Drawing Area-Proportional Euler Diagrams Representing Up To Three Sets

Peter Rodgers, Gem Stapleton, Jean Flower and John Howse

Abstract— Area-proportional Euler diagrams representing three sets are commonly used to visualize the results of medical experiments, business data, and information from other applications where statistical results are best shown using interlinking curves. Currently, there is no tool that will reliably visualize exact area-proportional diagrams for up to three sets. Limited success, in terms of diagram accuracy, has been achieved for a small number of cases, such as Venn-2 and Venn-3 where all intersections between the sets must be represented. Euler diagrams do not have to include all intersections and so permit the visualization of cases where some intersections have a zero value. This paper describes a general, implemented, method for visualizing all 40 Euler-3 diagrams in an area-proportional manner. We provide techniques for generating the curves with circles and convex polygons, analyze the drawability of data with these shapes, and give a mechanism for deciding whether such data can be drawn with circles. For the cases where non-convex curves are necessary, our method draws an appropriate diagram using non-convex polygons. Thus, we are now always able to automatically visualize data for up to three sets.

Index Terms— Information visualization; Venn diagrams; Euler diagrams.

1 INTRODUCTION

Euler diagrams are frequently used to visualize statistical data where a diagrammatic representation of the values of set intersections is required. This paper discusses techniques for automatically drawing such diagrams. An example of the output from our software can be seen in Fig. 1. Typically, curves (often circles) represent the sets, and the areas of the regions formed from the curve intersections are required to be given values. An example can be seen in Fig. 2, which visualizes the results of a medical study [21]. Other applications include crime control [7] and genetics [10].

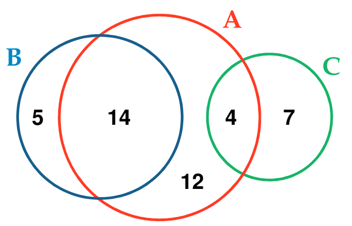


Fig. 1. An example area-proportional Euler-3 diagram.

Whilst Venn diagrams must include a region for every possible intersection of the curves in the diagram, Euler diagrams may omit some regions. Hence the diagram in Fig. 1 is an Euler diagram, rather than a Venn diagram. This is because some regions are missing. For instance, there is no region that is inside all three circles. We note that every Venn diagram is also an Euler diagram. Euler diagrams are called area-proportional when the relative region areas are taken to be of semantic importance.

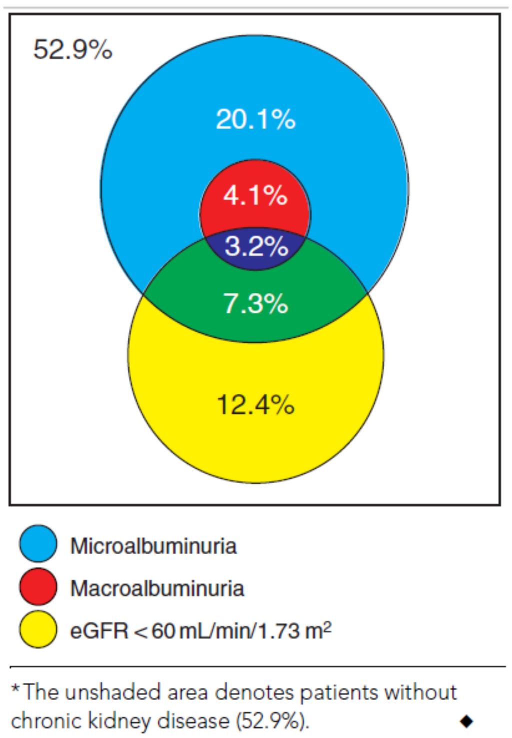


Fig. 2. The distribution of albuminuria and estimated glomerular filtration among patients with type 2 diabetes, obtained from [21].

- *Peter Rodgers is with the University of Kent, UK. E-mail: P.J.Rodgers@kent.ac.uk.*
- *Gem Stapleton is with the University of Brighton, UK. E-mail: G.E.Stapleton@brighton.ac.uk.*
- *Jean Flower is with Autodesk, UK.*
- *John Howse is with the University of Brighton, UK. E-mail: John.Howse@brighton.ac.uk.*

Manuscript received (insert date of submission if desired). Please note that all acknowledgments should be placed at the end of the paper, before the bibliography.

Euler diagrams [6] that represent three sets are frequently encountered. However, whilst some progress has been made to develop techniques to draw Venn-3 in an area-proportional manner [12], cases where not all regions appear have yet to be considered in any detail. We also note that cartograms, which are used in geographic visualization [5], superficially resemble area-proportional Euler diagrams. This resemblance is because, like Euler diagrams, the region areas of cartograms must be of the given values. However, Euler diagrams differ because they have curves that surround several regions, which are not found in cartograms. The curves impart extra information about set membership of the regions, hence visualizing area-proportional Euler diagrams requires ensuring that both the areas of regions are correct and that the curves surrounding regions are shown clearly. This is significantly different to visualizing cartograms where the areas of regions must be correct, but there is no requirement to show any curves.

DrawVenn is a method that draws Euler-3 with rectilinear shapes [4]. It can draw the diagrams that have an intersection between all three curves, but cannot draw diagrams that do not possess this triple intersection. Fig. 3 shows an Euler diagram with the same areas produced by both our method and DrawVenn. The numbers in the diagram indicate the areas. We argue that the lack of concurrency (curves overlapping) in the diagram on the left make it easier to interpret, whereas large sections of the border of the diagram on the right exhibit concurrency; for example, the curves labelled *A* and *B* run together on the bottom left and all curves run together on the bottom right. The curve labelled *A* is non-convex, making it relatively less easy to follow than the other curves in both diagrams. In some cases, the rectangular region shapes on the right are perhaps easier to compare by size (but this does not hold for all regions), whereas, on the left, the circular shapes for the curves makes overall comparison of the combined set sizes easier.

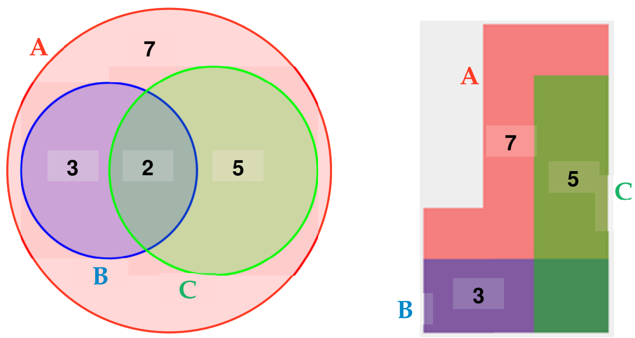


Fig. 3. Our method (left) compared against DrawVenn (right).

Venn-2 diagrams have been drawn exactly with circles [2] as have Venn-3 diagrams using convex polygons [12]. However, these techniques only draw the Venn case, ignoring all the other diagrams that can be drawn with three curves. A method for drawing a large (so-called monotonic) class of Euler diagrams with exact area-proportions has been implemented [2]. Unfortunately, many 3-curve diagrams cannot be drawn using this approach. Moreover, those diagrams that are drawn often unnecessarily break some wellformedness conditions (which are described later on in this section). Finally, a general exact area-proportional solution has been developed [18], but it has not been implemented; a restricted version of this method, that draws diagrams without regard for region areas, has been implemented but produces diagrams with poor layouts [19].

Other related work draws approximate area-proportional Euler diagrams with circles (or regular polygons) [3][10][22]. These techniques will often miss out required regions or will include regions not in the original specification, which is a significant shortcoming. In addition, these methods only approximate the required areas, so they fail to draw regions with the correct areas. These approximate methods do not ensure that the region areas are within any tolerance of the actual required areas. We acknowledge that under some circumstances it may be acceptable to approximate region areas. However it is not clear what is an acceptable error, which is likely to be difficult to define as it may vary by application area, shape of curve or shape of region. As an example of problems with these diagrams, an approximate diagram can be seen in Fig. 4. It shows the intersections of diagnosed asthma, chronic bronchitis, and emphysema within patients with obstructive lung disease. The intersection between Asthma and Emphysema, but not including Chronic Bronchitis, is clearly a larger area than that of the intersection of all three curves, but has a smaller value: 0.13% to 0.24%, leading to confusion on the part of those examining the diagram and there is potential for misinterpretation of the data.

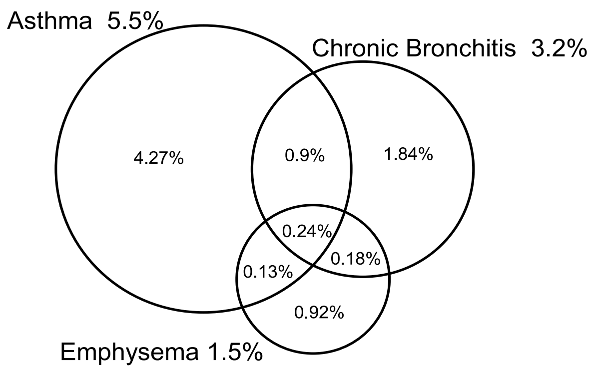


Fig. 4. An approximate area-proportional diagram, adapted from [15].

The aforementioned wellformedness properties include no concurrency, no triple points, the use of only simple curves and using only connected regions to represent set intersections - more detail is given in Section 2. The rectilinear diagram in Fig. 3 has concurrency and triple points. Breaking these properties is known to adversely affect user understanding [13] and so should be avoided where possible. There is also strong evidence that the shape of curves is considered important by users, as the majority of diagrams drawn by hand use circles rather than other forms of curve [22]. Evidence that users prefer smooth shapes over more jagged ones [1], alongside the notion that the use of simpler shapes means reduced cognitive load [14] compared to complex shapes, helps justify our preference for circles, followed by convex shapes.

The research contributions described this paper are:

1. to classify the 40 separate diagram descriptions drawable with three curves or fewer according to whether they can be drawn with circles or convex polygons;
2. to specify exactly when three-set area specifications can be drawn with circles;
3. to develop constructions for drawing an exact area-proportional Euler diagram for all 40 descriptions, so that any area specification for three sets or fewer can be drawn. These constructions minimize the well-formedness properties broken and attempt to minimize the number of vertices when polygons are used;
4. to describe a software implementation of these constructions, which prioritizes circles over convex polygons and uses non-convex shapes where necessary. The `Euler3` software is publically available: www.eulerdiagrams.com/Euler3.html.

A simplistic approach for drawing diagrams with fewer regions than Venn-3 would be to find a Venn-3 solution and shrink the unwanted regions to have no area. However, this leads to concurrency, triple points and non-simple (self-intersecting) curves. The more sophisticated approach taken in this paper is to examine the possible topological structures for the diagram, and to choose one that minimizes these effects. In fact, in many cases, no concurrency, triple points or self-intersecting curves need to appear at all.

