently Michael Friendly has developed and promoted a variant he calls "mosaic plots" to display observed frequencies upon which fitted model residuals are layered using colour (Friendly, 1994). The earliest use of an eikosogram (i.e. displaying probabilities) of which we are aware is by Edwards (1972, p. 47) where a single diagram appears with the unfortunate label of 'Venn diagram' (an example of how far the sense of a Venn diagram has been stretched). That such diagrams have been used and developed independently by many authors over time speaks to their naturalness and consequent value in describing and understanding probability.

4.1 Grounding probability.

From the eikosograms of Figure 15, it is clear that probability is a proportion – the probability that $Y = \mathsf{y}$ is the ratio of the shaded area to the whole area in Figure 15(a) and similarly in Figure 15(b). Taking the area of the square to be one yields a meaningful labelling of the axes which allows some probabilities to be read directly from them – e.g. in Figure 15(a) $Pr(Y = \mathsf{y}) = 1/3$ can be read directly off the vertical axis.

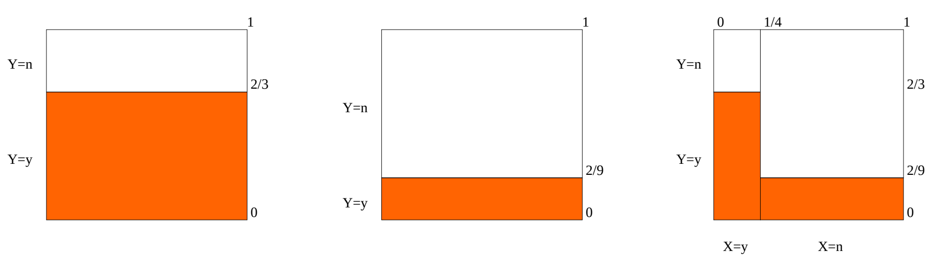
A meaningful physical analogy is had by imagining the eikosogram of Figure 15(a) lying flat on the ground in the rain; of those raindrops which hit the square, the proportion which strike the shaded region corresponds to the probability that $Y = \mathsf{y}$. (This could be easily simulated by Monte Carlo and displayed on a computer screen as, for example, in Oldford, 2001b). Figure 15(a) shows this proportion, and hence the probability, to be one third.

4.1.1 Conditional and joint probability

Conditional probability is introduced via Figure 15(b) by considering each vertical strip in turn. The leftmost strip fixes the condition $X = \mathsf{y}$. When we ask the question 'Of those raindrops which strike the leftmost strip, what proportion lands on the shaded area?', then we are asking for the probability that $Y = \mathsf{y}$ conditional on, or given that, $X = \mathsf{y}$, or symbolically for $Pr(Y = \mathsf{y}|X = \mathsf{y})$. The raindrop metaphor makes it clear that this conditional probability is the ratio of the area of the left shaded rectangle to the area of the entire leftmost strip. Again, the actual area of the leftmost strip does not matter; the ratio of the areas determine the proportion we call the conditional probability.

The horizontal axis allows easy determination of the probability for the various values of the conditioning variable – e.g. $Pr(X = \mathsf{y}) = 1/4$. The purpose of the vertical axis is now more clearly seen to be the determination of conditional probabilities – e.g. $Pr(Y = \mathsf{y}|X = \mathsf{y}) = 2/3$ and $Pr(Y = \mathsf{y}|X = \mathsf{n}) = 2/9$. In Figure 15(a), it is only because the background conditions are certain (i.e. having probability one) that the vertical axis determines the values of $Pr(Y)$.

In this way, all probabilities can be thought of as being conditional. Setting the total area to one, when the condition holds, shows focus and simplifies calculation. In the opposite direction, when each distinct condition is specified as a distinct value of a random variable, eikosograms can be combined according to the probabilities of that random variable. For example, Figure 16(a) shows the eikosogram for the probabilities of the values of the random variable Y



(a) When $X = \mathsf{y}$ holds (b) When $X = \mathsf{n}$ holds (c) Combined

Figure 16: The two conditional eikosograms of (a) and (b) combine, or mix, according to the probabilities 1/4 and 3/4, associated with $X = \mathsf{y}$ and $X = \mathsf{n}$, to produce the combined eikosogram for both Y and X.

when the condition that $X = \mathsf{y}$ holds, Figure 16(b) when $X = \mathsf{n}$ – each eikosogram provides a different focus. The two eikosograms are combined by adjusting each one's width to be proportional to the probability of its condition and aligning them side by side as in Figure 16(c). In this way, joint distributions can be built up as a mixture of conditional distributions, all the while preserving proportional areas and hence probability.

4.1.2 Joint and marginal probability

This particular picture, Figure 15(b) (or equivalently Figure 16(c)), of the joint distribution of Y and X is especially useful for understanding the transition between marginal and joint distributions, i.e. between Figure 15(a) and Figure 15(b).

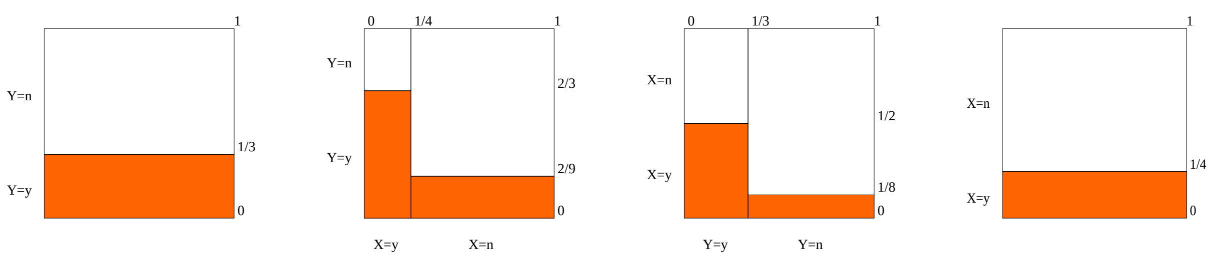
A different water analogy helps ground the concept. Think of the eikosogram of Figure 15(b) as a water container with the shaded areas corresponding to the level of water in each of two separate chambers: one being the left vertical strip with water filling 2/3 of the chamber, the other being the right vertical strip with water filling only 2/9 of this chamber. Imagine further that the line making the vertical division at 1/4 is actually a removable barrier which has created the separate chambers. Finding the marginal distribution of Y amounts to removing this barrier, i.e. removing the separate conditions, and having the water settle to some new level in the whole container as in Figure 15(a). This level determines the marginal probabilities for Y.

In the opposite direction, adding new conditions amounts to beginning with Figure 15(a), adding the barriers and redistributing the fixed amount of water to the separate sub-containers. In either direction, the probability (i.e. amount of water or area) is preserved.

4.1.3 Symmetry between variables

The assignment of one random variable to the vertical axis and another to the horizontal, or conditioning, axis will be natural in many contexts. However, in the grounding of probability it is important to realize that this assignment is arbitrary. Probability treats all variables symmetrically and distinguishes them only as to whether they are presently conditioned upon or not. A fuller understanding of the probabilistic underpinnings of two random variables is had by interchanging these roles for the variables in an eikosogram.

The various eikosograms of Figure 17 provide different insights about the probability relating two random vari-



(a) Marginal Y (b) Joint via Y given X (c) Joint via X given Y (d) Marginal X

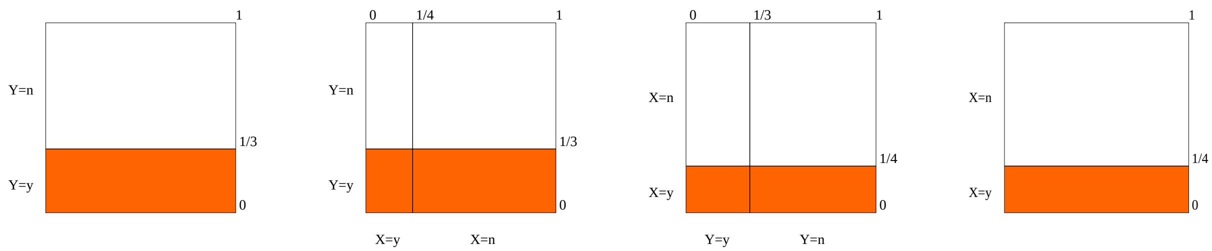
Figure 17: Different perspectives on the probability for Y and X. Probability is preserved under change of perspective.

ables. All probability information is contained in either Figure 17 (b) or (c); each simply provides a different perspective on the combined probabilities. Applying the water container metaphor to Figure 17(b) yields the marginal of Y in Figure 17(a); applying it to Figure 17(c) yields the marginal of Y in Figure 17(d).

It is important to realize that Figures 17(b) and (c) differ only in perspective and that *probability is preserved under change in perspective*. Each rectangular region in Figure 17(b) matches one in Figure 17(c) – e.g. the lower right rectangle in Figure 17(b) matches the top left one in Figure 17(c), both having $Y = \mathsf{y}$ and $X = \mathsf{n}$. Because the probability is preserved, the *areas* of matched regions must be equal. This consequence is called Bayes' Theorem.

4.1.4 Probabilistic independence

Probabilistic independence is readily apparent in, and naturally grounded by, eikosograms. Consider the eikosogram of Figure 18(b). Applying the raindrop metaphor we see that whether one focuses on the left side where $X = \mathsf{y}$ or



(a) Marginal Y (b) Joint via Y given X (c) Joint via X given Y (d) Marginal X

Figure 18: Independent random variables Y and X.

on the right side where $X = \mathsf{n}$, the proportion of the drops striking the shaded part where $Y = \mathsf{y}$ remains the same. That is, the conditional probability for each value of Y does not change when the value of the conditioning variable X changes – the conditional probability Y is independent of the value X.

Alternatively, applying the water container metaphor we see that the presence or absence of the vertical barrier at 1/4 in Figure 18 (b) has no effect on the water level anywhere – Figure 18(a) is essentially the same as Figure 18(b). Probabilities associated with the random variable Y remain the same whether one conditions on those of X or not.

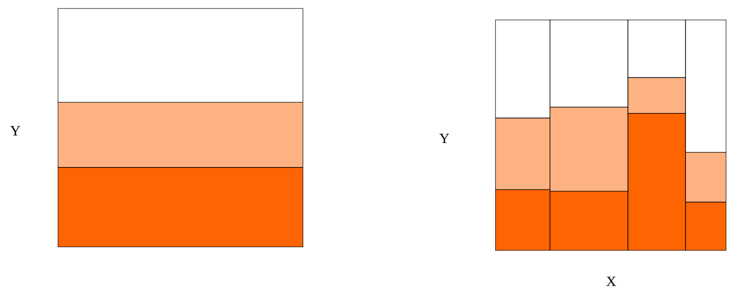
Either of these approaches could be used to define probabilistic independence, both, unlike Kolmogorov, being in terms of conditional probability. The overwhelming visual characteristic identified with the independence of two random variables is that of the eikosogram's flatness across all values of the conditioning variable. That the joint probability is, under these conditions, the product of the marginal probabilities is an artefact of this flatness; because of the flatness, the side of every rectangle in a row of Figure 18(b) has the same length and is easily read off the vertical axis, facilitating calculation of its area. Kolmogorov's definition of probabilistic independence corresponds to saying that the rectangles in the eikosogram align themselves in a checkerboard of rows and columns. Again, this seems more an artefact of independence than a natural starting point for its definition.

When more than two variables are involved, independence can exist between some variables and not others, between all pairs but not triples, between some variables conditional on the values of others, etc. Most of these possibilities are beyond the scope of this paper, although some will be explored later in Section 4.3; see Oldford (2003a) for an in-depth treatment of the case of three variables (or equivalently, three groups of variables). In every case, however, eikosograms which display flatness indicate some kind of independence and every kind of independence of random variables will exhibit itself as flatness in an eikosogram.

Finally, the symmetric nature of probability forces symmetry in probabilistic independence so that the eikosograms of Figures 18(c) and (d) would work as well to define independence. Indeed, under independence, the eikosograms of Figures 18(b) and (c) are essentially identical, being simple rigid transformations of one another – Kolmogorov's independence checkerboard.

4.1.5 More than two values per random variable

The intuition used throughout the preceding prevails however many distinct values the random variable Y can take, or, however many distinct conditions are considered. Figure 19 shows an example where Y takes three distinct values and X four. As before, probabilities correspond directly to areas, or ratios of areas in the case of conditional probabilities. For example, in Figure 19(b), numbering the possible values in left to right order for X and vertically for Y, the middle square of the leftmost strip has $Pr(Y = y_2, X = x_1)$ as its area, $Pr(Y = y_2|X = x_1)$ as its height, and $Pr(X = x_1)$ as its width. Both the raindrop and water container (though now with liquids of different density) metaphors continue to apply.



(a) Marginal probabilities for Y (b) Y given X

Figure 19: Eikosograms with multi-category random variables Y and X: Y taking three distinct values, X four.

4.2 Harmony with symbolic representation.

Eikosograms visually ground probability and are semantically consistent with its rules. The connection between the picture and a formal symbolic representation is as straightforward as that between Venn's diagrams and his symbolic logic.

Consider again the eikosogram of Figure 15(b) (for multiple values of X and Y, consider that of Figure 19(b)). Marginal probabilities of the conditioning variable determine the width of the strips, e.g. $Pr(X = \mathsf{y}) = 1/4$, and can be read off the horizontal axis – clearly these must sum to one. Conditional probabilities give the height of each rectangle within each strip, e.g. $Pr(Y = \mathsf{y}|X = \mathsf{y}) = 2/3$, and again must sum to one within each strip (i.e. $\sum_Y Pr(Y|X) = 1$); the location of the axis on the right even simplifies reading the conditional probability, as both picture and symbols follow a left to right order.

The joint probability $Pr(Y = \mathsf{y}, X = \mathsf{y})$, being the area of a rectangle, is seen to be $height \times width$ or $Pr(Y = \mathsf{y}|X = \mathsf{y}) \times Pr(X = \mathsf{y})$. That this holds whatever the value of Y or X is expressed more compactly as

$$Pr(Y, X) = Pr(Y|X) \times Pr(X)$$

and leads directly to Bayes theorem – expressed visually by the area equivalence of matching regions in Figures 17(b) and (c), and expressed symbolically as

$$Pr(X|Y) \times Pr(Y) = Pr(Y|X) \times Pr(X).$$

That the total area an eikosogram is one expresses the formal relation $\sum_Y \sum_X Pr(Y, X) = 1$; similarly the water container metaphor which produces marginal probabilities is simply $Pr(Y) = \sum_X Pr(Y, X)$.