The rest of the paper is organized as follows. Section 2 gives the definitions required for the following sections. Section 3 details our classification of Euler diagrams with three curves or fewer. Section 4 identifies when area specifications for these diagrams can be drawn with circles. When circles cannot be used, Section 5 discusses when convex curves can be used. Section 6 describes the software implementation, with more details given in the Appendix**.** Finally, Section 7 gives our conclusions and discusses further work.

2 Definitions

We define some of the terms we use throughout this paper. For a more formal approach, see [20].

Euler diagrams are collections of labelled simple closed curves. Note that we have incorporated the well-formedness condition of curve simplicity into our definition of Euler diagram. **Venn diagrams** are Euler diagrams where every possible intersection between curves is a non-empty, connected region. **Euler-*n*** is an Euler diagram containing exactly n curves. An Euler diagram, d, is **drawn with convex curves** if all of its curves are convex. Similarly, d is **drawn with circles** if all of its curves are circles.

A **triple point** is a point in the plane that is passed through at least three times by the curves in the diagram. A region is **disconnected** if it comprises more than one connected component. We also need to define concurrency. For our purposes, we need to distinguish two kinds of concurrency between curves. In particular, we define *complete concurrency* and *partial concurrency*. Two distinct curves are **completely concurrent** if they follow exactly the same path and they are **partially concurrent** if the two curves are not completely concurrent but segments of them follow the same path. The distinction is important because, as we shall see in Section 4, it can be possible to draw diagrams with complete concurrency using circles but not with partial concurrency. In Fig. 5, the diagram exhibits complete concurrency between B and C. Note that in this figure, the concurrent curves are shown slightly separated for clarity, an approach that is taken throughout this paper.

Given two completely concurrent curves, we can remove one of them, c say, from d, to give a diagram we call $d - c$ without altering the presence or absence of partial concurrency. Given d, we define a **concurrency reduced** diagram, denoted d_{cr}, where d_{cr} contains exactly one curve from each equivalence class under the completely concurrent equivalence relation. It should be obvious that any two concurrency reduced diagrams obtained from d are visually identical except for the choice of curve labels. If d_{cr} contains exactly n curves then we say d **reduces** to Euler-n. In Fig. 5, B and C are completely concurrent and removing either one of them from this diagram, d, yields a d_{cr} with two curves. Thus d reduces to Euler-2. Fig. 6 exhibits partial concurrency between the curves A and B.

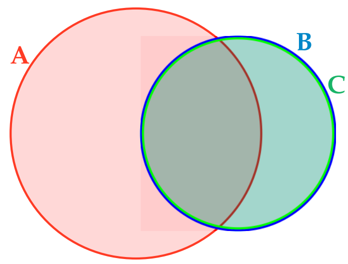


Fig. 5. Complete concurrency.

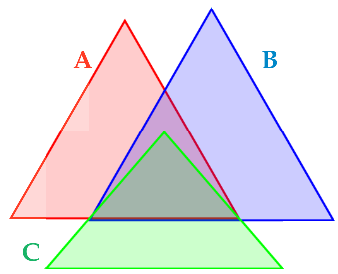


Fig. 6. Partial Concurrency

Minimal regions, or simply regions, are connected components of $\mathbb{R}^2$ less the (images of the) curves in the diagram and a **minimal region description,** or simply region description, is the set of labels of the curves that the minimal region is inside. In Fig. 5, there are four minimal regions: one is outside all three curves and has description $\emptyset$, a second is inside just A (so outside B and C) and has description $\{A\}$, a third is inside just B and C (so outside A) and has description $\{B, C\}$, and the fourth is inside all three curves and has description $\{A, B, C\}$. When drawing an area-proportional Euler diagram, we need to ensure that the minimal regions have the desired areas.

A **connected component** of a diagram is a maximal subset of the curves whose images form a single component in the plane [9]. Fig. 7 shows a diagram that contains two connected components, one comprises the curves A and B, the other comprises just curve C. The connectivity of a diagram is related to its drawability with given areas. For instance, in Fig. 7, we can arbitrarily alter the area inside the curve C, since this forms a connected component drawn 'independently' of A and B. In Fig. 8, even though C still forms a connected component, its area is constrained by the area of the overlap between A and B.

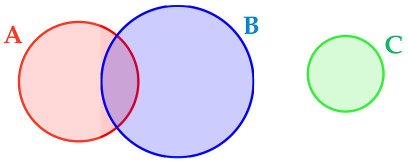


Fig. 7. An Euler diagram with 2 connected components.

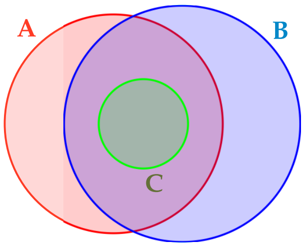


Fig. 8. An Euler diagram where C is contained by both A and B.

We are only considering Euler diagrams with three or fewer curves. The remainder of the definitions in this section will, where appropriate, be presented for this class of diagrams. We will use a fixed set of curve labels $\{A, B, C\}$. Then, for example, the minimal region inside three curves with these labels is described by $\{A, B, C\}$. When writing region descriptions we will abuse this notation and write, for example, ABC. Where it is convenient to do so, we blur the distinction between a minimal region and its description. For example, we will use the terminology BC for both the region description and the minimal region with that description. The unbounded 'outside' minimal region, $\emptyset$, is always present. In terms of the represented sets, in a diagram, AB represents $(A \cap B) - C$, A represents $A - (B \cup C)$, $\emptyset$ represents $U - (A \cup B \cup C)$, where U is the universal set, and so forth.

Given a diagram, it can be described by the descriptions of its minimal regions. For example, Fig. 1 shows a diagram with description $\{\emptyset, A, B, C, AB, AC\}$, we typically write this as $\emptyset\ A\ B\ C\ AB\ AC$. A **diagram description,** D, is defined to be a subset of $\mathbb{P}(\{A, B, C\})$ such that $\emptyset \in D$. A diagram, d, has description D whose elements are precisely the descriptions of the minimal regions in d. When there is more one diagram description under consideration, we will write D_d to mean the description of d.

An **area specification** is a function, w, where $w: \mathbb{P}\{\mathrm{A}, \mathrm{B}, \mathrm{C}\} - \{\emptyset\} \to \mathbb{R}^+ \cup \{0\}$. This indicates the required areas of the minimal regions in the diagram: given w, the **induced** diagram description, D_w, is

$$D_w = \{rd \in \mathbb{P}\{\mathrm{A}, \mathrm{B}, \mathrm{C}\}: w(rd) \neq 0 \vee rd = \emptyset\}.$$

In a diagram, the actual area of a minimal region, r, is denoted by $area(r)$. Given an area specification, w, we say that the diagram d **represents** w if:

- $D_d = D_w$, and
- for each minimal region, r, in d, $area(r) = w(r)$.

In Fig. 1, the numbers written in the minimal regions give the region areas, and the diagram represents the derivable area specification. Given an area specification, w, we say w can be **drawn with circles** if there exists a diagram that is drawn with circles and represents w. Similarly we can define when w is **drawn with convex curves**. *The goal of automatic layout of area-proportional Euler diagrams is to take an area specification and draw a diagram that represents it.*

3 Classification of Euler Diagrams with up to Three Curves

In [16], the number of Euler diagrams with particular numbers of curve labels was established. It was shown that there are essentially 40 different diagram descriptions with up to three curves, although there was no attempt at determining the actual descriptions. We only need to consider how to draw representative diagram descriptions from each equivalence class. Table 1 shows the 40 different diagram descriptions, with each row corresponding to an equivalence class. The Venn Diagram column shows a Venn diagram drawn with three circles where the shaded regions correspond to those not present in the diagram description.

In Section 4 we identify, for each diagram description, D, in Table 1 whether D is

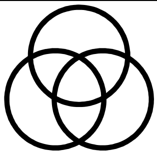
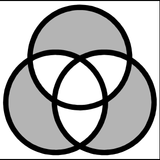
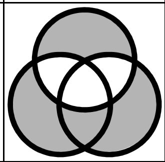
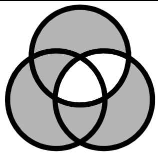
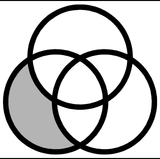
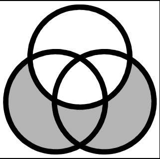
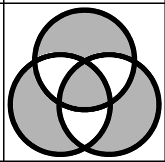
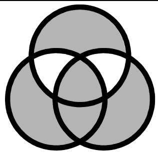
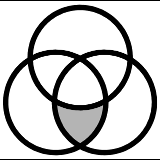
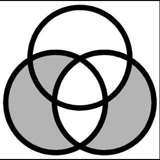
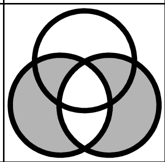
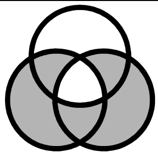
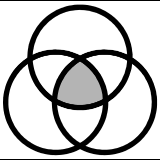
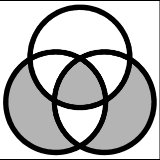
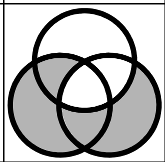
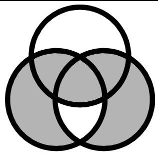
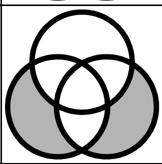
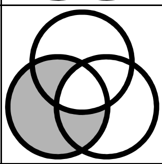
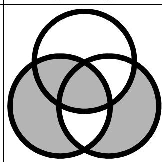
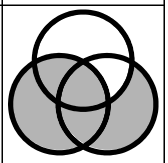
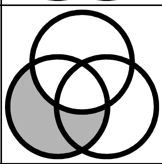
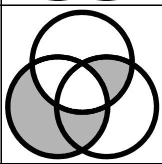
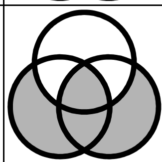
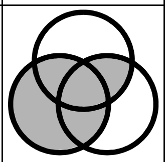
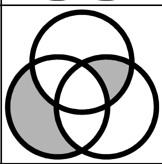
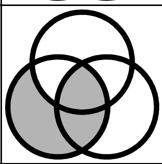
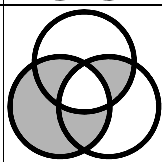
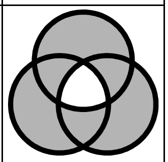
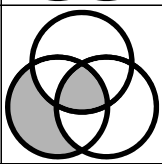
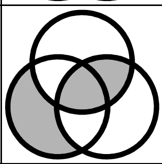
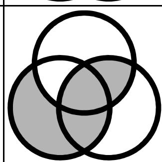
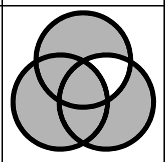
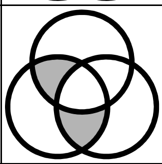
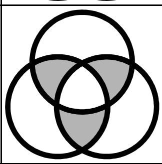
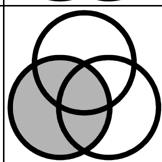
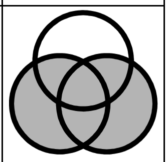
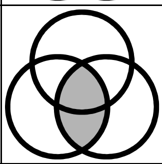
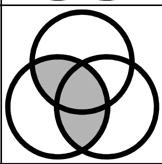
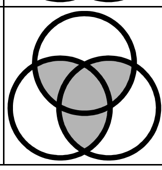
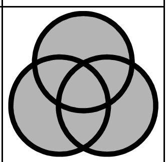
a) **always drawable:** given any area-specification, w, which induces D, w can be drawn with circles,
b) **drawable in a range**: some w that induces D can be drawn with circles and, when w can be drawn, any small change can be made to the area specification without impacting drawability,
c) **over-constrained:** some w that induces D can be drawn with circles but when w is drawable some small change made to the areas results in an undrawable case, and
d) **never drawable:** w is not drawable with circles.