The concept of probabilistic independence of variables is expressed visually by the flat of Figure 18, and symbolically as

$$Pr(Y|X) = Pr(Y).$$

The symbol[5] '$\perp\!\!\!\perp$' will denote independence as in $Y \perp\!\!\!\perp X$ for the independence of two random variables Y and X and $\perp\!\!\!\perp(Y, X, Z)$ for the mutual or complete independence of two or more random variables; an absence of independence will be indicated using the same symbol but with a stroke through it as in $Y \not\!\perp\!\!\!\perp X$.

Finally the inclusion-exclusion rule which appears so early in development via Venn diagrams is seen here to be a straightforward relation about overlapping areas of an eikosogram; from Figure 15(b) we have that

$$Pr(Y = \mathsf{y} \text{ or } X = \mathsf{y}) = Pr(Y = \mathsf{y}) + Pr(X = \mathsf{y}) - Pr(Y = \mathsf{y}, X = \mathsf{y})$$

which is easily extended. Similar rules about other overlapping regions could be developed as well.

All of the above rules for probability are semantically coincident with visual features of eikosograms. More are explored in Section 4.3.

[5]This symbol used as a binary operator is often mistakenly attributed to Dawid (1979), e.g. Whittaker (1990) page 31. However, Fine (1973, pp. 32-37) makes earlier use of the notation including as an $n-$ary operator and even develops axioms for independence which the usual probabilistic independence is seen to satisfy.

4.2.1 Axioms for probability

Eikosograms can also be used to motivate formal axioms for probability. Here we consider two different approaches: the first is based on intuition about marginal probabilities as shown for example in the eikosogram of Figure 19(a), the second based on intuition about conditional probabilities as shown for example in Figure 19(b).

Implicit in either case is the concept of a variable which takes on distinct values. We start with at least one such variable in hand and possibly many more. From these, other variables can be derived as make sense and we require probability axioms to hold for all such derived variables. Two such derivations are of particular interest.

First, a new variable could be defined such that each of its values matches one or more of another variable's values – a *many to one* mapping from the old variable's values *onto* the new variable's values. For example, suppose Y takes three values y_1, y_2, y_3; we now define Z to take value z_1 whenever Y takes either value y_1 or y_2 and to take value z_2 when Y takes value y_3. In the eikosogram of Figure 19(a), this amounts to combining the shaded regions to produce the single value z_1; the unshaded correspond to z_2. Note that a *many to many* mapping is prohibited because the values of the derived variable would no longer be distinct; the lightly shaded portions of Figure 19(a) cannot be combined with *both* the unshaded portions and with the darkly shaded portions *at once in the same variable* – although each such combination could be a value for different variables.

Second, given any pair of variables, Y and X, a new variable, Z, can always be constructed from the cross product $Y \times X$. That is each distinct value $z_{i,j}$ of Z corresponds to the pair (y_i, x_j) of values from the cross-product $Y \times X$. In the eikosogram of Figure 19(b), this amounts to forming Z by assigning a distinct value to each tile in the eikosogram.

Axioms for marginal probability, essentially the same as Kolmogorov's, can be motivated by considering eikosograms involving only a single variable. These are shown in Figure 20 and are motivated visually via Figure 19(a). That

M1	For every value y which a variable Y can take
	$Pr(Y = y) \geq 0$
M2	If Y has only a single value c, then
	$Pr(Y = c) = 1$
M3	If $y_i \neq y_j$ represent two distinct values of a variable Y then
	$Pr(Y = y_i \text{ or } Y = y_j) = Pr(Y = y_i) + Pr(Y = y_j)$

Figure 20: Kolomogorov like axioms based on marginal probability are motivated by single variable eikosograms.

they match Kolmogorov's is seen by noticing that an event in Kolmogorov's system is simply the realization of a value of a binary random variable – i.e. $[Y = \mathsf{y}]$ would be such an event. Of course just as with Kolmogorov, conditional probability will need to be defined separately. Eikosograms with two variables will be helpful in this definition and also in that of independence.

Alternatively, the structure of an eikosogram for two variables suggests axioms for conditional probability such as those shown in Figure 21. Here, a random variable without specified value indicates any possible value for that

C1	$Pr(Y\|Z) \geq 0$
C2	$Pr(Y\|Y) = 1$
C3	If Y takes on only two distinct values y and n, then
	$Pr(Y = \mathsf{y}\|Z) + Pr(Y = \mathsf{n}\|Z) = 1$
C4	$Pr(Y, X\|Z) = Pr(Y\|X, Z)Pr(X\|Z)$

Figure 21: Jaynes-Renyi like axioms based on conditional probability are motivated by two or more variable eikosograms.

variable provided it is the same wherever the variable appears. Axioms very much like these, but based on sets, were suggested by Renyi (1970, p. 38) (see also Fine, 1973, pp. 76-79). Jaynes (1996, Chapter 2) develops essentially these characteristics of probability as a consequence of reasonable postulates for a theory of plausibility; there, rather than sets, the variables represent logical propositions.

Marginal probability is now merely a notational convenience. One could take it to be the case that probabilities are always conditional and so a variable Z is always present. This is arguably more realistic in practice and complements the discussion surrounding Figure 16. When $Z = c$ represents the background condition common to all probability statements under consideration, it can be dropped to simplify notation. Alternatively, if the probability of Y is conditional on a random variable which takes only one value (such a variable can always be constructed) then it would be simpler to drop the condition rather than continually carry it around. With this notational simplification, it follows from axiom C3, for example, that without confusion we can write

$$Pr(Y, X) = Pr(Y|X)Pr(X)$$

as a matter of convenience.

If we consider the eikosograms first introduced in Figure 15 as describing the relevant probabilities given Z has taken some value, then the axioms of Figure 21 are straightforward observations. C1 asserts the non-negative nature of area, C2 a consistency axiom that given a condition holds, the probability that it is realized is 1, C3 a sum rule for the two regions in Figure 15(a), and C4 the product rule for the area of rectangular regions in Figure 15(b).

The Kolmogorov-like axioms of Figure 20 can be derived from these conditional axioms. The least obvious of these is M3 for which it is sufficient that the more general rule for inclusion-exclusion be derived. Obvious from an eikosogram, it is formally achieved from the axioms by repeated application of C3 and C4 as follows:

$$\begin{aligned} Pr(Y = \mathsf{y} \text{ or } X = \mathsf{y} \,|Z) &= 1 - Pr(Y = \mathsf{n}, X = \mathsf{n} \,|Z) \;=\; 1 - Pr(Y = \mathsf{n} \,|Z)Pr(X = \mathsf{n} \,|Y = \mathsf{n}, Z) \\ &= 1 - Pr(Y = \mathsf{n} \,|Z)(1 - Pr(X = \mathsf{y} \,|Y = \mathsf{n}, Z)) \\ &= Pr(Y = \mathsf{y} \,|Z) + Pr(Y = \mathsf{n}, X = \mathsf{y} \,|Z) \\ &= Pr(Y = \mathsf{y} \,|Z) + Pr(X = \mathsf{y} \,|Z) - Pr(Y = \mathsf{y}, X = \mathsf{y} \,|Z). \end{aligned}$$

The step from the third line to the fourth is a repetition of the steps from the first to the third, except focussed on changing the value of Y from n to y (this derivation essentially follows Jaynes, 1996, p. 210).