Subsequently, Section 5 classifies each D similarly, but in terms of convex polygons.

4 Three-Circle Analysis

Here we establish which area specifications for three or fewer sets can be represented by diagrams drawn with circles. In many diagrams that follow, we do not label the curves to reduce clutter and adopt the convention that the curve labelled A is drawn in red, B is drawn in blue and C is drawn in green.

TABLE 1
Classifying diagram descriptions for up to three sets

No.	Diagram Description	Venn Diagram	No.	Diagram Description	Venn Diagram	No.	Diagram Description	Venn Diagram	No.	Diagram Description	Venn Diagram
1	∅ *A B C AB AC BC ABC*		11	∅ *AB AC BC ABC*		21	∅ *AB AC ABC*		31	∅ *AB ABC*	
2	∅ *A B AB AC BC ABC*		12	∅ *A AB AC ABC*		22	∅ *AB AC BC*		32	∅ *AB AC*	
3	∅ *A B C AB AC ABC*		13	∅ *A AB BC ABC*		23	∅ *A BC ABC*		33	∅ *A ABC*	
4	∅ *A B C AB AC BC*		14	∅ *A AB AC BC*		24	∅ *A AB ABC*		34	∅ *A BC*	
5	∅ *A AB AC BC ABC*		15	∅ *A B AB ABC*		25	∅ *A AB BC*		35	∅ *A AB*	
6	∅ *A B AB AC ABC*		16	∅ *A B AC ABC*		26	∅ *A AB AC*		36	∅ *A B*	
7	∅ *A B AC BC ABC*		17	∅ *A B AB AC*		27	∅ *A B ABC*		37	∅ *ABC*	
8	∅ *A B AB AC BC*		18	∅ *A B AC BC*		28	∅ *A B AC*		38	∅ *AB*	
9	∅ *A B C AB ABC*		19	∅ *A B C ABC*		29	∅ *A B AB*		39	∅ *A*	
10	∅ *A B C AB AC*		20	∅ *A B C AB*		30	∅ *A B C*		40	∅	

In order to draw an area specification, w, with circles, we first need to know the required area of each circle. Given a curve label, l, we can compute the required area of the circle, c, to be labelled l, denoted $area(l)$, by adding up the values of $w(r)$ where r contains l. Thus, the required radius, denoted $radius(l)$, of c can be derived from $area(l)$.

All that remains is to determine whether the centres of the circles can be chosen so that the minimal regions thus formed have the required areas. To proceed with this, we note that given any pair of circles that are not completely concurrent, they either overlap, one contains the other, or they have completely disjoint interiors. Here, we focus on the overlapping case. Given a diagram description, D, two distinct labels l_1 and l_2 **form a Venn-2** (i.e. they overlap) in D if D contains

(a) a region description containing both l_1 and l_2,
(b) a region description containing l_1 but not l_2, and
(c) a region description containing l_2 but not l_1.

In other words, if we remove the third label, l_3, from each minimal region description in D then we obtain all minimal region descriptions in $\mathbb{P}(\{l_1, l_2\})$. Given a description D, a set of curve labels, L, **pairwise form a Venn-2** whenever each pair of distinct curve labels in L form a Venn-2 in D. For example, in Fig. 1, A and B form Venn-2, as do A and C. However, B and C do not form a Venn-2. Hence, the set of labels $\{A, B, C\}$ does not pairwise form a Venn-2. Whilst we have only defined the concepts of forming a Venn-2 and pairwise forming a Venn-2 on diagram descriptions, they have obvious analogies in drawn diagrams. We will use the terminology at both levels.

Given two labels l_1 and l_2 that form a Venn-2, the required distance between centres of the to-be-drawn circles, denoted $cd(l_1, l_2)$, can be derived numerically as we know the radii of the circles and the area of their intersection (which is the sum of the areas of the region descriptions containing both l_1 and l_2); details can be found in [4]. Fig. 9 shows an illustrative example.

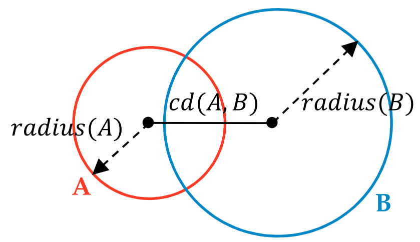


Fig. 9. For any given area specification, the radii and centre distance are fixed.

Relating this to drawability with circles, given two curves that pairwise form a Venn-2, the distances between their centres is fixed, given an area specification, w. This has the advantage that the layout of any two such curves is essentially unique. However, there is also a disadvantage: if the third curve needs to intersect with one, or both, of these two curves then it may not be possible to draw it as a circle in a manner that represents w, for example.

Sometimes the third curve needs to be drawn inside a minimal region, r, of the formed Venn-2. This is illustrated in Fig. 8, where C is drawn inside the minimal region AB of the Venn-2 comprising A and B. Whether C is drawable with the correct area depends on the *overlapping distance* between A and B. The **overlapping distance** of two curve labels, l_1 and l_2, in a diagram description D, denoted $od(l_1, l_2)$, is defined to be

$$od(l_1, l_2) = radius(l_1) + radius(l_2) - cd(l_1, l_2).$$

In a drawn diagram, the overlapping distance is defined to be the width of the lens formed from the overlapping region of the two circles. We require access to the overlapping distance in order to determine when a circle with a particular area can be drawn inside this lens.

Sometimes it is not just a single curve that we wish to draw inside a region, but a set of curves, as in Fig. 10. Here, B and C (each of which forms a connected component) are both drawn inside A. Clearly the areas of B and C are constrained by A. The sum of the diameters of B and C must be less than the diameter of A.

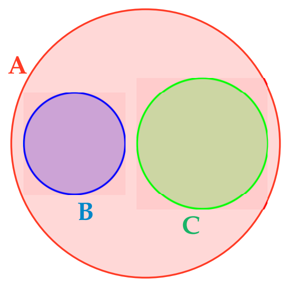


Fig. 10. An Euler diagram with B and C inside A.

Let d be a diagram and let r be a minimal region in d. An *r-component* is a non-empty set of connected components, K, of d that comprise, between them, all curves drawn properly inside the region that r becomes on removing all curves in K. More precisely, in d, an ***r*-component** is a non-empty set, K, of connected components, , of d such that all of the minimal region descriptions, r', in D_d that include a label of a curve in a component of K also:

(a) include the description (of) r, that is $r \subseteq r'$, and
(b) only include labels of curves in components of K or r, that is $r' \subseteq r \cup L(K)$, where $L(K)$ is the set of labels of the curves in the components of K.

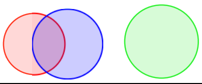
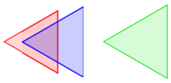
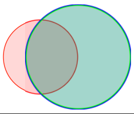
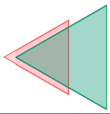
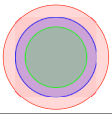
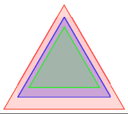
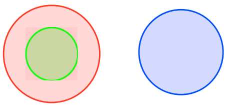
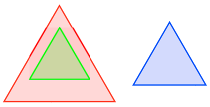
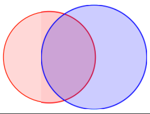
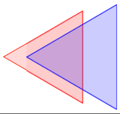
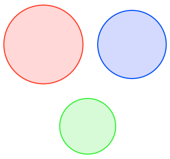
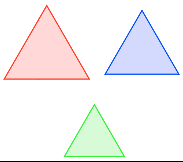
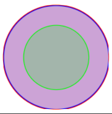
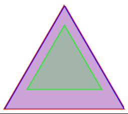
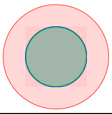
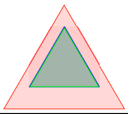
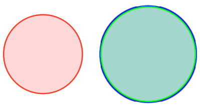
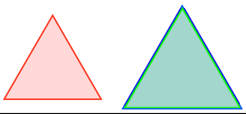
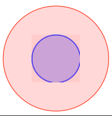
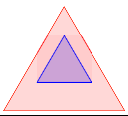
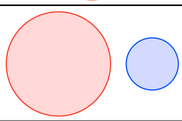
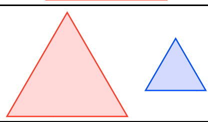
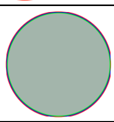
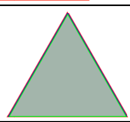
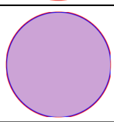
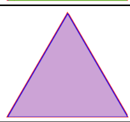
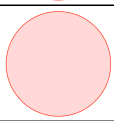
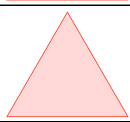
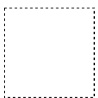
The r-components of a diagram descriptions, D, independently of a drawing of D, can be found using results in [17]. In Fig. 10 the circles B and C form an r-component where r is the minimal region with description A.

Given the concepts introduced here, we can now proceed to fully analyze the drawability of area specifications when using only circles.

4.1 Circles: Always Drawable

Here we identify the class of diagram descriptions that can always be drawn with circles. These are illustrated in Table 2. The diagrams shown in the columns Circle Construction and Polygon Construction are illustrative of those produced by our software tool (see Section 6).