There are two other interesting features of the conditional axioms. First, C2': $Pr(Y|Y) > 0$ could replace C2, which would then follow as a consequence of the axioms. Second, it would seem then that probability is bounded by 1 only because of C3. This seemingly arbitrary choice has been criticised in the past (e.g. Fine, 1973). Yet if a different bound in C3 were proposed, say 10, and eikosograms constructed with sides from 0 to 10, then joint probabilities would no longer be the areas of the rectangles for these areas could exceed the bound for probability. Instead, the eikosograms suggest that axiom C4 would have to become

$$10 \times Pr(Y, X|Z) = Pr(Y|X, Z)Pr(X|Z)$$

for the joint probability to remain properly bounded by 10. Far from being arbitrary, a maximum of one is the only value which makes eikosograms work in the sense of having probabilities directly determinable (marginal and conditional from axes, joint from areas). Moreover, the eikosograms show that changing the bound would require C3 and C4 to change in concert (a point made by Jaynes, 1996, p. A-2, as well); C2 would follow with the correct bound, were C2' to be adopted instead.

All the usual rules of probability follow from either set of axioms. To be mathematically complete, however, an additional continuity axiom needs to be added in order to deal with variables that take on an infinite number of distinct values. The following matches Kolmogorov's (1933, p. 14) but first requires some notation. Let X_0 be a variable which takes on a possibly infinite number of distinct values: x_0, x_1, x_2, ... Now, define a sequence of variables X_0, X_1, ... each composed from the one before where X_i takes distinct values y_i, x_{i+1}, x_{i+2}, ... with y_i being "y_{i-1} or x_i" and $y_0 = x_0$. In this way, the first value of X_i gathers up more and more of the values as i increases. Kolmogorov's continuity axiom amounts to

M4	For the above sequence of variables $X_0, X_1, \ldots, X_i, \ldots$ $\lim_{i\to\infty} Pr(X_i = y_i) = 1$
C5	For the above sequence of variables $X_0, X_1, \ldots, X_i, \ldots$ $\lim_{i\to\infty} Pr(X_i = y_i \,|Z) = 1$

If there are only a finite number of distinct values, then the limit follows from the previous axioms and this one is unnecessary. As Kolmogorov (1933, p. 15) pointed out

> "...the new axiom is essential for infinite fields of probability only, it is almost impossible to elucidate its empirical meaning, as has been done, for example, in the case of the ... [previous axioms] ... For, in

describing any observable random process we can obtain only finite fields of probability. Infinite fields of probability occur only as idealized models of real random processes."

The new axiom extends probability to the case where an eikosogram is chopped into infinitesimally small slices and, as such, leaves behind intuition based on direct experience. The semantics of probability depend on the earlier axioms. The last reminds us, as Jaynes (1996, chapter 15) was at pains to point out, that passage to a mathematical limit occurs only from probability determined first for the finite case.

4.2.2 Agnostic on other interpretation

The axioms just presented depend on random variables and need not make any use of the word event. They do not appeal to sets for their development nor to plausibility of propositions nor to personal preferences or decisions. They neither support nor discount any theory developed for other purposes which might happen to coincide with these rules. In this sense, the eikosograms and the axioms we derive from them are independent of their applicability in any other context. The semantics of probability need only be grounded visually.

4.3 Further explorations

4.3.1 Association between variables

Because eikosograms are constructed from conditional probabilities, association between two variables is automatically displayed by them. Figure 22 shows different levels of association between the two binary variables Y and X.

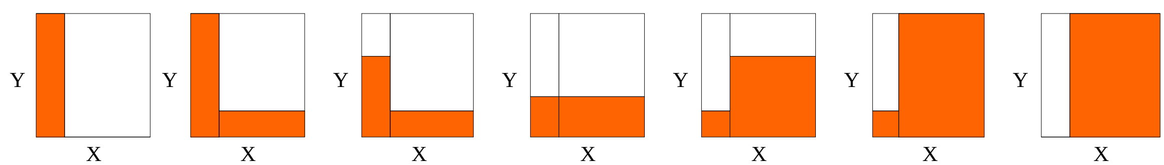


Figure 22: Binary associations left to right. (a) Perfect positive association: $Y = X$; (b) Part perfect positive association: $1 = Pr(Y = \mathsf{y}|X = \mathsf{y}) > Pr(Y = \mathsf{y}|X = \mathsf{n})$; (c) Positive association: $Pr(Y = \mathsf{y}|X = \mathsf{y}) > Pr(Y = \mathsf{y}|X = \mathsf{n})$; (d) Independence; (e) Negative association: $Pr(Y = \mathsf{y}|X = \mathsf{y}) < Pr(Y = \mathsf{y}|X = \mathsf{n})$; (f) Part perfect negative association; (g) Perfect negative association: Y and X are complementary.

Across these diagrams the marginal distribution of X is held fixed, that of Y is not. While association is easily seen in the pictures, providing a measure of association is not obvious; even the most commonly recommended one, the odds ratio, fails to distinguish 'Perfect' association (Figure 22a) from 'Part Perfect' association (Figure 22b) (or "Absolute" from "Complete" association as in Fienberg, 1977, pp. 18-19).

4.3.2 Events and variables

Events can be introduced in two ways if desired. First for any event, there is a corresponding binary random variable having value y when the event holds and value n when it does not. As a consequence, probability rules for events (e.g. as with Kolmogorov's for independence of events) follow from rules for binary random variables. Secondly, a random variable taking a particular value, say $Y = y_i$, can be thought of as an event $[Y = y_i]$, the construction of the corresponding binary variable being obvious. Square brackets are used to identify this interpretation as an event.

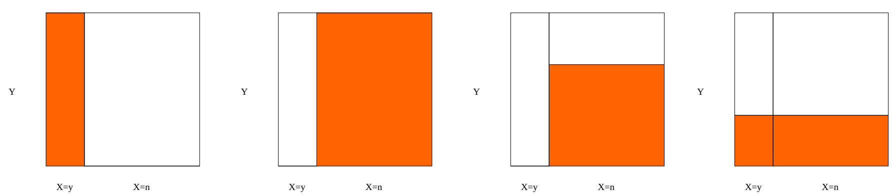
Figure 23 shows how eikosograms help distinguish the difference between independence of events and of random variables. Supposing X to take on more than two values, say $X = \mathsf{a}$, $X = \mathsf{b}$, or $X = \mathsf{c}$, then although the random variables Y and X are not independent, certain events defined by their values can be. Here, $[Y = \mathsf{y}]$ and $[X = \mathsf{a}]$ are independent events even though Y and X are dependent random variables. Symbolically, it is possible to have $[Y = \mathsf{y}] \perp\!\!\!\perp [X = \mathsf{a}]$ yet $Y \not\!\perp\!\!\!\perp X$.

(a) Dependent random variables Y and X (b) Independent events $[Y = \mathsf{y}]$ and $[X = \mathsf{a}]$

Figure 23: Some events of dependent random variables can themselves be independent.