TABLE 2: Always drawable with circles and convex polygons

Description	Circle Template	Polygon Template
20: ∅ *A B C AB*		
23: ∅ *A BC ABC*		
24: ∅ *A AB ABC*		
28: ∅ *A B AC*		
29: ∅ *A B AB*		
30: ∅ *A B C*		
31: ∅ *AB ABC*		
33: ∅ *A ABC*		
34: ∅ *A BC*		
35: ∅ *A AB*		
36: ∅ *A B*		
37: ∅ *ABC*		
38: ∅ *AB*		
39: ∅ *A*		
40: ∅		

Given a description, D, we say that D is **always drawable with circles** if it is listed in Table 2. We observe that the following conditions are true of the Circle Template diagrams in Table 2:

1. the connected components reduce to Euler-1 or Euler-2, and
2. for every non-outside region, r, (i.e. r is not described by $\emptyset$) that contains an r-component, K, we observe:
   a. the region that r becomes when K is removed has a boundary formed from a single curve, c_1,
   b. one of the curves, c_2, in K contains all other curves in K that are not completely concurrent with c_2.

We call these two conditions the **always drawable conditions.** Diagram descriptions that are not in Table 2 cannot be drawn as diagrams that satisfy these conditions. That is, these conditions completely describe the diagrams with up to three curves that are always drawable with circles. Extending our observations to area specifications, we have the following result:

> *given any area specification, w, whose induced description is always drawable with circles, w is drawable with circles.*

To justify this, we start by observing that the Circle Template representative drawings of the diagram descriptions in Table 2 satisfy the always drawable constraints. Thus, the representative drawings show that it is possible to draw the descriptions in Table 2 with circles. In all 15 cases it is trivial to verify that the required areas can be achieved by altering the radii and centre points of the circles. For instance, for description 20, draw circle A with area $area(A)$. Then compute $radius(B)$ and $cd(A,B)$ and use this to determine a centre of B. Finally, draw the circle C sufficiently far away from A and B, to ensure no overlap occurs. Arguments for the remaining cases are equally straightforward, observing that in Case 2a above it can be shown that c_1 can be drawn as a circle. We can then draw a circle, c_2, inside it; such a circle is given by the containing curve, c_2, in the contained r-component K.

In general, when automatically drawing these diagrams, the only significant issue is that of drawing a correct Venn-2 when it is present. The area-proportional layout for Venn-2 can be found using numeric methods [2].

We finish this subsection by noting our results imply that all area specifications for Euler-1 and Euler-2 diagrams are drawable with circles. The Euler-0, case (description 40) is also classified as drawable with circles. It is only when there are three curves that drawability problems arise. The following subsections determine whether the remaining cases are drawable in a range, are overconstrained, or never drawable with circles.

4.2 Circles: Drawable in a Range

This subsection identifies the class of diagram descriptions that can be drawn with circles in a practical subset of area specifications. These are illustrated in Table 3. Given a description, D, we say that D is **drawable in a range with circles** if it is listed in Table 3. We observe that the following conditions are true of the Circle Template diagrams in Table 3:

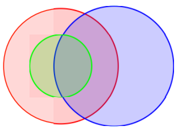
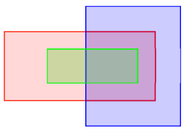
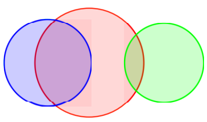
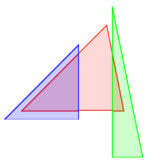
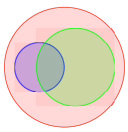
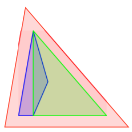
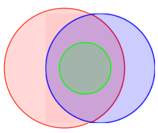
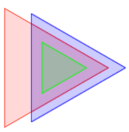
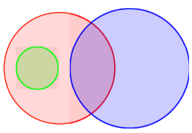
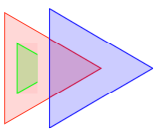
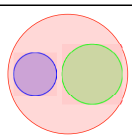
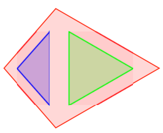
1. d is not always drawable with circles (i.e. it is not in Table 2),
2. d does not possess any partial concurrency, and
3. the three curve labels do not pairwise form a Venn-2.

We call these three conditions the **drawable in a range with circles conditions**. Diagram descriptions that are not in Table 3 cannot be drawn as diagrams that satisfy these conditions. As with the always drawable case, these conditions completely describe the diagrams with up to three curves that are drawable in a range with circles. We have the following result:

> *given any area specification, w, whose induced description is drawable in a range with circles, w is drawable with circles if and only if an additional constraint is satisfied.*

In each of the sections below, we will give such an additional constraint, which we call a *drawability constraint*, one for each of the six cases in Table 3; the drawability of w with circles is conditional on the relative circle radii and centres derived from the area specification we wish to represent. The conditions we give do not allow circles to touch in the limiting case, so that we avoid splitting the containing minimal region into two minimal regions.

TABLE 3: Drawable in a range with circles, and always with convex polygons

Description	Circle Template	Polygon Template
6: ∅ *A B AB AC ABC*		
10: ∅ *A B C AB AC*		
12: ∅ *A AB AC ABC*		
15: ∅ *A B AB ABC*		
17: ∅ *A B AB AC*		
26: ∅ *A AB AC*		

4.2.1 Description 6: ∅ A B AB AC ABC

Fig. 11 shows a representation of an area specification which induces description 6. Since the circles labelled A and B form Venn-2, the relative placement of their centre points is fixed. Essentially, this means that they cannot be moved relative to each other without changing the area specification. If the area inside the circle C is fixed, but we wish to enlarge the area of ABC (thus reducing the area of AC) then C would need to move to the right. But then we would have an unrequired minimal region, namely BC, and the other minimal regions would not all have the correct areas. Such a specification is undrawable with circles. Other alterations of the area specification are, however, sometimes drawable. Consider the case where we instead wish to reduce the area of ABC by a small amount whilst keeping the area of C fixed (thus increasing the area of AC). Here, C would need to move to the left in order to obtain the correct region areas, which is possible without introducing extra minimal regions and without making minimal region areas incorrect.

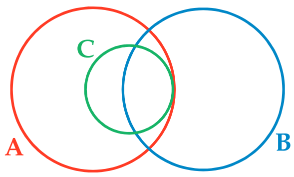


Fig. 11. ∅ A B AB AC ABC on the limit of drawability.

Given an area specification, w, which induces D_6, the areas for the required circles can be computed. The area of circle A is easily calculated by adding up the required areas for A, AB, AC and ABC. Similarly, the areas of B and C can be computed. In addition, we can numerically determine the distance between the centre points of A and B and between B and C. The distance between the centres of A and C must be within a specific range which ensures that C is drawn inside A. Using the notation for circle radius and overlapping distance as introduced at the beginning of Section 4, the drawability constraint on w is:

$$od(B,C) < od(A,B) \text{ and}$$
$$2radius(C) - od(B,C) < 2radius(A) - od(A,B).$$

4.2.2 Description 10: ∅ A B C AB AC

Fig. 12 shows a representation of an area specification which induces description 10. A change in the values of the minimal regions inside the A or C circles could make these circles overlap. The rightmost point of the B circle cannot be moved any further right and the leftmost point of the C circle cannot be moved any further left. Here, A forms a Venn-2 with each of B and C, but these two curves must not overlap. Thus, we have fixed distances between the centre of A and the centres of B and C and a lower bound on the distance between the centres of B and C. These constraints determine the drawability of w. The drawability constraint on w is:

$$2radius(A) > od(A,B) + od(B,C).$$

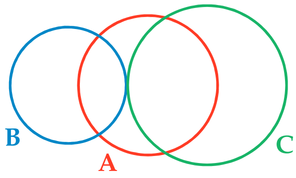


Fig. 12. $\emptyset$ A B C AB AC on the limit of drawability.

4.2.3 Description 12: $\emptyset$ A AB AC ABC

Fig. 13 shows a representation of an area specification which induces description 12. Here, any enlargement to the minimal regions inside A and B, or A and C, whilst keeping the rest fixed, would mean that B or C would go outside of A. Thus, such a change to the area specification is not drawable with circles. By contrast, reducing the areas of these minimal regions would not impact on drawability. Again, since B and C form a Venn-2, the distance between their centres is fixed. Each of B and C must have their centre within a certain distance of the centre of A, to ensure containment by A. Intuitively, the 'width' of the Venn-2 formed by B and C must be less than the diameter of A. The drawability constraint on w is:

$$2radius(A) > 2radius(B) + 2radius(C) - od(B, C).$$

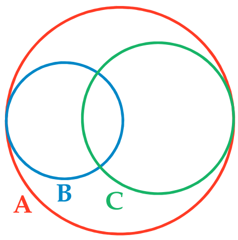


Fig. 13. $\emptyset$ A AB AC ABC on the limit of drawability.

4.2.4 Description 15: $\emptyset$ A B AB ABC

Fig. 14 shows a representation of an area specification which induces description 15. Since the relative layout of circles A and B is fixed, it is evident that we cannot make the circle C any larger without changing the diagram description. By contrast, we can make the area of C arbitrarily smaller without impacting on drawability. Here, the diameter of C must be less than the width of the overlap of A and B. The drawability constraint on w is:

$$od(A, B) > 2radius(C).$$

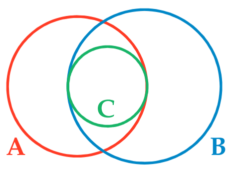


Fig. 14. $\emptyset$ A B AB ABC on the limit of drawability.

4.2.5 Description 17: $\emptyset$ A B AB AC

Fig. 15 shows a representation of an area specification which induces description 17. Since the relative layout of circles A and B is fixed, clearly we cannot make the circle C any larger without changing the diagram description, but we can make it smaller. The drawability constraint on w is:

$$2radius(C) < 2radius(A) - od(A, B).$$

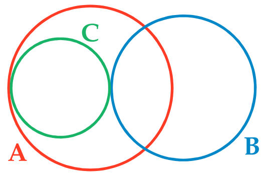


Fig. 15. $\emptyset$ A B AB AC on the limit of drawability.

4.2.6 Description 26: $\emptyset$ A AB AC

Fig. 15 shows a representation of an area specification which induces description 26. Here, the sum of the diameters of B and C cannot be larger than the diameter of A. The drawability constraint on w is:

$$radius(A) > radius(B) + radius(C).$$

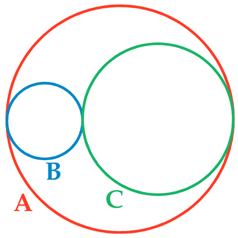


Fig. 15. $\emptyset$ A AB AC on the limit of drawability.