4.3.3 Disjoint, complementary, coincident, and independent events

One consequence of using Venn diagrams to introduce probability was the need to carefully distinguish disjoint from independent events. This need, to a large extent, disappears when probability is introduced via eikosograms. Disjoint events now appear simply as different distinct values of a single random variable as illustrated for example by the values of Y in Figure 19(a); there three disjoint events appear, namely $[Y = y_1]$, $[Y = y_2]$ and $[Y = y_3]$. Nevertheless, should it be desirable to distinguish disjoint from independent events, the point is easily made via eikosograms as in Figure 24.



(a) Coincident (b) Complementary (c) Exclusive (d) Independent

Figure 24: Important possibilities for events $[Y = \mathsf{y}]$ and $[X = \mathsf{y}]$.

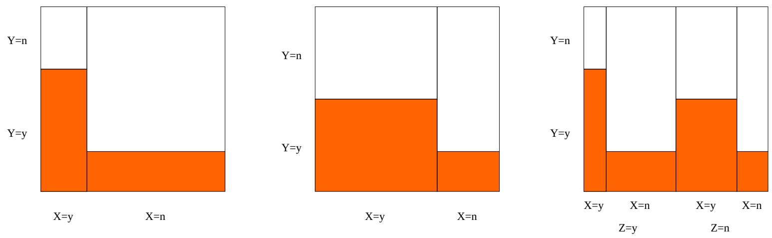
The first two of these appeared in Figure 22 to illustrate the two extremes of dependence relations. We now consider what they say about the nature of the events $[Y = \mathsf{y}]$ and $[X = \mathsf{y}]$. The first, Figure 24(a), shows two events which are coincident – when $[X = \mathsf{y}]$ occurs then $[Y = \mathsf{y}]$ must occur; when $[X = \mathsf{y}]$ does not occur then $[Y = \mathsf{y}]$ cannot occur. The second, Figure 24(b), shows the opposite, two events which are complementary – when $[X = \mathsf{y}]$ occurs then $[Y = \mathsf{y}]$ cannot; when $[X = \mathsf{y}]$ does not occur then $[Y = \mathsf{y}]$ must occur. This is the most extreme case of disjoint events; more typical are the disjoint events seen in Figure 24(c). There, again, the occurrence of one event excludes the possibility that the other can occur. However, unlike complementary events, one event not occurring need not ensure that the other must occur; it may be that neither event occurs. The final eikosogram, Figure 24(d), is the now familiar flat pattern which is the mark of independence.

The adjective ‘disjoint’, so strongly tied to non-overlapping rings in a Venn diagram, now seems out of place from the perspective of an eikosogram; the adjective ‘exclusive’ or ‘mutually exclusive’ seems better fitting. Even with many binary variables, or events, the mark of mutually exclusive events will be eikosograms with unshaded vertical strips everywhere except where all conditioning variables have value ‘n’ (i.e. the conditioning events do not occur); there the shading will occupy a portion of the vertical strip, all of it if the events are also complementary.

Confusing this visual pattern with that of independence is unlikely.

4.3.4 Three and more variables and conditional independence

In Figure 16, the eikosograms of one variable for each of the separate values of another were combined into a single eikosogram to display their joint distribution. In a similar fashion, in Figure 25 the eikosograms of Y given X for each value of a third variable Z are combined to give the eikosogram of all three variables in the form of Y given (X, Z). In this example, each value of Z has the same probability of $1/2$. As before, the width of individual strips gives the



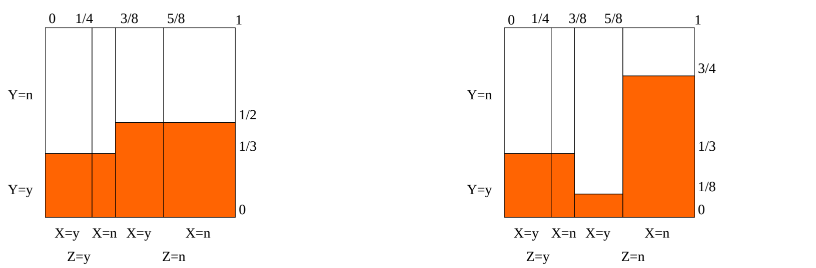
(a) When $Z = \mathsf{y}$ holds (b) When $Z = \mathsf{n}$ holds (c) Combined

Figure 25: The two joint eikosograms under conditions of (a) and (b) combine, or mix, according to the probabilities associated with $Z = \mathsf{y}$ and $Z = \mathsf{n}$, to produce (c) a combined eikosogram for Y given X and Z. In this example, the mixing probabilities are equal at 1/2.

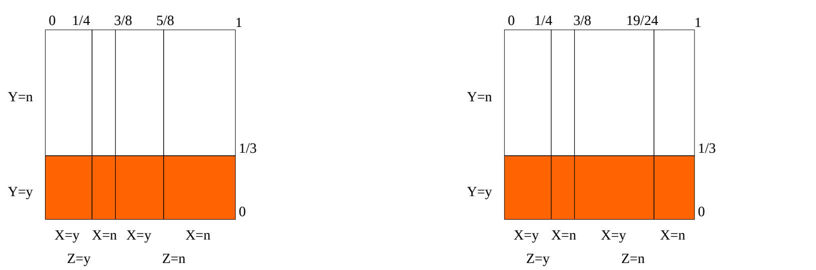
probability of each condition and the heights the corresponding conditional probabilities. This construction is general in that eikosograms for Y given the cross product of arbitrarily many conditioning variables can similarly be formed.

Figure 25(c) now provides a visual means to explore the joint distribution of Y, X, and Z. Those visual characteristics which were of interest in considering only two variables remain of interest conditionally when considering three random variables. For example, we notice that there seems to be a consistent, though not identical, positive association between Y and X conditional on either $Z = \mathsf{y}$ or $Z = \mathsf{n}$. The water container metaphor applies as before so that removing the barrier separating $X = \mathsf{y}$ from $X = \mathsf{n}$ for each value of Z will produce the eikosogram for Y given Z marginalized over X. Flat regions now indicate *conditional* independence of some sort.

Eikosograms displaying different sorts of conditional independence relations are shown in Figure 26. Note that a



(a) $Y \perp\!\!\!\perp X | Z$ (b) $Y \perp\!\!\!\perp X \mid [Z = \mathsf{y}]$ but $Y \not\!\perp\!\!\!\perp X \mid [Z = \mathsf{n}]$



(c) $Y \perp\!\!\!\perp X | Z$ and $Y \perp\!\!\!\perp Z | X$, but $X \not\!\perp\!\!\!\perp Z$ (d) Complete independence: $\perp\!\!\!\perp(Y, X, Z)$.