4.3 Circles: Over-Constrained

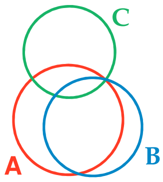
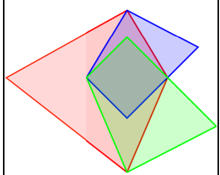
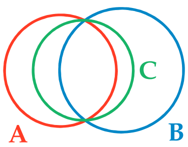
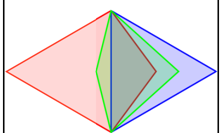
This subsection discusses the class of diagram descriptions that can be drawn with circles only for specially chosen area specifications. In practice, area specifications which induce such descriptions will not be drawable using circles. Given a description, D, we say that D is **over-constrained** if it is listed in Table 4a or Table 4b. We observe that the following conditions are true of the Circle Template diagrams in these tables:

1. d is not always drawable (i.e. it is not in Table 2),
2. d is not drawable in a range (i.e. it is not in Table 3), and
3. d does not possess any partial concurrency.

There are five diagram descriptions of this type: 1, 2, 3, 4 and 7. The circle drawing for these descriptions given in Tables 4a and 4b were not produced by our software as it is not sensible to give an automatic generation method for this class. This is because any input area specification is extremely unlikely to be drawable with circles. However, Tables 4a and 4b show hand-drawn circle versions for illustration.

All of the diagram descriptions in this class include three curve labels that pairwise form a Venn-2. This means that the relative layout of each pair of circles is essentially fixed and an appropriate diagram is only drawable under extreme conditions. For example see Fig. 17, which gives an example of how Venn-3 is constructed from pairs of Venn-2. When combined, the circle centres form a triangle with a fixed geometry, so the layout of the entire diagram is essentially fixed. Using this fixed relative positioning of the centres, all of the region areas in the Venn-2 diagrams are correct. This means that in a resultant combined diagram the areas of some regions are correct. For example the region formed from the union of minimal regions AB and ABC is correct. However, the area of each of the minimal regions (such as AB) is now fixed by how the Venn-2 minimal regions are divided by the curve not in the Venn-2 (such as C, which divides AB and ABC). This curve position is fixed in the diagram because of the distance of its centre from the centres of the other two curves. Hence, achieving the required areas in the final diagram is highly unlikely. Only if the area specification happens to coincide with the constructed centres and radii will the diagram be correct. In practice, this means that any real data set which induces one of these diagram descriptions will not be drawable with circles.

TABLE 4a: Over-constrained with circles,
always drawable with convex polygons

Description	Circle Template	Polygon Template
3: ∅ *A B C AB AC ABC*		
7: ∅ *A B AC BC ABC*		

In addition, it is entirely possible that the diagram formed from the Venn-2 sub-diagrams has the incorrect minimal regions present (this is almost a certainty when a triple point is required, (for example description 2, $\emptyset$ A B AB AC BC ABC) as the chance of the Venn-2 constructions leading to a point where all three curves intersect is vanishingly small, and an extra minimal region could be created.

4.4 Circles: Never Drawable

The class of diagram descriptions for which no area specifications exist that can be drawn with circles covers 14 Euler diagrams, as shown in tables 5a, 5b and 5c. Given a description, D, we say that D is **never drawable with circles** if it is in one of these three tables. We observe that any diagram, d, which represents one of these descriptions satisfies the following condition:

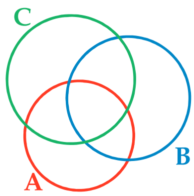
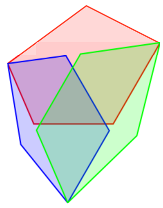
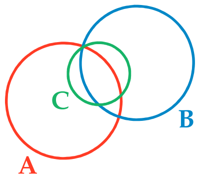
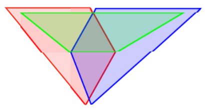
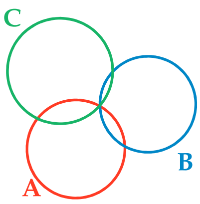
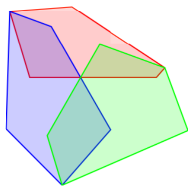
1. d possesses partial concurrency.

We call this condition the **never drawable with circles condition.** All of the cases that can be drawn without partial concurrency are covered in the previous classes. Extending our observations to area specifications, we have the following result:

> *given any area specification, w, whose induced description is never drawable with circles, w is not drawable with circles.*

To justify this, it is sufficient to observe that any pair of curves that must be drawn with partial concurrency cannot both be circles; here we note that the necessary presence of partial concurrency can be easily detected at the diagram description level for these cases, using the so-called connectivity conditions given in [8].

TABLE 4b: Over-constrained with circles,
sometimes drawable with convex polygons

Description	Circle Template	Polygon Template
1: ∅ *A B C AB AC BC ABC*		
2: ∅ *A B AB AC BC ABC*		
4: ∅ *A B C AB AC BC*		

5 Three-Polygon Analysis

As noted in Section 4, there are many 3-set area specifications that cannot be drawn with circles. Any practical 3-curve drawing system must have some way of dealing with these cases. Hence we introduce the notion of drawing area specifications with polygons. Whilst the usability of polygon based diagrams may be reduced compared to diagrams that only use circles, we can be confident of *always* generating accurate diagrams.

In this section, we derive drawability conditions when using convex polygons. As with the circle analysis of Section 4, we examine the diagrams in terms of when area specifications can be drawn with convex polygons. The analysis here is less complete than that for circles due to the variety of constructions, as each diagram description requires a tailored construction. The analysis is also made significantly more difficult by the increased degrees of freedom when drawing diagrams with convex polygons over circles. Where the area specification is not drawable with convex polygons, we use non-convex polygons. This means that any area specification can be drawn with our methods.

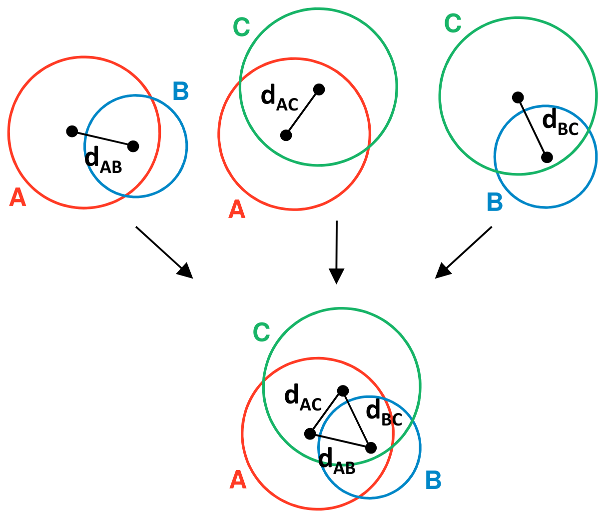


Fig. 17. How three Venn-2 pairs join to form Venn-3.

Clearly, there are a wide variety of ways for drawing each description with polygons. Hence we have produced a set of preferences that we use when deriving our constructions, in order of priority:

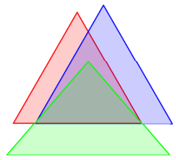
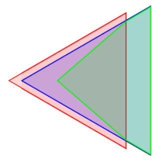
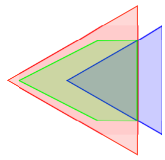
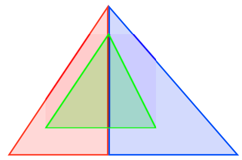
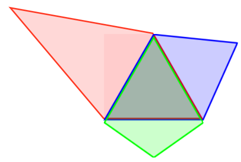
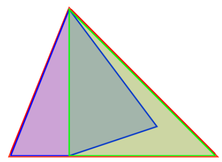
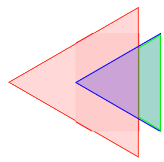
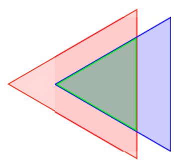
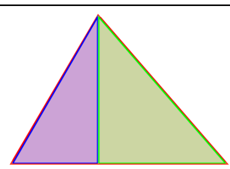
1. Use convex polygons wherever possible.
2. Minimize the number of concurrent polygon edges.
3. Minimize the number of polygon vertices.

However, we relax the third condition when there is some significant symmetry in the drawing (for example, description 22, $\emptyset\ AB\ AC\ BC$, could be drawn with fewer vertices, at the expense of symmetry, see Table 5c).

5.1 Convex Polygons: Always Drawable

In addition to the descriptions covered in Table 2, the other area-specifications always drawable with convex polygons are shown in tables 3, 4a and 5a. Justification for a diagram description being drawable with convex curves depends on the construction used to produce the diagram. The number of diagrams which are always drawable with convex curves (32) is larger than the number of circular diagrams (20) and so 12 individual drawability arguments need to be made. We do not provide formal arguments for all 12 cases, but note that each one follows a similar strategy: argue about how to draw a correct diagram using the Polygon Construction templates shown in the tables to achieve the required area specification, whilst maintaining convex curves.

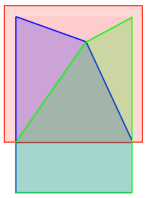
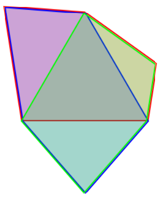
TABLE 5a: Never drawable with circles, always with convex polygons

Description	Polygon Construction
9: ∅ *A B C AB ABC*	
13: ∅ *A AB BC ABC*	
16: ∅ *A B AC ABC*	
18: ∅ *A B AC BC*	
19: ∅ *A B C ABC*	
21: ∅ *AB AC ABC*	
25: ∅ *A AB BC*	
27: ∅ *A B ABC*	
32: ∅ *AB AC*	

5.2 Convex Polygons: Drawable in a Range

The descriptions that are drawable in a range with convex polygons are those in Tables 4b and 5b; note that the polygon construction for diagram template 1, Venn-3, is has already been outlined in previous work [12]. Justification of drawability in a range is made by considering each of the cases. To illustrate the proof strategy, consider as an example description 5 in Table 5b. It is easy to see that it is possible to make some small changes to the area specification without impacting drawability. For instance, to enlarge the area of the minimal region A (inside only the red curve) just 'slide' one of the left, top, or right edges appropriately. Clearly the template diagram given does not allow all area specifications that induce description 5 to be drawn with convex curves. There is a lower bound on the area of the minimal region A: if we keep all other areas fixed, A cannot be reduced arbitrarily whilst maintaining convex curves.

TABLE 5b: Never drawable with circles,
sometimes with convex polygons

Description	Polygon Construction
5: ∅ *A AB AC BC ABC*	
11: ∅ *AB AC BC ABC*	

All of the drawable in a range diagram descriptions contain the three region descriptions that are formed of exactly two curve labels, namely AB, AC, and BC. For the descriptions that also include ABC, which are 1, 2, 5 and 11, a layout of these regions is like that shown on the left of Fig. 18. For description 4, the only remaining one in this class, the equivalent layout of regions AB, AC, and BC, without ABC is as shown on the right of Fig. 18. If we assume the convexity of the curves, it can be argued that the border of the depicted regions is not convex for some area specifications. Thus, if the area of a region inside a single curve (for example, the region A) is sufficiently small then at least one curve must be non-convex (e.g. make the area of the region A very small, so the curve labelled A becomes non-convex). This contradicts the convexity assumption. See Fig. 19 for a non-convex example of this type. This argument does not apply to diagram description 11, which does not include a region description containing exactly one curve label. However in this case the non-convexity of at least one curve for many area specifications is easy to establish. This is illustrated in Fig. 20. The only time all curves will be convex is when α, β and γ are all at least 180^0 (see Fig. 18).