Figure 26: Some of the conditional independencies possible for Y given X and Z.

completely flat region as in both Figures 26(c) and (d) does not imply complete independence of all three variables. Complete independence between three variables occurs if and only if the eikosogram is completely flat whatever variable appears on the vertical axis; that $X \not\!\perp\!\!\!\perp Z$ in Figure 26(c) but $X \perp\!\!\!\perp Z$ in Figure 26(d) would require calculation or be made immediately apparent from an eikosogram having either X or Z on its vertical axis. Note, however, that from either Figure 26(c) or (d), application of the water container metaphor gives visual proof to the following ‘flat water' result (see Oldford, 2003a, for others):

$$Y \perp\!\!\!\perp X | Z \text{ and } Y \perp\!\!\!\perp Z | X \implies Y \perp\!\!\!\perp X \text{ and } Y \perp\!\!\!\perp Z.$$

The three way eikosograms explored here are each but one of nine possible arrangements of the joint probabilities of Y, X, and Z. Not all nine are always required; for example in Figure 25(c) the common level when $X = \mathsf{n}$ indicates $Y \perp\!\!\!\perp Z \mid [X = \mathsf{n}]$. Exploring the other arrangements gives much insight into the relationships between three variables but is beyond the scope of the present paper. Oldford (2003a) explores these in detail including their value in interpreting graphical and log-linear models.

5 Grounding events

Like probability, eikosograms presume that random variables have already been provided. In many instances, this will be true – categorical variates (e.g. sex, education level, programme of study, colour, etc.) arise immediately from many statistical contexts and are meaningfully grounded within that context. In other contexts, however, variables may not be immediately available but will need to be derived from whatever fundamental outcomes are possible and of interest. Events can be defined in terms of these outcomes and binary variables defined to indicate whether the event occurs or not.

One might think that this would be the proper place to use Venn diagrams, to define the events on which probability operates. However, Venn diagrams are ideally suited to describe *logical relationships between existing events*; what is needed are diagrams which help *define events* in the first place.

As is often the case, turning to historical sources where concepts were first correctly formulated can provide insight into how best to teach those concepts. After all, those earlier struggles are akin to those of students and, like students, those first formulating the concepts look for aids, diagrammatic and otherwise, which help naturally to clarify the concept itself.

5.1 Outcome trees.

Trees are perhaps the earliest diagrams used in probability dating back to at least Christiaan Huygen's use in 1676 (see Shafer, 1996). They are natural when the outcomes lead one to another in time. Figure 27(a) shows a simple

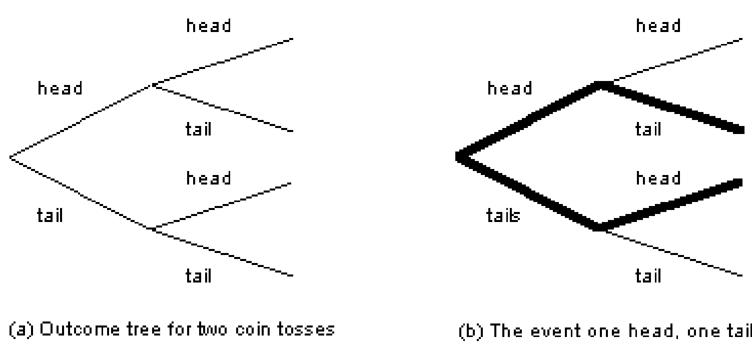


Figure 27: Defining events on an outcome tree.

tree describing two tosses of a coin. Branches at a point in the tree represent the mutually exclusive and exhaustive outcomes which could follow from that point.

While some notion of time is generally associated with movement from left to right across the tree, this is not strictly required. For some situations, the ordering of the tree branches might rather be one of convenience. For example, the tree of Figure 27 could also be used to provide a description for the simultaneous toss of two coins, with left and right components being labelled as "Coin 1" and "Coin 2".

Either way, the diagram provides a complete description of the situation under consideration in terms of all possible outcomes at each step – hence the name *outcome tree*.[6] If the branching probabilities were attached we would have the familiar *probability tree*. However, determining the probabilities is a separate stage in the probability modelling, and so it is best to spend some time with the outcome tree before moving on to this next stage.[7]

Events can now be defined by reference to the outcome tree. For example, the thick branches of Figure 27(b) show the event 'one head and one tail' without specifying which toss produced which. Similarly, if we were considering the event 'a head followed by a tail' only the topmost of the two thickly shaded paths would define the event; the bottommost of the two defines the event 'a tail followed by a head'. These two events combine to produce the first event of 'one head, one tail'.[8] The notion of *outcome space* (or more traditionally the *sample space*, a term we find to be less clear) could now be introduced as the set of all individual paths through the tree. An event, being a collection of paths, is simply a subset of the outcome space.

Outcome trees describe what can happen, step by step. The probability model is built on this structure by attaching conditional probabilities to each branch. The resulting probability tree will visually emphasize the conditional branching structure of the probability model whereas the corresponding eikosogram will visually emphasize the probability structure itself. One is easily constructed from the other since they contain the same information. The important difference is the different spatial priority each gives to the components of that information.

5.2 Outcome diagrams.

While outcome trees are often the most natural way to show *how* outcomes are possible, in some problems it is simpler just to show *what* outcomes are possible.

A notable early example of this approach is De Moivre's 1718 *Doctrine of Chances* in which he developed probability theory by addressing one problem after another. Although postdating Huygens (1676), no probability trees appear there. De Moivre did, however, find it convenient to completely enumerate all possible outcomes for some problems and, occasionally, to arrange these spatially in a table (e.g. De Moivre, 1756, p. 185). To each outcome, the number of 'chances' or frequency with which it can occur was attached and provided the information needed to determine the probability of any event composed from the listed outcomes.

In more modern times (dating to at least Fraser (1958) and predating standard use of Venn diagrams in probability books), it has been useful for teaching purposes to show all possible outcomes as spatially distinct points in a rectangular field as in Figure 28 (a). The spatial locations are arbitrary and so may be chosen so the events of interest

Figure 28: Defining events on an outcome diagram.

[6] Other authors, notably Edwards(1983) and following him Shafer (1996), prefer the name *event tree* for this diagram.

[7] Huygens's (1676) tree was not a probability tree in the modern sense. Huygens was interested in solving an early version of the gambler's ruin problem and labelled his branches with the 'hope' of winning (essentially the odds of winning at each stage) and the return due the gambler if the game were ended at that point. According to Shafer (1996, p.4) "[i]t was only after Jacob Bernoulli introduced the idea of mathematical probability in *Ars Conjectandi* that Huygens's methods became methods for finding 'the probability of winning'." (*Ars Conjectandi* was published posthumously in 1713.)

There are many interconnections between the players in this story. Jacob was the brother, teacher, and ultimately the mathematical rival of the Johann Bernoulli under whom Euler studied. Euler's father had attended Jacob's lectures and had lived with Johann at Jacob's house.

[8] This is the usual probabilistic use of the word event. Recently, in the development of a general theory for causal conjecture (one that depends heavily on the outcome tree description), Shafer has proposed calling such events *Moivrean events*. This then permits him to introduce what he calls *Humean events* to capture what common usage might consider to be a causal event in the tree structure. For example, the taking of a given branch might be considered the 'event' which 'caused' all that followed to be possible. The branch would be a Humean event whereas a Moivrean event must be one or more complete paths through the tree. With the introduction of Humean events for each branch, one can see why Shafer (1996) would choose to call these diagrams 'event trees'.

Since probability theory depends only on so-called 'Moivrean' events, we prefer 'outcome trees' to 'event trees'.

easily display as regions encompassing those outcomes which make up the event. In Figure 28(b) there are three non-overlapping regions which cover the entire field illustrating three mutually exclusive and exhaustive events. In Figure 28(c) two overlapping regions are drawn indicating two different events which have some outcomes in common.[9] In this figure, the unenclosed outcomes seem to constitute an event of no intrinsic interest; if they were of interest they would be best enclosed in a separate third region.