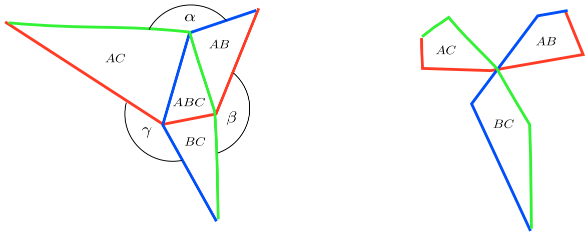


Fig. 18. Sections of two diagrams that have to be drawn with at least one non-convex curve.

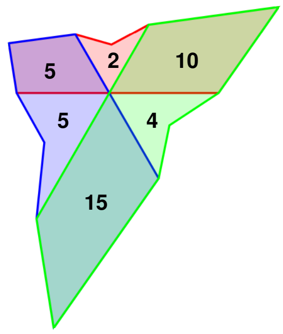


Fig. 19. A diagram template 4 with an area specification that results in non-convex curves.

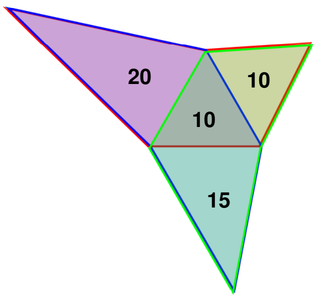


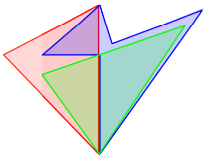
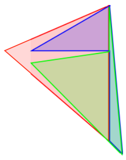
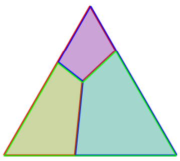
Fig. 20. Diagram template 11 with an area specification that results in non-convex curves.

Whilst we have attempted to find a constructions that produce convex curves for most area specifications, there may be alternative constructions for each of the diagram descriptions that allow area specifications to be drawn in a convex manner where they are not with our current constructions. Following on from this, it might be that some of these drawable in a range with convex polygon descriptions are, in fact, always drawable with convex curves but we conjecture otherwise. Establishing the non-drawability of particular area specifications in the convex polygon case is difficult due to the number of degrees of freedom. We note, though, that our constructions can be used to produce a correct drawing of each area specification that induce descriptions in this class, if non-convex curves are used.

5.3 Convex Polygons: Never Drawable

Unlike circles, the notion of over-constrained does not apply to convex polygons because of the greater degrees of freedom when using polygons. The class of diagrams that can never be drawn with convex curves consists of the three diagrams shown in Table 5c. These diagrams share the feature that the triple intersection, ABC, is not present. In addition, all of the double intersections, AB, AC, and BC, are present and there is at least one single intersection missing. We justify that description 8 is never drawable with convex polygons; in fact, our proof does not actually require that we use polygons: convex curves suffice for the truth of the result.

TABLE 5c: Never drawable with circles, never with convex polygons

Description	Polygon Construction
8: ∅ *A B AB AC BC*	
14: ∅ *A AB AC BC*	
22: ∅ *AB AC BC*	

Suppose, for a contradiction, that description 8 is drawable with convex polygons. Then, in particular, the curve C is drawable with a convex polygon. Choose such a polygon, P_C. Now, since the only two minimal regions inside P_C are those with descriptions AC and BC, we see that any choice of convex polygons for A and B must run concurrently inside P_C in a straight line (otherwise one of them would not be convex). This is illustrated in Fig. 21. However, to form the minimal region with description AB, it is then obvious that one of the polygons for A and B is necessarily non-convex, contradicting our assumption of drawability with convex curves.

It should be obvious that this argument readily applies to the other cases in Table 5c. Hence, all three descriptions in Table 5c are never drawable with convex polygons (or convex curves). However, the templates shown can be adjusted to achieve a drawing of any area specification that induces one of these three descriptions.

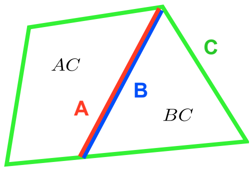


Fig. 21. If C is convex then either A or B must be non-convex.

6 Implementation

The drawing methods outlined above have been implemented in Java and draws area specifications with circles when possible and with convex or non-convex polygons otherwise. Typically, algebraic methods are used to find the polygon vertices. However some numeric solutions are required (for instance, finding a solution for Venn-2 drawn with circles) and more general search has been implemented in a few cases (for instance Description 22, ∅ AB AC BC, where the central intersection point is derived through a hill climbing search process). To ease the calculation of region areas, circles are approximated by 20 sided shapes. The software is quick enough so that all diagrams draw without any apparent delay. The software is available online: www.eulerdiagrams.com/Euler3.html. A screenshot of the tool can be seen in Fig. 22. More details on the implementation are given in the Appendix.

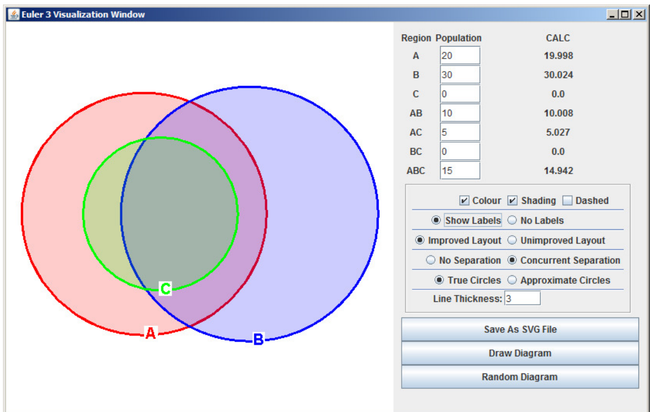


Fig. 22. A screenshot of the `Euler3` software.

7 Conclusions

In this paper we have shown how to draw all area specifications requiring up to three curves by using:

a) circles where possible,
b) convex polygons where possible, and
c) with non-convex polygons otherwise.

Further work in this area could take several directions. Firstly, examining other shapes is likely to be a profitable line of attack. Ellipses are both a desirable shape and have more degrees of freedom than circles (having a centre, major axis, a minor axis and a rotation). Using ellipses instead of circles should allow many more area specifications to be drawn with nice geometric shapes. Alternative constructions using polygons with a greater number of vertices may also prove fruitful, for instance to generate diagrams with rectangular regions, which may be easier to compare than the region shapes presented here. Tangential to this, is the notion of dynamic data sets, where the input data changes over time. It would be interesting to explore which constructions are best for maintaining the user's mental model of the diagram as the data varies over time.

Other work involves extending this research beyond three sets. An individual analysis of each case for larger diagrams is likely to be infeasible. However, general principles can be derived from results in this paper (such as the reducibility of diagrams and presence of partial concurrency), which can be used when extending the layout techniques to Euler-n. For instance, we trivially see that any area specification which reduces to a diagram description of an Euler-3 class can now be drawn. A little more subtly, we can detect the presence of sub-descriptions that are over-constrained or not drawable with circles on the basis of our results. For example, if a diagram description, D, contains three curve labels that form an over-constrained sub-description then D will either be over-constrained or undrawable with circles. We also note work in Bayesian reasoning with Euler diagrams [11] which uses diagram-text hybrids. This provides inspiration for further work in mixing notations, along with a discussion of more fundamental issues regarding the perception of these kinds of diagram.

Finally, we wish to establish when approximate drawings of area specifications are effective, which will be helpful when exact drawings are not possible. From a user perspective, an approximate result that still communicates the required information might be acceptable, particularly if that approximate result is drawn with preferred geometric shapes over convoluted curve shapes. Understanding what is acceptable as an approximate result (perhaps through empirical research) could help with this decision.

REFERENCES

[1] F. Benoy and P. Rodgers. Evaluating the Comprehension of Euler Diagrams. Proc. IEEE IV 2007. pp 771-778. 2007.

[2] S. Chow. Generating and Drawing Area-Proportional Venn and Euler Diagrams. PhD Thesis. University of Victoria, Canada. 2008.

[3] S. Chow and P. Rodgers. Constructing Area-Proportional Venn and Euler Diagrams With Three Circles. Euler Diagrams Workshop 2005.

[4] S. Chow and F. Ruskey, Drawing Area-Proportional Venn and Euler Diagrams, Graph Drawing 2003, LNCS 2912, pp. 466–477, Springer.

[5] D. Dorling. Area Cartograms: their use and creation. The Map Reader pp. 252-260. John Wiley and Sons. 1996.

[6] L. Euler, Lettres à une Princesse d'Allemagne, vol. 2., letters no. 102-108, 1761.

[7] G. Farrell, W. Sousa, Repeat Victimization and Hotspots: the Overlap and its Implication for Crime control and Problem-Oriented Policing, Crime Prevention Studies 12:221–240. 2001.

[8] J. Flower, J. Howse, Generating Euler diagrams, Proc. Diagrams 2002 LNCS 2317, Springer, pp.61–75. 2002.

[9] J. Flower, J. Howse, J. Taylor, Nesting in Euler diagrams: syntax, semantics and construction. Software and Systems Modelling 3:55–67. 2004.

[10] H.A. Kestler, A. Müller, J.M. Kraus, M. Buchholz, T.M. Gress, H. Liu, D.W. Kane, B.R. Zeeberg, J.N. Weinstein, Venn-Master: Area-Proportional Euler Diagrams for Functional GO Analysis of Microarrays, BMC Bioinformatics, 9(67). 2008.

[11] L. Micallef, P. Dragicevic, J.-D. Fekete, Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing. IEEE Trans. on Visualization and Computer Graphics, 18(12):2536-2545,. 2012.

[12] P. Rodgers, J. Flower, G. Stapleton, J. Howse, Drawing Area-Proportional Venn-3 Diagrams with Convex Polygons, Proc. Diagrams 2010, LNCS 6170, pp. 54-68, Springer. 2010.

[13] P. Rodgers, L. Zhang, H. Purchase. Wellformedness properties in Euler diagrams: which should be used?. IEEE Trans. on Visualization and Computer Graphics, 18(7):1089-1100, July 2012.

[14] M. Scaifeand, Y. Rogers. External Cognition: How Do Graphical Representations Work? Int. J. of Human-Computer Studies, 45: 185–213, 1996.

[15] J. Soriano, K. D. B. Coleman, G. Visick, et. al. The Proportional Venn Diagram of Obstructive Lung Disease. Chest, 124:474-481, 2003.