As with outcome trees, probabilities are missing from the outcome diagram. It is necessary to add them (usually to each individual outcome) in order to complete the probability model. Once outcome probabilities and events are in hand, any eikosogram for the events can be determined, although with more work than from a probability tree. Note however that, unlike probability trees, it will not generally be possible to construct an outcome diagram (and possibilities) from an eikosogram; at best only the construction of a Venn diagram (and attendant probabilities) will be possible.

6 Concluding remarks

Diagrams convey meaning. Throughout their history, ring diagrams have consistently been used to make concrete the concepts that abstract entities can be distinct from one another, have parts in common, or one necessarily entailing another. It seems unlikely, especially given their religious training and backgrounds, that either Euler or Venn would be unaware of this use of ringed diagrams in other contexts. Moreover, this knowledge would have preceded either man's original contribution to their use in logic: Euler to *describe* logical statements, Venn to *derive* them. Indeed, the relationship between the contributions of the background of ringed diagrams which prevailed at the time, Euler's use in explaining logic, and Venn's use in developing logic could be captured diagrammatically by three mutually intersecting rings.

Diagrams are important in learning any material, provided the diagram is well matched to that material. The eikosogram is just such a diagram for the introduction, definition, and exploration of probability and its attendant concepts such as conditional, marginal, and joint distributions as well as the more subtle concepts of probabilistic dependence and independence both unconditionally and conditionally.

Eikosograms obey Venn's dictum to match features of the diagram directly to the symbolic expression of the ideas. They fulfill Wittgenstein's notion of an 'ostensive definition' in that they can be used directly to define what is meant by these probability concepts. What eikosograms do not do is say how to use probability to model the real world.

This focus entirely on the mathematical abstraction of probability is a strength. Eikosograms permit a fundamental understanding of probability concepts to be had unclouded by the inherent difficulty of probability modelling. They do so by providing a definitive diagrammatic grounding for the symbolic expressions rather than one which appeals to some putatively natural application. Not only is the simultaneous introduction of probability and its application (often a source of confusion to many students) easily avoided but the important distinction between probability and model can be made early and more easily maintained thereafter.

If Venn's diagrams are to play a role in teaching probability it must be one considerably diminished from their present role. Outcome trees and probability trees have greater value for understanding events and the structure of a probability model. Eikosograms are coincident with probability. And outcome diagrams do much of the rest. Because of their inherent weaknesses for teaching probablity, it might be best at this time to avoid Venn diagrams altogether.

It is true that the intersecting ring diagrams are not original to Venn. But neither are they to Euler. The history of the diagrams, particularly in Christian symbolism, has shown them to be long associated with the demonstration of things separate and common to one another. This association is ostensibly inseparable from the diagrams. Given the religious training of both Euler and Venn, as well as the time periods in which these men lived, it seems likely that both men would have been aware of the vesica piscis and of the Christian symbolism associated with the two and three ring diagrams.

Euler's innovation was to use two-ring diagrams to demonstrate Aristotle's four fundamental propositions and to use more rings to illustrate the *known* outcomes of the syllogisms of Aristotelian logic. Venn, well aware of Euler's use, took the idea of intersecting rings (and of intersecting ellipses) to build a diagram which could be used to derive the consequence of possibly complex syllogisms as the logical information became available.[10] Each was an important

[9] Figure 28(c) is also a diagram which would be useful to ground Venn's diagrams in an application and is often used for that purpose. It is a mistake, however common, to call Figure 28(c) a Venn diagram.

[10] Venn even describes how to construct a physical apparatus based on the four ellipse diagram which can be used to carry out the logical calculations – foreshadowing today's digital, but electronic, computer.

and innovative use in its own right.

Historically and conceptually, eikosograms are direct descendants from Venn diagrams (e.g. Edwards, 1972). Their information content is that of probability and is easily organized and conveyed. Eikosograms should play a central role in teaching probability. Venn diagrams can be safely set aside, their value replaced by outcome trees and outcome diagrams.

References

Burkhardt, T. (1967). *Sacred Art in East and West*, Perennial, London.

Cleveland, W.S. (1985). *The Elements of Graphing Data*, Wadsworth, Monterey California.

de Finetti, B. (1937) "La prevision:ses lois logiques, ses sources subjectives", Ann. Inst. H. Poincaré, 7, pp. 1-68. English translation in Kyburg and Smokler (1981).

de Finetti, B. (1974) *Theory of Probability: A critical introductory treatment*, Vol. 1, John Wiley and Sons, New York.

DeMoivre, A. (1756). *The Doctrine of Chances, or, A Method of Calculating Probabilities of Events in Play: 3rd and Final Edition*, Reprinted 1967. Chelsea, New York.

Dewdney, A.K. (1999). *A Mathematical Mystery Tour*, John Wiley and Sons, New York.

Dunham, W. (1994). *The Mathematical Universe*, John Wiley and Sons, New York.

Edwards, A.W.F. (1972). *Likelihood*. Cambridge University Press, Cambridge.

Edwards, A.W.F. (1983). "Pascal's problem:the 'gambler's ruin'." *International Statistical Review*, 51, pp. 73-79.

Euler, L. (1761). Letters CII through CVIII of *Lettres à une Princesse d'Allemagne* dated Feb. 14 through March 7, 1761 appearing in English translation by D. Brewster as *Letters of Euler on different subjects in Natural Philosophy addressed to a German Princess, Vol. I* 1840, Harper & Brothers, N.Y. pp. 337 -366.

Feller, W. (1950). *An introduction to probability theory and its application* (1st Ed.), Wiley, NY.

Fienberg, S.E. (1977). *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA, MIT Press.

Fine, T.L. (1973). *Theories of Probability: An Examination of Foundations*.

Fraser, D.A.S. (1958). *Statistics: An Introduction*, Wiley, NY.

Friendly, M. (1994). "Mosaic displays for Multi-way contingency tables" *Journal of the American Statistical Association*, 89, pp. 190-200.

Friendly, M. (2002). "A brief history of the mosaic display." *Journal of Computational and Graphical Statistics*, 11. pp. 89-107.

Glock, H-J. (1996). *A Wittgenstein Dictionary*. Blackwell, Oxford U.K.

Halley, E. (1693). "An estimate of the degrees of mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw, with an attempt to ascertain the price of annuities on lives." *Philosophical Transactions*, 17, pp. 596-610.

Heath, T.L. (1908). *The Thirteen Books of Euclid's Elements, Volume I (Introduction and Books, I, II)*, Cambridge University Press, Cambridge, U.K.

Huygens, C. (1676). "My last question on those matters that I published on Reckoning in Games of Chance; once proposed by Pascal" in English translation by G. Shafer in Appendix A. of *The Art of Causal Conjecture.*

Jaynes, E.T. (1994). *Probability Theory: The Logic of Science*, Fragmentary Edition of June 1994, http://omega.albany.edu:8008/JaynesBook

Kolmogorov, A.N. (1933). *Grundgriffe der Wahrscheinlichkeitsrechnung*, Ergebneisse der Math. (2), Berlin. English Translation (1950), *Foundations of the Theory of Probability*, Chelsea Publishing Co., New York.