[16] G. Stapleton, A. Fish, P. Rodgers, Abstract Euler Diagram Isomorphism. Proc. 14th DMS (Visual Languages and Computing), pp. 310-317. 2008.

[17] G. Stapleton, J. Howse, P. Rodgers. A Graph Theoretic Approach to General Euler Diagram Drawing. TCS 411(1):91-112, 2010.

[18] G. Stapleton, P. Rodgers, J. Howse. A General Method for Drawing Area-Proportional Euler Diagrams, J. Visual Languages and Computing. 22(6):426-442. 2011.

[19] G. Stapleton, P. Rodgers, J. Howse, L. Zhang. Inductively Generating Euler Diagrams IEEE Trans. Visualization and Computer Graphics, 16(1):88-100. 2011.

[20] G. Stapleton, P. Rodgers, J. Howse, J. Taylor, Properties of Euler diagrams, Proc. Layout of (Software) Enginerring Diagrams, EASST Vol. 7, 2007.

[21] M.C. Thomas, et. al., The Burden of Chronic Kidney Disease in Australian Patients with Type 2 Diabetes (the NEFRON study). The Medical Journal of Australia 185 (3): 140-144. 2006.

[22] L. Wilkinson, Exact and Approximate Area-proportional Circular Venn and Euler Diagrams, IEEE Trans. Visualization and Computer Graphics. 18(2):321-331. 2012.

Peter Rodgers is a Reader in the School of Computing at the University of Kent, UK. He has a PhD in Computer Science (1997) and a First Class degree in Computer Science with Mathematics (1991). He has published papers in topics such as automatic layout of diagrams, drawing schematic networks and novel visual languages. He was general chair of Diagrams 2012 and sits on a number of conference program committees. He has been principal investigator on several significant EPSRC projects and is currently the recipient of a Royal Society Industrial Fellowship.

Gem Stapleton has a PhD in Computer Science (2004) and a First Class BSc (Hons) degree in Mathematics (2001). She is a Reader at the University of Brighton. Dr Stapleton is an Associate Editor for the Journal of Visual Languages and Computing and Chairs the Steering Committee for the Diagrams conference series. She was awarded Best Paper at Diagrams 2004, AOW 2009 and VLC 2012, having around 75 publications to-date, and has held research grants from the EPSRC and the Leverhulme Trust. Her research interests encompass diagrammatic logics and automated visualization.

Jean Flower has a PhD in Mathematics (1995), an MSc in Software Engineering (2000), obtained Distinction for the Part-III Mathematics degree at Cambridge, where she also obtained her BA (hons) Mathematics degree. She is employed by Autodesk as a Senior Principal Engineer. Dr Flower serves on the Program Committee for the Diagrams conference series. She was awarded Best Paper at Diagrams 2002 and Diagrams 2004, and has received funding from the EPSRC. Her current research interests include automated diagram drawing and 3D geometric modelling.

John Howse is Professor of Mathematics and Computation at the University of Brighton, UK. He has received a Most Influential Paper Award in 2009 from IEEE VL/HCC. Prof. Howse is on the steering committees of VL/HCC and the Diagrams conference series and on the program committees of a number of conferences and workshops. He has received significant funding from the UK EPSRC funding body for his pioneering research on diagrammatic reasoning. He has around 100 publications and his current interests encompass formal diagrammatic modelling and visualization.

Appendix for Drawing Area-Proportional Euler Diagrams Representing Up To Three Sets

Peter Rodgers, Gem Stapleton, Jean Flower and John Howse

A. Diagram Generation Methods

Here we detail the methods used to generate the diagrams, implemented in the code available at: www.eulerdiagrams.com/Euler3.html. Several approaches are evident. Some can be drawn with analytical methods where the curves can be placed directly using geometric operations. Others require the use of a numerical approach. Yet others have a more general search technique applied to position various points.

A.1 Analytical Methods

The diagram types in this section are drawn using geometric operations to either compute the centre and radius of circles, or the points of the polygons. This includes diagram types that use circles which do not include a pair of circles that form a Venn-2. We note that the calculation of circle area is simply the addition of region areas for a single set, from which the circle radius can be calculated. Analytical methods are also used for the majority of polygon layouts.

Throughout this appendix, where circle methods can be applied for a range of diagrams, the decision to use circles or polygons is dependent on the constraints given in Section 4.2. If the constraints hold, the circle type can be used, otherwise the polygon type is used.

Table A1a shows the circle types that can be drawn with analytical methods. The circle centre points in all these cases are trivial to determine.

TABLE A1a: Analytical circle diagram types

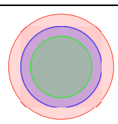
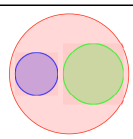
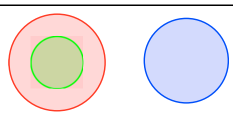
Description	Construction
24: $\emptyset$ ***A AB ABC***	
26: $\emptyset$ ***A AB AC***	
28: $\emptyset$ ***A B AC***	

TABLE A1a (continued)

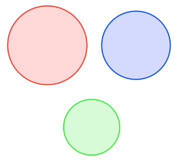
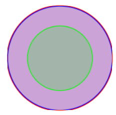
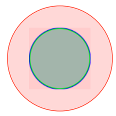
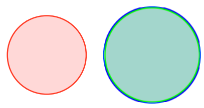
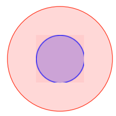
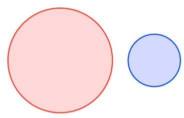
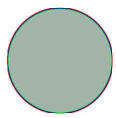
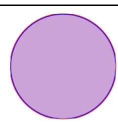
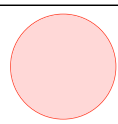
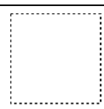
30: $\emptyset$ ***A B C***	
31: $\emptyset$ ***AB ABC***	
33: $\emptyset$ ***A ABC***	
34: $\emptyset$ ***A BC***	
35: $\emptyset$ ***A AB***	
36: $\emptyset$ ***A B***	
37: $\emptyset$ ***ABC***	
38: $\emptyset$ ***AB***	
39: $\emptyset$ ***A***	
40: $\emptyset$	

Table A1b shows the polygon types that can be drawn using analytical methods. This includes all the polygon equivalents of the analytic circle types (cases **24**, **26**, **28**, **30**, **31**, **33**, **34**, **35**, **36**, **37**, **38**, **39**, **40**). In these cases the triangles are equilateral, so their edge lengths are determined by the area required.

More interesting cases in Table A1b are those that can be drawn using circles but not with analytical methods, but can be drawn with polygons using analytical methods. These are:

- **Case 6** Rectangles are used because triangles cannot be guaranteed to work when some of the regions are very small. Here the region combining regions AB and ABC is drawn as a square, and the remaining region positions are calculated so that lines have even distance separation.
- **Case 8** The largest of AB and AC is formed into an equilateral triangle. The smaller of the two has the same height and so its base can then be calculated. The bases of these two triangles form the base of the triangle formed from AB and AC. The leftmost points on AB and AC are moved vertically so that they are distributed evenly between the upper right and lower right edges of curve A. The triangle BC has the same base as AC, so its height can be calculated. The rightmost point is placed on a line that extends the top left edge of AC. If possible, a top right point of the curve B is placed to extend the top left edge of curve A. However, if edge crossings occur an extra point is added so that curve B can route towards the centre of the diagram, following the other curves more closely.
- **Case 10** Non-regular triangles are used. First the triangle representing curve A is defined. The value for region A is split using a calculation that divides it based on the ratio of AB to AC. The splits are added to AB or AC as appropriate, and the largest value is taken. Let's assume this is AB. The value can be used to form a right angled triangle with two equal sides (base and height). This allows the calculation of AC with its share of the value of region A, as it must have the same height. The base of the triangle for curve A is then the base of the two triangles. Triangle forming regions AB and AC can then be derived, and the height of these extended to form triangles for curves B and C, with a suitable padding so that the extension above and below the A triangle are equal and the hypotenuses are parallel to the lines on the A triangle.
- **Case 12** Two congruent triangles are used. One is for the curve A, the other contains all the non-empty regions except region A. It is this second one that is calculated first, by applying the method used for polygon case **32** (given below) on the values for region AB and $AC + ABC$. The height of this triangle is used for the base of the inner triangle for region ABC. The right point of ABC is placed on the appropriate point of the line formed from the left point of AC and a point half way between the two rightmost points of AC. The triangle for A is placed at the same centre point as the inner triangle.
- **Case 14** An equilateral triangle is formed for regions A, AB and AC. The ratio of AB to AC is used to decide the point where the bases of the two triangles meet. The height of each can then be calculated, and the leftmost points are placed on a point on the line formed from the left point of curve A and the midpoint of their respective bases. The height of BC is calculated, and placed on a point that continues the left bottom edge of the AC triangle.
- **Case 15** This is formed from the triangle construction of diagram type number **29** (given below) and adding the inner equilateral triangle into the centre of the middle intersection for the curve C.
- **Case 17** This uses the construction of case **29** (given below). A congruent polygon to the left region is added into the middle of that region for the curve C.
- **Case 20** This is constructed from Case 29, given below, but with an extra equilateral triangle.
- **Case 23** This the same construction as Case 29, given below, but with a concurrent triangle.
- **Case 23** This the same construction as Case 29, given below, but with an extra equilateral triangle forming region ABC.
- **Case 29** Equilateral triangles are used, so the sizes and positions of all triangles is trivially derivable from the height of the triangles.

Some diagram types that cannot be drawn with circles can be drawn with polygons using analytical methods:

- **Case 3** The triple intersection is formed from a square. The diagonal of the square forms the base for triangles that have an area calculated from half the triple intersection area plus one of the double intersections, hence the height of these triangles can be derived. The height of these two triangles forms the base for the triangle including region A and half the double and triple intersections. The base and heights of triangles including regions B and C can also be derived by similar means.
- **Case 5** The base and height of the triple intersection are that of an equilateral triangle of the same area. The actual position of the top of the triangle is placed so the top of AB and AC are equal. BC is placed below the triple intersection triangle. If the diagram can be drawn in a convex manner, an equal padding from the left right and top of the other points is calculated for the curve A. If curve A must be non-convex, then an extra point to 'dent' curve A at the top is created, and the points around curve A are placed so that there is an even separation.
- **Case 7** An equilateral triangle is formed as the perimeter of regions A and AC. All the remaining points are formed from triangles with the same base as the equilateral triangle.
- **Case 9** The A and B curves are formed from equi-

lateral triangles and places appropriately. The height of the ABC triangle can be found as its base is the same as the base of the triangle formed from the intersection of curves A and B. This triangle is then extended downwards by a suitable amount to form the B region.