Kyburg, H.E. and H.E. Smokler (1958). *Studies in Subjective Probability* 2nd Edition, John Wiley & Sons, NY.

Liungman, C.G. (1991). *Dictionary of Symbols.* ABC-CLIO, Santa Barbara, CA (also on-line at www.symbols.com)

Mann, A.T. (1993). *Sacred Architecture*, Element Books, Dorset UK.

O'Connor, J.J. and E.F. Robertson (2001) "The MacTutor History of Mathematics Archive", http://www-history.mcs.st-andrews.ac.uk/history/, School of Mathematics and Statstics, University of St. Andrews, Scotland.

Oldford, R.W. (1995). "A physical device for demonstrating confounding, blocking, and the role of randomization in uncovering a causal relationship". *The American Statistician*, 49 (2), pp. 210-216.

Oldford, R.W. (2001a). "Theorem of Pythagoras", http://www.math.uwaterloo.ca/navigation/ideas/grains/pythagoras.shtml.

Oldford, R.W. (2001b). "Constant from Randomness (estimating pi)", http://www.math.uwaterloo.ca/navigation/ideas/grains/pi.shtml.

Oldford, R.W. (2003a). "Understanding probabilistic independence and its modelling via Eikosograms and graphs.". *(submitted).*

Oldford, R.W. (2003b). "Probability, problems, and paradoxes pictured by eikosograms". *(submitted).*

Pennick, N. (1980). *Sacred geometry: symbolism and purpose in religious structures*, Turnstone Press, Wellingborough

Plato (4th C. BCE). "The Meno Dialogue" (translated by W.K.C. Guthrie), pp. 353-384 of *Plato: The Collected Dialogues (including the Letters)* edited by E. Hamilton and H. Cairns, Princeton University Press, Princeton, NJ.

Santalo, L.A. (1976). “Integral Geometry and Geometric probability”, *Encyclopedia of Mathematics and its Applications*, Vol. 1, Addison-Wesley, Reading MA.
Shafer, G. (1996). *The Art of Causal Conjecture*. MIT Press, Cambridge MA.
Shafer, G. (2001). “The Notion of Event in Probability and Causality: Situating myself relative to Bruno de Finetti”. Presentation given in Pisa and Bologna in March 2001, unpublished Working Paper # 4, http://www.glennshafer.com/assets/downloads/italy.pdf
Venn, J. (1866). *The Logic of Chance*, (2nd Edition, 1876, 3rd 1888) Macmillan and Co., London.
Venn, J. (1880). “On the Diagrammatic and Mechanical Representation of Propositions and Reasonings.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. X (Fifth Series) July, pp. 1- 18.
Venn, J. (1881). *Symbolic Logic*, (2nd Edition, 1894) as reprinted 1971 by Chelsea Publishing Co., New York.
Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester UK, John Wiley & Sons.
Whittle, P. (1970). *Probability*, Penguin, Harmondsworth U.K.
Wittgenstein, L. (1964). *Philosophical Remarks*, University of Chicago Press, Chicago, IL.

Appendix: Use of Venn diagrams in probability texts

Judging by today's texts, one might have thought that Venn diagrams had been used in expositions of probability for well over 100 years since Venn first wrote about them, or at least dating back to the beginnings of the use of an axiomatic set theoretic approach to probability. But as the following table shows, this doesn't seem to be the case. The table summarizes the presence or absence of Venn diagrams for several books. Many authors used no diagrams or used their own diagrams. Some, like Gnedenko (a student of Kolmogorov) used Venn diagrams without calling them such. In any case use of the diagrams in probability seems to have been rare and certainly not popular until about a 100 years after Venn promoted them for symbolic logic.

Author	Date	Title	Use of Venn Diagrams
LaPlace	1812	*Théorie Analytique des Probabilités*	None
Venn, J.	1876	*Logic of Chance*	None
		3rd Edition 1888	None
Venn, J.	1881	*Symbolic Logic*	Introduction and extensive use
Venn, J.	1889	*The Principles of Empirical or Inductive Logic*	None
		2nd Edition 1907	None
Woodward, R.S.	1906	*Probability and the Theory of Errors*	None
Poincaré , H.	1912	*Calcul des Probabilités*	None
Keynes, J.M.	1921	*A Treatise on Probability*	None
Burnside, W.	1928	*Theory of Probability*	None
Kolmogorov, A.N.	1933	*Grundbegriffe der Wahrscheinlichkeitrechnung*; English (1950, 1956) *Foundations of the Theory of Probability*	None
Jeffreys, H.	1939	*Theory of Probability* (1st Ed.)	None
	1960	*Theory of Probability* (3rd Ed.)	None
Carnap, R.	1950	*Logical Foundations of Probability* (2nd Edition, 1962)	None
Feller, W.	1950	*An introduction to probability theory and its application*	**Yes**
Lévy, P.	1954	*Théorie de L'Addition des Variables Aléatoire*	None
Savage, L.J.	1954	*The Foundations of Statistics*	**Yes**. Savage also remarks (p. 11) that the set theoretic approach is "not ordinarily taught except in connection with logic or relatively advanced mathematics."
Loeve, M.M.	1955	*Probability Theory: Foundations, random sequences*	None
Cramér, H.	1955	*The Elements of Probability Theory and Some of its Applications*	None
Rényi, A.	1957	*Calcul des Probabilités*	None, but uses his own series of concentric circle diagrams to illustrate sets, their intersection and union
Fraser, D.A.S.	1957	*Nonparametric Methods in Statistics*	None
Fraser, D.A.S.	1958	*Statistics: An Introduction*	No. Instead he uses what we call *outcome diagrams* though he doesn't name them.
Dugue, D.	1958	*Ensembles Mesurables et Probabilisables*	None, but shows a (noncircular) set B nested within a larger (noncircular) set A
Derman, C.	1959	*Probability and Statistical Inference for Engineers*	None, even though it begins with a set theoretic approach
David, F.N. and D.E. Barton	1962	*Combinatorial Chance*	None
Hodges, J.L. Jr. and E.L. Lehmann	1965	*Elements of Finite Probability*	Uses *outcome diagrams* and a Venn-like diagram
Gnedenko, B.V.	1966	*Theory of Probability* (3rd Ed.)	Yes, but doesn't name them
	1968	*Theory of Probability* (4th Ed.)	Yes, but now introduced with quotes as: "so-called Venn diagrams"
Johnston, J.B., G.B. Price, and F.S. Van Kleck	1968	*Sets, Functions, and Probability*	Extensive use. Also uses non-overlapping rectangular regions for partitions because they: "afford a more realistic picture of this situation."
Lindley, D.	1969	*Introduction to Probability and Statistics from a Bayesian Viewpoint*	Not really, uses overlapping rectangular boxes for motivating axioms but curiously not for his conditional probability axiom
De Finetti, B.	1970	*Teoria Delle Probabilità*	Yes.
	1974	English: *Theory of Probability: A critical introductory treatment*	
Whittle, P.	1970	*Probability*	None
De Finetti, B.	1972	*Probability, Induction and Statistics: The Art of Guessing*	No use, but a related diagram is used to discuss limiting properties
Fine, T.L.	1973	*Theories of Probability: An Examination of Foundations*	None