- **Case 13** The construction of case **29** (given above) is used for the values of regions A, B and $AC + ABC$. The height of the ABC triangle can be derived as its base is the intersection between the A and C curves
- **Case 16** The construction of case **29** (given above) is used for the values of regions A, B and ABC. Then, the extra points in curve C are formed, first by finding an appropriate point on the line between the left curve C point and halfway along the gaps on the left of curve C where there is no concurrency. If this cannot produce a diagram because the points are to the left of the other curve C points, the C curve is drawn as a triangle. This works because, in that case, the AC value is small.
- **Case 18** The method used for polygon case **32** (given below) is applied on the values for regions AC and BC. Taking the smallest of $A + AC$ and $B + BC$, it is made congruent to the triangle it surrounds and placed so that its vertical edge is on top of the vertical edge of the other two diagrams. This forms the height of the larger triangle, so that its base can then be calculated.
- **Case 19** The triple intersection is drawn as an equilateral triangle, the edges of which form the bases of the other triangles, whose height can then be calculated. They point directly away from the centre of the equilateral triangle.
- **Case 21** The method used for polygon diagram type number **32** (given below) is applied on the values for region AB and $AC + ABC$. The height of this triangle is used for the base of the inner triangle for region ABC. The right point of ABC is placed on the appropriate point of the line formed from the left point of AC and a point half way between the two rightmost points of AC.
- **Case 25** This uses the construction of case **29** (given above) for the three regions. A polygon for the curve C is added that fills the rightmost region.
- **Case 32** The equilateral triangle height and base of the largest of AB or AC is calculated. The base also forms the base of the smaller triangle, so its height can be calculated. The left and rightmost points of the triangles are placed in a position so that the base of the larger triangle can be formed.

TABLE A1b: Analytical polygon diagram types

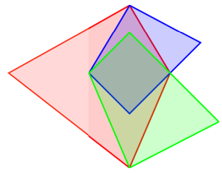
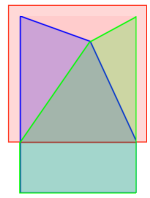
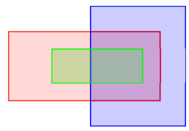
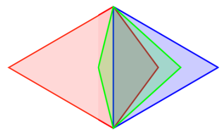
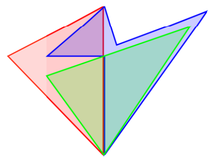
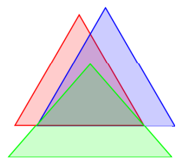
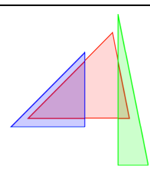
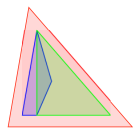
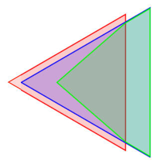
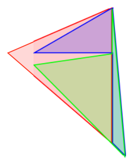
Description	Construction
3: ***∅ A B C AB AC ABC***	
5: ***∅ A AB AC BC ABC***	
6: ***∅ A B AB AC ABC***	
7: ***∅ A B AC BC ABC***	
8: ***∅ A B AB AC BC***	
9: ***∅ A B C AB ABC***	
10: ***∅ A B C AB AC***	
12: ***∅ A AB AC ABC***	
13: ***∅ A AB BC ABC***	
14: ***∅ A AB AC BC***	

TABLE A1b (continued)

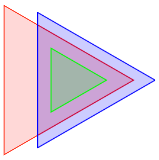
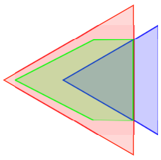
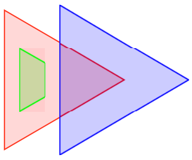
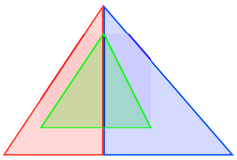
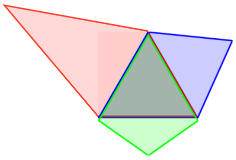
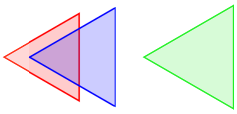
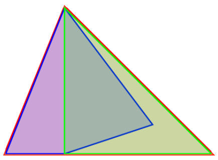
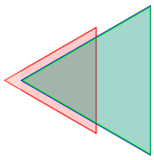
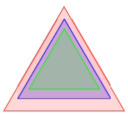
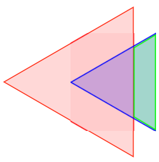
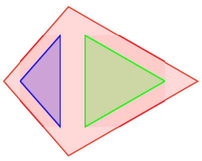
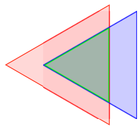
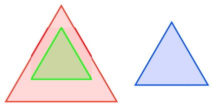
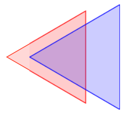
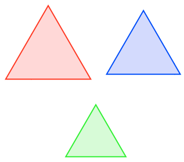
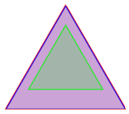
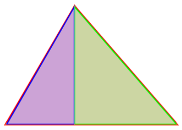
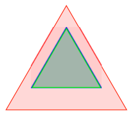
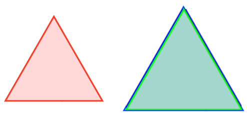
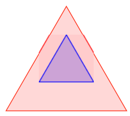
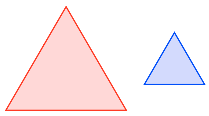
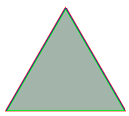
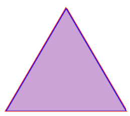
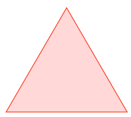
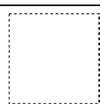
15: ∅ *A B AB ABC*	
16: ∅ *A B AC ABC*	
17: ∅ *A B AB AC*	
18: ∅ *A B AC BC*	
19: ∅ *A B C ABC*	
20: ∅ *A B C AB*	
21: ∅ *AB AC ABC*	
23: ∅ *A BC ABC*	
24: ∅*A AB ABC*	
25: ∅ *A AB BC*	
26: ∅ *A AB AC*	

TABLE A1b (continued)

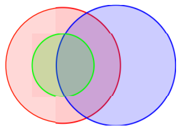
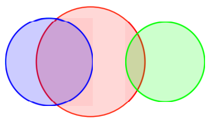
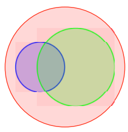
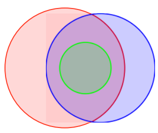
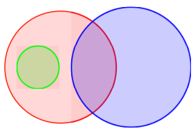
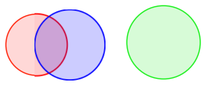
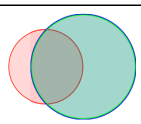
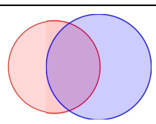
27: ∅ *A B ABC*	
28: ∅ *A B AC*	
29: ∅ *A B AB*	
30: ∅ *A B C*	
31: ∅ *AB ABC*	
32: ∅ *AB AC*	
33: ∅ *A ABC*	
34: ∅ *A BC*	
35: ∅ *A AB*	
36: ∅ *A B*	
37: ∅ *ABC*	
38: ∅ *AB*	
39: ∅ *A*	
40: ∅	

A.2 Numeric approximation

The diagram types in this section require a numeric approximation. In this case, bisection is used.

One important case is where two circles that form a Venn-2 are present. Here, numerical approximation has to be used to find the distance of the circle centres. These are shown in Table A2a. In all these cases, the y coordinate of the circle centre is the same for all circles in the diagram, so the position of the circle centres is easily defined. Where two pairs of circles form a Venn-2 (cases 6 and 10), two numerical approximations are applied.

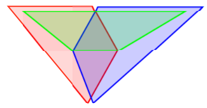
TABLE A2a: Numeric approximation circle diagram types

Description	Construction
6: ∅ *A B AB AC ABC*	
10: ∅ *A B C AB AC*	
12: ∅ *A AB AC ABC*	
15: ∅ *A B AB ABC*	
17: ∅ *A B AB AC*	
20: ∅ *A B C AB*	
23: ∅ *A BC ABC*	
29: ∅ *A B AB*	

One polygon diagram type makes use of a numerical approximation (Table A2b). This is:

- **Case 2** The region ABC is drawn as a equilateral triangle. Triangles are formed for AC and BC that have a base which is an edge of the middle triangle. They are placed so that their top edges are in a straight line. A numerical approximation is used for placing the bottom point of the AB triangle. It is positioned so that the two angles connecting the AB triangle to AC and BC are equal, maximizing the chances of a convex diagram. If a convex diagram is possible then curves A and B are extended by a suitable amounts and joined up at the far left and right. If a convex diagram is not possible then curves A and B are extended by a set amount and an extra point added between either or both of the left A point and bottom A point or between the right B point and the bottom B point.

TABLE A2b: Numeric approximation polygon diagram type

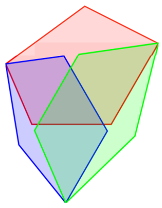
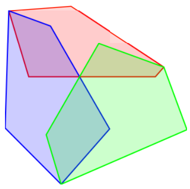
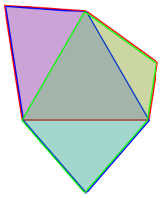
Description	Construction
2: ∅ *A B AB AC BC ABC*	

A.3 Search based methods

Here, some aspect of the drawing process needs a more general search. A hill-climbing approach is used. The diagrams are given in Table A3:

- **Case 1** This diagram type is described in the previous literature [11]. The region ABC is formed as an equilateral triangle and a search method moves it around, and consequently the points on the double intersections, namely AB, AC and BC, are moved until the 'cut-outs' in the core are minimized. If the diagram can be drawn in a convex manner, then the curves are drawn with one additional point. Otherwise, if the diagram is non-convex, three additional points are required to form each curve, so that the curves can dip inside the cut-outs.
- **Case 4** A triangle is formed for the three intersection points where two curves meet. The lines crossing the middle can then be placed to make the two set areas correct. A search, varying the triangle is made until the 'cut-outs' are of equal size. The outside segments of the curves can then be added, with one extra point if the diagram is convex, and three, dipping into the cut outs, added if the diagram is non-convex.
- **Case 11** The triple intersection is an equilateral triangle. The outer points each form a triangle with an edge of the inner triangle as their base. They are placed by a search based method that test the position of all three along lines that maintain the correct height from their base to achieve convex curves if possible.
- **Case 22** An equilateral triangle is formed from AB+AC+BC. The middle point is found by searching the space inside the triangle until a position is found where all three of the regions have the correct area.

TABLE A3: Numeric approximation polygon diagram types

Description	Construction
1: ∅ *A B C AB AC BC ABC*	
4: ∅ *A B C AB AC BC*	
11: ∅ *AB AC BC ABC*	
22: ∅ *AB AC BC*